

21st Edition

# HARRISON'S® PRINCIPLES OF INTERNAL MEDICINE

LOSCALZO

FAUCI

KASPER

HAUSER

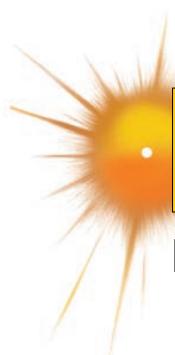
LONGO

JAMESON

VOLUME 1

Mc  
Graw  
Hill

21st Edition



**HARRISON'S®**

PRINCIPLES OF

**INTERNAL  
MEDICINE**

The title features a stylized sunburst graphic on the left side. The word "HARRISON'S" is in large yellow capital letters, with a registered trademark symbol (®) at the top right. Below it, "PRINCIPLES OF" is in smaller black capital letters. The main title, "INTERNAL MEDICINE", is in large red capital letters.

Yusuf Yazici

365 Inflammatory Myopathies.....	2819
<i>Steven A. Greenberg, Anthony A. Amato</i>	
366 Relapsing Polychondritis.....	2826
<i>Carol A. Langford</i>	
367 Sarcoidosis.....	2829
<i>Robert P. Baughman, Elyse E. Lower</i>	
368 IgG4-Related Disease.....	2837
<i>John H. Stone</i>	
369 Familial Mediterranean Fever and Other Hereditary Autoinflammatory Diseases.....	2840
<i>Daniel L. Kastner</i>	

## SECTION 3 Disorders of the Joints and Adjacent Tissues

370 Approach to Articular and Musculoskeletal Disorders.....	2844
<i>John J. Cush</i>	
371 Osteoarthritis.....	2854
<i>David T. Felson, Tuhina Neogi</i>	
372 Gout and Other Crystal-Associated Arthropathies.....	2862
<i>Hyon K. Choi</i>	
373 Fibromyalgia.....	2868
<i>Leslie J. Crofford</i>	
374 Arthritis Associated with Systemic Disease, and Other Arthritis.....	2871
<i>Carol A. Langford, Brian F. Mandell</i>	
375 Periarticular Disorders of the Extremities.....	2878
<i>Carol A. Langford</i>	

## PART 12 Endocrinology and Metabolism

### SECTION 1 Endocrinology

376 Approach to the Patient with Endocrine Disorders .....	2881
<i>J. Larry Jameson</i>	
377 Mechanisms of Hormone Action .....	2884
<i>J. Larry Jameson</i>	
378 Physiology of Anterior Pituitary Hormones .....	2891
<i>Shlomo Melmed, J. Larry Jameson</i>	
379 Hypopituitarism .....	2896
<i>Shlomo Melmed, J. Larry Jameson</i>	
380 Pituitary Tumor Syndromes .....	2902
<i>Shlomo Melmed, J. Larry Jameson</i>	
381 Disorders of the Neurohypophysis .....	2918
<i>Gary L. Robertson, Daniel G. Bichet</i>	
382 Thyroid Gland Physiology and Testing.....	2926
<i>J. Larry Jameson, Susan J. Mandel, Anthony P. Weetman</i>	
383 Hypothyroidism.....	2933
<i>J. Larry Jameson, Susan J. Mandel, Anthony P. Weetman</i>	
384 Hyperthyroidism and Other Causes of Thyrotoxicosis .....	2938
<i>J. Larry Jameson, Susan J. Mandel, Anthony P. Weetman</i>	
385 Thyroid Nodular Disease and Thyroid Cancer.....	2946
<i>J. Larry Jameson, Susan J. Mandel, Anthony P. Weetman</i>	
386 Disorders of the Adrenal Cortex.....	2955
<i>Wiebke Arlt</i>	
387 Pheochromocytoma.....	2976
<i>Hartmut P. H. Neumann</i>	
388 Multiple Endocrine Neoplasia Syndromes.....	2983

R. V. Thakker

389 Autoimmune Polyendocrine Syndromes .....	2992
<i>Peter A. Gottlieb, Aaron W. Michels</i>	

### SECTION 2 Sex- and Gender-Based Medicine

390 Sex Development .....	2997
<i>Courtney Finlayson, J. Larry Jameson, John C. Achermann</i>	
391 Disorders of the Testes and Male Reproductive System.....	3006
<i>Shalender Bhasin, J. Larry Jameson</i>	
392 Disorders of the Female Reproductive System .....	3027
<i>Janet E. Hall, Anuja Dokras</i>	
393 Menstrual Disorders and Pelvic Pain .....	3033
<i>Janet E. Hall, Anuja Dokras</i>	
394 Hirsutism .....	3039
<i>David A. Ehrmann</i>	
395 Menopause and Postmenopausal Hormone Therapy .....	3043
<i>JoAnn E. Manson, Shari S. Bassuk</i>	
396 Infertility and Contraception .....	3050
<i>Anuja Dokras, Janet E. Hall</i>	
397 Sexual Dysfunction .....	3055
<i>Kevin T. McVay</i>	
398 Women's Health.....	3063
<i>Emily Nosova, Andrea Dunaff</i>	
399 Men's Health .....	3069
<i>Shalender Bhasin</i>	
400 Lesbian, Gay, Bisexual, and Transgender (LGBT) Health .....	3078
<i>Baligh R. Yehia, Zachary B. R. McClain</i>	

### SECTION 3 Obesity, Diabetes Mellitus, and Metabolic Syndrome

401 Pathobiology of Obesity.....	3080
<i>Stephen O'Rahilly, I. Sadaf Farooqi</i>	
402 Evaluation and Management of Obesity.....	3087
<i>Robert F. Kushner</i>	
403 Diabetes Mellitus: Diagnosis, Classification, and Pathophysiology .....	3094
<i>Alvin C. Powers, Kevin D. Niswender, Carmella Evans-Molina</i>	
404 Diabetes Mellitus: Management and Therapies.....	3104
<i>Alvin C. Powers, Michael J. Fowler, Michael R. Rickels</i>	
405 Diabetes Mellitus: Complications .....	3120
<i>Alvin C. Powers, John M. Stafford, Michael R. Rickels</i>	
406 Hypoglycemia .....	3129
<i>Stephen N. Davis, Philip E. Cryer</i>	
407 Disorders of Lipoprotein Metabolism.....	3135
<i>Daniel J. Rader</i>	
408 The Metabolic Syndrome.....	3150
<i>Robert H. Eckel</i>	

### SECTION 4 Disorders of Bone and Mineral Metabolism

409 Bone and Mineral Metabolism in Health and Disease .....	3157
<i>F. Richard Brinthurst, Henry M. Kronenberg, Eva S. Liu</i>	
410 Disorders of the Parathyroid Gland and Calcium Homeostasis .....	3169
<i>John T. Potts, Jr., Harald Jüppner</i>	
411 Osteoporosis.....	3191

<i>Robert Lindsay, Blossom Samuels</i>	
412 Paget's Disease and Other Dysplasias of Bone .....	3209
<i>Rajesh K. Jain, Tamara J. Vokes</i>	
<b>SECTION 5 Disorders of Intermediary Metabolism</b>	
413 Heritable Disorders of Connective Tissue.....	3217
<i>Joan C. Marini, Fransiska Malfait</i>	
414 Hemochromatosis.....	3230
<i>Lawrie W. Powell, David M. Frazer</i>	
415 Wilson's Disease .....	3235
<i>Stephen G. Kaler</i>	
416 The Porphyrias .....	3237
<i>Robert J. Desnick, Manisha Balwani</i>	
417 Disorders of Purine and Pyrimidine Metabolism .....	3248
<i>John N. Mecchella, Christopher M. Burns</i>	
418 Lysosomal Storage Diseases .....	3254
<i>Robert J. Hopkin, Gregory A. Grabowski</i>	
419 Glycogen Storage Diseases and Other Inherited Disorders of Carbohydrate Metabolism.....	3261
<i>Priya S. Kishnani</i>	
420 Inherited Disorders of Amino Acid Metabolism in Adults.....	3268
<i>Nicola Longo</i>	
421 Inherited Defects of Membrane Transport .....	3274
<i>Nicola Longo</i>	

## PART 13 Neurologic Disorders

### SECTION 1 Diagnosis of Neurologic Disorders

422 Approach to the Patient with Neurologic Disease .....	3277
<i>Daniel H. Lowenstein, S. Andrew Josephson, Stephen L. Hauser</i>	
423 Neuroimaging in Neurologic Disorders .....	3282
<i>William P. Dillon</i>	
424 Pathobiology of Neurologic Diseases.....	3293
<i>Stephen L. Hauser, Arnold R. Kriegstein, Stanley B. Prusiner</i>	

### SECTION 2 Diseases of the Central Nervous System

425 Seizures and Epilepsy .....	3305
<i>Vikram R. Rao, Daniel H. Lowenstein</i>	
426 Introduction to Cerebrovascular Diseases .....	3324
<i>Wade S. Smith, S. Claiborne Johnston, J. Claude Hemphill, III</i>	
427 Ischemic Stroke .....	3335
<i>Wade S. Smith, S. Claiborne Johnston, J. Claude Hemphill, III</i>	
428 Intracranial Hemorrhage.....	3348
<i>Wade S. Smith, J. Claude Hemphill, III, S. Claiborne Johnston</i>	
429 Subarachnoid Hemorrhage .....	3353
<i>J. Claude Hemphill, III, Wade S. Smith, Daryl R. Gress</i>	
430 Migraine and Other Primary Headache Disorders.....	3357
<i>Peter J. Goadsby</i>	
431 Alzheimer's Disease .....	3370
<i>Gil D. Rabinovici, William W. Seeley, Bruce L. Miller</i>	
432 Frontotemporal Dementia .....	3378
<i>William W. Seeley, Bruce L. Miller</i>	
433 Vascular Dementia .....	3381
<i>Steven M. Greenberg, William W. Seeley</i>	
434 Dementia with Lewy Bodies.....	3385

<i>Irene Litvan, William W. Seeley, Bruce L. Miller</i>	
435 Parkinson's Disease .....	3386
<i>C. Warren Olanow, Anthony H.V. Schapira</i>	
436 Tremor, Chorea, and Other Movement Disorders .....	3400
<i>C. Warren Olanow, Christine Klein</i>	
437 Amyotrophic Lateral Sclerosis and Other Motor Neuron Diseases .....	3410
<i>Robert H. Brown, Jr.</i>	
438 Prion Diseases .....	3416
<i>Stanley B. Prusiner, Michael Geschwind</i>	
439 Ataxic Disorders .....	3422
<i>Roger N. Rosenberg</i>	
440 Disorders of the Autonomic Nervous System .....	3427
<i>Richard J. Barohn, John W. Engstrom</i>	
441 Trigeminal Neuralgia, Bell's Palsy, and Other Cranial Nerve Disorders .....	3436
<i>Vanja C. Douglas, Stephen L. Hauser</i>	
442 Diseases of the Spinal Cord .....	3445
<i>Stephen L. Hauser</i>	
443 Concussion and Other Traumatic Brain Injuries .....	3456
<i>Geoffrey T. Manley, Benjamin L. Brett, Michael McCrea</i>	
444 Multiple Sclerosis .....	3462
<i>Bruce A. C. Cree, Stephen L. Hauser</i>	
445 Neuromyelitis Optica .....	3477
<i>Bruce A. C. Cree, Stephen L. Hauser</i>	

### SECTION 3 Nerve and Muscle Disorders

446 Peripheral Neuropathy .....	3480
<i>Anthony A. Amato, Richard J. Barohn</i>	
447 Guillain-Barré Syndrome and Other Immune-Mediated Neuropathies.....	3501
<i>Stephen L. Hauser, Anthony A. Amato</i>	
448 Myasthenia Gravis and Other Diseases of the Neuromuscular Junction .....	3509
<i>Anthony A. Amato</i>	
449 Muscular Dystrophies and Other Muscle Diseases .....	3516
<i>Anthony A. Amato, Robert H. Brown, Jr.</i>	

### SECTION 4 Myalgic Encephalomyelitis/Chronic Fatigue Syndrome

450 Myalgic Encephalomyelitis/Chronic Fatigue Syndrome .....	3532
<i>Elizabeth R. Unger, Jin-Mann S. Lin, Jeanne Bertolli</i>	

### SECTION 5 Psychiatric and Addiction Disorders

451 Biology of Psychiatric Disorders .....	3534
<i>Robert O. Messing, Eric J. Nestler, Matthew W. State</i>	
452 Psychiatric Disorders .....	3540
<i>Victor I. Reus</i>	
453 Alcohol and Alcohol Use Disorders .....	3556
<i>Marc A. Schuckit</i>	
454 Nicotine Addiction .....	3563
<i>David M. Burns</i>	
455 Marijuana and Marijuana Use Disorders .....	3567
<i>Nora D. Volkow, Aidan Hampson, Ruben Baler</i>	
456 Opioid-Related Disorders .....	3569
<i>Thomas R. Kosten, Colin N. Haile</i>	
457 Cocaine, Other Psychostimulants, and Hallucinogens.....	3573
<i>Karran A. Phillips, Wilson M. Compton</i>	

## Section 1 Endocrinology

**376**

### Approach to the Patient with Endocrine Disorders

J. Larry Jameson



The management of endocrine disorders requires a broad understanding of intermediary metabolism, reproductive physiology, bone metabolism, and growth. Accordingly, the practice of endocrinology is intimately linked to a conceptual framework for understanding hormone secretion, hormone action, and principles of feedback control (**Chap. 377**). The endocrine system is evaluated primarily by measuring hormone concentrations, arming the clinician with valuable diagnostic information. Most disorders of the endocrine system are amenable to effective treatment once the correct diagnosis is established. Endocrine deficiency disorders are treated with physiologic hormone replacement; hormone excess conditions, which usually are caused by benign glandular adenomas, are managed by removing tumors surgically or reducing hormone levels medically.

#### SCOPE OF ENDOCRINOLOGY

Classically, the specialty of endocrinology encompasses the study of glands and the hormones they produce. Over time, the field has expanded because of the discovery of hormones and growth factors produced by the brain, gastrointestinal (GI) tract, musculoskeletal system, and other nonglandular organs. The term *endocrine* was coined by Starling to contrast the actions of hormones secreted internally (*endocrine*) with those secreted externally (*exocrine*) or into a lumen, such as the GI tract. The term *hormone*, derived from a Greek phrase meaning “to set in motion,” aptly describes the dynamic actions of hormones as they elicit cellular responses and regulate physiologic processes through feedback mechanisms.

Unlike many other specialties in medicine, it is not possible to define endocrinology strictly along anatomic lines. The classic endocrine glands—pituitary, thyroid, parathyroid, pancreatic islets, adrenals, and gonads—communicate broadly with other organs through the nervous system, hormones, cytokines, and growth factors. In addition to its traditional synaptic functions, the brain produces a vast array of peptide hormones, and this has led to the discipline of neuroendocrinology. Through the production of hypothalamic releasing factors, the central nervous system (CNS) exerts a major regulatory influence over pituitary hormone secretion (**Chap. 378**). The peripheral nervous system stimulates the adrenal medulla. The immune and endocrine systems are also intimately intertwined. The adrenal hormone cortisol is a powerful immunosuppressant. Cytokines and interleukins (ILs) have profound effects on the functions of the pituitary, adrenal, thyroid, and gonads. Common endocrine diseases such as autoimmune thyroid disease and type 1 diabetes mellitus are caused by dysregulation of immune surveillance and tolerance. Less common diseases such as polyglandular failure, Addison’s disease, and lymphocytic hypophysitis also have an immunologic basis. Immune therapies for cancer and various autoimmune diseases can initiate autoimmune endocrine disease as a side effect of treatment.

The interdigitation of endocrinology with physiologic processes in other specialties sometimes blurs the role of hormones. For example, hormones play an important role in maintenance of blood pressure, intravascular volume, and peripheral resistance in the cardiovascular system. Vasoactive substances such as catecholamines, angiotensin II, endothelin, and nitric oxide are involved in dynamic changes of

vascular tone in addition to their multiple roles in other tissues. The heart is the principal source of atrial natriuretic peptide, which acts in classic endocrine fashion to induce natriuresis at a distant target organ (the kidney). Erythropoietin, a traditional circulating hormone, is made in the kidney and stimulates erythropoiesis in bone marrow (**Chap. 63**). The kidney is also integrally involved in the renin-angiotensin axis (**Chap. 386**) and is a primary target of several hormones, including parathyroid hormone (PTH), mineralocorticoids, fibroblast growth factor 23 (FGF23), and vasopressin. The GI tract produces a vast array of peptide hormones, such as glucagon-like peptide 1 (GLP1), cholecystokinin, ghrelin, gastrin, secretin, and vasoactive intestinal peptide, among many others. Carcinoid and islet tumors can secrete excessive amounts of these hormones, leading to specific clinical syndromes (**Chap. 84**). Many of these GI hormones are also produced in the CNS, where their functions are poorly understood. Adipose tissue produces leptin, which acts centrally to control appetite, along with adiponectin, resistin, and other hormones that regulate metabolism. As hormones such as inhibin, ghrelin, and leptin are discovered, they become integrated into the science and practice of medicine on the basis of their functional roles rather than their tissues of origin.

Characterization of hormone receptors frequently reveals unexpected relationships to factors in nonendocrine disciplines. The growth hormone (GH) and leptin receptors, for example, are members of the cytokine receptor family. The G protein-coupled receptors (GPCRs), which mediate the actions of many peptide hormones, are used in numerous physiologic processes, including vision, smell, and neurotransmission.

#### PATHOLOGIC MECHANISMS OF ENDOCRINE DISEASE

Endocrine diseases can be divided into three major types of conditions: (1) hormone excess, (2) hormone deficiency, and (3) hormone resistance (**Table 376-1**).

#### CAUSES OF HORMONE EXCESS

Syndromes of hormone excess can be caused by neoplastic growth of endocrine cells, autoimmune disorders, and excess hormone administration. Benign endocrine tumors, including parathyroid, pituitary, and adrenal adenomas, often retain the capacity to produce hormones, reflecting the fact that these tumors are relatively well differentiated. Many endocrine tumors exhibit subtle defects in their “set points” for feedback regulation. For example, in Cushing’s disease, impaired feedback inhibition of adrenocorticotrophic hormone (ACTH) secretion is associated with autonomous function. However, the tumor cells are not completely resistant to feedback, as evidenced by ACTH suppression by higher doses of dexamethasone (e.g., high-dose dexamethasone test) (**Chap. 386**). Similar set point defects are also typical of parathyroid adenomas and autonomously functioning thyroid nodules.

The molecular basis of some endocrine tumors, such as the multiple endocrine neoplasia (MEN) syndromes (MEN1, 2A, 2B), has provided important insights into tumorigenesis (**Chap. 388**). MEN1 is characterized primarily by the triad of parathyroid, pancreatic islet, and pituitary tumors. MEN2 predisposes to medullary thyroid carcinoma, pheochromocytoma, and hyperparathyroidism. The *MEN1* gene, located on chromosome 11q13, encodes a tumor-suppressor gene, menin. Analogous to the paradigm first described for retinoblastoma, the affected individual inherits a mutant copy of the *MEN1* gene, and tumorigenesis ensues after a somatic “second hit” leads to loss of function of the normal *MEN1* gene (through deletion or point mutations).

In contrast to inactivation of a tumor-suppressor gene, as occurs in MEN1 and most other inherited cancer syndromes, MEN2 is caused by activating mutations in a single allele. In this case, activating mutations of the *RET* protooncogene, which encodes a receptor tyrosine kinase, leads to thyroid C-cell hyperplasia in childhood before the development of medullary thyroid carcinoma. Elucidation of this pathogenic mechanism has allowed early genetic screening for *RET* mutations in

TABLE 376-1 Causes of Endocrine Dysfunction

Type of Endocrine Disorder	Examples
<b>Hyperfunction</b>	
Neoplastic	
Benign	Pituitary adenomas, hyperparathyroidism, autonomous thyroid or adrenal nodules
Malignant	Adrenal cancer, medullary thyroid cancer, carcinoid
Ectopic	Ectopic ACTH, SIADH secretion
Multiple endocrine neoplasia (MEN)	MEN1, MEN2
Autoimmune	Graves' disease
Iatrogenic	Cushing's syndrome, hypoglycemia
Infectious/inflammatory	Subacute thyroiditis
Activating receptor mutations	LH, TSH, $\text{Ca}^{2+}$ , PTH receptors, $G_s\alpha$
<b>Hypofunction</b>	
Autoimmune	Hashimoto's thyroiditis, type 1 diabetes mellitus, Addison's disease, polyglandular failure
Iatrogenic	Radiation-induced hypopituitarism, hypothyroidism, surgical
Infectious/inflammatory	Adrenal insufficiency, hypothalamic sarcoidosis
Hormone mutations	GH, LH $\beta$ , FSH $\beta$ , vasopressin
Enzyme defects	21-Hydroxylase deficiency
Developmental defects	Kallmann's syndrome, Turner's syndrome, transcription factors
Nutritional/vitamin deficiency	Vitamin D deficiency, iodine deficiency
Hemorrhage/infarction	Sheehan's syndrome, adrenal insufficiency
<b>Hormone Resistance</b>	
Receptor mutations	
Membrane	GH, vasopressin, LH, FSH, ACTH, GnRH, GHRH, PTH, leptin, $\text{Ca}^{2+}$
Nuclear	AR, TR, VDR, ER, GR, PPAR $\gamma$
Signaling pathway mutations	Albright's hereditary osteodystrophy
Postreceptor	Type 2 diabetes mellitus, leptin resistance

**Abbreviations:** ACTH, adrenocorticotrophic hormone; AR, androgen receptor; ER, estrogen receptor; FSH, follicle-stimulating hormone; GH, growth hormone; GHRH, growth hormone-releasing hormone; GnRH, gonadotropin-releasing hormone; GR, glucocorticoid receptor; LH, luteinizing hormone; PPAR, peroxisome proliferator activated receptor; PTH, parathyroid hormone; SIADH, syndrome of inappropriate antidiuretic hormone; TR, thyroid hormone receptor; TSH, thyroid-stimulating hormone; VDR, vitamin D receptor.

individuals at risk for MEN2, permitting identification of those who may benefit from prophylactic thyroidectomy and biochemical screening for pheochromocytoma and hyperparathyroidism.

Mutations that activate hormone receptor signaling have been identified in several GPCRs. For example, activating mutations of the luteinizing hormone (LH) receptor cause a dominantly transmitted form of male-limited precocious puberty, reflecting premature stimulation of testosterone synthesis in Leydig cells (Chap. 391). Activating mutations in these GPCRs are located predominantly in the transmembrane domains and induce receptor coupling to  $G_s$  even in the absence of hormone. Consequently, adenylyl cyclase is activated, and cyclic adenosine monophosphate (AMP) levels increase in a manner that mimics hormone action. A similar phenomenon results from activating mutations in  $G_s$ . When these mutations occur early in development, they cause McCune-Albright syndrome. When they occur only in somatotropes, the activating  $G_s$  mutations cause GH-secreting tumors and acromegaly (Chap. 380).

In autoimmune Graves' disease, antibody interactions with the thyroid-stimulating hormone (TSH) receptor mimic TSH action, leading to hormone overproduction (Chap. 382). Analogous to the effects of activating mutations of the TSH receptor, these stimulating autoantibodies induce conformational changes in the TSH receptor that release it from a constrained state, thereby triggering receptor coupling to G proteins.

### CAUSES OF HORMONE DEFICIENCY

Most examples of hormone deficiency states can be attributed to glandular destruction caused by autoimmunity, surgery, infection, inflammation, infarction, hemorrhage, or tumor infiltration (Table 376-1). Autoimmune damage to the thyroid gland (Hashimoto's thyroiditis) and pancreatic islet cells (type 1 diabetes mellitus) are examples of relatively common endocrine diseases. Mutations in a number of hormones, hormone receptors, transcription factors, enzymes, and channels can also lead to hormone deficiencies.

### HORMONE RESISTANCE

Most severe hormone resistance syndromes are due to inherited defects in membrane receptors, nuclear receptors, or the pathways that transduce receptor signals. These disorders are characterized by defective hormone action despite the presence of increased hormone levels. In complete androgen resistance, for example, mutations in the androgen receptor result in a female phenotypic appearance in genetic (XY) males, even though LH and testosterone levels are increased (Chap. 388). In addition to these relatively rare genetic disorders, more common acquired forms of functional hormone resistance include insulin resistance in type 2 diabetes mellitus, leptin resistance in obesity, and GH resistance in catabolic states. The pathogenesis of functional resistance involves receptor downregulation and postreceptor desensitization of signaling pathways; functional forms of resistance are generally reversible.

### CLINICAL EVALUATION OF ENDOCRINE DISORDERS

Because most glands are relatively inaccessible, the physical examination usually focuses on the manifestations of hormone excess or deficiency as well as direct examination of palpable glands, such as the thyroid and gonads. For these reasons, it is important to evaluate patients in the context of their presenting symptoms, review of systems, family and social history, and exposure to medications that may affect the endocrine system. Astute clinical skills are required to detect subtle symptoms and signs suggestive of underlying endocrine disease. For example, a patient with Cushing's syndrome may manifest specific findings, such as central fat redistribution, skin striae, and proximal muscle weakness, in addition to features seen commonly in the general population, such as obesity, plethora, hypertension, and glucose intolerance. Similarly, the insidious onset of hypothyroidism—with mental slowing, fatigue, dry skin, and other features—can be difficult to distinguish from similar, nonspecific findings in the general population. Clinical judgment that is based on knowledge of disease prevalence

and pathophysiology is required to decide when to embark on more extensive evaluation of these disorders. Laboratory testing plays an essential role in endocrinology by allowing quantitative assessment of hormone levels and dynamics. Radiologic imaging tests such as computed tomography (CT) scan, magnetic resonance imaging (MRI), thyroid scan, and ultrasound are also used for the diagnosis of endocrine disorders. However, these tests generally are employed only after a hormonal abnormality has been established by biochemical testing.

## HORMONE MEASUREMENTS AND ENDOCRINE TESTING

Immunoassays are the most important diagnostic tool in endocrinology, as they allow sensitive, specific, and quantitative determination of steady-state and dynamic changes in hormone concentrations. Immunoassays use antibodies to detect specific hormones. For many peptide hormones, these measurements are now configured to use two different antibodies to increase binding affinity and specificity. There are many variations of these assays; a common format involves using one antibody to capture the antigen (hormone) onto an immobilized surface and a second antibody, coupled to a chemiluminescent (immunochemiluminescent assay [ICMA]) or radioactive (immunoradiometric assay [IRMA]) signal, to detect the antigen. These assays are sensitive enough to detect plasma hormone concentrations in the picomolar to nanomolar range, and they can readily distinguish structurally related proteins, such as PTH from PTH-related peptide (PTHRP). A variety of other techniques are used to measure specific hormones, including mass spectroscopy, various forms of chromatography, and enzymatic methods; bioassays are now used rarely.

Most hormone measurements are based on plasma or serum samples. However, urinary hormone determinations remain useful for the evaluation of some conditions. Urinary collections over 24 h provide an integrated assessment of the production of a hormone or metabolite, many of which vary during the day. It is important to ensure complete collections of 24-h urine samples; simultaneous measurement of creatinine provides an internal control for the adequacy of collection and can be used to normalize some hormone measurements. A 24-h urine-free cortisol measurement largely reflects the amount of unbound cortisol, thus providing a reasonable index of biologically available hormone. Other commonly used urine determinations include 17-hydroxycorticosteroids, 17-ketosteroids, vanillylmandelic acid, metanephrine, catecholamines, 5-hydroxyindoleacetic acid, and calcium.

The value of quantitative hormone measurements lies in their correct interpretation in a clinical context. The normal range for most hormones is relatively broad, often varying by a factor of two- to tenfold. The normal ranges for many hormones are sex- and age-specific. Thus, using the correct normative database is an essential part of interpreting

hormone tests. The pulsatile nature of hormones and factors that can affect their secretion, such as sleep, meals, and medications, must also be considered. Cortisol values increase fivefold between midnight and dawn; reproductive hormone levels vary dramatically during the female menstrual cycle.

For many endocrine systems, much information can be gained from basal hormone testing, particularly when different components of an endocrine axis are assessed simultaneously. For example, low testosterone and elevated LH levels suggest a primary gonadal problem, whereas a hypothalamic-pituitary disorder is likely if both LH and testosterone are low. Because TSH is a sensitive indicator of thyroid function, it is generally recommended as a first-line test for thyroid disorders. An elevated TSH level is almost always the result of primary hypothyroidism, whereas a low TSH is most often caused by thyrotoxicosis. These predictions can be confirmed by determining the free thyroxine level. In the less common circumstance when free thyroxine and TSH are both low, it is important to consider secondary hypopituitarism caused by hypothalamic-pituitary disease. Elevated calcium and PTH levels suggest hyperparathyroidism, whereas PTH is suppressed in hypercalcemia caused by malignancy or granulomatous diseases. A suppressed ACTH in the setting of hypercortisolism, or increased urine free cortisol, is seen with hyperfunctioning adrenal adenomas.

It is not uncommon, however, for baseline hormone levels associated with pathologic endocrine conditions to overlap with the normal range. In this circumstance, dynamic testing is useful to separate the two groups further. There are a multitude of dynamic endocrine tests, but all are based on principles of feedback regulation, and most responses can be rationalized based on principles that govern the regulation of endocrine axes. *Suppression tests* are used in the setting of suspected endocrine hyperfunction. An example is the dexamethasone suppression test used to evaluate Cushing's syndrome (Chaps. 380 and 386). *Stimulation tests* generally are used to assess endocrine hypofunction. The ACTH stimulation test, for example, is used to assess the adrenal gland response in patients with suspected adrenal insufficiency. Other stimulation tests use hypothalamic-releasing factors such as corticotropin-releasing hormone (CRH) and growth hormone-releasing hormone (GHRH) to evaluate pituitary hormone reserve (Chap. 380). Insulin-induced hypoglycemia evokes pituitary ACTH and GH responses. Stimulation tests based on reduction or inhibition of endogenous hormones are now used infrequently. Examples include metyrapone inhibition of cortisol synthesis and clomiphene inhibition of estrogen feedback.

## SCREENING AND ASSESSMENT OF COMMON ENDOCRINE DISORDERS

Many endocrine disorders are prevalent in the adult population (Table 376-2) and can be diagnosed and managed by general internists, family practitioners, or other primary health care providers.

**TABLE 376-2 Examples of Prevalent Endocrine and Metabolic Disorders in the Adult**

DISORDER	APPROXIMATE PREVALENCE IN ADULTS <sup>a</sup>	SCREENING/TESTING RECOMMENDATIONS <sup>b</sup>	CHAPTER(S)
Obesity	40% Obese, BMI 30 70% Overweight, BMI 25	Calculate BMI Measure waist circumference Exclude secondary causes Consider comorbid complications	402
Type 2 diabetes mellitus	>10%	Beginning at age 45, screen every 3 years, or earlier in high-risk groups: FPG >126 mg/dL Random plasma glucose >200 mg/dL An elevated HbA <sub>1c</sub> Consider comorbid complications	403
Hyperlipidemia	20–25%	Cholesterol screening at least every 5 years; more often in high-risk groups Lipoprotein analysis (LDL, HDL) for increased cholesterol, CAD, diabetes Consider secondary causes	407

(Continued)

**TABLE 376-2 Examples of Prevalent Endocrine and Metabolic Disorders in the Adult (Continued)**

DISORDER	APPROXIMATE PREVALENCE IN ADULTS <sup>a</sup>	SCREENING/TESTING RECOMMENDATIONS <sup>b</sup>	CHAPTER(S)
Metabolic syndrome	35%	Measure waist circumference, FPG, BP, lipids	408
Hypothyroidism	5–10%, women 0.5–2%, men	TSH; confirm with free T <sub>4</sub>	384
Graves' disease	1–3%, women 0.1%, men	TSH, free T <sub>4</sub>	383
Thyroid nodules and neoplasia	2–5% palpable >25% by ultrasound	Physical examination or ultrasound of thyroid Fine-needle aspiration biopsy	385
Osteoporosis	5–10%, women 2–5%, men	Bone mineral density measurements in women >65 years or in postmenopausal women or men at risk Exclude secondary causes	411
Hyperparathyroidism	0.1–0.5%, women > men	Serum calcium PTH, if calcium is elevated Assess comorbid conditions	410
Infertility	10%, couples	Investigate both members of couple Semen analysis in male Assess ovulatory cycles in female Specific tests as indicated	391,392
Polycystic ovarian syndrome	5–10%, women	Free testosterone, DHEAS Consider comorbid conditions	392
Hirsutism	5–10%	Free testosterone, DHEAS Exclude secondary causes Additional tests as indicated	394
Menopause	Median age, 51	FSH	395
Hyperprolactinemia	15% in women with amenorrhea or galactorrhea	PRL level MRI, if not medication-related	380
Erectile dysfunction	10–25%	Careful history, PRL, testosterone Consider secondary causes (e.g., diabetes)	397
Hypogonadism, male	1–2%	Testosterone, LH	391
Gynecomastia	15%	Often, no tests are indicated Consider Klinefelter's syndrome Consider medications, hypogonadism, liver disease	391
Klinefelter's syndrome	0.2%, men	Karyotype Testosterone	390
Vitamin D deficiency	10%	Measure serum 25-OH vitamin D Consider secondary causes	409
Turner's syndrome	0.03%, women	Karyotype Consider comorbid conditions	390

<sup>a</sup>The prevalence of most disorders varies among ethnic groups and with aging. Data based primarily on U.S. population. <sup>b</sup>See individual chapters for additional information on evaluation and treatment. Early testing is indicated in patients with signs and symptoms of disease and in those at increased risk.

**Abbreviations:** BMI, body mass index; BP, blood pressure; CAD, coronary artery disease; DHEAS, dehydroepiandrosterone; FPG, fasting plasma glucose; FSH, follicle-stimulating hormone; HbA<sub>1c</sub>, hemoglobin A<sub>1c</sub>; HDL, high-density lipoprotein; LDL, low-density lipoprotein; LH, luteinizing hormone; MRI, magnetic resonance imaging; PRL, prolactin; PTH, parathyroid hormone; TSH, thyroid-stimulating hormone.

The high prevalence and clinical impact of certain endocrine diseases justify vigilance for features of these disorders during routine physical examinations; laboratory screening is indicated in selected high-risk populations.

### FURTHER READING

Endocrine Society: The Endocrine Society Clinical Practice Guidelines. Available from <https://www.endocrine.org/clinical-practice-guidelines>.

Golden SH et al: Health disparities in endocrine disorders: Biological, clinical, and nonclinical factors—an Endocrine Society Scientific Statement. *J Clin Endocrinol Metab* 97:E1579, 2012.

Jameson JL, DeGroot LJ (eds): *Endocrinology: Adult and Pediatric*, 7th ed. Philadelphia, Elsevier, 2016.

Loriaux DL: *A Biographical History of Endocrinology*. Hoboken, Wiley Blackwell, 2016.

## 377

### Mechanisms of Hormone Action

J. Larry Jameson



Hormones function to integrate physiologic systems in the body. The endocrine system, composed of various glands and the hormones they produce, interacts with essentially every organ to regulate growth, metabolism, homeostasis, and reproduction. Because hormones circulate and act via receptors in target tissues, they serve to coordinate physiologic responses to external or internal cues. For example, the light-dark cycle, sensed through the visual system, modulates hypothalamic corticotropin-releasing hormone (CRH), which increases pituitary adrenocorticotropin hormone (ACTH) production, leading to increased adrenal cortisol production before the time of waking

in the morning. Increased cortisol, in turn, circulates throughout the body, acting via the nuclear glucocorticoid receptor, to activate numerous genetic programs that influence metabolism, the cardiovascular system, behavior, and the immune system. This chapter provides an overview of the different types of hormones and how they function at the cellular level to control myriad physiologic processes.

## CLASSES OF HORMONES

Hormones can be divided into five major types: (1) *amino acid derivatives* such as dopamine, catecholamine, and thyroid hormone; (2) *small neuropeptides* such as gonadotropin-releasing hormone (GnRH), thyrotropin-releasing hormone (TRH), somatostatin, and vasopressin; (3) *large proteins* such as insulin, luteinizing hormone (LH), and parathyroid hormone (PTH); (4) *steroid hormones* such as cortisol and estrogen that are synthesized from cholesterol-based precursors; and (5) *vitamin derivatives* such as retinoids (vitamin A) and vitamin D. A variety of *peptide growth factors*, such as insulin-like growth factor 1 (IGF1), share actions with hormones but often act more locally. As a rule, amino acid derivatives and peptide hormones interact with cell-surface membrane receptors. Steroids, thyroid hormones, vitamin D, and retinoids are lipid-soluble and bind to intracellular nuclear receptors, although many also interact with membrane receptors or intracellular signaling proteins as well.

## HORMONE AND RECEPTOR FAMILIES

Hormones and receptors can be grouped into families, reflecting structural similarities and evolutionary origins (Table 377-1). The evolution of these families generates diverse but highly selective pathways of hormone action. Recognition of these relationships has proven useful for extrapolating information gleaned from one hormone or receptor to other family members.

The glycoprotein hormone family, consisting of thyroid-stimulating hormone (TSH), follicle-stimulating hormone (FSH), LH, and human chorionic gonadotropin (hCG), illustrates many features of evolutionarily related hormones. The glycoprotein hormones are heterodimers that share the  $\alpha$  subunit in common; the  $\beta$  subunits are distinct and

confer specific biologic actions. The overall three-dimensional architecture of the  $\beta$  subunits is similar, reflecting the locations of conserved disulfide bonds that constrain protein conformation. The cloning of the  $\beta$ -subunit genes from multiple species suggests that this family arose from a common ancestral gene, probably by gene duplication and subsequent divergence to evolve new biologic functions.

As hormone families enlarge and diverge, their receptors have co-evolved to create new biologic functions. Related G protein-coupled receptors (GPCRs), for example, have evolved for each of the glycoprotein hormones. These receptors are also structurally similar, and each is coupled predominantly to the  $G_s$  signaling pathway. However, there is minimal overlap of hormone binding. For example, TSH binds with high specificity to the TSH receptor but interacts minimally with the LH or FSH receptors. Nonetheless, there can be subtle physiologic consequences of hormone cross-reactivity with other receptors. Very high levels of hCG during pregnancy stimulate the TSH receptor and increase thyroid hormone levels, resulting via feedback inhibition in a compensatory decrease in TSH.

IGF1 and IGF2 have structural similarities that are most apparent when precursor forms of the proteins are compared. In contrast to the high degree of specificity seen with the glycoprotein hormones, there is moderate cross-talk among the members of the insulin/IGF family. High concentrations of an IGF2 precursor produced by certain tumors (e.g., sarcomas) can cause hypoglycemia, partly because of binding to insulin and IGF1 receptors (Chap. 410). High concentrations of insulin also bind to the IGF1 receptor, perhaps accounting for some of the clinical manifestations seen in conditions with chronic hyperinsulinemia.

Another important example of receptor cross-talk is seen with PTH and parathyroid hormone-related peptide (PTHRP) (Chap. 410). PTH is produced by the parathyroid glands, whereas PTHR P is expressed at high levels during development and by a variety of tumors (Chap. 93). These hormones have amino acid sequence similarity, particularly in their amino-terminal regions. Both hormones bind to the PTH1R receptor that is expressed in bone and kidney. Hypercalcemia and hypophosphatemia therefore may result from excessive production of either hormone, making it difficult to distinguish hyperparathyroidism from hypercalcemia of malignancy solely on the basis of serum chemistries. However, sensitive and specific assays for PTH and PTHR P now allow these disorders to be distinguished more readily.

Based on their specificities for DNA-binding sites, the nuclear receptor family can be subdivided into type 1 receptors (glucocorticoid receptor, mineralocorticoid receptor, androgen receptor, estrogen receptor, progesterone receptor) that bind steroids and type 2 receptors (thyroid hormone receptor, vitamin D receptor, retinoic acid receptor, peroxisome proliferator activated receptor) that bind thyroid hormone, vitamin D, retinoic acid, or lipid derivatives, respectively. Certain functional domains in nuclear receptors, such as the zinc finger DNA-binding domains, are highly conserved. However, selective amino acid differences within this domain confer DNA sequence specificity. The hormone-binding domains are more variable, providing great diversity in the array of small molecules that bind to different nuclear receptors. With few exceptions, hormone binding is highly specific for a single type of nuclear receptor. One exception involves the glucocorticoid and mineralocorticoid receptors. Because the mineralocorticoid receptor also binds glucocorticoids with high affinity, an enzyme (11 $\beta$ -hydroxysteroid dehydrogenase) in renal tubular cells inactivates glucocorticoids, allowing selective responses to mineralocorticoids such as aldosterone. However, when very high glucocorticoid concentrations occur, as in Cushing's syndrome, the glucocorticoid degradation pathway becomes saturated, allowing excessive cortisol levels to bind mineralocorticoid receptors leading to sodium retention and potassium wasting. This phenomenon is particularly pronounced in ectopic ACTH syndromes (Chap. 386). Another example of relaxed nuclear receptor specificity involves the estrogen receptor, which can bind an array of compounds, some of which have little apparent structural similarity to the high-affinity ligand estradiol. This feature of the estrogen receptor makes it susceptible to activation by "environmental estrogens" such as resveratrol, octylphenol, and many other aromatic hydrocarbons. However, this lack of specificity provides an opportunity to

**TABLE 377-1 Examples of Membrane Receptor Families and Signaling Pathways**

RECEPTORS	EFFECTORS	SIGNALING PATHWAYS
<b>G Protein-Coupled Seven-Transmembrane Receptor (GPCR)</b>		
$\beta$ -Adrenergic, LH, FSH, TSH	$G_{\alpha}$ , adenylate cyclase	Stimulation of cyclic AMP production, protein kinase A
Glucagon, PTH, PTHrP, ACH, MSH, GHRH, CRH	$Ca^{2+}$ channels	Calmodulin, $Ca^{2+}$ -dependent kinases
$\alpha$ -Adrenergic, somatostatin	$G_{\alpha}$	Inhibition of cyclic AMP production Activation of $K^+$ , $Ca^{2+}$ channels
TRH, GnRH	$G_q/G_{11}$	Phospholipase C, diacylglycerol, $IP_3$ , protein kinase C, voltage-dependent $Ca^{2+}$ channels
<b>Receptor Tyrosine Kinase</b>		
Insulin, IGF-I EGF, NGF	Tyrosine kinases, IRS Tyrosine kinases, ras	MAP kinases, PI 3-kinase; AKT Raf, MAP kinases, RSK
<b>Cytokine Receptor-Linked Kinase</b>		
GH, PRL	JAK, tyrosine kinases	STAT, MAP kinase, PI 3-kinase, IRS-1
<b>Serine Kinase</b>		
Activin, TGF- $\beta$ , MIS	Serine kinase	Smads

*Abbreviations:* IP<sub>3</sub>, inositol triphosphate; IRS, insulin receptor substrates; MAP, mitogen-activated protein; MSH, melanocyte-stimulating hormone; NGF, nerve growth factor; PI, phosphatidylinositol; RSK, ribosomal S6 kinase; TGF- $\beta$ , transforming growth factor  $\beta$ . For all other abbreviations, see text. Note that most receptors interact with multiple effectors and activate networks of signaling pathways.

synthesize clinically useful antagonists (e.g., tamoxifen) and selective estrogen response modulators (SERMs) such as raloxifene. These compounds generate distinct estrogen receptor conformations that alter receptor interactions with components of the transcription machinery (see below), thereby conferring their unique actions.

## HORMONE SYNTHESIS AND PROCESSING

The synthesis of peptide hormones and their receptors occurs through a classic pathway of gene expression: transcription → mRNA → protein → posttranslational protein processing → intracellular sorting followed by membrane integration or secretion.

Many hormones are embedded within larger precursor polypeptides that are proteolytically processed to yield the biologically active hormone. Examples include proopiomelanocortin (POMC) → ACTH; proglucagon → glucagon; proinsulin → insulin; and pro-PTH → PTH, among others. In many cases, such as POMC and proglucagon, these precursors generate multiple biologically active peptides. For example, proglucagon generates glucagon, as well as glucagon-like peptide 1 (GLP1), among other peptides. It is provocative that hormone precursors are typically inactive, presumably adding an additional level of control through peptide processing. Prohormone conversion occurs not only for peptide hormones but also for certain steroids (testosterone → dihydrotestosterone) and thyroid hormone ( $T_4 \rightarrow T_3$ ).

Peptide precursor processing is intimately linked to intracellular sorting pathways that transport proteins to appropriate vesicles and enzymes, resulting in specific cleavage steps, followed by protein folding and translocation to secretory vesicles. Hormones destined for secretion are translocated across the endoplasmic reticulum guided by an amino-terminal signal sequence that subsequently is cleaved. Cell-surface receptors are inserted into the membrane via short segments of hydrophobic amino acids that remain embedded within the lipid bilayer. During translocation through the Golgi and endoplasmic reticulum, hormones and receptors are subject to a variety of post-translational modifications, such as glycosylation and phosphorylation, which can alter protein conformation, modify circulating half-life, and alter biologic activity.

Synthesis of most steroid hormones is based on modifications of the precursor, cholesterol. Multiple regulated enzymatic steps are required for the synthesis of testosterone (Chap. 391), estradiol (Chap. 392), cortisol (Chap. 386), and vitamin D (Chap. 409). This large number of synthetic steps predisposes to multiple genetic and acquired disorders of steroidogenesis.

Endocrine genes contain regulatory DNA elements similar to those found in many other genes, but their exquisite control by hormones reflects the presence of specific hormone response elements. For example, the TSH genes are repressed directly by thyroid hormones acting through the thyroid hormone receptor (TR), a member of the nuclear receptor family. Steroidogenic enzyme gene expression requires specific transcription factors, such as steroidogenic factor 1 (SF1), acting in conjunction with signals transmitted by trophic hormones (e.g., ACTH or LH). Once activated, SF1 functions as a master regulator, inducing a large array of genes required for steroidogenic and metabolic pathways required for steroid synthesis. For some hormones, substantial regulation occurs at the level of translational efficiency. Insulin biosynthesis, although it requires ongoing gene transcription, is regulated primarily at the translational and secretory levels in response to the levels of glucose or amino acids.

## HORMONE SECRETION, TRANSPORT, AND DEGRADATION

The circulating level of a hormone is determined by its rate of secretion and its half-life. After protein processing, peptide hormones (e.g., GnRH, insulin, growth hormone [GH]) are stored in secretory granules. As these granules mature, they are poised beneath the plasma membrane for imminent release into the circulation. In most instances, the stimulus for hormone secretion is a releasing factor or neural signal that induces rapid changes in voltage-gated channel activity or intracellular calcium concentrations, leading to secretory granule fusion with the plasma membrane and release of its contents into the extracellular

environment and bloodstream. Steroid hormones, in contrast, diffuse into the circulation as they are synthesized. Thus, their secretory rates are closely aligned with rates of synthesis. For example, ACTH and LH induce steroidogenesis by stimulating the activity of the steroidogenic acute regulatory (STAR) protein, which transports cholesterol into the mitochondrion, along with other rate-limiting steps (e.g., cholesterol side-chain cleavage enzyme, CYP11A1) in the steroidogenic pathway.

Hormone transport and degradation dictate the rapidity with which a hormonal signal decays. Some hormone signals are evanescent (e.g., somatostatin), whereas others are longer-lived (e.g., TSH). Because somatostatin exerts effects in virtually every tissue, a short half-life allows its concentrations and actions to be controlled locally. Structural modifications that impair somatostatin degradation have been useful for generating long-acting therapeutic analogues such as octreotide (Chap. 380). In contrast, the actions of TSH are highly specific for the thyroid gland. Its prolonged half-life generates relatively constant serum levels even though TSH is secreted in discrete pulses.

An understanding of circulating hormone half-life is important for achieving physiologic hormone replacement, as the frequency of dosing and the time required to reach steady state are intimately linked to rates of hormone decay.  $T_4$ , for example, has a circulating half-life of 7 days. Consequently, >1 month is required to reach a new steady state, and single daily doses are sufficient to achieve constant hormone levels.  $T_3$ , in contrast, has a half-life of 1 day. Its administration is associated with more dynamic serum levels, and it must be administered two to three times per day. Similarly, synthetic glucocorticoids vary widely in their half-lives; those with longer half-lives (e.g., dexamethasone) are associated with greater suppression of the hypothalamic-pituitary-adrenal (HPA) axis. Most protein hormones (e.g., ACTH, GH, prolactin [PRL], PTH, LH) have relatively short half-lives (<20 min), leading to sharp peaks of secretion and decay. The only accurate way to profile the pulse frequency and amplitude of these hormones is to measure levels in frequently sampled blood (every 10 min or less) over long durations (8–24 h). Because this is not practical in a clinical setting, an alternative strategy is to pool three to four blood samples drawn at about 30-min intervals or interpret the results in the context of a relatively wide normal range. Rapid hormone decay is useful in certain clinical settings. For example, the short half-life of PTH allows the use of intraoperative PTH levels to confirm successful removal of a parathyroid adenoma. This is particularly valuable diagnostically when there is a possibility of multicentric disease or parathyroid hyperplasia, as occurs with multiple endocrine neoplasia (MEN) or renal insufficiency.

Many hormones circulate in association with serum-binding proteins. Examples include (1)  $T_4$  and  $T_3$  binding to thyroxine-binding globulin (TBG), albumin, and thyroxine-binding prealbumin (TBPA); (2) cortisol binding to cortisol-binding globulin (CBG); (3) androgen and estrogen binding to sex hormone-binding globulin (SHBG); (4) IGF1 and IGF2 binding to multiple IGF-binding proteins (IGFBPs); (5) GH interactions with GH-binding protein (GHBPs), a circulating fragment of the GH receptor extracellular domain; and (6) activin binding to follistatin. These interactions provide a hormone reservoir, prevent otherwise rapid degradation of unbound hormones, restrict hormone access to certain sites (e.g., IGFBPs), and modulate the levels of unbound, or “free,” hormone concentrations. Although a variety of binding protein abnormalities have been identified, most have little clinical consequence aside from creating diagnostic problems. For example, TBG deficiency can reduce total thyroid hormone levels greatly, but the free concentrations of  $T_4$  and  $T_3$  remain normal. Liver disease and certain medications can also influence binding protein levels (e.g., estrogen increases TBG) or cause displacement of hormones from binding proteins (e.g., salsalate displaces  $T_4$  from TBG). In general, only unbound hormone is available to interact with receptors and thus elicit a biologic response. Short-term perturbations in binding proteins change the free hormone concentration, which in turn induces compensatory adaptations through feedback loops. SHBG changes in women are an exception to this self-correcting mechanism. When SHBG decreases because of insulin resistance or androgen excess, the unbound testosterone concentration is increased, potentially contributing to hirsutism in women with polycystic ovary syndrome (PCOS).

**(Chap. 394).** The increased unbound testosterone level does not result in an adequate compensatory feedback correction because estrogen, not testosterone, is the primary regulator of the reproductive axis.

An additional exception to the unbound hormone hypothesis involves megalin, a member of the low-density lipoprotein (LDL) receptor family that serves as an endocytic receptor for thyroglobulin, carrier-bound vitamins A and D, and SHBG-bound androgens and estrogens. After internalization, the carrier proteins are degraded in lysosomes and release their bound ligands within the cells. Other membrane transporters have also been identified for thyroid hormones.

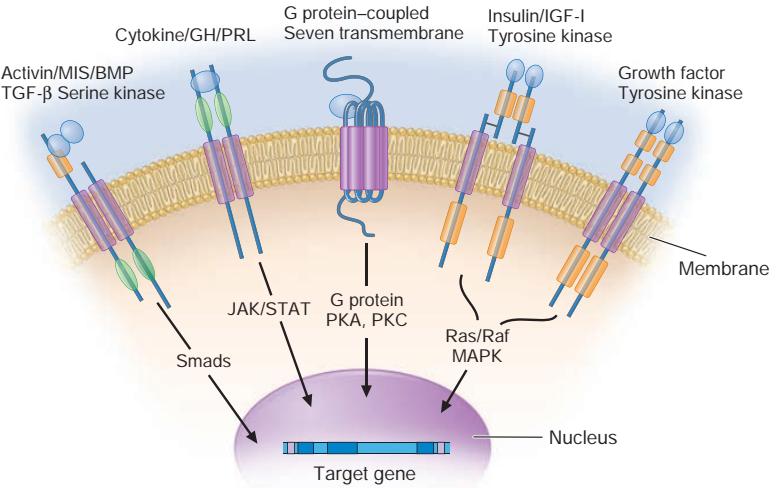
Hormone degradation can be an important mechanism for regulating concentrations locally. As noted above, 11-hydroxysteroid dehydrogenase inactivates glucocorticoids in renal tubular cells, preventing actions through the mineralocorticoid receptor. Thyroid hormone deiodinases convert T<sub>4</sub> to T<sub>3</sub> and can inactivate T<sub>3</sub>. During development, degradation of retinoic acid by Cyp26b1 prevents primordial germ cells in the male from entering meiosis, as occurs in the female ovary.

## HORMONE ACTION THROUGH RECEPTORS

Receptors for hormones are divided into two major classes: membrane and nuclear. **Membrane receptors** primarily bind peptide hormones and catecholamines. **Nuclear receptors** bind small molecules that can diffuse across the cell membrane, such as steroids and vitamin D. Certain general principles apply to hormone-receptor interactions regardless of the class of receptor. Hormones bind to receptors with specificity and an affinity that generally coincides with the dynamic range of circulating hormone concentrations. Low concentrations of free hormone (usually 10<sup>-12</sup> to 10<sup>-9</sup> M) rapidly associate and dissociate from receptors in a bimolecular reaction such that the occupancy of the receptor at any given moment is a function of hormone concentration and the receptor's affinity for the hormone. Receptor numbers vary greatly in different target tissues, providing one of the major determinants of tissue-specific responses to circulating hormones. For example, ACTH receptors are located almost exclusively in the adrenal cortex, and LH receptors are found predominantly in the gonads. In contrast, insulin and TRs are widely distributed, reflecting the need for metabolic responses in all tissues.

## MEMBRANE RECEPTORS

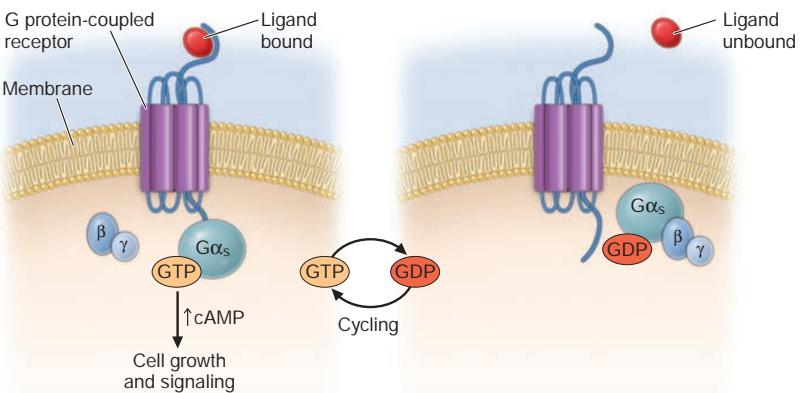
Membrane receptors for hormones can be divided into several major groups: (1) seven transmembrane GPCRs, (2) tyrosine kinase receptors, (3) cytokine receptors, and (4) serine kinase receptors (Fig. 377-1). The *seven transmembrane GPCR family* binds a huge array of hormones, including large proteins (e.g., LH, PTH), small peptides (e.g., TRH, somatostatin), catecholamines (epinephrine, dopamine), and even minerals (e.g., calcium). The extracellular domains of GPCRs vary widely in size and are the major binding site for large hormones. The transmembrane-spanning regions are composed of hydrophobic  $\alpha$ -helical domains that traverse the lipid bilayer. Like some channels, these domains are thought to circularize and form a hydrophobic pocket into which certain small ligands fit. Hormone binding induces



**FIGURE 377-1** Membrane receptor signaling. MAPK, mitogen-activated protein kinase; PKA, C, protein kinase A, C; TGF, transforming growth factor. For other abbreviations, see text.

conformational changes in these domains, transducing structural changes to the intracellular domain, which is a docking site for G proteins.

The large family of *G proteins*, so named because they bind guanine nucleotides (guanosine triphosphate [GTP], guanosine diphosphate [GDP]), provides great diversity for coupling receptors to different signaling pathways. G proteins form a heterotrimeric complex that is composed of various  $\alpha$ ,  $\beta$ , and  $\gamma$  subunits (Fig. 377-2). The  $\alpha$  subunit contains the guanine nucleotide-binding site and an intrinsic GTPase that hydrolyzes GTP to GDP. The  $\beta$  and  $\gamma$  subunits are tightly associated and modulate the activity of the  $\alpha$  subunit as well as mediating their own effector signaling pathways. G protein activity is regulated by a cycle that involves GTP hydrolysis and dynamic interactions between the  $\alpha$  and  $\beta\gamma$  subunits. Hormone binding to the receptor induces GDP



**FIGURE 377-2** G protein signaling. G protein-coupled receptors (GPCRs) signal via the family of G proteins, so named because they bind guanyl nucleotides. In the example shown, a GPCR bound to a ligand induces GDP dissociation, allowing G $\alpha$ s to bind GTP and dissociate from the  $\beta\gamma$  complex. GTP-bound G $\alpha$ s increases cAMP production by adenyl cyclase and activates the protein kinase A pathway. Not shown are separate signaling pathways activated by the  $\beta\gamma$  complex. When GTP is converted to GDP by an intrinsic GTPase, the  $\beta\gamma$  subunits reassociate with GDP-bound G $\alpha$ s and the complex returns to an inactive state. As noted in the text, mutations in G $\alpha$ s that eliminate GTPase activity result in constitutive activation of receptor signaling pathways because GTP-bound G $\alpha$ s cannot be converted to its GDP-bound inactive state. cAMP, cyclic adenosine 5'-monophosphate; GDP, guanosine diphosphate; G $\alpha$ , G protein  $\alpha$ ; GTP, guanosine triphosphate.

dissociation, allowing G<sub>i</sub> to bind GTP and dissociate from the complex. Under these conditions, the G<sub>i</sub> subunit is activated and mediates signal transduction through various enzymes, such as adenylate cyclase and phospholipase C. GTP hydrolysis to GDP allows reassociation with the G<sub>α</sub> subunits and restores the inactive state. G proteins interact with other cellular proteins, including kinases, channels, G protein-coupled receptor kinases (GRKs), and arrestins, that mediate signaling as well as receptor desensitization and recycling.

A variety of endocrinopathies result from mutations in GPCRs that alter their interactions with G proteins (Table 377-2). Loss-of-function mutations are generally recessive and inactivate the relevant hormone signaling pathway. Because many of these receptors are important for development as well as signaling, patient presentations resemble glandular failure syndromes (e.g., mutations in LH-R, FSH-R, TSH-R). Gain-of-function (GOF) mutations are more complex. Selected GOF mutations induce conformational changes in the GPCR that mimic the activated state normally induced by hormone binding. These GOF mutations result in a constitutively active state in which G protein coupling stimulates cell signaling pathways, most commonly via cyclic adenosine 5'-monophosphate (cAMP) and protein kinase A. When mutations occur in the germline, the conditions are heritable and

present in early life (e.g., LH-R, TSH-R). Sporadic, somatic mutations can also occur and result in clonal expansion of hyperfunctioning cells.

Mutations in the TSH-R illustrate the range of possible clinical consequences of GPCR mutations. Recessive inactivating mutations in the TSH-R cause congenital hypothyroidism with thyroid gland hypoplasia and resistance to TSH. Clinically, the hormone profile resembles primary hypothyroidism with low T<sub>4</sub> and high TSH. On the other hand, germline activating mutations cause congenital hyperthyroidism. The disorder is autosomal dominant because an activating mutation of one TSH-R allele is sufficient to induce cellular hyperfunction and disease. Because the TSH-R is activated in every cell of the thyroid, there is hyperplastic growth and hyperfunction that resembles the pathology seen in Graves' disease. This unusual disorder presents in infancy and must be distinguished from the more common clinical circumstance in which maternal antibodies in women with active or previously treated Graves' disease cross the placenta and stimulate the thyroid gland of the fetus. If an activating TSH-R mutation occurs later in life, in the somatic tissue, there is clonal expansion of the thyrocyte harboring the mutation, ultimately leading to an autonomous hyperfunctioning thyroid nodule. Of note, a similar condition can be caused by somatic mutations in G<sub>s</sub>. In this case, the G<sub>s</sub> GTPase is inactivated and GTP cannot be converted to GDP. Consequently, the G<sub>s</sub> signaling pathway in this particular cell is constitutively active, mimicking chronic TSH stimulation and again leading to clonal expansion and an autonomous hyperfunctioning thyroid nodule. About one-third of hyperfunctioning "hot" thyroid nodules harbor sporadic mutations in either the TSH-R or G<sub>s</sub>. (TSH-R mutations are more common).

G<sub>s</sub> mutations in tissues other than the thyroid can also cause endocrine disease. For example, G<sub>s</sub> mutations in pituitary somatotropes mimic activation of the growth hormone-releasing hormone (GHRH) pathway and lead to GH-producing adenomas and acromegaly. Rarely, mutations in other components of the protein kinase A pathway in somatotropes can also cause GH-producing adenomas. G<sub>s</sub> mutations that occur early in development (typically mosaic) cause McCune-Albright syndrome (Chap. 412), and the clinical features are manifest because the activated G protein pathway mimics the actions of various hormones (PTH, melanocyte-stimulating hormone [MSH], TSH, GHRH) in different tissues. Germline inactivating G<sub>s</sub> mutations cause a range of disorders that are transmitted and expressed in a complex manner because the locus is imprinted (Chap. 410). These conditions include Albright's hereditary osteodystrophy (AHO), pseudopseudohypoparathyroidism (PPHP), and pseudohypoparathyroidism types 1b, 1c, and 2.

The tyrosine kinase receptors transduce signals for insulin and a variety of growth factors, such as IGF1, epidermal growth factor (EGF), nerve growth factor, platelet-derived growth factor, and fibroblast growth factors. The cysteine-rich extracellular domains contain binding sites for the growth factors. After ligand binding, this class of receptors undergoes autophosphorylation, inducing interactions with intracellular adaptor proteins such as Shc and insulin receptor substrates (IRS). In the case of the insulin receptor, multiple kinases are activated, including the Raf-Ras-MAPK and the Akt/protein kinase B pathways. The tyrosine kinase receptors play a prominent role in cell growth and differentiation as well as in intermediary metabolism.

The GH and PRL receptors belong to the cytokine receptor family. Analogous to the tyrosine kinase receptors, ligand binding induces receptor interaction with intracellular kinases—the Janus kinases (JAKs), which phosphorylate members of the signal transduction and activators of transcription (STAT) family—as well as with other signaling pathways (Ras, PI3-K, MAPK). The activated STAT proteins translocate to the nucleus and stimulate expression of target genes.

The serine kinase receptors mediate the actions of activins, transforming growth factor  $\beta$ , müllerian-inhibiting substance (MIS; also known as anti-müllerian hormone [AMH]), and bone morphogenic proteins (BMPs). This family of receptors (consisting of type I and II subunits) signals through proteins termed smads (fusion of terms for *Caenorhabditis elegans* *sma* + mammalian *mad*). Like the STAT proteins, the smads serve a dual role of transducing the receptor signal and acting as transcription factors. The pleomorphic actions of these growth factors dictate that they act primarily in a local (paracrine or autocrine)

**TABLE 377-2 Genetic Causes of G protein Receptor Disorders**

RECEPTOR	DISORDER	GENETICS
LH	Leydig cell hypoplasia (male)	AR, inactivating
	Primary amenorrhea, resistance to LH (female)	AR, inactivating
	Familial male precocious puberty (male)	AD, activating
	Leydig cell adenoma, precocious puberty (male)	Sporadic, activating
FSH	Hypergonadotropic ovarian failure (female)	AR, inactivating
	Hypospermia (male)	AR, inactivating
	Ovarian hyperstimulation (female)	Sporadic, activating
TSH	Congenital hypothyroidism, TSH resistance	AR, AD, inactivating
	Nonautoimmune familial hyperthyroidism	AD, activating
	Hyperfunctioning thyroid adenoma	Sporadic, activating
GnRH	Hypogonadotropic hypogonadism	AR, inactivating
Kisspeptin	Hypogonadotropic hypogonadism	AR, inactivating
	Precocious puberty	AD, activating
Prokineticin	Precocious puberty	Sporadic, activating
TRH	Central hypothyroidism	AR, inactivating
GHRH	GH deficiency	AR, inactivating
PTH	Blomstrand chondrodysplasia	AR, inactivating
	Jansen metaphyseal chondrodysplasia	AD, activating
Calcium sensing receptor	Familial hypocalciuric hypercalcemia	AD, inactivating
	Neonatal severe hyperparathyroidism	AR, inactivating
	Familial hypocalcemic hypercalcemia	AD, activating
Arginine vasopressin receptor 2	Nephrogenic diabetes insipidus	XL, inactivating
	Nephrogenic SIADH	XL, activating
ACTH	Familial ACTH resistance	AR, inactivating
	ACTH-independent Cushing syndrome	Sporadic, activating
Melanocortin 4	Severe obesity	Codominant, inactivating

*Abbreviations:* ACTH, adrenocorticotropin hormone; AD, autosomal dominant; AR, autosomal recessive; FSH, follicle-stimulating hormone; GH, growth hormone; GHRH, growth hormone-releasing hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; PTH, parathyroid hormone; SIADH, syndrome of inappropriate antidiuretic hormone secretion; TRH, thyrotropin-releasing hormone; TSH, thyroid-stimulating hormone; XL, X-linked.

manner. Binding proteins such as follistatin (which binds activin and other members of this family) function to inactivate the growth factors and restrict their distribution.

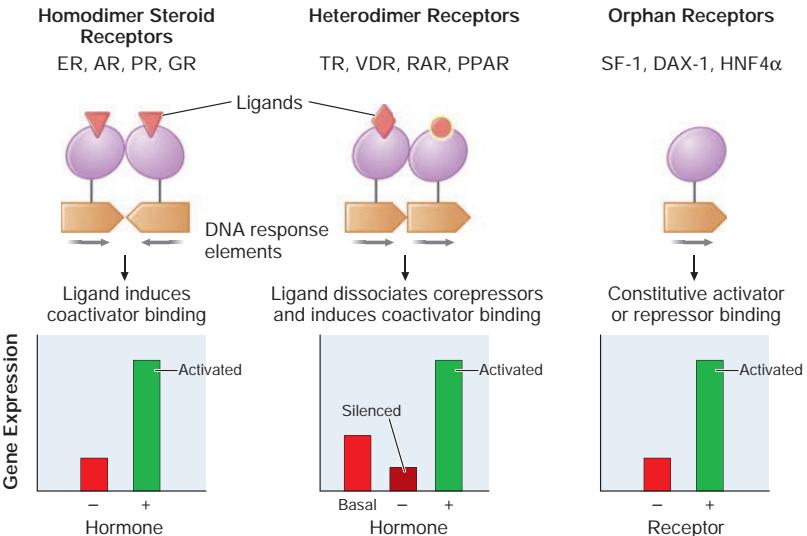
Disease-causing mutations also occur in each of these classes of receptors. For example, insulin receptor mutations cause an extreme form of insulin resistance. GH receptor mutations cause Laron-type dwarfism, characterized by low IGF1 and high GH. AMH receptor mutations cause persistent müllerian duct syndrome. These hormone resistance syndromes are autosomal recessive and relatively uncommon. Unlike the GPCRs, activating mutations are unusual, although they do occur for the RET tyrosine kinase receptor, which causes the autosomal dominant disorder MEN type 2 (MEN2) (Chap. 388).

### NUCLEAR RECEPTORS

The family of nuclear receptors has grown to nearly 100 members, many of which are still classified as orphan receptors because their ligands, if they exist, have not been identified (Fig. 377-3). Otherwise, most nuclear receptors are classified on the basis of their ligands. Although all nuclear receptors ultimately act to increase or decrease gene transcription, some (e.g., glucocorticoid receptor) reside primarily in the cytoplasm, whereas others (e.g., TR) are located in the nucleus. After ligand binding, the cytoplasmically localized receptors translocate to the nucleus. There is growing evidence that certain nuclear receptors (e.g., glucocorticoid, estrogen) can also act at the membrane or in the cytoplasm to activate or repress signal transduction pathways, providing a mechanism for cross-talk between membrane and nuclear receptors.

The structures of nuclear receptors have been studied extensively, including by x-ray crystallography. The DNA-binding domain, consisting of two zinc fingers, contacts specific DNA recognition sequences in target genes. Most nuclear receptors bind to DNA as dimers. Consequently, each monomer recognizes an individual DNA motif, referred to as a "half-site." The steroid receptors, including the glucocorticoid, estrogen, progesterone, and androgen receptors, bind to DNA as homodimers. Consistent with this twofold symmetry, their DNA recognition half-sites are palindromic. The thyroid, retinoid, peroxisome proliferator activated, and vitamin D receptors bind to DNA preferentially as heterodimers in combination with retinoid X receptors (RXRs). Their DNA half-sites are typically arranged as direct repeats.

The carboxy-terminal hormone-binding domains mediate transcriptional control. For type II receptors such as TR and retinoic acid receptor (RAR), co-repressor proteins bind to the receptor in the absence of ligand and silence gene transcription. Hormone binding induces conformational changes, triggering the release of co-repressors and the recruitment of coactivators that stimulate transcription. Thus, these receptors are capable of mediating dynamic changes in the level of gene activity. Disease states can be associated with defective regulation of these events. For example, in promyelocytic leukemia, fusion of RAR to other nuclear proteins causes aberrant gene silencing that prevents normal cellular differentiation. Treatment with retinoic acid reverses this repression and allows cellular differentiation and apoptosis to occur. Most type I steroid receptors interact weakly with co-repressors, but ligand binding still induces interactions with an array of coactivators. X-ray crystallography shows that various SERMs induce distinct estrogen receptor conformations. The tissue-specific responses caused by these agents in breast, bone, and uterus appear to reflect distinct interactions with various coactivators. The receptor-coactivator complex stimulates gene transcription by several pathways, including (1) recruitment of enzymes (histone acetyl transferases) that modify chromatin structure, (2) interactions with additional transcription factors on the target gene, and (3) direct interactions with components



**FIGURE 377-3 Nuclear receptor signaling.** AR, androgen receptor; DAX, dosage-sensitive sex-reversal, adrenal hypoplasia congenita, X chromosome; ER, estrogen receptor; GR, glucocorticoid receptor; HNF4 $\alpha$ , hepatic nuclear factor 4 $\alpha$ ; PPAR, peroxisome proliferator activated receptor; PR, progesterone receptor; RAR, retinoic acid receptor; SF-1, steroidogenic factor-1; TR, thyroid hormone receptor; VDR, vitamin D receptor.

of the general transcription apparatus to enhance the rate of RNA polymerase II-mediated transcription. Studies of nuclear receptor-mediated transcription reveal relatively rapid (e.g., 30–60 min) cycling of transcription complexes on any specific target gene.

Nuclear receptor mutations are an important cause of endocrine disease. Androgen receptor mutations cause androgen insensitivity syndrome (AIS) (Chap. 390). Because the androgen receptor is located on the X chromosome, phenotypic expression is more commonly manifest than with other nuclear receptor disorders. Affected individuals with AIS are XY phenotypic females with retained testes and male-range testosterone levels. Tissue insensitivity to androgens varies based on the severity of the mutation. Müllerian structures are absent because Sertoli cells of the testis produce AMH during development. Female carriers of androgen receptor mutations are phenotypically normal. Recessive mutations of the estrogen, glucocorticoid, and vitamin D receptors occur but are rare.

Thyroid hormone receptor (TR) mutations have an unusual pathophysiology. They are autosomal dominant and function via a "dominant negative" mechanism to cause resistance to thyroid hormone (RTH) (Chap. 382). The mutations occur in selected regions of the TR hormone-binding domain and preserve the ability of the mutant receptor to heterodimerize with RXR, interact with co-repressors, and bind to DNA regulatory sites. The mutant receptors function as antagonists of receptors from the normal copy of the TR gene. Affected patients have high T<sub>4</sub> and T<sub>3</sub> and inappropriately elevated (unsuppressed) TSH, reflecting impaired feedback regulation of the hypothalamic-pituitary-thyroid axis. Organ systems are variably resistant to thyroid hormones based on the relative expression of TR and TR. Mutations in the genes encoding TR and PPAR can also cause disease by functioning in an analogous dominant negative manner.

### FUNCTIONS OF HORMONES

The functions of individual hormones are described in detail in subsequent chapters. Nevertheless, it is useful to illustrate how most biologic responses require the integration of several different hormone pathways. The physiologic functions of hormones can be divided into three general types: (1) growth and differentiation, (2) maintenance of homeostasis, and (3) reproduction.

### GROWTH

Multiple hormones and nutritional factors mediate the complex phenomenon of growth (Chap. 378). Short stature may be caused by GH

deficiency, hypothyroidism, Cushing's syndrome, precocious puberty, malnutrition, chronic illness, or genetic abnormalities that affect the epiphyseal growth plates (e.g., *FGFR3* and *SHOX* mutations). Many factors (GH, IGF1, thyroid hormones) stimulate growth, whereas others (sex steroids) lead to epiphyseal closure. Understanding these hormonal interactions is important in the diagnosis and management of growth disorders. For example, delaying exposure to high levels of sex steroids may enhance the efficacy of GH treatment.

### MAINTENANCE OF HOMEOSTASIS

Although virtually all hormones affect homeostasis, the most important among them are the following:

1. Thyroid hormone—controls ~25% of basal metabolism in most tissues.
2. Cortisol—exerts a permissive action for many hormones in addition to its own direct effects.
3. PTH—regulates calcium and phosphorus levels.
4. Vasopressin—regulates serum osmolality by controlling renal free-water clearance.
5. Mineralocorticoids—control vascular volume and serum electrolyte ( $\text{Na}^+$ ,  $\text{K}^+$ ) concentrations.
6. Insulin—maintains euglycemia in the fed and fasted states.

The defense against hypoglycemia is an impressive example of integrated hormone action (Chap. 406). In response to the fasting state and falling blood glucose, insulin secretion is suppressed, resulting in decreased glucose uptake and enhanced glycogenolysis, lipolysis, proteolysis, and gluconeogenesis to mobilize fuel sources. If hypoglycemia develops (usually from insulin administration or sulfonylureas), an orchestrated counterregulatory response occurs—glucagon and epinephrine rapidly stimulate glycogenolysis and gluconeogenesis, whereas GH and cortisol act over several hours to raise glucose levels and antagonize insulin action.

Although free-water clearance is controlled primarily by vasopressin, cortisol and thyroid hormone are also important for facilitating renal tubular responses to vasopressin (Chap. 381). PTH and vitamin D function in an interdependent manner to control calcium metabolism (Chap. 409). PTH stimulates renal synthesis of 1,25-dihydroxyvitamin D, which increases calcium absorption in the gastrointestinal tract and enhances PTH action in bone. Increased calcium, along with vitamin D, feeds back to suppress PTH, thus maintaining calcium balance.

Depending on the severity of a specific stress and whether it is acute or chronic, multiple endocrine and cytokine pathways are activated to mount an appropriate physiologic response. In severe acute stress such as trauma or shock, the sympathetic nervous system is activated, and catecholamines are released, leading to increased cardiac output and a primed musculoskeletal system. Catecholamines also increase mean blood pressure and stimulate glucose production. Multiple stress-induced pathways converge on the hypothalamus, stimulating several hormones, including vasopressin and CRH. These hormones, in addition to cytokines (tumor necrosis factor  $\alpha$ , interleukin [IL] 2, IL-6), increase ACTH and GH production. ACTH stimulates the adrenal gland, increasing cortisol, which in turn helps sustain blood pressure and dampen the inflammatory response. Increased vasopressin acts to conserve free water.

### REPRODUCTION

The stages of reproduction include (1) sex determination during fetal development (Chap. 390); (2) sexual maturation during puberty (Chaps. 391 and 392); (3) conception, pregnancy, lactation, and child rearing (Chap. 392); and (4) cessation of reproductive capability at menopause (Chap. 395). Each of these stages involves an orchestrated interplay of multiple hormones, a phenomenon well illustrated by the dynamic hormonal changes that occur during each 28-day menstrual cycle. In the early follicular phase, pulsatile secretion of LH and FSH stimulates the progressive maturation of the ovarian follicle. This results in gradually increasing estrogen and progesterone levels, leading to enhanced pituitary sensitivity to GnRH, which, when combined with accelerated GnRH secretion, triggers the LH surge and rupture of the mature follicle. Inhibin, a protein produced by the granulosa cells,

enhances follicular growth and feeds back to the pituitary to selectively suppress FSH without affecting LH. Growth factors such as EGF and IGF1 modulate follicular responsiveness to gonadotropins. Vascular endothelial growth factor and prostaglandins play a role in follicle vascularization and rupture.

During pregnancy, the increased production of PRL, in combination with placentally derived steroids (e.g., estrogen and progesterone), prepares the breast for lactation. Estrogens induce the production of progesterone receptors, allowing for increased responsiveness to progesterone. In addition to these and other hormones involved in lactation, the nervous system and oxytocin mediate the suckling response and milk release.

### HORMONAL FEEDBACK REGULATORY SYSTEMS

*Feedback control*, both negative and positive, is a fundamental feature of endocrine systems. Each of the major hypothalamic-pituitary-hormone axes is governed by negative feedback, a process that maintains hormone levels within a relatively narrow range (Chap. 378). Examples of hypothalamic-pituitary negative feedback include (1) thyroid hormones on the TRH-TSH axis, (2) cortisol on the CRH-ACTH axis, (3) gonadal steroids on the GnRH-LH/FSH axis, and (4) IGF1 on the GHRH-GH axis (Fig. 377-4). These regulatory loops include both positive (e.g., TRH, TSH) and negative (e.g.,  $\text{T}_4$ ,  $\text{T}_3$ ) components, allowing for exquisite control of hormone levels. As an example, a small reduction of thyroid hormone triggers a rapid increase of TRH and TSH secretion, resulting in thyroid gland stimulation and increased thyroid hormone production. When thyroid hormone reaches a normal level, it feeds back to suppress TRH and TSH, and a new steady state is attained. Feedback regulation also occurs for endocrine systems that do not involve the pituitary gland, such as calcium feedback on PTH, glucose inhibition of insulin secretion, and leptin feedback on the hypothalamus. An understanding of feedback regulation provides important insights into endocrine testing paradigms (see below).

Positive feedback control also occurs but is not well understood. The primary example is estrogen-mediated stimulation of the midcycle LH surge. Although chronic low levels of estrogen are inhibitory, gradually rising estrogen levels stimulate LH secretion. This effect, which is

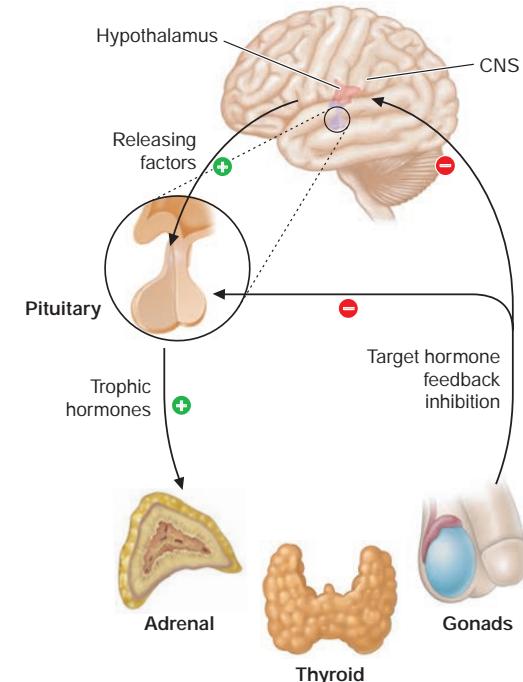


FIGURE 377-4 Feedback regulation of endocrine axes. CNS, central nervous system.

illustrative of an endocrine rhythm (see below), involves activation of the hypothalamic GnRH pulse generator. In addition, estrogen-primed gonadotropes are extraordinarily sensitive to GnRH, leading to amplification of LH release.

## PARACRINE AND AUTOCRINE CONTROL

The previously mentioned examples of feedback control involve classic endocrine pathways in which hormones are released by one gland and act on a distant target gland. However, local regulatory systems, often involving growth factors, are increasingly recognized. *Paracrine regulation* refers to factors released by one cell that act on an adjacent cell in the same tissue. For example, somatostatin secretion by pancreatic islet cells inhibits insulin secretion from nearby cells. *Autocrine regulation* describes the action of a factor on the same cell from which it is produced. IGF1 acts on many cells that produce it, including chondrocytes, breast epithelium, and gonadal cells. Unlike endocrine actions, paracrine and autocrine control are difficult to document because local growth factor concentrations cannot be measured readily.

Anatomic relationships of glandular systems also greatly influence hormonal exposure: the physical organization of islet cells enhances their intercellular communication; the portal vasculature of the hypothalamic-pituitary system exposes the pituitary to high concentrations of hypothalamic releasing factors; testicular seminiferous tubules gain exposure to high testosterone levels produced by the interdigitated Leydig cells; the pancreas receives nutrient information and local exposure to peptide hormones (incretins) from the gastrointestinal tract; and the liver is the proximal target of insulin action because of portal drainage from the pancreas.

## HORMONAL RHYTHMS

The feedback regulatory systems described above are superimposed on hormonal rhythms that are used for adaptation to the environment. Seasonal changes, the daily occurrence of the light-dark cycle, sleep, meals, and stress are examples of the many environmental events that affect hormonal rhythms. The *menstrual cycle* is repeated on average every 28 days, reflecting the time required to follicular maturation, ovulation, and potential implantation (Chap. 392). Essentially all pituitary hormone rhythms are entrained to sleep and to the *circadian cycle*, generating reproducible patterns that are repeated approximately every 24 h. The HPA axis, for example, exhibits characteristic peaks of ACTH and cortisol production in the early morning, with a nadir during the night. Recognition of these rhythms is important for endocrine testing and treatment. Patients with Cushing's syndrome characteristically exhibit increased midnight cortisol levels compared with normal individuals (Chap. 386). In contrast, morning cortisol levels are similar in these groups, as cortisol is normally high at this time of day in normal individuals. The HPA axis is more susceptible to suppression by glucocorticoids administered at night as they blunt the early-morning rise of ACTH. Understanding these rhythms allows glucocorticoid replacement that mimics diurnal production by administering larger doses in the morning than in the afternoon. Disrupted sleep rhythms can alter hormonal regulation. For example, sleep deprivation causes mild insulin resistance, food craving, and hypertension, which are reversible, at least in the short term. Emerging evidence indicates that circadian clock pathways not only regulate sleep-wake cycles but also play important roles in virtually every cell type. For example, tissue-specific deletion of clock genes alters rhythms and levels of gene expression, as well as metabolic responses in liver, adipose, and other tissues.

Other endocrine rhythms occur on a more rapid time scale. Many peptide hormones are secreted in discrete bursts every few hours. LH and FSH secretion are exquisitely sensitive to GnRH pulse frequency. Intermittent pulses of GnRH are required to maintain pituitary sensitivity, whereas continuous exposure to GnRH causes pituitary gonadotrope desensitization. This feature of the hypothalamic-pituitary-gonadotrope axis forms the basis for using long-acting GnRH agonists to treat central precocious puberty or to decrease testosterone levels in the management of prostate cancer. It is important to be aware of the pulsatile nature of hormone secretion and the rhythmic patterns of hormone production in relating serum hormone measurements

to normal values. For some hormones, integrated markers have been developed to circumvent hormonal fluctuations. Examples include 24-h urine collections for cortisol, the measurement of IGF1 as a biologic marker of GH action, and HbA<sub>1c</sub> as an index of long-term (weeks to months) blood glucose control.

Often, one must interpret endocrine data only in the context of other hormones. For example, PTH levels typically are assessed in combination with serum calcium concentrations. A high serum calcium level in association with elevated PTH is suggestive of hyperparathyroidism, whereas a suppressed PTH in the setting of hypercalcemia is more likely to be caused by hypercalcemia of malignancy, or other causes of hypercalcemia. Similarly, when T<sub>4</sub> and T<sub>3</sub> concentrations are low, TSH should be elevated, reflecting reduced feedback inhibition. When this is not the case, it is important to consider secondary hypothyroidism, which is caused by a defect at the level of the pituitary.

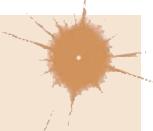
## FURTHER READING

- Evans RM, Mangelsdorf DJ: Nuclear receptors, RXR, and the big bang. *Cell* 157:255, 2014.
- Fukami M et al: Gain-of-function mutations in G-protein-coupled receptor genes associated with human endocrine disorders. *Clin Endocrinol* 88:351, 2018.
- Jameson JL, De Groot LJ (eds): *Endocrinology: Adult and Pediatric*, 7th ed. Philadelphia, Elsevier, 2016.
- Kim YH, Lazar MA: Transcriptional control of circadian rhythms and metabolism: A matter of time and space. *Endocr Rev* 41:707, 2020.
- Smith RL et al: Metabolic flexibility as an adaptation to energy resources and requirements in health and disease. *Endocr Rev* 39:489, 2018.

# 378

## Physiology of Anterior Pituitary Hormones

Shlomo Melmed, J. Larry Jameson



The anterior pituitary often is referred to as the "master gland" because, together with the hypothalamus, it orchestrates the complex regulatory functions of many other endocrine glands. The anterior pituitary gland produces six major hormones: (1) prolactin (PRL), (2) growth hormone (GH), (3) adrenocorticotrophic hormone (ACTH), (4) luteinizing hormone (LH), (5) follicle-stimulating hormone (FSH), and (6) thyroid-stimulating hormone (TSH) (Table 378-1). Pituitary hormones are secreted in a pulsatile manner, reflecting regulation by an array of specific hypothalamic releasing factors. Each of these pituitary hormones elicits specific trophic responses in peripheral target tissues including the adrenal, thyroid, and gonads, as well as tissues involved in metabolism (e.g., liver, breast, bone). Elicited hormonal products of peripheral glands, in turn, exert feedback control at the level of the hypothalamus and pituitary to modulate pituitary function (Fig. 378-1). Pituitary tumors cause characteristic hormone excess syndromes. Hormone deficiency may be inherited or acquired. Fortunately, there are efficacious treatments for many pituitary hormone excess and deficiency syndromes. Nonetheless, these diagnoses are often elusive; this emphasizes the importance of recognizing subtle clinical manifestations and performing the correct laboratory diagnostic tests. **For discussion of disorders of the posterior pituitary or neurohypophysis, see Chap. 381.**

## ANATOMY AND DEVELOPMENT

### ANATOMY

The pituitary gland weighs ~600 mg and is located within the sella turcica ventral to the diaphragma sella; it consists of anatomically and functionally distinct anterior and posterior lobes. The bony sella is

**TABLE 378-1** Anterior Pituitary Hormone Expression and Regulation

CELL	CORTICOTROPE	SOMATOTROPE	LACTOTROPE	THYROTROPE	GONADOTROPE
Tissue-specific transcription factor	T-Pit	Prop-1, Pit-1	Prop-1, Pit-1	Prop-1, Pit-1, TEF	SF-1, DAX-1
Fetal appearance	6 weeks	8 weeks	12 weeks	12 weeks	12 weeks
Hormone	POMC	GH	PRL	TSH	FSH, LH
Protein	Polypeptide	Polypeptide	Polypeptide	Glycoprotein $\alpha$ , $\beta$ subunits	Glycoprotein $\alpha$ , $\beta$ subunits
Amino acids	266 (ACTH 1–39)	191	198	211	210, 204
Stimulators	CRH, AVP, gp-130 cytokines	GHRH, ghrelin	Estrogen, TRH, VIP	TRH	GnRH, activins, estrogen
Inhibitors	Glucocorticoids	Somatostatin, IGF-1	Dopamine	$T_3$ , $T_4$ , dopamine, somatostatin, glucocorticoids	Sex steroids, inhibin
Target gland	Adrenal	Liver, bone, other tissues	Breast, other tissues	Thyroid	Ovary, testis
Trophic effect	Steroid production	IGF-1 production, growth induction, insulin antagonism	Milk production	$T_4$ synthesis and secretion	Sex steroid production, follicle growth, germ cell maturation
Normal range	ACTH, 4–22 pg/L	<0.5 $\mu$ g/L <sup>a</sup>	M <15 $\mu$ g/L; F <20 $\mu$ g/L	0.1–5 mU/L	M, 5–20 IU/L; F (basal), 5–20 IU/L

<sup>a</sup>Hormone secretion integrated over 24 h.

Abbreviations: F, female; M, male. For other abbreviations, see text.

Source: Adapted with permission from Melmed S: Hypothalamic-pituitary regulation, in P Conn (ed): *Conn's Translational Neuroscience*. San Diego, CA: Elsevier; 2017.

contiguous to vascular and neurologic structures, including the cavernous sinuses, cranial nerves, and optic chiasm. Thus, expanding intrasellar pathologic processes may have significant central mass effects in addition to their endocrinologic impact.

Hypothalamic neural cells synthesize specific releasing and inhibiting hormones that are secreted directly into the portal vessels of the pituitary stalk. Blood supply of the pituitary gland comes from the superior and inferior hypophyseal arteries (Fig. 378-2). The hypothalamic-pituitary portal plexus provides the major blood source for the anterior pituitary, allowing reliable transmission of hypothalamic peptide pulses without significant systemic dilution; consequently, anterior pituitary cells are exposed to specific releasing or inhibiting factors and in turn release their respective hormones as discrete pulses into the systemic circulation (Fig. 378-3).

The posterior pituitary is supplied by the inferior hypophyseal arteries. In contrast to the anterior pituitary, the posterior lobe is directly innervated by hypothalamic neurons (supraopticohypophyseal and tuberohypophyseal nerve tracts) via the pituitary stalk (Chap. 381). Thus, posterior pituitary production of vasopressin (antidiuretic hormone [ADH]) and oxytocin is particularly sensitive to neuronal damage by lesions that affect the pituitary stalk or hypothalamus.

### PITUITARY DEVELOPMENT

The embryonic differentiation and maturation of anterior pituitary cells have been elucidated in considerable detail. Pituitary development from Rathke's pouch involves a complex interplay of lineage-specific transcription factors expressed in pluripotent Sox2-expressing precursor cells and gradients of locally produced growth factors (Table 378-1). The transcription factor Prop-1 induces pituitary development of Pit-1-specific lineages as well as gonadotropes. The transcription factor Pit-1 determines cell-specific expression of GH, PRL, and TSH in somatotropes, lactotropes, and thyrotropes. Expression of high levels of estrogen receptors in cells that contain Pit-1 favors PRL expression, whereas thyrotrope embryonic factor (TEF) induces TSH expression. Pit-1 binds to GH, PRL, and TSH gene regulatory elements, providing a mechanism for determining specific pituitary hormone phenotypic stability. Gonadotrope cell development is further defined by the cell-specific expression of the nuclear receptors steroidogenic factor (SF-1) and dosage-sensitive sex reversal, adrenal hypoplasia critical region, on chromosome X, gene 1 (DAX-1). Development of corticotrope cells, which express the proopiomelanocortin (POMC) gene, requires the T-Pit transcription factor. Abnormalities of pituitary development can be caused by inherited mutations of developmental transcription

factors including Pit-1, Prop-1, SF-1, DAX-1, and T-Pit, resulting in selective or combined pituitary hormone deficit syndromes.

## ANTERIOR PITUITARY HORMONES

Each anterior pituitary hormone is under unique control, and each exhibits highly specific normal and dysregulated secretory characteristics.

### PROLACTIN

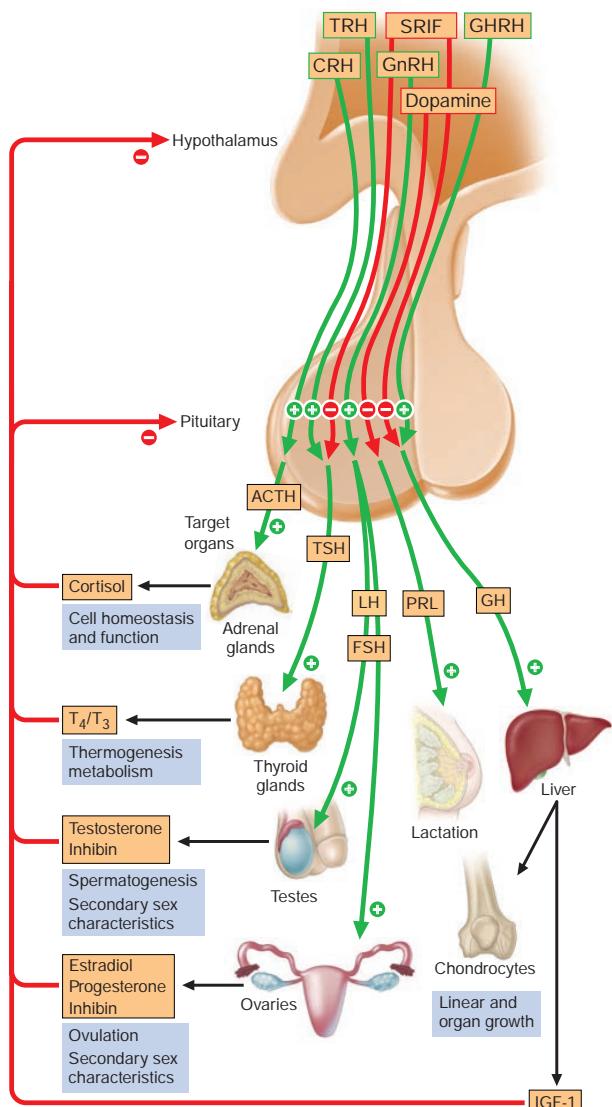
**Synthesis** PRL consists of 198 amino acids and has a molecular mass of 21,500 kDa; it is weakly homologous to GH and human placental lactogen (hPL), reflecting the duplication and divergence of a common GH-PRL-hPL precursor gene. PRL is synthesized in lactotropes, which constitute ~20% of anterior pituitary cells. Lactotropes and somatotropes are derived from a common precursor cell that may give rise to a tumor that secretes both PRL and GH. Marked lactotrope cell hyperplasia develops during pregnancy and the first few months of lactation. These transient functional changes in the lactotrope population are induced by estrogen to increase PRL production.

**Secretion** Normal adult serum PRL levels are about 10–25  $\mu$ g/L in women and 10–20  $\mu$ g/L in men. PRL secretion is pulsatile, with the highest secretory peaks occurring during non-rapid eye movement (non-REM) sleep. Peak serum PRL levels (up to 30  $\mu$ g/L) occur between 4:00 and 6:00 a.m. The circulating half-life of PRL is ~50 min.

PRL is unique among the pituitary hormones in that the predominant hypothalamic control mechanism is inhibitory, reflecting tonic dopamine-mediated suppression of PRL release. This regulatory pathway accounts for the spontaneous PRL hypersecretion that occurs with pituitary stalk section, often a consequence of head trauma or compressive mass lesions at the skull base. Pituitary dopamine type 2 ( $D_2$ ) receptors mediate inhibition of PRL synthesis and secretion. Targeted disruption (gene knockout) of the murine  $D_2$  receptor in mice results in hyperprolactinemia and lactotrope proliferation. As discussed below, dopamine agonists play a central role in the management of hyperprolactinemic disorders.

Thyrotropin-releasing hormone (TRH) (pyro Glu-His-Pro-NH<sub>2</sub>) is a hypothalamic tripeptide that elicits PRL release within 15–30 min after intravenous injection. TRH primarily regulates TSH, and the physiologic relevance of TRH for PRL regulation is unclear (Chap. 382).

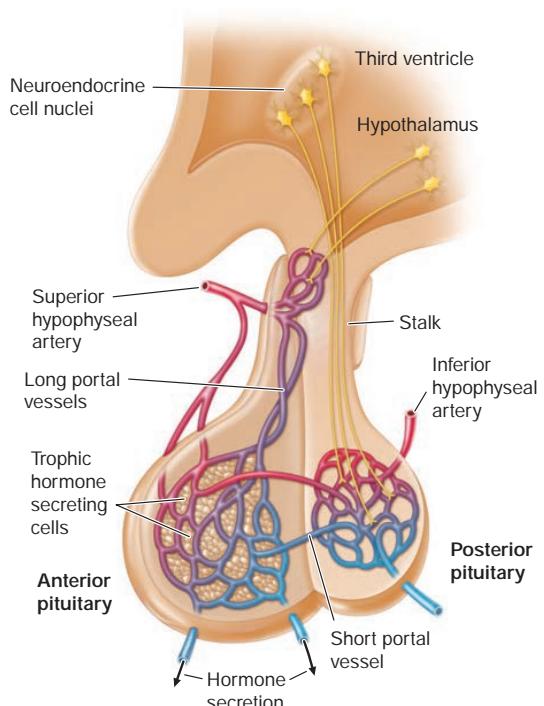
Serum PRL levels rise transiently after exercise, meals, sexual intercourse, minor surgical procedures, general anesthesia, chest wall injury, acute myocardial infarction, and other forms of acute stress.



**FIGURE 378-1** Diagram of pituitary axes. Hypothalamic hormones regulate anterior pituitary trophic hormones that in turn determine target gland secretion. Peripheral hormones feed back to regulate hypothalamic and pituitary hormones. For abbreviations, see text.

PRL levels increase markedly (about tenfold) during pregnancy and decline rapidly within 2 weeks of parturition. If breast-feeding is initiated, basal PRL levels remain elevated; suckling stimulates transient reflex increases in PRL levels that last for ~30–45 min. Breast suckling activates afferent neural pathways in the hypothalamus that induce PRL release. With time, suckling-induced responses diminish and interfeeding PRL levels return to normal.

**Action** The PRL receptor is a member of the type I cytokine receptor family that also includes GH and interleukin (IL) 6 receptors. Ligand binding induces receptor dimerization and intracellular signaling by Janus kinase (JAK), which stimulates translocation of the signal transduction and activators of transcription (STAT) family to activate target genes. Mutations of the PRL receptor result in PRL insensitivity, hyperprolactinemia, and oligomenorrhea. When homozygous, PRL receptor mutations cause agalactia, demonstrating that PRL action is necessary for lactation. In the breast, the lobuloalveolar epithelium proliferates in response to PRL, placental lactogens, estrogen,



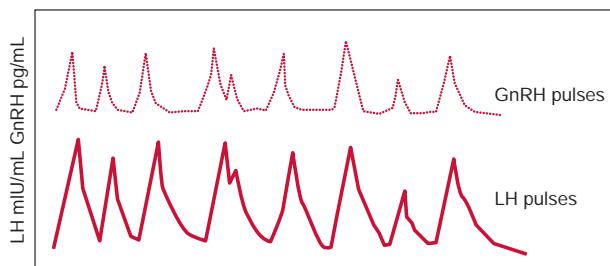
**FIGURE 378-2** Diagram of hypothalamic-pituitary vasculature. The hypothalamic nuclei produce hormones that traverse the portal system and impinge on anterior pituitary cells to regulate pituitary hormone secretion. Posterior pituitary hormones are derived from direct neural extensions.

progesterone, and local paracrine growth factors, including insulin-like growth factor 1 (IGF-1).

PRL acts to induce and maintain lactation and to suppress both reproductive function and sexual drive. These functions are geared toward ensuring that maternal lactation is sustained and not interrupted by pregnancy. PRL inhibits reproductive function by suppressing hypothalamic gonadotropin-releasing hormone (GnRH) and pituitary gonadotropin secretion and by impairing gonadal steroidogenesis in both women and men. In the ovary, PRL blocks folliculogenesis and inhibits granulosa cell aromatase activity, leading to hypoestrogenism and anovulation. PRL also has a luteolytic effect, generating a shortened, or inadequate, luteal phase of the menstrual cycle. In men, attenuated LH secretion leads to low testosterone levels and decreased spermatogenesis. These hormonal changes decrease libido and reduce fertility in patients with hyperprolactinemia.

## GROWTH HORMONE

**Synthesis** GH is the most abundant anterior pituitary hormone, and GH-secreting somatotrope cells constitute up to 50% of the total



**FIGURE 378-3** Hypothalamic gonadotropin-releasing hormone (GnRH) pulses induce secretory pulses of luteinizing hormone (LH).

anterior pituitary cell population. Mammosomatotrope cells, which coexpress PRL with GH, can be identified by using double immunostaining techniques. Somatotrope development and GH transcription are determined by expression of the cell-specific Pit-1 nuclear transcription factor. Five distinct genes encode GH and related proteins. The pituitary GH gene (*hGH-N*) produces two alternatively spliced products that give rise to 22-kDa GH (191 amino acids) and a less abundant 20-kDa GH molecule with similar biologic activity. Placental syncytiotrophoblast cells express a GH variant (*hGH-V*) gene; the related hormone human chorionic somatotropin (HCS) is expressed by distinct members of the gene cluster.

**Secretion** GH secretion is controlled by complex hypothalamic and peripheral factors. *GH-releasing hormone* (GHRH) is a 44-amino-acid hypothalamic peptide that stimulates GH synthesis and release. Ghrelin, an octanoylated gastric-derived peptide, and synthetic agonists of the *GHS-R* induce GHRH and also directly stimulate GH release. *Somatostatin* (somatotropin-release inhibiting factor [SRIF]) is synthesized in the medial preoptic area of the hypothalamus and inhibits GH secretion. GHRH is secreted in discrete spikes that elicit GH pulses, whereas SRIF sets basal GH secretory tone. SRIF also is expressed in many extrahypothalamic tissues, including the central nervous system (CNS), gastrointestinal tract, and pancreas, where it also acts to inhibit islet hormone secretion. *IGF-1*, the peripheral target hormone for GH, feeds back to inhibit GH; estrogen induces GH, whereas chronic glucocorticoid excess suppresses GH release, leading to growth delay in children.

Surface receptors on the somatotrope regulate GH synthesis and secretion. The GHRH receptor is a G protein-coupled receptor (GPCR) that signals through the intracellular cyclic AMP pathway to stimulate somatotrope cell proliferation as well as GH production. Inactivating mutations of the GHRH receptor cause profound dwarfism. A distinct surface receptor for ghrelin, the gastric-derived GH secretagogue, is expressed in both the hypothalamus and pituitary. Somatostatin binds to five distinct receptor subtypes (SST<sub>1</sub> to SST<sub>5</sub>); SST<sub>2</sub> and SST<sub>5</sub> subtypes preferentially suppress GH (and TSH) secretion, while SST<sub>5</sub> predominantly signals to suppress ACTH secretion.

GH secretion is pulsatile, with highest peak levels occurring at night, generally correlating with sleep onset. GH secretory rates decline markedly with age so that hormone levels in middle age are ~15% of pubertal levels. These changes are paralleled by an age-related decline in lean muscle mass. GH secretion is also reduced in obese individuals, although IGF-1 levels may not be suppressed, suggesting a change in the setpoint for feedback control. Elevated GH levels occur within an hour of deep sleep onset as well as after exercise, physical stress, and trauma and during sepsis. Integrated 24-h GH secretion is higher in women and is also enhanced by estrogen replacement, likely reflective of increased peripheral GH resistance. Using standard assays, random GH measurements are undetectable in ~50% of daytime samples obtained from healthy subjects and are also undetectable (<1 µg/L) in most obese and elderly subjects. Thus, single random GH measurements do not distinguish patients with adult GH deficiency from those with GH levels in the normal range.

GH secretion is profoundly influenced by nutritional factors. Using ultrasensitive GH assays with a sensitivity of 0.002 µg/L, a glucose load suppresses GH to <0.7 µg/L in women and to <0.07 µg/L in men. Increased GH pulse frequency and peak amplitudes occur with chronic malnutrition or prolonged fasting. GH is stimulated by oral ghrelin receptor agonists, intravenous l-arginine, dopamine, and apomorphine (a dopamine receptor agonist), as well as by -adrenergic pathways. -Adrenergic blockade induces basal GH and enhances GHRH- and insulin-evoked GH release.

**Action** The pattern of GH secretion may affect tissue responses. The higher GH pulsatility observed in men compared with the relatively continuous basal GH secretion in women may be an important biologic determinant of linear growth patterns and liver enzyme induction.

The 70-kDa peripheral GH receptor protein has structural homology with the cytokine/hematopoietic superfamily. A fragment of the receptor extracellular domain generates a soluble GH binding protein (GHB<sup>P</sup>) that binds to circulating GH. The liver and cartilage express the greatest number of GH receptors. GH binding to preformed receptor dimers is followed by internal rotation and subsequent signaling through the JAK/STAT pathway. Activated STAT proteins translocate to the nucleus, where they modulate expression of GH-regulated target genes. GH analogues that bind to the receptor but are incapable of mediating receptor signaling are potent antagonists of GH action. A GH receptor antagonist (pegvisomant) is approved for treatment of acromegaly.

GH induces protein synthesis and nitrogen retention and also impairs glucose tolerance by antagonizing insulin action. GH also stimulates lipolysis, leading to increased circulating fatty acid levels, reduced omental fat mass, and enhanced lean body mass. GH promotes sodium, potassium, and water retention and elevates serum levels of inorganic phosphate. Linear bone growth occurs as a result of complex hormonal and growth factor actions, including those of IGF-1. GH stimulates epiphyseal prechondrocyte differentiation. These precursor cells produce IGF-1 locally, and their proliferation is also responsive to the growth factor.

**Insulin-Like Growth Factors** Although GH exerts direct effects in target tissues, many of its physiologic effects are mediated indirectly through IGF-1, a potent growth and differentiation factor. The liver is the major source of circulating IGF-1. In peripheral tissues, IGF-1 also exerts local paracrine actions that appear to be both dependent on and independent of GH. Thus, GH administration induces circulating IGF-1 as well as stimulating local IGF-1 production in multiple tissues.

Both IGF-1 and IGF-2 are bound to high-affinity circulating IGF-binding proteins (IGFBPs) that regulate IGF availability and bioactivity. Levels of IGFBP3 are GH dependent, and it serves as the major carrier protein for circulating IGF-1. GH deficiency and malnutrition usually are associated with low IGFBP3 levels. IGFBP1 and IGFBP2 regulate local tissue IGF action but do not bind appreciable amounts of circulating IGF-1.

Serum IGF-1 concentrations are profoundly affected by physiologic factors. Levels increase during puberty, peak at 16 years, and subsequently decline by >80% during the aging process. IGF-1 concentrations are higher in women than in men. Because GH is the major determinant of hepatic IGF-1 synthesis, abnormalities of GH synthesis or action (including pituitary failure, GHRH receptor defect, GH receptor defect, or pharmacologic GH receptor blockade) lead to reduced IGF-1 levels. Hypocaloric states are associated with GH resistance; IGF-1 levels are therefore low with cachexia, malnutrition, and sepsis. In acromegaly, IGF-1 levels are invariably high and reflect a log-linear relationship with circulating GH concentrations.

**IGF-1 PHYSIOLOGY** Injected IGF-1 (100 µg/kg) induces hypoglycemia, and lower doses improve insulin sensitivity in patients with severe insulin resistance and diabetes. In cachectic subjects, IGF-1 infusion (12 µg/kg per h) enhances nitrogen retention and lowers cholesterol levels. Longer-term subcutaneous IGF-1 injections enhance protein synthesis and are anabolic. Although bone formation markers are induced, bone turnover also may be stimulated by IGF-1. IGF-1 is approved for use in patients with GH-resistance syndromes.

IGF-1 side effects are dose dependent, and overdose may result in hypoglycemia, hypotension, fluid retention, temporomandibular jaw pain, and increased intracranial pressure, all of which are reversible. Retinal damage and avascular femoral head necrosis have been reported. Chronic excess IGF-1 administration presumably would result in features of acromegaly.

## ADRENOCORTICOTROPIC HORMONE (See also Chap. 386)

**Synthesis** ACTH-secreting corticotrope cells constitute ~20% of the pituitary cell population. ACTH (39 amino acids) is derived from the POMC precursor protein (266 amino acids) that also generates several

other peptides, including -lipotropin, -endorphin, met-enkephalin, -melanocyte-stimulating hormone (-MSH), and corticotropin-like intermediate lobe protein (CLIP). The *POMC* gene is potently suppressed by glucocorticoids and induced by corticotropin-releasing hormone (CRH), arginine vasopressin (AVP), and proinflammatory cytokines, including IL-6, as well as leukemia inhibitory factor.

CRH, a 41-amino-acid hypothalamic peptide synthesized in the paraventricular nucleus as well as in higher brain centers, is the predominant stimulator of ACTH synthesis and release. The CRH receptor is a GPCR that is expressed on the corticotrope and signals to induce *POMC* transcription.

**Secretion** ACTH secretion is pulsatile and exhibits a characteristic circadian rhythm, peaking at about 6:00 a.m. and reaching a nadir about midnight. Adrenal glucocorticoid secretion, which is driven by ACTH, follows a parallel diurnal pattern. ACTH circadian rhythmicity is determined by variations in secretory pulse amplitude rather than changes in pulse frequency. Superimposed on this endogenous rhythm, ACTH levels are increased by physical and psychological stress, exercise, acute illness, and insulin-induced hypoglycemia.

Glucocorticoid-mediated negative regulation of the hypothalamic-pituitary-adrenal (HPA) axis occurs as a consequence of both hypothalamic CRH suppression and direct attenuation of pituitary *POMC* gene expression and ACTH release. In contrast, loss of cortisol feedback inhibition, as occurs in primary adrenal failure, results in extremely high ACTH levels.

Acute inflammatory or septic insults activate the HPA axis through the integrated actions of proinflammatory cytokines, bacterial toxins, and neural signals. The overlapping cascade of ACTH-inducing cytokines (tumor necrosis factor [TNF]; IL-1, -2, and -6; and leukemia inhibitory factor) activates hypothalamic CRH and AVP secretion, pituitary *POMC* gene expression, and local pituitary paracrine cytokine networks. The resulting cortisol elevation restrains the inflammatory response and enables host protection. Concomitantly, cytokine-mediated central glucocorticoid receptor resistance impairs glucocorticoid suppression of the HPA. Thus, the neuroendocrine stress response reflects the net result of highly integrated hypothalamic, intrapituitary, and peripheral hormone and cytokine signals acting to regulate cortisol secretion.

**Action** The major function of the HPA axis is to maintain metabolic homeostasis and mediate the neuroendocrine stress response. ACTH induces adrenocortical steroidogenesis by sustaining adrenal cell proliferation and function. The receptor for ACTH, designated *melanocortin-2 receptor*, is a GPCR that induces steroidogenesis by stimulating a cascade of steroidogenic enzymes (Chap. 386).

## GONADOTROPINS: FSH AND LH

**Synthesis and Secretion** Gonadotrope cells constitute ~10% of anterior pituitary cells and produce two gonadotropin hormones—LH and FSH. Like TSH and human chorionic gonadotropin, LH and FSH are glycoprotein hormones that comprise α and β subunits. The β subunit is common to these glycoprotein hormones; specificity of hormone function is conferred by the α subunits, which are expressed by separate genes.

Gonadotropin synthesis and release are dynamically regulated. This is particularly true in women, in whom rapidly fluctuating gonadal steroid levels vary throughout the menstrual cycle. Hypothalamic GnRH, a 10-amino-acid peptide, regulates the synthesis and secretion of both LH and FSH. Brain kisspeptin, a product of the *KISS1* gene, regulates hypothalamic GnRH release. GnRH is secreted in discrete pulses every 60–120 min, and the pulses in turn elicit LH and FSH pulses (Fig. 378-3). The pulsatile mode of GnRH input is essential to its action; pulses prime gonadotrope responsiveness, whereas continuous GnRH exposure induces desensitization. Based on this phenomenon, long-acting GnRH agonists are used to suppress gonadotropin levels in children with precocious puberty and in men with prostate cancer (Chap. 87) and are used in some ovulation-induction protocols to reduce levels

of endogenous gonadotropins (Chap. 392). Estrogens act at both the hypothalamus and the pituitary to modulate gonadotropin secretion. Chronic estrogen exposure is inhibitory, whereas rising estrogen levels, as occur during the preovulatory surge, exert positive feedback to increase gonadotropin pulse frequency and amplitude. Progesterone slows GnRH pulse frequency but enhances gonadotropin responses to GnRH. Testosterone feedback in men also occurs at the hypothalamic and pituitary levels and is mediated in part by its conversion to estrogens.

Although GnRH is the main regulator of LH and FSH secretion, FSH synthesis is also under distinct control by the gonadal peptides inhibin and activin, members of the transforming growth factor (TGF-) family. Inhibin selectively suppresses FSH, whereas activin stimulates FSH synthesis (Chap. 392).

**Action** The gonadotropin hormones interact with their respective GPCRs expressed in the ovary and testis, evoking germ cell development and maturation and steroid hormone biosynthesis. In women, FSH regulates ovarian follicle development and stimulates ovarian estrogen production. LH mediates ovulation and maintenance of the corpus luteum. In men, LH induces Leydig cell testosterone synthesis and secretion, and FSH stimulates seminiferous tubule development and regulates spermatogenesis.

## THYROID-STIMULATING HORMONE

**Synthesis and Secretion** TSH-secreting thyrotrope cells constitute 5% of the anterior pituitary cell population. TSH shares a common subunit with LH and FSH but contains a specific TSH subunit. TRH is a hypothalamic tripeptide (pyroglutamyl histidylprolinamide) that acts through a pituitary GPCR to stimulate TSH synthesis and secretion; it also stimulates the lactotrope cell to secrete PRL. TSH secretion is stimulated by TRH, whereas thyroid hormones, dopamine, somatostatin, and glucocorticoids suppress TSH by overriding TRH induction.

Thyrotrope cell proliferation and TSH secretion are both induced when negative feedback inhibition by thyroid hormones is removed. Thus, thyroid damage (including surgical thyroidectomy), radiation-induced hypothyroidism, chronic thyroiditis, and prolonged goitrogen exposure are associated with increased TSH levels. Long-standing untreated hypothyroidism can lead to elevated TSH levels, which may be associated with thyrotrope hyperplasia and pituitary enlargement and may sometimes be evident on magnetic resonance imaging.

**Action** TSH is secreted in pulses, although the excursions are modest in comparison to other pituitary hormones because of the low amplitude of the pulses and the relatively long half-life of TSH. Consequently, single determinations of TSH suffice to precisely assess its circulating levels. TSH binds to a GPCR on thyroid follicular cells to stimulate thyroid hormone synthesis and release (Chap. 382).

## FURTHER READING

- Bernard V et al: Prolactin: A pleiotropic factor in health and disease. *Nat Rev Endocrinol* 15:356, 2019.
- Cheung LYM et al: Single-cell RNA sequencing reveals novel markers of male pituitary stem cells and hormone-producing cell types. *Endocrinology* 159:3910, 2018.
- Das N, Kumar TR: Molecular regulation of follicle-stimulating hormone synthesis, secretion and action. *J Mol Endocrinol* 60:R131, 2018.
- Langlais D et al: Adult pituitary cell maintenance: Lineage-specific contribution of self-duplication. *Mol Endocrinol* 27:1103, 2013.
- Le Tissier P et al: The process of anterior pituitary hormone pulse generation. *Endocrinology* 159:3524, 2018.
- Murray PG et al: 60 years of neuroendocrinology: The hypothalamo-GH axis: The past 60 years. *J Endocrinol* 226:T123, 2015.
- Ranke MB, Wit JM: Growth hormone: Past, present and future. *Nat Rev Endocrinol* 14:285, 2018.



Deficient production of anterior pituitary hormones leads to features of hypopituitarism. Impaired production of one or more of the anterior pituitary trophic hormones can result from inherited disorders; more commonly, adult hypopituitarism is acquired and reflects the compressive mass effects of tumors or the consequences of local pituitary or hypothalamic traumatic, autoimmune, inflammatory, or vascular damage. These processes also may impair synthesis or secretion of hypothalamic hormones, with resultant pituitary failure (**Table 379-1**).

**TABLE 379-1 Etiology of Hypopituitarism<sup>a</sup>**

Development/structural	
Midline cerebral defect syndromes	
Pituitary dysplasia/aplasia	
Primary empty sella	
Congenital hypothalamic disorders (septo-optic dysplasia, Prader-Willi syndrome, Bardet-Biedl syndrome, Kallmann syndrome)	
Congenital central nervous system mass, encephalocele	
Genetic	
Combined pituitary hormone deficiencies	
Isolated primary hormone deficiencies	
Traumatic	
Surgical resection	
Radiotherapy damage	
Head injuries	
Neoplastic	
Pituitary adenoma	
Parasellar mass (germinoma, ependymoma, glioma)	
Rathke's cyst	
Craniopharyngioma	
Hypothalamic hamartoma, gangliocytoma	
Pituitary metastases (breast, lung, colon carcinoma)	
Lymphoma and leukemia	
Meningioma	
Infiltrative/inflammatory	
Lymphocytic hypophysitis	
Hemochromatosis	
Sarcoidosis	
Histiocytosis X	
Granulomatous hypophysitis	
Transcription factor antibodies	
Immunotherapy	
Vascular	
Pituitary apoplexy	
Pregnancy-related (infarction with diabetes; postpartum necrosis)	
Subarachnoid hemorrhage	
Sickle cell disease	
Arteritis	
Infections	
Fungal (histoplasmosis)	
Parasitic (toxoplasmosis)	
Tuberculosis	
<i>Pneumocystis jirovecii</i>	

<sup>a</sup>Trophic hormone failure associated with pituitary compression or destruction usually occurs sequentially: growth hormone > follicle-stimulating hormone > luteinizing hormone > thyroid-stimulating hormone > adrenocorticotrophic hormone. During childhood, growth retardation is often the presenting feature, and in adults, hypogonadism is the earliest symptom.

## DEVELOPMENTAL CAUSES OF HYPOPITUITARISM

Pituitary dysplasia may result in aplastic, hypoplastic, or ectopic pituitary gland development. Because pituitary development follows midline cell migration from the nasopharyngeal Rathke's pouch, midline craniofacial disorders may be associated with pituitary dysplasia. Acquired pituitary failure in the newborn also can be caused by birth trauma, including cranial hemorrhage, asphyxia, and breech delivery.

A large number of transcription factors and growth factors are critical for the development of the hypothalamus and pituitary gland and the function of differentiated anterior pituitary cell lineages. mutations have been described in the *HESX1*, *SOX2*, *SOX3*, *LHX3*, *LHX4*, *OTX*, *GLI2*, *PAX6*, *BMP4*, *ARNT2*, *FGF8*, *FGFR1*, *SHH*, *PROKR2*, *GPR161*, *IGSF1*, *PITX2*, and *CHD7* genes, among others. Heterozygous loss-of-function or autosomal recessive mutations disrupt hypothalamic and pituitary development at different developmental stages, causing a wide array of phenotypes ranging from severe syndromic midline and other defects to combined pituitary hormone defects or isolated hormone deficiencies. Depending on the gene involved, the pituitary may be hypoplastic, hyperplastic, or ectopic. Midline defects include variable combinations of abnormal development of the eyes, corpus colossum, vertebrae, and genital systems. Pituitary dysfunction ranges from isolated hormone deficiency to combined pituitary hormone deficiency (CPHD) and diabetes insipidus (DI).

In addition to these syndromic developmental disorders, some mutations affect specific pituitary cell lineages. For example, Pit-1 mutations cause combined growth hormone (GH), prolactin (PRL), and thyroid-stimulating hormone (TSH) deficiencies. These patients usually present with growth failure and varying degrees of hypothyroidism. The pituitary may appear hypoplastic on magnetic resonance imaging (MRI). Prop-1 is expressed early in pituitary development and appears to be required for Pit-1 function. Familial and sporadic *PROP1* mutations result in combined GH, PRL, TSH, and gonadotropin deficiency. Over 80% of these patients have growth retardation; by adulthood, all are deficient in TSH and gonadotropins, and a small minority later develop adrenocorticotrophic hormone (ACTH) deficiency. Because of gonadotropin deficiency, these individuals do not enter puberty spontaneously. In some cases, the pituitary gland appears enlarged on MRI. *TPIT* mutations result in ACTH deficiency associated with hypocortisolism. Mutations in *NR5A1* (also known as steroidogenic factor 1 [SF1]) impair development of gonadotrope cells, as well as adrenal/gonadal development.

## HYPOTHALAMIC ENDOCRINE DYSFUNCTION

Hypothalamic disorders can affect temperature regulation, appetite, sleep-wake cycles, autonomic systems, behavior, and memory, as well as multiple endocrine systems. Selected examples of hypothalamic disorders that affect the endocrine system are described below.

**Kallmann Syndrome** Kallmann syndrome results from defective hypothalamic gonadotropin-releasing hormone (GnRH) synthesis and is associated with anosmia or hyposmia due to olfactory bulb agenesis or hypoplasia (**Chap. 391**). Classically, the syndrome may also be associated with color blindness, optic atrophy, nerve deafness, cleft palate, renal abnormalities, cryptorchidism, and neurologic abnormalities such as mirror movements. The initial genetic cause was the X-linked *KAL* gene, mutations of which impair embryonic migration of GnRH neurons from the hypothalamic olfactory placode to the hypothalamus. Since then, more than a dozen additional genetic abnormalities, in addition to *KAL* mutations, have been found to cause isolated GnRH deficiency. Autosomal recessive (i.e., *GPR54*, *KISS1*) and dominant (i.e., *FGFR1*) modes of transmission have been described, and there is a growing list of genes associated with GnRH deficiency (including *GNRH1*, *PROK2*, *PROKR2*, *CHD7*, *PCSK1*, *FGF8*, *NELF*, *WDR11*, *TAC3*, *TACR3*, and *SEMA3E*). Some patients have oligogenic mutations in which mutations in a combination of different genes lead to the phenotype. Associated clinical features, in addition to GnRH deficiency, vary depending on the genetic cause. GnRH deficiency prevents progression through puberty. Males present with delayed puberty and pronounced hypogonadal features, including micropenis, probably the

result of low testosterone levels during infancy. Females present with primary amenorrhea and failure of secondary sexual development.

Kallmann syndrome and other causes of congenital GnRH deficiency are characterized by low luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels and low concentrations of sex steroids (testosterone or estradiol). In sporadic cases of isolated gonadotropin deficiency, the diagnosis is often one of exclusion after other known causes of hypothalamic-pituitary dysfunction have been eliminated. Repetitive GnRH administration restores normal pituitary gonadotropin responses, pointing to a hypothalamic defect in these patients.

Long-term treatment of males with human chorionic gonadotropin (hCG) or testosterone restores pubertal development and secondary sex characteristics; women can be treated with cyclic estrogen and progestin. Fertility may be restored by the administration of gonadotropins or by using a portable infusion pump to deliver subcutaneous, pulsatile GnRH.

**Bardet-Biedl Syndrome** This very rare genetically heterogeneous disorder is characterized by intellectual disability, renal abnormalities, obesity, and hexadactyly, brachydactyly, or syndactyly. Central DI may or may not be associated. GnRH deficiency occurs in 75% of males and half of affected females. Retinal degeneration begins in early childhood, and most patients are blind by age 30. Numerous subtypes of Bardet-Biedl syndrome (BBS) have been identified, with genetic linkage to at least nine different loci. Several of the loci encode genes involved in basal body cilia function, and this may account for the diverse clinical manifestations.

**Leptin and Leptin Receptor Mutations** Deficiencies of leptin or its receptor cause a broad spectrum of hypothalamic abnormalities, including hyperphagia, obesity, and central hypogonadism (Chap. 401). Decreased GnRH production in these patients results in attenuated pituitary FSH and LH synthesis and release.

**Prader-Willi Syndrome** This is a contiguous gene syndrome that results from deletion of the paternal copies of the imprinted *SNRPN* gene, the *NECDIN* gene, and possibly other genes on chromosome 15q. Prader-Willi syndrome is associated with hypogonadotropic hypogonadism, hyperphagia-obesity, chronic muscle hypotonia, mental retardation, and adult-onset diabetes mellitus. Multiple somatic defects also involve the skull, eyes, ears, hands, and feet. Diminished hypothalamic oxytocin- and vasopressin-producing nuclei have been reported. Deficient GnRH synthesis is suggested by the observation that chronic GnRH treatment restores pituitary LH and FSH release.

## ACQUIRED HYPOPITUITARISM

Hypopituitarism may be caused by accidental or neurosurgical trauma; vascular events such as apoplexy; pituitary or hypothalamic neoplasms, craniopharyngioma, lymphoma, or metastatic tumors; inflammatory disease such as lymphocytic hypophysitis; autoimmune hypophysitis associated with checkpoint inhibitor cancer immunotherapy; infiltrative disorders such as sarcoidosis, hemochromatosis (Chap. 414), and tuberculosis; or irradiation.

Patients with brain injury, including from contact sports trauma, motor vehicle accidents, explosive causes, subarachnoid hemorrhage, and irradiation, can experience transient or long-term hypopituitarism. Long-term periodic endocrine follow-up is indicated because hypothalamic or pituitary dysfunction will develop in 25–40% of these patients.

**Hypothalamic Infiltration Disorders** Sarcoidosis, histiocytosis X, amyloidosis, and hemochromatosis frequently involve both hypothalamic and pituitary neuronal and neurochemical tracts. Consequently, DI is a common presentation, reported in half of patients with these disorders. Growth retardation is seen if attenuated GH secretion occurs before puberty. Hypogonadotropic hypogonadism and hyperprolactinemia are also common.

**Inflammatory Lesions** Pituitary damage and subsequent secretory dysfunction can be seen with chronic site infections such as

tuberculosis, with opportunistic fungal infections associated with AIDS, and in tertiary syphilis. Other inflammatory processes, such as granulomas and sarcoidosis, should be considered in the differential diagnosis of imaging studies suggestive of a pituitary adenoma. These lesions may cause extensive hypothalamic and pituitary damage, leading to hormone deficiencies.

**Cranial Irradiation** Cranial irradiation may result in long-term hypothalamic and pituitary dysfunction, especially in children and adolescents, as they are more susceptible to damage after whole-brain or head and neck therapeutic irradiation. The development of subsequent hormonal abnormalities correlates strongly with irradiation dosage and the time interval after completion of radiotherapy. Up to two-thirds of patients ultimately develop hormone insufficiency after a median dose of 50 Gy (5000 rad) directed at the skull base. The development of hypopituitarism occurs over 5–15 years and usually reflects hypothalamic damage rather than primary destruction of pituitary cells. Although the pattern of hormone loss is variable, GH deficiency is most common, followed by gonadotropin, thyroid, and ACTH deficiency. When deficiency of one or more hormones is documented, the possibility of diminished reserve of other hormones is likely. Accordingly, anterior pituitary function should be continually evaluated over the long term in previously irradiated patients, and replacement therapy instituted when appropriate (see below).

**Lymphocytic Hypophysitis** This occurs most often in postpartum women; it usually presents with hyperprolactinemia and MRI evidence of a prominent pituitary mass that often resembles an adenoma, with mildly elevated PRL levels. Pituitary failure caused by diffuse lymphocytic infiltration may be transient or permanent but requires immediate evaluation and treatment. Rarely, isolated pituitary hormone deficiencies have been described, suggesting a selective autoimmune process targeted to specific cell types. Most patients manifest symptoms of progressive mass effects with headache and visual disturbance. The erythrocyte sedimentation rate often is elevated. Because it may be indistinguishable from a pituitary adenoma on MRI, hypophysitis should be considered in a postpartum woman with a newly diagnosed pituitary mass before an unnecessary surgical intervention is undertaken. The inflammatory process often resolves after several months of glucocorticoid treatment, and pituitary function may be restored, depending on the extent of damage.

**Immunotherapy and Hypophysitis** Pituitary cells express cytotoxic T lymphocyte antigen-4 (CTLA-4), and up to 20% of patients receiving cancer immunotherapy with CTLA-4 inhibitors (e.g., ipilimumab) may develop hypophysitis with heterogeneously associated thyroid, adrenal, islet, and gonadal failure. Hypophysitis is also reported with PD-1/PD-L1 inhibitors (e.g., pembrolizumab and nivolumab) and may show delayed presentation. Pituitary hormone replacement, with or without high-dose glucocorticoids, may be safely tolerated with continued immunotherapy.

**Pituitary Apoplexy** Acute intrapituitary hemorrhagic vascular events can cause substantial damage to the pituitary and surrounding sellar structures. Pituitary apoplexy may occur spontaneously in a preexisting pituitary adenoma; postpartum (Sheehan's syndrome); or in association with diabetes, hypertension, sickle cell anemia, or acute shock. The hyperplastic enlargement of the pituitary, which occurs normally during pregnancy, increases the risk for hemorrhage and infarction. Apoplexy is an endocrine emergency that may result in severe hypoglycemia, hypotension and shock, central nervous system (CNS) hemorrhage, and death. Acute symptoms may include severe headache with signs of meningeal irritation, bilateral visual changes, ophthalmoplegia, and, in severe cases, cardiovascular collapse and loss of consciousness. Pituitary computed tomography (CT) or MRI may reveal signs of intratumoral or sellar hemorrhage, with pituitary stalk deviation and compression of pituitary tissue.

Patients with no evident visual loss or impaired consciousness can be observed and managed conservatively with high-dose glucocorticoids. Those with significant or progressive visual loss, cranial nerve

palsy, or loss of consciousness require urgent surgical decompression. Visual recovery after sellar surgery is inversely correlated with the length of time after the acute event. Therefore, severe ophthalmoplegia or visual deficits are indications for early surgery. Hypopituitarism is common after apoplexy.

**Empty Sella** A partial or apparently totally empty sella is often an incidental MRI finding and may sometimes be associated with intracranial hypertension. These patients usually have normal pituitary function, implying that the surrounding rim of pituitary tissue is fully functional. Hypopituitarism, however, may develop insidiously. Pituitary adenomas also may undergo clinically silent infarction and involution with development of a partial or totally empty sella by cerebrospinal fluid (CSF) filling the dural herniation. Rarely, small but functional pituitary adenomas may arise within the rim of normal pituitary tissue, and they are not always visible on MRI.

### PRESENTATION AND DIAGNOSIS

The clinical manifestations of hypopituitarism depend on which hormones are lost and the extent of the hormone deficiency (see below). GH deficiency causes growth disorders in children and leads to abnormal body composition in adults. Gonadotropin deficiency causes menstrual disorders and infertility in women and decreased sexual function, infertility, and loss of secondary sexual characteristics in

men. TSH and ACTH deficiencies usually develop later in the course of pituitary failure. TSH deficiency causes growth retardation in children and features of hypothyroidism in children and adults. The secondary form of adrenal insufficiency caused by ACTH deficiency leads to hypocortisolism with relative preservation of mineralocorticoid production. PRL deficiency causes failure of lactation. When lesions involve the posterior pituitary, polyuria and polydipsia reflect loss of vasopressin secretion. In patients with long-standing pituitary damage, epidemiologic studies document an increased mortality rate, primarily from increased cardiovascular and cerebrovascular disease. Previous head or neck irradiation is also a determinant of increased mortality rates in patients with hypopituitarism, especially from cerebrovascular disease.

### LABORATORY INVESTIGATION

Biochemical diagnosis of pituitary insufficiency is made by demonstrating low levels of respective pituitary trophic hormones in the setting of low levels of target organ hormones. For example, low free thyroxine in the setting of a low or inappropriately normal TSH level suggests secondary hypothyroidism. Similarly, a low testosterone level without elevation of gonadotropins suggests hypogonadotropic hypogonadism. Provocative tests may be required to assess pituitary reserve (**Table 379-2**). GH responses to insulin-induced hypoglycemia, arginine,

**TABLE 379-2 Tests of Pituitary Sufficiency**

HORMONE	TEST	BLOOD SAMPLES	INTERPRETATION
Growth hormone (GH)	Insulin tolerance test: Regular insulin (0.05–0.15 U/kg IV)	-30, 0, 30, 60, 120 min for glucose and GH	Glucose <40 mg/dL; GH should be >3 µg/L
	GHRH test: 1 µg/kg IV	0, 15, 30, 45, 60, 120 min for GH	Normal response is GH >3 µg/L
	L-Arginine test: 30 g IV over 30 min	0, 30, 60, 120 min for GH	Normal response is GH >3 µg/L
	L-Dopa test: 500 mg PO	0, 30, 60, 120 min for GH	Normal response is GH >3 µg/L
Prolactin	TRH test: 200–500 µg IV	0, 20, and 60 min for TSH and PRL	Normal prolactin is >2 µg/L and increase >200% of baseline
ACTH	Insulin tolerance test: regular insulin (0.05–0.15 U/kg IV)	-30, 0, 30, 60, 90 min for glucose and cortisol	Glucose <40 mg/dL Cortisol should increase by >7 µg/dL or to >20 µg/dL
	CRH test: 1 µg/kg ovine CRH IV at 8 A.M.	0, 15, 30, 60, 90, 120 min for ACTH and cortisol	Basal ACTH increases 2- to 4-fold and peaks at 20–100 pg/mL Cortisol levels >20–25 µg/dL
	Metyrapone test: Metyrapone (30 mg/kg) at midnight	Plasma 11-deoxycortisol and cortisol at 8 A.M.; ACTH can also be measured	Plasma cortisol should be <4 g/dL to assure an adequate response Normal response is 11-deoxycortisol >7.5 µg/dL or ACTH >75 pg/mL
	Standard ACTH stimulation test: ACTH 1-24 (cosyntropin), 0.25 mg IM or IV	0, 30, 60 min for cortisol and aldosterone	Normal response is cortisol >21 g/dL and aldosterone response >4 ng/dL above baseline
	Low-dose ACTH test: ACTH 1-24 (cosyntropin), 1 µg IV	0, 30, 60 min for cortisol	Cortisol should be >21 µg/dL
	3-day ACTH stimulation test consists of 0.25 mg ACTH 1-24 given IV over 8 h each day		Cortisol >21 µg/dL
TSH	Basal thyroid function tests: $T_4$ , $T_3$ , TSH	Basal measurements	Low free thyroid hormone levels in the setting of TSH levels that are not appropriately increased indicate pituitary insufficiency
	TRH test: 200–500 µg IV	0, 20, 60 min for TSH and PRL <sup>a</sup>	TSH should increase by >5 mU/L unless thyroid hormone levels are increased
LH, FSH	LH, FSH, testosterone, estrogen	Basal measurements	Basal LH and FSH should be increased in postmenopausal women Low testosterone levels in the setting of low LH and FSH indicate pituitary insufficiency
	GnRH test: GnRH (100 µg) IV	0, 30, 60 min for LH and FSH	In most adults, LH should increase by 10 IU/L and FSH by 2 IU/L Normal responses are variable
Multiple hormones	Combined anterior pituitary test: GHRH (1 µg/kg), CRH (1 µg/kg), GnRH (100 µg), TRH (200 µg) are given IV	-30, 0, 15, 30, 60, 90, 120 min for GH, ACTH, cortisol, LH, FSH, and TSH	Combined or individual releasing hormone responses must be elevated in the context of basal target gland hormone values and may not be uniformly diagnostic (see text)

<sup>a</sup>Evoked PRL response indicates lactotrope integrity.

Abbreviations:  $T_3$ , triiodothyronine;  $T_4$ , thyroxine; TRH, thyrotropin-releasing hormone. For other abbreviations, see text.

glucagon, l-dopa, growth hormone-releasing hormone (GHRH), or growth hormone-releasing orally active ghrelin receptor agonist macimorelin can be used to assess GH reserve. Corticotropin-releasing hormone (CRH) administration induces ACTH release, and administration of synthetic ACTH (cosyntropin) evokes adrenal cortisol release as an indirect indicator of pituitary ACTH reserve (Chap. 386). ACTH reserve is most reliably assessed by measuring ACTH and cortisol levels during insulin-induced hypoglycemia. However, this test should be performed cautiously in patients with suspected adrenal insufficiency because of enhanced susceptibility to hypoglycemia and hypotension. Administering insulin to induce hypoglycemia is contraindicated in patients with active coronary artery disease or known seizure disorders.

## TREATMENT

### Hypopituitarism

Hormone replacement therapy, including glucocorticoids, thyroid hormone, sex steroids, GH, and vasopressin, is usually safe and free of complications. Treatment regimens that mimic physiologic hormone production allow for maintenance of satisfactory clinical homeostasis. Effective dosage schedules are outlined in Table 379-3. Patients in need of glucocorticoid replacement require especially careful dose adjustments during stressful events such as acute illness, dental procedures, trauma, and hospitalization.

## DISORDERS OF GROWTH AND DEVELOPMENT

**Skeletal Maturation and Somatic Growth** The growth plate is dependent on a variety of hormonal stimuli, including GH, insulin-like growth factor (IGF)-1, sex steroids, thyroid hormones, paracrine and circulating growth factors (e.g., fibroblast growth factor family), and cytokines. The growth-promoting process also requires caloric energy, amino acids, vitamins, and trace metals and consumes ~10% of normal energy production. Malnutrition impairs chondrocyte activity, increases GH resistance, and leads to reduced circulating IGF-1 and IGF binding protein (IGFBP)-3 levels.

Linear bone growth rates are very high in infancy and are pituitary dependent. Mean growth velocity is ~6 cm/year in later childhood and

usually is maintained within a given range on a standardized percentile chart. Peak growth rates occur during midpuberty when bone age is 12 (girls) or 13 (boys). Secondary sexual development is associated with elevated sex steroids that cause progressive epiphyseal growth plate closure. *Bone age* is delayed in patients with all forms of true GH deficiency or GH receptor defects that result in attenuated GH action.

*Short stature* may occur as a result of constitutive intrinsic growth defects or because of acquired extrinsic factors that impair growth. In general, delayed bone age in a child with short stature is suggestive of a hormonal or systemic disorder, whereas normal bone age in a short child is more likely to be caused by a genetic cartilage dysplasia or growth plate disorder (Chap. 413).

**GH Deficiency in Children** Isolated GH deficiency is characterized by short stature, micropenis, increased fat, high-pitched voice, and a propensity to hypoglycemia due to relatively unopposed insulin action. Familial modes of inheritance are seen in at least one-third of these individuals and may be autosomal dominant, recessive, or X-linked. About 10% of children with GH deficiency have mutations in the *GH-N* gene, including gene deletions and a wide range of point mutations. Mutations in transcription factors Pit-1 and Prop-1, which control somatotrope development (see above), result in GH deficiency in combination with other pituitary hormone deficiencies, which may become manifest only in adulthood. The diagnosis of *idiopathic GH deficiency* should be made only after known molecular defects have been rigorously excluded.

**GHRH RECEPTOR MUTATIONS** Recessive mutations of the GHRH receptor gene in subjects with severe proportionate dwarfism are associated with low basal GH levels that cannot be stimulated by exogenous GHRH, GH-releasing peptide, or insulin-induced hypoglycemia, as well as anterior pituitary hypoplasia. The syndrome exemplifies the importance of the GHRH receptor for determining somatotrope cell proliferation and hormonal responsiveness.

**GH INSENSITIVITY** This is caused by defects of GH receptor structure or signaling. Homozygous or heterozygous mutations of the GH receptor are associated with partial or complete GH insensitivity and growth failure (*Laron syndrome*). The diagnosis is based on normal or high GH levels, with decreased circulating GH-binding protein (GHBP), and low IGF-1 levels. Very rarely, defective IGF-1, IGF-1 receptor, or IGF-1 signaling defects are also encountered. *STAT5B* mutations result in both immunodeficiency as well as abrogated GH signaling, leading to short stature with normal or elevated GH levels and low IGF-1 levels. Circulating GH receptor antibodies may rarely cause peripheral GH insensitivity.

**NUTRITIONAL SHORT STATURE** Caloric deprivation and malnutrition, uncontrolled diabetes, and chronic renal failure represent secondary causes of abrogated GH receptor function. These conditions also stimulate production of proinflammatory cytokines, which act to exacerbate the block of GH-mediated signal transduction. Children with these conditions typically exhibit features of acquired short stature with normal or elevated GH and low IGF-1 levels.

**PSYCHOSOCIAL SHORT STATURE** Emotional and social deprivation lead to growth retardation accompanied by delayed speech, discordant hyperphagia, and an attenuated response to administered GH. A nurturing environment restores growth rates.

## PRESENTATION AND DIAGNOSIS

Short stature is commonly encountered in clinical practice, and the decision to evaluate these children requires clinical judgment in association with auxologic data and family history. Short stature should be evaluated comprehensively if a patient's height is >3 standard deviations below the mean for age or if the growth rate has decelerated. Skeletal maturation is best evaluated by measuring a radiologic bone age, which is based mainly on the degree of wrist bone growth plate fusion. Final height can be predicted using standardized scales (Bayley-Pinneau or Tanner-Whitehouse) or estimated by adding 6.5 cm (boys) or subtracting 6.5 cm (girls) from the midparental height.

**TABLE 379-3 Hormone Replacement Therapy for Adult Hypopituitarism<sup>a</sup>**

HORMONE DEFICIT	HORMONE REPLACEMENT
ACTH	Hydrocortisone (10–20 mg/d in divided doses)
	Cortisone acetate (15–25 mg/d in divided doses)
	Prednisone (5 mg A.M.)
TSH	L-Thyroxine (0.075–0.15 mg daily)
FSH/LH	Males
	Testosterone gel (5–10 g/d)
	Testosterone skin patch (5 mg/d)
	Testosterone enanthate (200 mg IM every 2 weeks)
	Females
	Conjugated estrogen (0.65–1.25 mg qd for 25 days)
	Progesterone (5–10 mg qd) on days 16–25
	Estradiol skin patch (0.025–0.1 mg every week), adding progesterone on days 16–25 if uterus intact
	For fertility: menopausal gonadotropins, human chorionic gonadotropins
GH	Adults: Somatotropin (0.1–1.25 mg SC qd)
	Children: Somatotropin (0.02–0.05 mg/kg per day)
Vasopressin	Intranasal desmopressin (5–20 g twice daily) Oral 300–600 µg qd

<sup>a</sup>All doses shown should be individualized for specific patients and should be reassessed during stress, surgery, or pregnancy. Male and female fertility requirements should be managed as discussed in Chaps. 391 and 392.

Note: For abbreviations, see text.

Because GH secretion is pulsatile, GH deficiency is best assessed by examining the response to provocative stimuli, including exercise, insulin-induced hypoglycemia, and other pharmacologic tests that normally increase GH to  $>7 \mu\text{g/L}$  in children. Random GH measurements do not distinguish normal children from those with true GH deficiency. Adequate adrenal and thyroid hormone replacement should be assured before testing. Age-matched IGF-1 levels are not sufficiently sensitive or specific to make the diagnosis but can be useful to confirm GH deficiency. Pituitary MRI may reveal pituitary mass lesions or structural defects. Molecular analyses for known mutations should be undertaken when the cause of short stature remains cryptic or when additional clinical features suggest a genetic cause.

## TREATMENT

### Disorders of Growth and Development

Replacement therapy with recombinant GH (0.02–0.05 mg/kg per day SC) restores growth velocity in GH-deficient children to  $\sim 10 \text{ cm/year}$ . If pituitary insufficiency is documented, other associated hormone deficits should be corrected, especially adrenal steroids. In selected situations, GH treatment may be combined with strategies to delay puberty (e.g., GnRH agonist) or reduce sex steroids (e.g., aromatase inhibitors) as a means to mitigate sex steroid effect on epiphyseal closure. GH treatment is also moderately effective for accelerating growth rates in children with Turner syndrome and chronic renal failure. Treating psychosocial or constitutional (idiopathic) short stature with GH is not uniformly recommended as these children may only experience modest additive growth, which should be weighed against GH cost and side effect profiles.

In patients with GH insensitivity and growth retardation due to mutations of the GH receptor, treatment with IGF-1 bypasses the dysfunctional GH receptor.

## ADULT GH DEFICIENCY

Adult GH deficiency (AGHD) usually is caused by acquired hypothalamic or pituitary somatotrope damage. Acquired pituitary hormone deficiency follows a typical pattern in which loss of adequate GH reserve foreshadows subsequent hormone deficits. The sequential order of hormone loss is usually GH  $\rightarrow$  FSH/LH  $\rightarrow$  TSH  $\rightarrow$  ACTH. Patients previously diagnosed with childhood-onset GH deficiency should be retested as adults to affirm the diagnosis.

### PRESENTATION AND DIAGNOSIS

The clinical features of AGHD include changes in body composition, lipid metabolism, and quality of life and cardiovascular dysfunction (**Table 379-4**). Body composition changes are common and include reduced lean body mass, increased fat mass with selective deposition of intraabdominal visceral fat, and increased waist-to-hip ratio. Hyperlipidemia, left ventricular dysfunction, hypertension, and increased plasma fibrinogen levels also may be present. Bone mineral content is reduced, with resultant increased fracture rates. Patients may experience social isolation, depression, and difficulty maintaining gainful employment. Adult hypopituitarism is associated with a threefold increase in cardiovascular mortality rates in comparison to age- and sex-matched controls, and this may be due to GH deficiency, as patients in these studies were replaced with other deficient pituitary hormones.

### LABORATORY INVESTIGATION

AGHD is rare, and in light of the nonspecific nature of associated clinical symptoms, patients appropriate for testing should be selected carefully on the basis of well-defined criteria. With few exceptions, testing should be restricted to patients with the following predisposing factors: (1) pituitary surgery, (2) pituitary or hypothalamic tumor or granulomas, (3) history of cranial irradiation, (4) radiologic evidence of a pituitary lesion, and (5) childhood requirement for GH replacement therapy. The transition of a GH-deficient adolescent to

**TABLE 379-4 Features of Adult Growth Hormone Deficiency**

#### Clinical

Impaired quality of life  
Decreased energy and drive  
Poor concentration  
Low self-esteem  
Social isolation

Body composition changes  
Increased body fat mass  
Central fat deposition  
Increased waist-to-hip ratio  
Decreased lean body mass

Reduced exercise capacity  
Reduced maximum  $O_2$  uptake  
Impaired cardiac function  
Reduced muscle mass

Cardiovascular risk factors  
Impaired cardiac structure and function  
Abnormal lipid profile  
Decreased fibrinolytic activity  
Atherosclerosis  
Omental obesity

#### Imaging

Pituitary: mass or structural damage  
Bone: reduced bone mineral density  
Abdomen: excess omental adiposity

#### Laboratory

Evoked GH  $<3 \text{ ng/mL}$   
IGF-1 and IGFBP3 low or normal  
Increased LDL cholesterol  
Concomitant gonadotropin, TSH, and/or ACTH reserve deficits may be present

Abbreviation: LDL, low-density lipoprotein. For other abbreviations, see text.

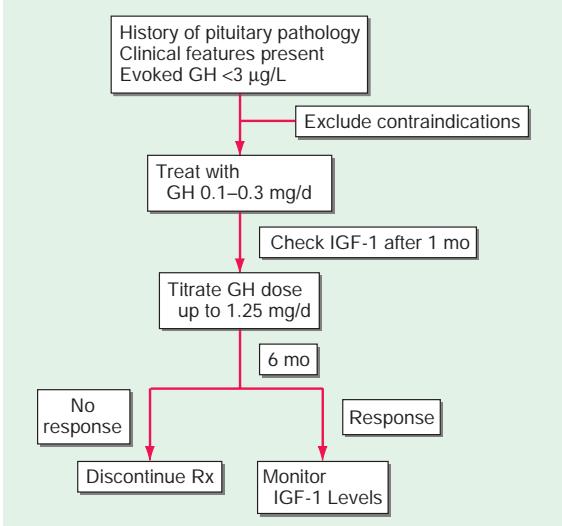
adulthood requires retesting to document subsequent AGHD. Up to 20% of patients previously treated for childhood-onset GH deficiency are found to be GH sufficient on repeat testing as adults.

A significant proportion (~25%) of truly GH-deficient adults have low-normal IGF-1 levels. Thus, as in the evaluation of GH deficiency in children, valid age-matched IGF-1 measurements provide a useful index of therapeutic responses but are not sufficiently precise for diagnostic purposes. The most validated test to distinguish pituitary-sufficient patients from those with AGHD is insulin-induced (0.05–0.1 U/kg) hypoglycemia. After glucose reduction to  $\sim 40 \text{ mg/dL}$ , most individuals experience neuroglycopenic symptoms (**Chap. 406**), and peak GH release occurs at 60 min and remains elevated for up to 2 h. About 90% of healthy adults exhibit GH responses  $>5 \mu\text{g/L}$ ; AGHD is defined by a peak GH response to hypoglycemia of  $<3 \mu\text{g/L}$ . Although insulin-induced hypoglycemia is safe when performed under appropriate supervision, it is contraindicated in patients with diabetes, ischemic heart disease, cerebrovascular disease, or epilepsy and in elderly patients. Alternative stimulatory tests include intravenous arginine (30 g), GHRH (1  $\mu\text{g/kg}$ ), oral ghrelin receptor agonist (0.5 mg/kg), and glucagon (1 mg). Combinations of these tests may evoke GH secretion in subjects who are not responsive to a single test.

## TREATMENT

### Adult GH Deficiency

Once the diagnosis of AGHD is unequivocally established, replacement of GH may be indicated. Contraindications to therapy include the presence of an active neoplasm, intracranial hypertension, and uncontrolled diabetes and retinopathy. The starting adult dose of 0.1–0.2 mg/d should be titrated (up to a maximum of 1.25 mg/d) to maintain IGF-1 levels in the mid-normal range for age- and



**FIGURE 379-1** Management of adult growth hormone (GH) deficiency. IGF, insulin-like growth factor; Rx, treatment.

sex-matched controls (Fig. 379-1). Women require higher doses than men, and elderly patients require less GH. Long-term GH maintenance sustains normal IGF-1 levels and is associated with persistent body composition changes (e.g., enhanced lean body mass and lower body fat). High-density lipoprotein cholesterol increases, but total cholesterol and insulin levels may not change significantly. Lumbar spine bone mineral density increases, but this response is gradual (>1 year). Many patients note significant improvement in quality of life when evaluated by standardized questionnaires. The effect of GH replacement on mortality rates in GH-deficient patients is currently the subject of long-term prospective investigation. Recently approved long-acting GH preparations for patients with AGHD require weekly injections. Ideally, dosing should be titrated to achieve normal but not supra-normal IGF-1 levels. Early reports indicate that side effects appear similar to subcutaneous formulations.

About 30% of patients exhibit reversible dose-related fluid retention, joint pain, and carpal tunnel syndrome, and up to 40% exhibit myalgias and paresthesia. Patients receiving insulin require careful monitoring for dosing adjustments, as GH is a potent counterregulatory hormone for insulin action. Patients with type 2 diabetes mellitus may initially develop further insulin resistance. However, glycemic control usually improves with the sustained loss of abdominal fat associated with long-term GH replacement. Headache, increased intracranial pressure, hypertension, and tinnitus occur rarely. Pituitary tumor regrowth and progression of skin lesions or other tumors have not been encountered in long-term surveillance programs with appropriate replacement doses.

## ACTH DEFICIENCY

### PRESENTATION AND DIAGNOSIS

Secondary adrenal insufficiency occurs as a result of pituitary ACTH deficiency. It is characterized by fatigue, weakness, anorexia, nausea, vomiting, and, occasionally, hypoglycemia. In contrast to primary adrenal failure, hypocortisolism associated with pituitary failure usually is not accompanied by hyperpigmentation or mineralocorticoid deficiency.

ACTH deficiency is commonly due to glucocorticoid withdrawal after treatment-associated suppression of the hypothalamic-pituitary-adrenal (HPA) axis. Isolated ACTH deficiency may occur after surgical resection of an ACTH-secreting pituitary adenoma that has suppressed the HPA axis; this phenomenon is in fact suggestive of

a surgical cure. The mass effects of other pituitary adenomas or sellar lesions may lead to ACTH deficiency, usually in combination with other pituitary hormone deficiencies. Partial ACTH deficiency may be unmasked in the presence of an acute medical or surgical illness, when clinically significant hypocortisolism reflects diminished ACTH reserve. Rarely, *TPIT* or *POMC* mutations result in primary ACTH deficiency.

### LABORATORY DIAGNOSIS

Inappropriately low ACTH levels in the setting of low cortisol levels are characteristic of diminished ACTH reserve. Low basal serum cortisol levels are associated with blunted cortisol responses to ACTH stimulation and impaired cortisol response to insulin-induced hypoglycemia or testing with metyrapone or CRH. **For a description of provocative ACTH tests, see Chap. 386.**

### TREATMENT

#### ACTH Deficiency

Glucocorticoid replacement therapy improves most features of ACTH deficiency. The total daily dose of hydrocortisone replacement preferably should generally not exceed 20 mg daily, divided into two or three doses. Prednisone (5 mg each morning) is longer acting and has fewer mineralocorticoid effects than hydrocortisone. Some authorities advocate lower maintenance doses in an effort to avoid cushingoid side effects. Doses should be increased severalfold during periods of acute illness or stress. Patients should wear medical alert bracelets and/or carry identification cards with information about their glucocorticoid requirements.

## GONADOTROPIN DEFICIENCY

Hypogonadism is the most common presenting feature of adult hypopituitarism even when other pituitary hormones are also deficient. It is often a harbinger of hypothalamic or pituitary lesions that impair GnRH production or delivery through the pituitary stalk. As noted below, hypogonadotropic hypogonadism is a common presenting feature of hyperprolactinemia.

A variety of inherited and acquired disorders are associated with *isolated hypogonadotropic hypogonadism* (Chap. 391). Hypothalamic defects associated with GnRH deficiency include Kallmann syndrome and mutations in more than a dozen genes that regulate GnRH neuron migration, development, and function (see above). Mutations in *CPR54*, *DAX1*, *NR5A1*, kisspeptin, the GnRH receptor, and the LH or FSH subunit genes also cause pituitary gonadotropin deficiency. Acquired forms of GnRH deficiency leading to hypogonadotropism are seen in association with anorexia nervosa, stress, starvation, and extreme exercise but also may be idiopathic. Hypogonadotropic hypogonadism in these disorders is reversed by removal of the stressful stimulus or by caloric replenishment.

### PRESENTATION AND DIAGNOSIS

In premenopausal women, hypogonadotropic hypogonadism presents as diminished ovarian function leading to oligomenorrhea or amenorrhea, infertility, decreased vaginal secretions, decreased libido, and breast atrophy. In hypogonadal adult men, secondary testicular failure is associated with decreased libido and potency, infertility, decreased muscle mass with weakness, reduced beard and body hair growth, soft testes, and characteristic fine facial wrinkles. Osteoporosis occurs in both untreated hypogonadal women and men.

### LABORATORY INVESTIGATION

Central hypogonadism is associated with low or inappropriately normal serum gonadotropin levels in the setting of low sex hormone concentrations (testosterone in men, estradiol in women). Because gonadotropin secretion is pulsatile, valid assessments may require repeated measurements or the use of pooled serum samples. Men have reduced sperm counts.

Intravenous GnRH (100 µg) stimulates gonadotropes to secrete LH (which peaks within 30 min) and FSH (which plateaus during the ensuing 60 min). Normal responses vary according to menstrual cycle stage, age, and sex of the patient. Generally, LH levels increase about threefold, whereas FSH responses are less pronounced. In the setting of gonadotropin deficiency, a normal gonadotropin response to GnRH indicates intact pituitary gonadotrope function and suggests a hypothalamic abnormality. An absent response, however, does not reliably distinguish pituitary from hypothalamic causes of hypogonadism. For this reason, GnRH testing usually adds little to the information gained from baseline evaluation of the hypothalamic-pituitary-gonadotrope axis except in cases of isolated GnRH deficiency (e.g., Kallmann syndrome).

MRI examination of the sellar region and assessment of other pituitary functions usually are indicated in patients with documented central hypogonadism.

## TREATMENT

### Gonadotropin Deficiency

In males, testosterone replacement is necessary to achieve and maintain normal growth and development of the external genitalia, secondary sex characteristics, male sexual behavior, and androgenic anabolic effects, including maintenance of muscle function and bone mass. Testosterone may be administered by intramuscular injections every 1–4 weeks or by using skin patches or testosterone gels (Chap. 391). Gonadotropin injections (hCG or human menopausal gonadotropin [hMG]) over 12–18 months are used to restore fertility. Pulsatile GnRH therapy (25–150 ng/kg every 2 h), administered by a subcutaneous infusion pump, is also effective for treatment of hypothalamic hypogonadism when fertility is desired.

In premenopausal women, cyclical replacement of estrogen and progesterone maintains secondary sexual characteristics and integrity of genitourinary tract mucosa and prevents premature osteoporosis (Chap. 392). Gonadotropin therapy is used for ovulation induction. Follicular growth and maturation are initiated using hMG or recombinant FSH; hCG or human luteinizing hormone (hLH) is subsequently injected to induce ovulation. As in men, pulsatile GnRH therapy can be used to treat hypothalamic causes of gonadotropin deficiency.

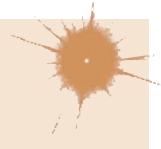
## DIABETES INSIPIDUS

See Chap. 381 for diagnosis and treatment of DI.

### FURTHER READING

- Chanson P et al: Adrenal insufficiency: Screening methods and confirmation of diagnosis. Ann Endocrinol (Paris) 78:495, 2017.
- Fleseriu M et al: Hormonal replacement in hypopituitarism in adults: An Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 101:3888, 2016.
- Garcia JM et al: Macimorelin as a diagnostic test for adult GH deficiency. J Clin Endocrinol Metab 103:3083, 2018.
- Higham CE et al: Hypopituitarism. Lancet 388:2403, 2016.
- Melmed S: Pathogenesis and diagnosis of growth hormone deficiency in adults. N Engl J Med 380:2551, 2019.
- Miller BS et al: Long-acting growth hormone preparations-current status and future considerations. J Clin Endocrinol Metab 105:e2121, 2020.
- Tanriverdi F et al: Pituitary dysfunction after traumatic brain injury: A clinical and pathophysiological approach. Endocr Rev 36:305, 2015.
- Xatzipasali M et al: Congenital hypopituitarism: Various genes, various phenotypes. Horm Metab Res 51:81, 2019.
- Yamamoto M et al: Autoimmune pituitary disease: New concepts with clinical implications. Endocr Rev 41:261, 2020.

Shlomo Melmed, J. Larry Jameson



## HYPOTHALAMIC, PITUITARY, AND OTHER SELLAR MASSES

### EVALUATION OF SELLAR MASSES

**Local Mass Effects** Clinical manifestations of sellar lesions vary, depending on the anatomic location of the mass and the direction of its extension (Table 380-1). The dorsal sellar diaphragm presents the least resistance to soft tissue expansion from the sella; consequently, pituitary adenomas frequently extend in a suprasellar direction. Bony invasion may occur as well, especially through the sellar floor to the sphenoid sinus (Fig. 380-1).

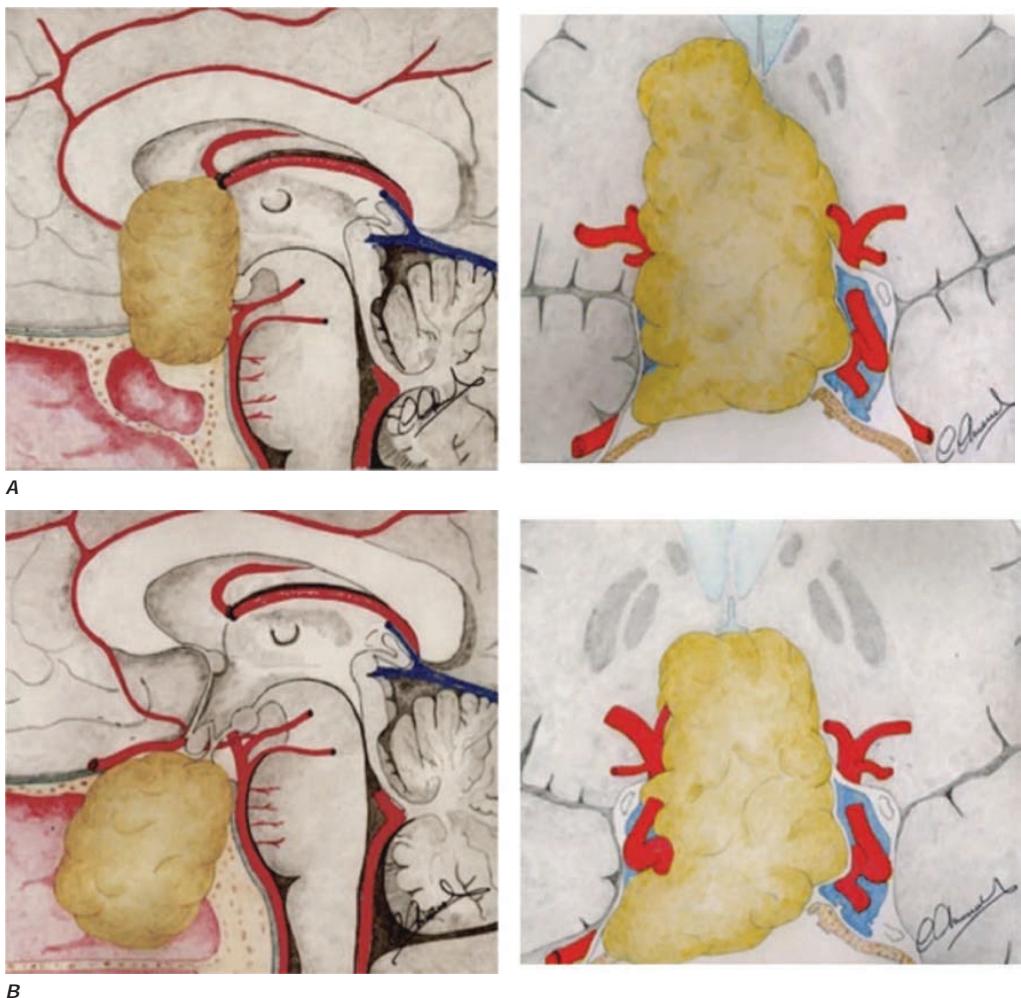
Headaches are common features of small intrasellar tumors, even with no demonstrable suprasellar extension. Because of the confined nature of the pituitary, small changes in intrasellar pressure stretch the dural plate; however, headache severity correlates poorly with adenoma size or extension.

Suprasellar extension can lead to visual loss by several mechanisms, the most common being compression of the optic chiasm. Rarely, direct invasion of the optic nerves or obstruction of cerebrospinal fluid (CSF) flow leading to secondary visual disturbances can occur. Pituitary stalk compression by a hormonally active or inactive intrasellar mass may compress the portal vessels, disrupting pituitary access to

TABLE 380-1 Features of Sellar Mass Lesions<sup>a</sup>

IMPACTED STRUCTURE	CLINICAL IMPACT
Pituitary	Hypogonadism Hypothyroidism Growth failure, adult growth hormone deficiency Hypoadrenalinism Hyperprolactinemia (stalk compression)
Optic chiasm	Loss of red perception Bitemporal hemianopia Superior or bitemporal field defect Scotoma Blindness
Hypothalamus	Temperature dysregulation Appetite and thirst disorders Obesity Diabetes insipidus Sleep disorders Behavioral dysfunction Autonomic dysfunction
Cavernous sinus	Ophthalmoplegia with or without ptosis or diplopia Facial numbness
Frontal lobe	Personality disorder Anosmia
Brain	Headache Hydrocephalus Psychosis Dementia Laughing seizures

<sup>a</sup>As the intrasellar mass expands, it first compresses intrasellar pituitary tissue, then usually invades dorsally through the dura to lift the optic chiasm or laterally to the cavernous sinuses. Bony erosion is rare, as is direct brain compression. Microadenomas may present with headache.



**FIGURE 380-1 Expanding pituitary mass.** Pituitary mass expansion may (A) impinge vital soft tissue structures and (B) invade the sphenoid sinus. (Reproduced with permission from P Cappabianca et al: Size does not matter. The intrigue of giant adenomas: a true surgical challenge. *Acta Neurochir (Wien)* 156:2217, 2014.)

hypothalamic hormones and dopamine; this results in early hyperprolactinemia and later concurrent loss of other pituitary hormones. This “stalk section” phenomenon may also be caused by trauma, whiplash injury with posterior clinoid stalk compression, or skull base fractures. Lateral mass invasion may impinge on the cavernous sinus and compress its neural contents, leading to cranial nerve III, IV, and VI palsies as well as effects on the ophthalmic and maxillary branches of the fifth cranial nerve (Chap. 441). Patients may present with diplopia, ptosis, ophthalmoplegia, and decreased facial sensation, depending on the extent of neural damage. Extension into the sphenoid sinus indicates that the pituitary mass has eroded through the sellar floor (Fig. 380-1). Aggressive tumors rarely invade the palatal roof and cause nasopharyngeal obstruction, infection, and CSF leakage. Temporal and frontal lobe involvement may rarely lead to uncinate seizures, personality disorders, and anosmia. Direct hypothalamic encroachment by an invasive pituitary mass may cause important metabolic sequelae, including precocious puberty or hypogonadism, diabetes insipidus, sleep disturbances, dysthermia, and appetite disorders.

**Magnetic Resonance Imaging** Sagittal and coronal T1-weighted magnetic resonance imaging (MRI) before and after administration of gadolinium allows precise visualization of the pituitary gland with clear delineation of the hypothalamus, pituitary stalk, pituitary tissue

and surrounding suprasellar cisterns, cavernous sinuses, sphenoid sinus, and optic chiasm. Pituitary gland height ranges from 6 mm in children to 8 mm in adults; during pregnancy and puberty, the height may reach 10–12 mm. The upper aspect of the adult pituitary is flat or slightly concave, but in adolescent and pregnant individuals, this surface may be convex, reflecting physiologic pituitary enlargement. The stalk should be midline and vertical.

Anterior pituitary gland soft tissue consistency is slightly heterogeneous on MRI, and signal intensity resembles that of brain matter on T1-weighted imaging (Fig. 380-2). Adenoma density is usually lower than that of surrounding normal tissue on T1-weighted imaging, and the signal intensity increases with T2-weighted images. Computed tomography (CT) scan is reserved to define the extent of bony erosion or the presence of calcification.

Sellar masses are encountered commonly as incidental findings on MRI, and most are pituitary adenomas (incidentalomas). In the absence of hormone hypersecretion, these small intrasellar lesions can be monitored safely with MRI, which is performed annually and then less often if there is no evidence of further growth. Resection should be considered for incidentally discovered larger macroadenomas, because about one-third become invasive or cause local pressure effects. If hormone hypersecretion is identified, specific therapies are indicated as described below. When larger masses (>1 cm) are encountered, they



**FIGURE 380-2 Pituitary adenoma.** Coronal T1-weighted postcontrast magnetic resonance image shows a homogeneously enhancing mass (arrowheads) in the sella turcica and suprasellar region compatible with a pituitary adenoma; the small arrows outline the carotid arteries.

should also be distinguished from nonadenomatous lesions. Meningiomas often are associated with bony hyperostosis; craniopharyngiomas may have calcifications and are usually hypodense, whereas gliomas are hyperdense on T2-weighted images.

**Ophthalmologic Evaluation** Because optic tracts may be contiguous to an expanding pituitary mass, reproducible visual field assessment using perimetry techniques should be performed on all patients with sellar mass lesions that impinge the optic chiasm (*Chap. 32*). Bitemporal hemianopia, often more pronounced superiorly, is observed classically. It occurs because nasal ganglion cell fibers, which cross in the optic chiasm, are especially vulnerable to compression of the ventral optic chiasm. Occasionally, homonymous hemianopia occurs from postchiasmatic compression or monocular temporal field loss from prechiasmatic compression. Invasion of the cavernous sinus can produce diplopia from ocular motor nerve palsy. Early diagnosis reduces the risk of optic atrophy, vision loss, or eye misalignment.

**Laboratory Investigation** The presenting clinical features of functional pituitary adenomas (e.g., acromegaly, prolactinoma, or Cushing's disease) should guide the laboratory studies (**Table 380-2**). However, for a sellar mass with no obvious clinical features of hormone excess, laboratory studies are geared toward determining the nature of the tumor and assessing the possible presence of hypopituitarism. When a pituitary adenoma is suspected based on MRI, initial hormonal evaluation usually includes (1) basal prolactin (PRL); (2) insulin-like growth factor (IGF)-1; (3) 24-h urinary free cortisol (UFC) and/or overnight oral dexamethasone (1 mg) suppression test; (4)  $\alpha$  subunit, follicle-stimulating hormone (FSH), and luteinizing hormone (LH); and (5) thyroid function tests. Additional hormonal evaluation may be indicated based on the results of these tests. Pending more detailed assessment of hypopituitarism, a menstrual history, measurement of testosterone and 8 a.m. cortisol levels, and thyroid function tests usually identify patients with pituitary hormone deficiencies that require hormone replacement before further testing or surgery (*Chap. 379*).

**Histologic Evaluation** Immunohistochemical staining of pituitary tumor specimens obtained at transsphenoidal surgery for hormones as well as cell-type specific transcription factors confirms clinical and laboratory studies and provides a histologic diagnosis when hormone studies are equivocal and in cases of clinically nonfunctioning tumors.

**TABLE 380-2 Screening Tests for Functional Pituitary Adenomas**

TEST	COMMENTS
Acromegaly	Serum IGF-1 Oral glucose tolerance test with GH obtained at 0, 30, and 60 min Normal subjects should suppress growth hormone to <1 $\mu\text{g/L}$
Prolactinoma	Serum PRL Exclude medications MRI of the sella should be ordered if PRL is elevated
Cushing's disease	24-h urinary free cortisol Dexamethasone (1 mg) at 11 P.M. and fasting plasma cortisol measured at 8 A.M. Late night salivary cortisol ACTH assay Ensure urine collection is total and accurate Normal subjects suppress to <5 $\mu\text{g/dL}$ Distinguishes adrenal adenoma (ACTH suppressed) from ectopic ACTH or Cushing's disease (ACTH normal or elevated)
Gonadotropinoma	Baseline FSH, LH, free $\alpha$ subunit, ovarian hyperstimulation, estrogen (females), testosterone (males) TRH stimulation test with assays for LH, FSH, free $\alpha$ subunit, free LH $\beta$ , free FSH $\beta$ subunits Rare; more commonly nonfunctioning adenomas Consider screening for hypopituitarism Some gonadotropinomas exhibit an inappropriate gonadotropin response to TRH
TSH-producing adenoma	Free $\text{T}_4$ , free $\text{T}_3$ , TSH, free $\alpha$ subunit Key feature is an inappropriately normal or high TSH in the setting of elevated free $\text{T}_4$ and $\text{T}_3$

*Abbreviations:* ACTH, adrenocorticotropin hormone; FSH, follicle-stimulating hormone; GH, growth hormone; IGF-I, insulin-like growth factor I; LH, luteinizing hormone; MRI, magnetic resonance imaging; PRL, prolactin; TSH, thyroid-stimulating hormone.

## TREATMENT

### Hypothalamic, Pituitary, and Other Sellar Masses

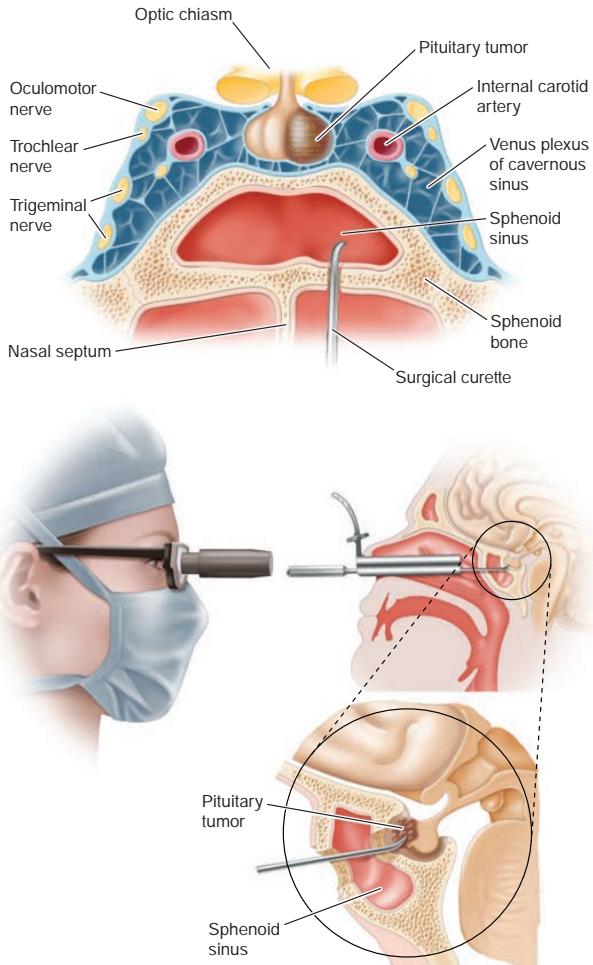
#### OVERVIEW

Successful management of sellar masses requires accurate diagnosis as well as selection of optimal therapeutic modalities. Most pituitary tumors are benign and slow growing. Clinical features result from local mass effects and hormonal hyper- or hyposecretion syndromes caused directly by the adenoma or occurring as a consequence of treatment. Thus, lifelong management and follow-up are necessary for these patients.

MRI with gadolinium enhancement for pituitary visualization, new advances in transsphenoidal surgery and in stereotactic radiotherapy, and novel therapeutic agents have improved pituitary tumor management. The goals of pituitary tumor treatment include normalization of excess pituitary secretion, amelioration of symptoms and signs of hormonal hypersecretion syndromes, and shrinkage or ablation of large tumor masses with relief of adjacent structure compression. Residual anterior pituitary function should be preserved during treatment and sometimes can be restored by removing the tumor mass. Ideally, adenoma recurrence should be prevented.

#### TRANSSPHENOIDAL SURGERY

Transsphenoidal resection is the desired surgical approach for pituitary tumors, except for the rare invasive suprasellar mass surrounding the frontal or middle fossa or the optic nerves or invading posteriorly behind the clivus, which may require transcranial approaches. Intraoperative microscopy facilitates visual distinction between adenomatous and normal pituitary tissue as well as microdissection of small tumors that may not be visible by MRI (**Fig. 380-3**). Endoscopic techniques with three-dimensional



**FIGURE 380-3** Transsphenoidal resection of pituitary mass via the endonasal approach.

intraoperative localization enable better visualization and access to tumor tissue. Transsphenoidal surgery also avoids cranial invasion and manipulation of brain tissue required by subfrontal surgical approaches. Individual surgical experience is a major determinant of outcome efficacy with these techniques.

In addition to correction of hormonal hypersecretion, pituitary surgery is indicated for mass lesions that impinge on surrounding structures. Surgical decompression and resection are required for an expanding pituitary mass, which may be asymptomatic or accompanied by persistent headache, progressive visual field defects, cranial nerve palsies, hydrocephalus, and, occasionally, intrapituitary hemorrhage and apoplexy. Transsphenoidal surgery rarely is used for pituitary tissue biopsy to establish a histologic diagnosis. Whenever possible, the pituitary mass lesion should be selectively excised; normal pituitary tissue should be manipulated or resected only when critical for effective mass dissection. Non-selective hemihypophysectomy or total hypophysectomy may be indicated if no hypersecreting mass lesion is clearly discernible, multifocal lesions are present, or the remaining nontumorous pituitary tissue is obviously necrotic. This strategy, however, increases the likelihood of postoperative hypopituitarism and the need for lifelong hormone replacement.

Preoperative mass effects, including visual field defects and compromised pituitary function, may be reversed by surgery, particularly when the deficits are not long-standing. For large and invasive tumors, it is necessary to determine the optimal balance between

maximal tumor resection and preservation of anterior pituitary hormonal function, especially for preserving growth and reproductive function in younger patients. Tumor invasion outside the sella is rarely amenable to surgical cure, and the surgeon must judge the risk-versus-benefit ratio of extensive tumor resection.

**Side Effects** Tumor size, the degree of invasiveness, and experience of the surgeon largely determine the incidence of surgical complications. Operative mortality rate is ~1%. Transient diabetes insipidus and hypopituitarism occur in up to 20% of patients. Permanent diabetes insipidus, cranial nerve damage, nasal septal perforation, or visual disturbances may be encountered in up to 10% of patients. CSF leaks occur in 4% of patients. Less common complications include carotid artery injury, loss of vision, hypothalamic damage, and meningitis. Permanent side effects are rare after surgery for microadenomas.

### RADIATION

Radiation is used either as a primary therapy for pituitary or parasellar masses or, more commonly, as an adjunct to surgery or medical therapy. Focused megavoltage irradiation is achieved by precise MRI localization, using a high-voltage linear accelerator and accurate isocentric rotational arcing. A major determinant of accurate irradiation is reproduction of the patient's head position during multiple visits and maintenance of absolute head immobility. A total of <50 Gy (5000 rad) is given as 180-cGy (180-rad) fractions divided over ~6 weeks. Stereotactic radiosurgery delivers a large single high-energy dose from a cobalt-60 source (Gamma Knife), linear accelerator, or cyclotron. Long-term effects of Gamma Knife surgery appear to be similar to those encountered with conventional radiation. Proton beam therapy is available in some centers and provides concentrated radiation doses within a localized region.

The role of radiation therapy in pituitary tumor management depends on the nature and anatomic location of the tumor, the age of the patient, and the availability of surgical and radiation expertise. Because of its relatively slow onset of action, radiation therapy is usually reserved for postsurgical management. As an adjuvant to surgery, radiation is used to treat residual tumor in an attempt to prevent persistent growth or recurrence. Irradiation offers the only means for potentially ablating significant postoperative residual nonfunctioning tumor tissue. By contrast, PRL-, growth hormone (GH)-, adrenocorticotropin hormone (ACTH)-, and thyrotropin (thyroid-stimulating hormone [TSH])-secreting residual tumor tissues are amenable to medical therapy.

**Side Effects** In the short term, radiation may cause transient nausea and weakness. Alopecia and loss of taste and smell may be more long-lasting. Failure of pituitary hormone synthesis is common in patients who have undergone head and neck or pituitary-directed irradiation. More than 50% of patients develop loss of GH, ACTH, TSH, and/or gonadotropin secretion within 10 years, usually due to hypothalamic damage. Lifelong follow-up with testing of anterior pituitary hormone reserve is therefore required after radiation treatment. Optic nerve damage with impaired vision due to optic neuritis is reported in ~2% of patients who undergo pituitary irradiation. Cranial nerve damage is uncommon now that radiation doses are <2 Gy (200 rad) at any one treatment session and the maximum dose is <50 Gy (5000 rad). The use of stereotactic radiotherapy reduces the risk of damage to adjacent structures. Conventional radiotherapy for pituitary tumors has been associated with adverse mortality rates, mainly from cerebrovascular disease. The cumulative risk of developing a secondary tumor after conventional radiation is 1.3% after 10 years and 1.9% after 20 years.

### MEDICAL

Medical therapy for pituitary tumors is highly specific and depends on tumor type. For prolactinomas, dopamine agonists are the treatment of choice. For acromegaly, somatostatin receptor ligands (SRLs) and a GH receptor antagonist are indicated. For TSH-secreting tumors, SRLs and occasionally dopamine agonists

are indicated. ACTH-secreting tumors may respond to SRLs, and adrenal-directed therapy may also be of benefit. Nonfunctioning tumors are generally not responsive to medications and require surgery and/or irradiation.

## SELLAR MASSES

Sellar masses may arise from brain, hypothalamic, or pituitary tissues. Each exhibit features related to the lesion location but also unique to the specific etiology. Unique MRI characteristics inform the differential diagnosis of pituitary masses (Fig. 380-4).

Lesions involving the anterior and preoptic hypothalamic regions cause paradoxical vasoconstriction, tachycardia, and hyperthermia. Acute hyperthermia usually is due to a hemorrhagic insult, but poikilothermia may also occur. Central disorders of thermoregulation result from posterior hypothalamic damage. The *periodic hypothermia syndrome* is characterized by episodic attacks of rectal temperatures <30°C (86°F), sweating, vasodilation, vomiting, and bradycardia (Chap. 464). Damage to the ventromedial hypothalamic nuclei by craniopharyngiomas, hypothalamic trauma, or inflammatory disorders may be associated with *hyperphagia* and *obesity*. This region appears to contain an energy-satiety center where melanocortin receptors are influenced by leptin, insulin, proopiomelanocortin (POMC) products, and gastrointestinal peptides (Chap. 401). Polydipsia and hypodipsia are associated with damage to central osmoreceptors located in preoptic nuclei (Chap. 381). Slow-growing hypothalamic lesions can cause increased somnolence and disturbed sleep cycles as well as obesity, hypothermia, and emotional outbursts. Lesions of the central hypothalamus may stimulate sympathetic neurons, leading to elevated serum catecholamine and cortisol levels. These patients are predisposed to cardiac arrhythmias, hypertension, and gastric erosions.

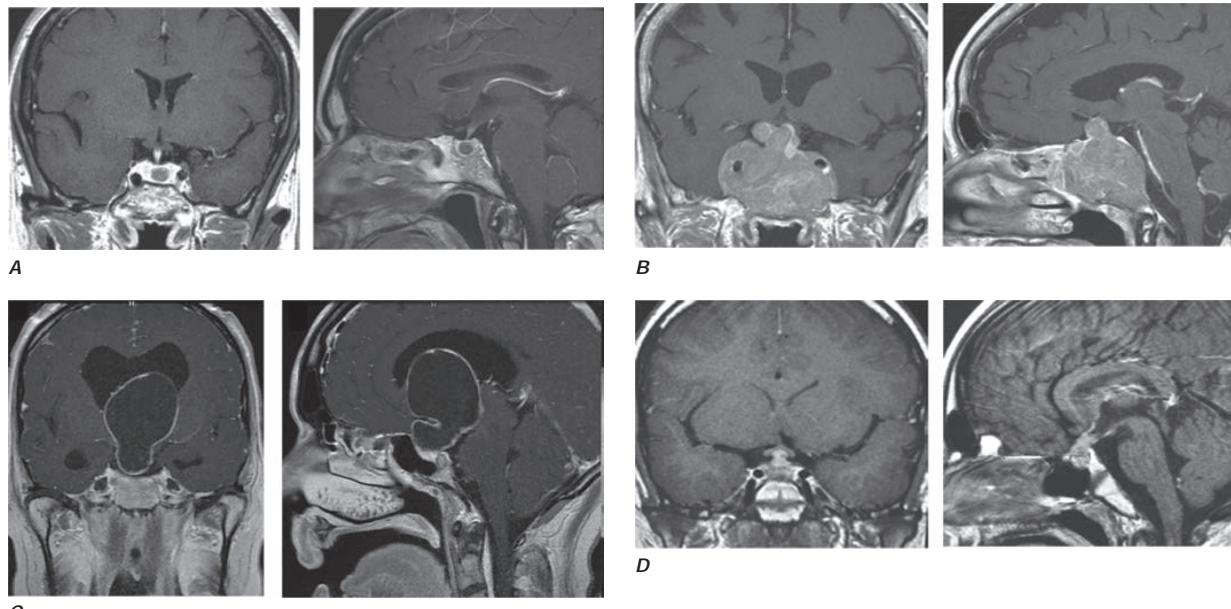
*Craniopharyngiomas* are benign, suprasellar cystic masses that present with headaches, visual field deficits, and variable degrees of hypopituitarism. They are derived from Rathke's pouch and arise near the pituitary stalk, commonly extending into the suprasellar cistern. Craniopharyngiomas are often large, cystic, and locally invasive. Many are partially calcified, exhibiting a characteristic appearance on skull x-ray and CT images. More than half of all patients present before age 20, usually with signs of increased intracranial pressure, including

headache, vomiting, papilledema, and hydrocephalus. Associated symptoms include visual field abnormalities, personality changes and cognitive deterioration, cranial nerve damage, sleep difficulties, and weight gain accompanied by features of the metabolic syndrome. Hypopituitarism is documented in ~90%, and diabetes insipidus occurs in ~10% of patients. About half of affected children present with growth retardation. MRI is generally superior to CT for evaluating cystic structure and tissue components of craniopharyngiomas. CT is useful to define calcifications and evaluate invasion into surrounding bony structures and sinuses.

Treatment usually involves transcranial or transsphenoidal surgical resection followed by postoperative radiation of residual tumor. Surgery alone is curative in less than half of patients because of recurrences due to adherence to vital structures or because of small tumor deposits in the hypothalamus or brain parenchyma. The goal of surgery is to remove as much tumor as possible without risking complications associated with efforts to remove firmly adherent or inaccessible tissue. In the absence of radiotherapy, ~75% of craniopharyngiomas recur, and 10-year survival is <50%. In patients with incomplete resection, radiotherapy improves 10-year survival to 70–90% but is associated with increased risk of secondary malignancies. Most patients require lifelong pituitary hormone replacement. As some craniopharyngiomas (particularly papillary) are associated with activated BRAF V600E mutations, use of BRAF inhibitors (dabrafenib or vemurafenib) either alone or in combination with MEK inhibitors (trametinib or cobimetinib) has resulted in long-term growth responses in some patients.

Developmental failure of Rathke's pouch obliteration may lead to *Rathke's cysts*, which are small (<5 mm) cysts entrapped by squamous epithelium and are found in ~20% of individuals at autopsy. Although Rathke's cleft cysts do not usually grow and are often diagnosed incidentally, about a third present in adulthood with compressive symptoms, diabetes insipidus, and hyperprolactinemia due to stalk compression. Rarely, hydrocephalus develops. The diagnosis is suggested preoperatively by visualizing the cyst wall on MRI, which distinguishes these lesions from craniopharyngiomas. Cyst contents range from CSF-like fluid to mucoid material. *Arachnoid cysts* are rare and generate an MRI image that is isointense with CSF.

*Sella chordomas* usually present with bony clival erosion, local invasiveness, and, on occasion, calcification. Normal pituitary tissue may be visible on MRI, distinguishing chordomas from aggressive



**FIGURE 380-4** Imaging differential diagnosis of sellar masses. **A.** Microadenoma. **B.** Macroadenoma. **C.** Craniopharyngioma. **D.** Hypophysitis with stalk thickening. (*C*. Reproduced with permission from Muller HL: Childhood craniopharyngioma. Recent advances in diagnosis, treatment and follow-up. Horm Res 69:193, 2008. *A, B, D*: Used with permission from Vivien Bonert, MD.)

pituitary adenomas. Mucinous material may be obtained by fine-needle aspiration.

*Meningiomas* arising in the sellar region may be difficult to distinguish from nonfunctioning pituitary adenomas. Meningiomas typically enhance on MRI and may show evidence of calcification or bony erosion. Meningiomas may cause compressive symptoms.

*Histiocytosis X* includes a variety of syndromes associated with foci of eosinophilic granulomas. Diabetes insipidus, exophthalmos, and punched-out lytic bone lesions (*Hand-Schüller-Christian disease*) are associated with granulomatous lesions visible on MRI, as well as a characteristic axillary skin rash. Rarely, the pituitary stalk may be involved.

*Pituitary metastases* occur in ~3% of cancer patients. Bloodborne metastatic deposits are found almost exclusively in the posterior pituitary. Accordingly, diabetes insipidus can be a presenting feature of lung, gastrointestinal, breast, and other pituitary metastases. About half of pituitary metastases originate from breast cancer; ~25% of patients with metastatic breast cancer have such deposits. Rarely, pituitary stalk involvement results in anterior pituitary insufficiency. The MRI diagnosis of a metastatic lesion may be difficult to distinguish from an aggressive pituitary adenoma; the diagnosis may require histologic examination of excised tumor tissue. Primary or metastatic lymphoma, leukemias, and plasmacytomas also occur within the sella.

*Hypothalamic hamartomas* and *gangliocytomas* may arise from astrocytes, oligodendrocytes, and neurons with varying degrees of differentiation. These tumors may overexpress hypothalamic neuropeptides, including gonadotropin-releasing hormone (GnRH), growth hormone-releasing hormone (GHRH), and corticotropin-releasing hormone (CRH). With GnRH-producing tumors, children present with precocious puberty, psychomotor delay, and laughing-associated seizures. Medical treatment of GnRH-producing hamartomas with long-acting GnRH analogues effectively suppresses gonadotropin secretion and controls premature pubertal development. Rarely, hamartomas also are associated with craniofacial abnormalities; imperforate anus; cardiac, renal, and lung disorders; and pituitary failure as features of *Pallister-Hall syndrome*, which is caused by mutations in the carboxy terminus of the *GLI3* gene. Hypothalamic hamartomas are often contiguous with the pituitary, and preoperative MRI diagnosis may not be possible. Histologic evidence of hypothalamic neurons in tissue resected at transsphenoidal surgery may be the first indication of a primary hypothalamic lesion.

*Hypothalamic gliomas* and *optic gliomas* occur mainly in childhood and usually present with visual loss. Adults have more aggressive tumors; about a third are associated with neurofibromatosis.

*Brain germ cell tumors* may arise within the sellar region. They include *dysgerminomas*, which frequently are associated with diabetes insipidus and visual loss. They rarely metastasize. *Germinomas*, *embryonal carcinomas*, *teratomas*, and *choriocarcinomas* may arise in the parasellar region and produce human chorionic gonadotropin (hCG). These germ cell tumors present with precocious puberty, diabetes insipidus, visual field defects, and thirst disorders. Many patients are GH deficient with short stature.

## PITUITARY ADENOMAS AND HYPERSECRETION SYNDROMES

Pituitary adenomas are the most common cause of pituitary hormone hypersecretion and hyposecretion syndromes in adults. They account for ~15% of all intracranial neoplasms and have been identified with a population prevalence of ~80/100,000. At autopsy, up to one-quarter of all pituitary glands harbor an unsuspected microadenoma (<10 mm diameter). Similarly, pituitary imaging detects small clinically inapparent pituitary lesions in at least 10% of individuals.

**Pathogenesis** Pituitary adenomas are benign neoplasms that arise from one of the five anterior pituitary cell types. The clinical and biochemical phenotypes of pituitary adenomas depend on the cell type from which they are derived. Thus, tumors arising from lactotrope (PRL), somatotrope (GH), corticotrope (ACTH), thyrotrope (TSH), or gonadotrope (LH, FSH) cells hypersecrete their respective hormones (Table 380-3). Plurihormonal tumors express various combinations of

TABLE 380-3 Classification of Pituitary Adenomas<sup>a</sup>

ADENOMA CELL ORIGIN	HORMONE PRODUCT	CLINICAL SYNDROME
Lactotrope	PRL	Hypogonadism, galactorrhea
Gonadotrope	FSH, LH, subunits	Silent, ovarian hyperstimulation, hypogonadism
Somatotrope	GH	Acromegaly/gigantism
Corticotrope	ACTH/none	Cushing's disease or silent
Mixed growth hormone and prolactin cell	GH, PRL	Acromegaly, hypogonadism, galactorrhea
Other plurihormonal cell	Any	Mixed
Acidophil stem cell	PRL, GH	Hypogonadism, galactorrhea, acromegaly
Mammosomatotrope	PRL, GH	Hypogonadism, galactorrhea, acromegaly
Thyrotrope	TSH	Thyrototoxicosis
Null cell	None	Hypopituitarism/none
Oncocytoma	None	Hypopituitarism/none

<sup>a</sup>Hormone-secreting tumors are listed in decreasing order of frequency. All tumors may cause local pressure effects, including visual disturbances, cranial nerve palsy, and headache.

Note: For abbreviations, see text.

Source: Adapted with permission from S Melmed: Pathogenesis of pituitary tumors. Nat Rev Endocrinol 7:257, 2011.

GH, PRL, TSH, ACTH, or the glycoprotein hormone or subunits. They may be diagnosed by careful immunocytochemistry of specific hormone and transcription factor expression or may manifest as clinical syndromes that combine features of these hormonal hypersecretory syndromes. Morphologically, these tumors may arise from a single polysecreting cell type or include cells with mixed function within the same tumor.

Hormonally active tumors are characterized by autonomous hormone secretion with diminished feedback responsiveness to physiologic inhibitory pathways. Hormone production does not always correlate with tumor size. Small hormone-secreting adenomas may cause significant clinical perturbations, whereas larger adenomas that produce less hormone may be clinically silent and remain undiagnosed (if no central compressive effects occur). About one-third of all adenomas are clinically nonfunctioning and produce no distinct clinical hypersecretory syndrome. Most of them arise from gonadotrope cells and may secrete small amounts of - and -glycoprotein hormone subunits or, very rarely, intact circulating gonadotropins. True pituitary carcinomas with documented extracranial metastases are exceedingly rare.

Almost all pituitary adenomas are monoclonal in origin, implying the acquisition of one or more somatic mutations that confer a selective growth advantage. Consistent with their clonal origin, complete surgical resection of small pituitary adenomas usually cures hormone hypersecretion. Nevertheless, hypothalamic hormones such as GHRH and CRH also enhance mitotic activity of their respective pituitary target cells in addition to their role in pituitary hormone regulation. Thus, patients who harbor rare abdominal or chest tumors that elaborate ectopic GHRH or CRH may present with somatotrope or corticotrope hyperplasia with GH or ACTH hypersecretion.

Several etiologic genetic events have been implicated in the development of pituitary tumors. The pathogenesis of sporadic forms of acromegaly has been particularly informative as a model of tumorigenesis. GHRH, after binding to its G protein-coupled somatotrope receptor, uses cyclic adenosine monophosphate (AMP) as a second messenger to stimulate GH secretion and somatotrope proliferation. A subset (~35%) of GH-secreting pituitary tumors contains sporadic mutations in  $G_s$ . These mutations attenuate intrinsic GTPase activity, resulting in constitutive elevation of cyclic AMP, Pit-1 induction, and activation of cyclic AMP response element binding protein (CREB), thereby promoting somatotrope cell proliferation and GH secretion.

	GENE MUTATED	CLINICAL FEATURES
Multiple endocrine neoplasia 1 (MEN 1)	<i>MEN1</i> (11q13)	Hyperparathyroidism Pancreatic neuroendocrine tumors Foregut carcinoids Adrenal adenomas Skin lesions Pituitary adenomas (40%)
Multiple endocrine neoplasia 4 (MEN 4)	<i>CDKNIB</i> (12p13)	Hyperparathyroidism Pituitary adenomas Other tumors
Carney complex	<i>PRKAR1A</i> (17q23-24)	Pituitary hyperplasia and adenomas (10%) Atrial myomas Schwannomas Adrenal hyperplasia Lentigines
Familial pituitary adenomas	<i>AIP</i> (11q13.2)	Acromegaly/gigantism (-15% of afflicted families)

Growth factors may also promote pituitary tumor proliferation. Basic fibroblast growth factor (bFGF) is abundant in the pituitary and stimulates pituitary cell mitogenesis, whereas epidermal growth factor receptor (EGFR) signaling induces both hormone synthesis and cell proliferation. Mutations of *USP8* may result in overexpressed EGFR in a subset of ACTH-secreting tumors. Other factors involved in initiation and promotion of pituitary tumors include loss of negative-feedback inhibition (as seen with primary hypothyroidism or hypogonadism) and estrogen-mediated or paracrine angiogenesis. Growth characteristics and neoplastic behavior also may be influenced by activated oncogenes, including *RAS* and pituitary tumor transforming gene (*PTTG*), or inactivation of growth suppressor genes, including *MEG3*. Pituitary adenomas exhibit lineage-specific features of cell-cycle disruption, including cellular senescence, with chromosomal instability and copy number alterations as well as elevated levels of CDK inhibitors. These features underlie the invariably benign nature of these adenomas.

**Genetic Syndromes Associated with Pituitary Tumors** Several familial syndromes are associated with pituitary tumors, and the genetic mechanisms for some of them have been unraveled (Table 380-4).

**Multiple endocrine neoplasia (MEN) 1** is an autosomal dominant syndrome characterized primarily by a genetic predisposition to parathyroid, pancreatic islet, and pituitary adenomas (Chap. 388). MEN 1 is caused by inactivating germline mutations in *MEN1*, a constitutively expressed tumor-suppressor gene located on chromosome 11q13. Loss of heterozygosity or a somatic mutation of the remaining normal *MEN1* allele leads to tumorigenesis. About half of affected patients develop prolactinomas; acromegaly and Cushing's disease are less commonly encountered.

**Carney complex** is characterized by spotty skin pigmentation, myomas, and endocrine tumors, including testicular, adrenal, and pituitary adenomas. Acromegaly occurs in ~20% of these patients. A subset of patients has mutations in the R1 regulatory subunit of protein kinase A (*PRKAR1A*).

**McCune-Albright syndrome** consists of polyostotic fibrous dysplasia, pigmented skin patches, and a variety of endocrine disorders, including acromegaly, adrenal adenomas, and autonomous ovarian function (Chap. 412). Hormonal hypersecretion results from constitutive cyclic AMP production caused by inactivation of the GTPase activity of  $G_s$ . The  $G_s$  mutations occur postzygotically, leading to a mosaic pattern of mutant expression.

**Familial acromegaly** is a rare disorder in which family members may manifest either acromegaly or gigantism. A subset of families with a predisposition for familial pituitary tumors, especially acromegaly,

has been found to harbor germline mutations in the *AIP* gene, which encodes the aryl hydrocarbon receptor interacting protein.

## HYPERPROLACTINEMIA

**Etiology** Hyperprolactinemia is the most common pituitary hormone hypersecretion syndrome in both men and women. PRL-secreting pituitary adenomas (prolactinomas) are the most common cause of PRL levels >200 µg/L (see below). Less pronounced PRL elevation can also be seen with microprolactinomas but is more commonly caused by drugs, pituitary stalk compression, hypothyroidism, or renal failure (Table 380-5).

Pregnancy and lactation are the important physiologic causes of hyperprolactinemia. Sleep-associated hyperprolactinemia reverts to normal within an hour of awakening. Nipple stimulation and sexual orgasm also may increase PRL. Chest wall stimulation or trauma (including chest surgery and herpes zoster) invokes the reflex suckling arc with resultant hyperprolactinemia. Chronic renal failure elevates PRL by decreasing peripheral clearance. Primary hypothyroidism is associated with mild hyperprolactinemia, probably because of compensatory TRH secretion. Mutation of the PRL receptor is a rare cause of hyperprolactinemia.

Lesions of the hypothalamic-pituitary region that disrupt hypothalamic dopamine synthesis, portal vessel delivery, or lactotropin responses are associated with hyperprolactinemia. Thus, hypothalamic tumors, cysts, infiltrative disorders, and radiation-induced damage cause elevated PRL levels, usually in the range of 30–100 µg/L. Plurihormonal adenomas (including GH and ACTH tumors) may hypersecrete PRL directly. Pituitary masses, including clinically non-functioning pituitary tumors, may compress the pituitary stalk to cause hyperprolactinemia.

Drug-induced inhibition or disruption of dopaminergic receptor function is a common cause of hyperprolactinemia (Table 380-5). Thus, antipsychotics and antidepressants are a relatively common cause of mild hyperprolactinemia. Most patients receiving risperidone have elevated PRL levels, sometimes exceeding 200 µg/L. Methyldopa inhibits dopamine synthesis, and verapamil blocks dopamine release, also leading to hyperprolactinemia. Hormonal agents that induce PRL include estrogens and thyrotropin-releasing hormone (TRH).

**Presentation and Diagnosis** Amenorrhea, galactorrhea, and infertility are the hallmarks of hyperprolactinemia in women. If hyperprolactinemia develops before menarche, primary amenorrhea results. More commonly, hyperprolactinemia develops later in life and leads to oligomenorrhea and ultimately to amenorrhea. If hyperprolactinemia is sustained, vertebral bone mineral density can be reduced compared with age-matched controls, particularly when it is associated with pronounced hypoestrogenemia. Galactorrhea is present in up to 80% of hyperprolactinemic women. Although usually bilateral and spontaneous, it may be unilateral or expressed only manually. Patients also may complain of decreased libido, weight gain, and mild hirsutism.

In men with hyperprolactinemia, diminished libido, infertility, and visual loss (from optic nerve compression) are the usual presenting symptoms. Gonadotropin suppression leads to reduced testosterone, impotence, and oligospermia. True galactorrhea is uncommon in men with hyperprolactinemia. If the disorder is long-standing, secondary effects of hypogonadism are evident, including osteopenia, reduced muscle mass, and decreased beard growth.

The diagnosis of idiopathic hyperprolactinemia is made by exclusion of known causes of hyperprolactinemia in the setting of a normal pituitary MRI. Some of these patients may harbor small microadenomas below visible MRI sensitivity (~2 mm).

## GALACTORRHEA

**Galactorrhea**, the inappropriate discharge of milk-containing fluid from the breast, is considered abnormal if it persists longer than 6 months after childbirth or discontinuation of breast-feeding. Postpartum galactorrhea associated with amenorrhea is a self-limiting disorder usually associated with moderately elevated PRL levels. Galactorrhea may occur spontaneously, or it may be elicited by nipple

**TABLE 380-5 Etiology of Hyperprolactinemia**

<b>I. Physiologic hypersecretion</b>
Pregnancy
Lactation
Chest wall stimulation
Sleep
Stress
<b>II. Hypothalamic-pituitary stalk damage</b>
Tumors
Craniopharyngioma
Suprasellar pituitary mass
Meningioma
Dysgerminoma
Metastases
Empty sella
Lymphocytic hypophysitis
Adenoma with stalk compression
Granulomas
Rathke's cyst
Irradiation
Trauma
Pituitary stalk section
Suprasellar surgery
<b>III. Pituitary hypersecretion</b>
Prolactinoma
Acromegaly
<b>IV. Systemic disorders</b>
Chronic renal failure
Hypothyroidism
Cirrhosis
Pseudocyesis
Epileptic seizures
<b>V. Drug-induced hypersecretion</b>
Dopamine receptor blockers
Atypical antipsychotics: risperidone
Phenothiazines: chlorpromazine, perphenazine
Butyrophenones: haloperidol
Thioxanthenes
Metoclopramide
Dopamine synthesis inhibitors
$\alpha$ -Methyldopa
Catecholamine depletors
Reserpine
Opiates
$H_2$ antagonists
Cimetidine, ranitidine
Imipramines
Amitriptyline, amoxapine
Serotonin reuptake inhibitors
Fluoxetine
Calcium channel blockers
Verapamil
Estrogens
Thyrotropin-releasing hormone

*Note:* Hyperprolactinemia >200 µg/L almost invariably is indicative of a prolactin-secreting pituitary adenoma. Physiologic causes, hypothyroidism, and drug-induced hyperprolactinemia should be excluded before extensive evaluation.

pressure. In both men and women, galactorrhea may vary in color and consistency (transparent, milky, or bloody) and arise either unilaterally or bilaterally. Mammography or ultrasound is indicated for bloody discharges (particularly from a single nipple), which may be caused by breast cancer. Galactorrhea is commonly associated with hyperprolactinemia caused by any of the conditions listed in Table

380-5. Acromegaly is associated with galactorrhea in about one-third of patients. Treatment of galactorrhea usually involves managing the underlying disorder (e.g., replacing T<sub>4</sub> for hypothyroidism, discontinuing a medication, treating prolactinoma).

**Laboratory Investigation** Basal, fasting morning PRL levels (normally <20 µg/L) should be measured to assess hypersecretion. Both false-positive and false-negative results may be encountered. In patients with markedly elevated PRL levels (>1000 µg/L), reported results may be falsely lowered because of assay artifacts; sample dilution is required to measure these high values accurately. Falsely elevated values may be caused by aggregated forms of circulating PRL, which are usually biologically inactive (macroprolactinemia). Hypothyroidism should be excluded by measuring TSH and T<sub>4</sub> levels.

## TREATMENT

### Hyperprolactinemia

Treatment of hyperprolactinemia depends on the cause of elevated PRL levels. Regardless of the etiology, however, treatment should be aimed at normalizing PRL levels to alleviate suppressive effects on gonadal function, halt galactorrhea, and preserve bone mineral density. Dopamine agonists are effective for most causes of hyperprolactinemia (see the treatment section for prolactinoma, below) regardless of the underlying cause.

If the patient is taking a medication known to cause hyperprolactinemia, the drug should be withdrawn, if possible. For psychiatric patients who require neuroleptic agents, supervised dose titration or the addition of a dopamine agonist can help restore normoprolactinemia and alleviate reproductive symptoms. However, dopamine agonists may worsen the underlying psychiatric condition, especially at high doses. Hyperprolactinemia usually resolves after adequate thyroid hormone replacement in hypothyroid patients or after renal transplantation in patients undergoing dialysis. Resection of hypothalamic or sellar mass lesions can reverse hyperprolactinemia caused by stalk compression and reduced dopamine tone. Granulomatous infiltrates occasionally respond to glucocorticoid administration. In patients with irreversible hypothalamic damage, no treatment may be warranted. In up to 30% of patients with hyperprolactinemia—usually without a visible pituitary microadenoma—the condition may resolve spontaneously.

## PROLACTINOMA

**Etiology and Prevalence** Tumors arising from lactotrope cells account for about half of all functioning pituitary tumors, with a population prevalence of ~10/100,000 in men and ~30/100,000 in women. Mixed tumors that secrete combinations of GH and PRL, ACTH and PRL, and rarely TSH and PRL are also seen. These plurihormonal tumors are usually recognized by immunohistochemistry, sometimes without apparent clinical manifestations from the production of additional hormones. Microadenomas are classified as <1 cm in diameter and usually do not invade the parasellar region. Macroadenomas are >1 cm in diameter and may be locally invasive and impinge on adjacent structures. The female-to-male ratio for microprolactinomas is 20:1, whereas the sex ratio is near 1:1 for macroadenomas. Tumor size generally correlates directly with PRL concentrations; values >250 µg/L usually are associated with macroadenomas. Men tend to present with larger tumors than women, possibly because the features of male hypogonadism are less readily evident. PRL levels remain stable in most patients, reflecting the slow growth of these tumors. About 5% of microadenomas progress in the long term to macroadenomas.

**Presentation and Diagnosis** Women usually present with amenorrhea, infertility, and galactorrhea. If the tumor extends outside the sella, visual field defects or other mass effects may be seen. Men often present with impotence, loss of libido, infertility, or signs of central nervous system (CNS) compression, including headaches and visual defects. Assuming that physiologic and medication-induced causes of

**2910** hyperprolactinemia are excluded (Table 380-5), the diagnosis of prolactinoma is likely with a PRL level  $>200 \mu\text{g/L}$ . PRL levels  $<100 \mu\text{g/L}$  may be caused by microadenomas, other sellar lesions that decrease dopamine inhibition, or nonneoplastic causes of hyperprolactinemia. For this reason, an MRI should be performed in all patients with hyperprolactinemia. It is important to remember that hyperprolactinemia caused secondarily by the mass effects of nonlactotrope lesions is also corrected by treatment with dopamine agonists despite failure to shrink the underlying mass. Consequently, PRL suppression by dopamine agonists does not necessarily indicate that the underlying lesion is a prolactinoma.

## TREATMENT

### Prolactinoma

Because microadenomas rarely progress to become macroadenomas, no treatment may be needed if patients are asymptomatic and fertility is not desired; these patients should be monitored by regular serial PRL measurements and MRI scans. For symptomatic microadenomas, therapeutic goals include control of hyperprolactinemia, reduction of tumor size, restoration of menses and fertility, and resolution of galactorrhea. Dopamine agonist doses should be titrated to achieve maximal PRL suppression and restoration of reproductive function (Fig. 380-5). A normalized PRL level does not ensure reduced tumor size. However, tumor shrinkage usually is not seen in those who do not respond with lowered PRL levels. For macroadenomas, formal visual field testing should be performed before initiating dopamine agonists. MRI and visual fields should be assessed at 6- to 12-month intervals until the mass shrinks and annually thereafter until maximum size reduction has occurred.

Oral dopamine agonists (cabergoline and bromocriptine) are the mainstay of therapy for patients with micro- or macroadenomas. Dopamine agonists suppress PRL secretion and synthesis as well as lactotrope cell proliferation. In patients with microadenomas who have achieved normoprolactinemia and significant reduction of tumor mass, the dopamine agonist may be withdrawn after 2 years. These patients should be monitored carefully for evidence of prolactinoma recurrence. About 20% of patients (especially males) are resistant to dopaminergic treatment; these adenomas may exhibit decreased  $D_2$  dopamine receptor numbers

or a postreceptor defect.  $D_2$  receptor gene mutations in the pituitary have not been reported.

**Cabergoline** An ergoline derivative, cabergoline is a long-acting dopamine agonist with high  $D_2$  receptor affinity. The drug effectively suppresses PRL for  $>14$  days after a single oral dose and induces prolactinoma shrinkage in most patients. Cabergoline (0.5–1.0 mg twice weekly) achieves normoprolactinemia and resumption of normal gonadal function in ~80% of patients with microadenomas; galactorrhea improves or resolves in 90% of patients. Cabergoline normalizes PRL and shrinks ~70% of macroadenomas. Mass effect symptoms, including headaches and visual disorders, usually improve dramatically within days after cabergoline initiation; improvement of sexual function requires several weeks of treatment but may occur before complete normalization of PRL levels. MRI should be repeated within 16 weeks after initial therapy of macroadenomas as shrinkage of invasive adenomas may be striking (Fig. 380-6). After initial control of PRL levels has been achieved, cabergoline should be reduced to the lowest effective maintenance dose. In ~5% of treated patients harboring a microadenoma, hyperprolactinemia may resolve and not recur when dopamine agonists are discontinued after long-term treatment. Cabergoline also may be effective in patients resistant to bromocriptine. Adverse effects and drug intolerance are encountered less commonly than with bromocriptine.

**Bromocriptine** The ergot alkaloid bromocriptine mesylate is a dopamine receptor agonist that suppresses PRL secretion. Because it is short-acting, the drug is preferred when pregnancy is desired. Therapy is initiated by administering a low bromocriptine dose (0.625–1.25 mg) at bedtime with a snack, followed by gradually increasing the dose. Most patients are controlled with a daily dose of  $<7.5$  mg (2.5 mg tid).

### SIDE EFFECTS

Side effects of dopamine agonists include constipation, nasal stuffiness, dry mouth, nightmares, insomnia, and vertigo; decreasing the dose usually alleviates these problems. Nausea, vomiting, and postural hypotension with faintness may occur in ~25% of patients after the initial dose. These symptoms may persist in some patients. In general, fewer side effects are reported with cabergoline. For the ~15% of patients who are intolerant of oral bromocriptine,

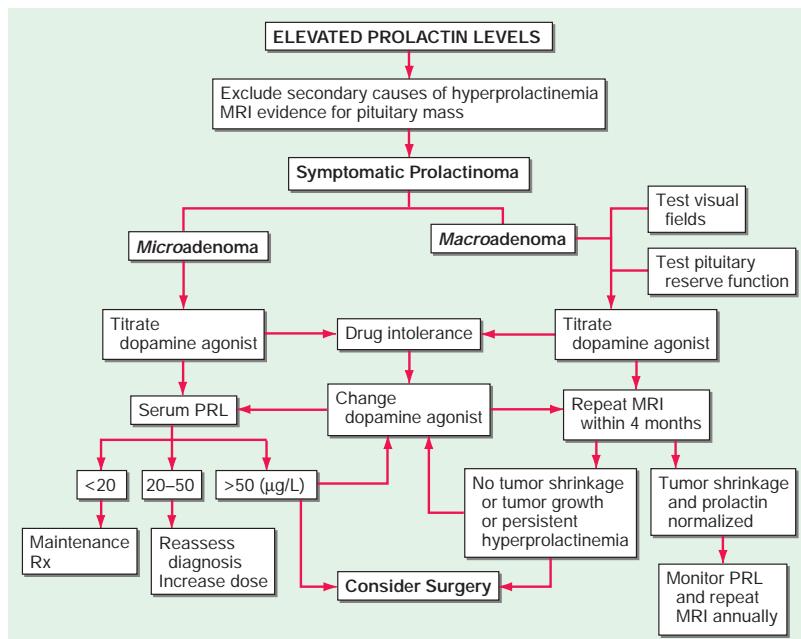
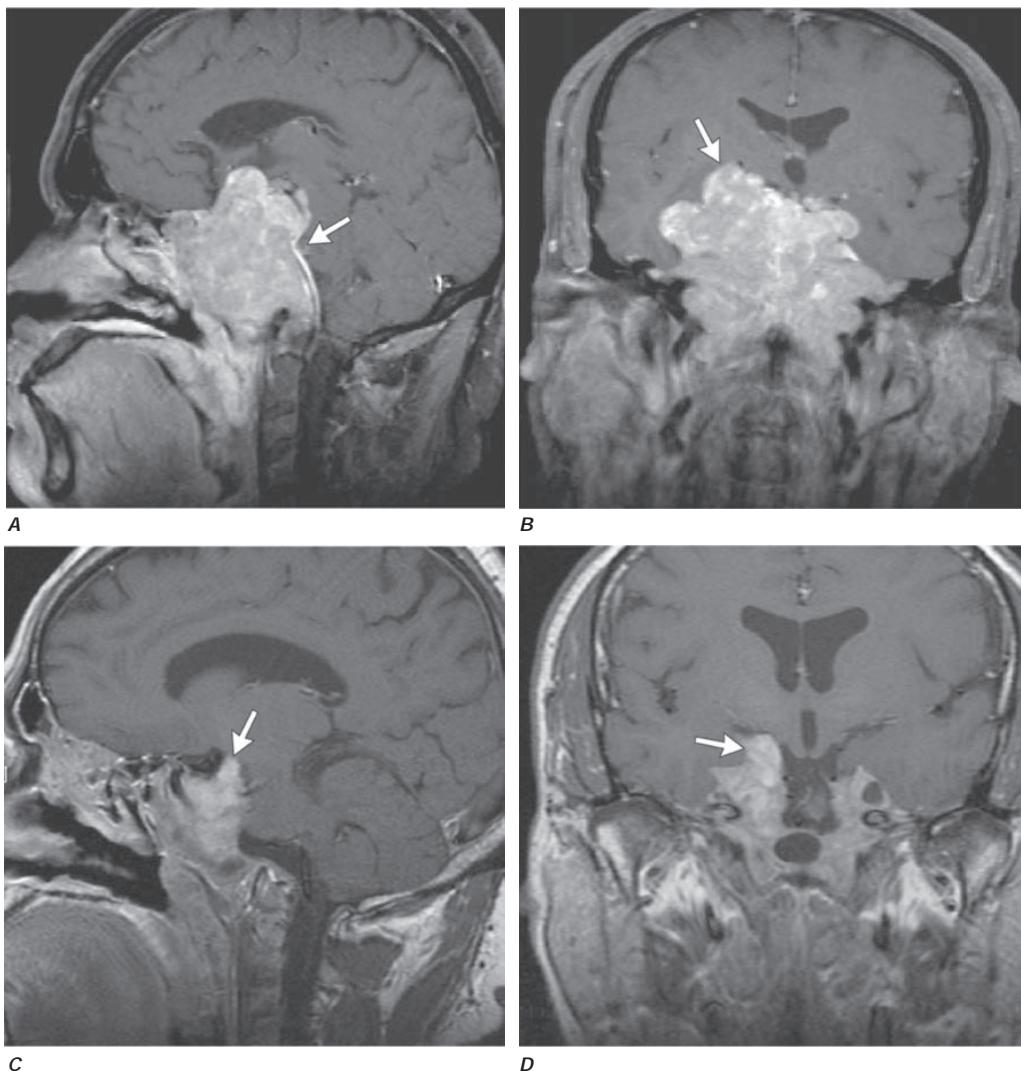


FIGURE 380-5 Management of prolactinoma. MRI, magnetic resonance imaging; PRL, prolactin.



**FIGURE 380-6** Large invasive prolactinoma successfully treated with cabergoline. **A–B.** Prolactin-secreting macroadenoma in a 32-year-old male measuring  $5.6 \times 6.9$  cm invading the skull base. PRL level was  $122,260 \mu\text{g/L}$ . Four days after cabergoline was started, PRL was  $10,823 \mu\text{g/L}$  and dropped to  $772 \mu\text{g/L}$  after 3 weeks. **C–D.** Substantial tumor regression after 40 months of treatment, with PRL levels stable at  $25 \mu\text{g/L}$ . (Reproduced with permission from M Ahmed, O Al-Nozha: *Images in clinical medicine: Large prolactinoma*. N Engl J Med 363:177, 2010.)

cabergoline may be better tolerated. Intravaginal administration of bromocriptine is often efficacious in patients with intractable gastrointestinal side effects. Auditory hallucinations, delusions, and mood swings have been reported in up to 5% of patients and may be due to the dopamine agonist properties or to the lysergic acid derivative of the compounds. Rare reports of leukopenia, thrombocytopenia, pleural fibrosis, cardiac arrhythmias, and hepatitis have been described. Patients with Parkinson disease who receive at least 3 mg of cabergoline daily have been reported to be at risk for development of cardiac valve regurgitation. Studies analyzing >500 prolactinoma patients receiving recommended doses of cabergoline (up to 2 mg weekly) have shown no evidence for an increased incidence of valvular disorders. Nevertheless, because no controlled prospective studies in pituitary tumor patients are available, it is prudent to perform echocardiograms before initiating standard-dose cabergoline therapy.

**Surgery** Rarely, surgical adenoma debulking may be indicated for dopamine resistance or intolerance as well as the presence of an invasive macroadenoma with compromised vision that fails to improve after drug treatment. Initial PRL normalization is achieved

in ~70% of microprolactinomas after surgical resection, but only 30% of macroadenomas can be resected successfully. Follow-up studies have shown that hyperprolactinemia recurs in up to 20% of patients within the first year after surgery; long-term recurrence rates may exceed 50% for macroadenomas. Radiotherapy for prolactinomas is reserved for patients with aggressive tumors that do not respond to maximally tolerated dopamine agonists and/or surgery.

#### PREGNANCY

The pituitary increases in size during pregnancy, reflecting the stimulatory effects of estrogen and perhaps other growth factors on pituitary vascularity and lactotrope cell hyperplasia. About 5% of microadenomas significantly increase in size, but 15–30% of macroadenomas grow during pregnancy. Bromocriptine has been used for >30 years to restore fertility in women with hyperprolactinemia, without evidence of teratogenic effects. Nonetheless, most authorities recommend strategies to minimize fetal exposure to the drug. For women taking bromocriptine who desire pregnancy, mechanical contraception should be used through three regular menstrual cycles to allow for conception timing. When pregnancy

is confirmed, bromocriptine should be discontinued and PRL levels followed serially, especially if headaches or visual symptoms occur. For women harboring macroadenomas, regular visual field testing is recommended, and the drug should be reinstated if tumor growth is apparent. Although pituitary MRI may be safe during pregnancy, this procedure should be reserved for symptomatic patients with severe headache and/or visual field defects. Surgical decompression may be indicated if vision is threatened. Although comprehensive data support the efficacy and relative safety of bromocriptine-facilitated fertility, patients should be advised of potential unknown deleterious effects and the risk of tumor growth during pregnancy. Because cabergoline is long-acting with a high D<sub>2</sub>-receptor affinity, it is not recommended for use in women when fertility is desired.

## ACROMEGALY

**Etiology** GH hypersecretion is usually the result of a somatotrope adenoma but may rarely be caused by extrapituitary lesions (Table 380-6). In addition to the more common GH-secreting somatotrope adenomas, mixed mammosomatotrope tumors and acidophilic stem cell adenomas secrete both GH and PRL. In patients with acidophilic stem cell adenomas, features of hyperprolactinemia (hypogonadism and galactorrhea) predominate over the less clinically evident signs of acromegaly. Occasionally, mixed plurihormonal tumors are encountered that also secrete ACTH, the glycoprotein hormone subunit, or TSH in addition to GH. Patients with partially empty sellae may present with GH hypersecretion due to a small GH-secreting adenoma within the compressed rim of pituitary tissue; some of these may reflect the spontaneous necrosis of tumors that were previously larger. GH-secreting tumors rarely arise from ectopic pituitary tissue remnants in the nasopharynx or midline sinuses.

There are case reports of ectopic GH secretion by tumors of pancreatic, ovarian, lung, or hematopoietic origin. Rarely, excess GHRH production may cause acromegaly because of chronic stimulation of somatotropes. These patients present with classic features of acromegaly, elevated GH levels, pituitary enlargement on MRI, and pathologic characteristics of pituitary hyperplasia. The most common cause of GHRH-mediated acromegaly is a chest or abdominal carcinoid tumor.

TABLE 380-6 Causes of Acromegaly

	PREVALENCE, %
<b>Excess Growth Hormone Secretion</b>	
Pituitary	98
Densely or sparsely granulated GH cell adenoma	60
Mixed GH cell and PRL cell adenoma	25
Mammosomatotrope cell adenoma	10
Plurihormonal adenoma	10
GH cell carcinoma or metastases	10
Multiple endocrine neoplasia 1 (GH cell adenoma)	10
McCune-Albright syndrome	10
Ectopic sphenoid or parapharyngeal sinus pituitary adenoma	10
Extrapituitary tumor	<1
Pancreatic islet cell tumor	10
Lymphoma	10
<b>Excess Growth Hormone-Releasing Hormone Secretion</b>	
Central	<1
Hypothalamic hamartoma, choristoma, ganglion neuroma	<1
Peripheral	<1
Bronchial carcinoid, pancreatic islet cell tumor, small-cell lung cancer, adrenal adenoma, medullary thyroid carcinoma, pheochromocytoma	<1

Abbreviations: GH, growth hormone; PRL, prolactin.

Source: Data from S Melmed: Medical progress: Acromegaly. N Engl J Med 355:2558, 2006.

Although these tumors usually express positive GHRH immunoreactivity, clinical features of acromegaly are evident in only a minority of patients with carcinoid disease. Excessive GHRH also may be elaborated by hypothalamic tumors, usually choristomas or neuroomas.

**Presentation and Diagnosis** Protean manifestations of GH and IGF-1 hypersecretion are indolent and often are not clinically diagnosed for 10 years or more. Acral bony overgrowth results in frontal bossing, increased hand and foot size, mandibular enlargement with prognathism, and widened space between the lower incisor teeth. In children and adolescents, initiation of GH hypersecretion before epiphyseal long bone closure is associated with development of pituitary gigantism (Fig. 380-7). Soft tissue swelling results in increased heel pad thickness, increased shoe or glove size, ring tightening, characteristic coarse facial features, and a large fleshy nose. Other commonly encountered clinical features include hyperhidrosis, a deep and hollow-sounding voice, oily skin, arthropathy, kyphosis, carpal tunnel syndrome, proximal muscle weakness and fatigue, acanthosis nigricans, and skin tags. Generalized visceromegaly occurs, including cardiomegaly, macroglossia, and thyroid gland enlargement.

The most significant clinical impact of GH excess occurs with respect to the cardiovascular system. Cardiomyopathy with arrhythmias, left ventricular hypertrophy, decreased diastolic function, and hypertension ultimately occur in most patients if untreated. Upper airway obstruction with sleep apnea occurs in >60% of patients and is associated with both soft tissue laryngeal airway obstruction and central sleep dysfunction. Diabetes mellitus develops in 25% of patients with acromegaly, and most patients are intolerant of a glucose load (as GH counteracts the action of insulin). Acromegaly is associated with an increased risk of colon polyps and mortality from colonic malignancy; polyps are diagnosed in up to one-third of patients. Overall mortality is increased about threefold and is due primarily to cardiovascular and cerebrovascular disorders and respiratory disease. Unless GH levels are controlled, survival is reduced by an average of 10 years compared with an age-matched control population.

**Laboratory Investigation** Age-matched serum IGF-1 levels are elevated in acromegaly. Consequently, an IGF-1 level provides a useful laboratory screening measure when clinical features raise the possibility of acromegaly. Owing to the pulsatility of GH secretion, measurement of a single random GH level is not useful for the diagnosis or exclusion of acromegaly and does not correlate with disease severity. The diagnosis of acromegaly is confirmed by demonstrating the failure of GH suppression to <0.4 µg/L within 1–2 h of an oral glucose load (75 g). When ultrasensitive GH assays are used, normal nadir GH levels are even lower (<0.05 µg/L). About 20% of patients exhibit a paradoxical GH rise after glucose. PRL should be measured, as it is elevated in ~25% of patients with acromegaly. Thyroid function, gonadotropins, and sex steroids may be attenuated because of tumor mass effects. Because most patients will undergo surgery with glucocorticoid coverage, tests of ACTH reserve in asymptomatic patients are more efficiently deferred until after surgery.

## TREATMENT

### Acromegaly

The goal of treatment is to control GH and IGF-1 hypersecretion, ablate or arrest tumor growth, ameliorate comorbidities, restore mortality rates to normal, and preserve pituitary function.

Surgical resection of GH-secreting adenomas is the initial treatment for most patients (Fig. 380-8). SRLs are used as adjuvant treatment for preoperative shrinkage of large invasive macroadenomas, immediate relief of debilitating symptoms, and reduction of GH hypersecretion; in frail patients experiencing morbidity; and in patients who decline surgery or when surgery fails to achieve biochemical control. Irradiation or repeat surgery may be required for patients who cannot tolerate or do not respond to adjunctive medical therapy. The high rate of late hypopituitarism and the slow rate (5–15 years) of biochemical response are the main disadvantages of



**FIGURE 380-7 Features of acromegaly/gigantism.** A 22-year-old man with gigantism due to excess growth hormone is shown to the left of his identical twin. The increased height and prognathism (**A**) and enlarged hand (**B**) and foot (**C**) of the affected twin are apparent. Their clinical features began to diverge at the age of ~13 years. (Reproduced with permission from RF Gagel, IE McCutcheon. *Images in clinical medicine. Pituitary gigantism*. N Engl J Med 340:524, 1999.)

radiotherapy. Irradiation is also relatively ineffective in normalizing IGF-1 levels. Stereotactic ablation of GH-secreting adenomas by Gamma Knife radiotherapy is promising, but long-term results and side effects appear similar to those observed with conventional radiation. SRLs may be required while awaiting the full benefits of radiotherapy. Systemic comorbid sequelae of acromegaly, including cardiovascular disease, diabetes, and arthritis, should be managed aggressively. Mandibular surgical repair may be indicated.

### SURGERY

Transsphenoidal surgical resection by an experienced surgeon is the preferred primary treatment for both microadenomas (remission rate ~70%) and macroadenomas (<50% in remission). Soft tissue swelling improves immediately after tumor resection. GH levels return to normal within an hour, and IGF-1 levels are normalized within 3–4 days. In ~10% of patients, acromegaly may recur several years after apparently successful surgery; hypopituitarism develops in up to 15% of patients after surgery.

### SOMATOSTATIN RECEPTOR LIGANDS

SRLs exert their therapeutic effects through SST<sub>2</sub> and SST<sub>5</sub> receptor subtypes, both expressed by GH-secreting tumors.

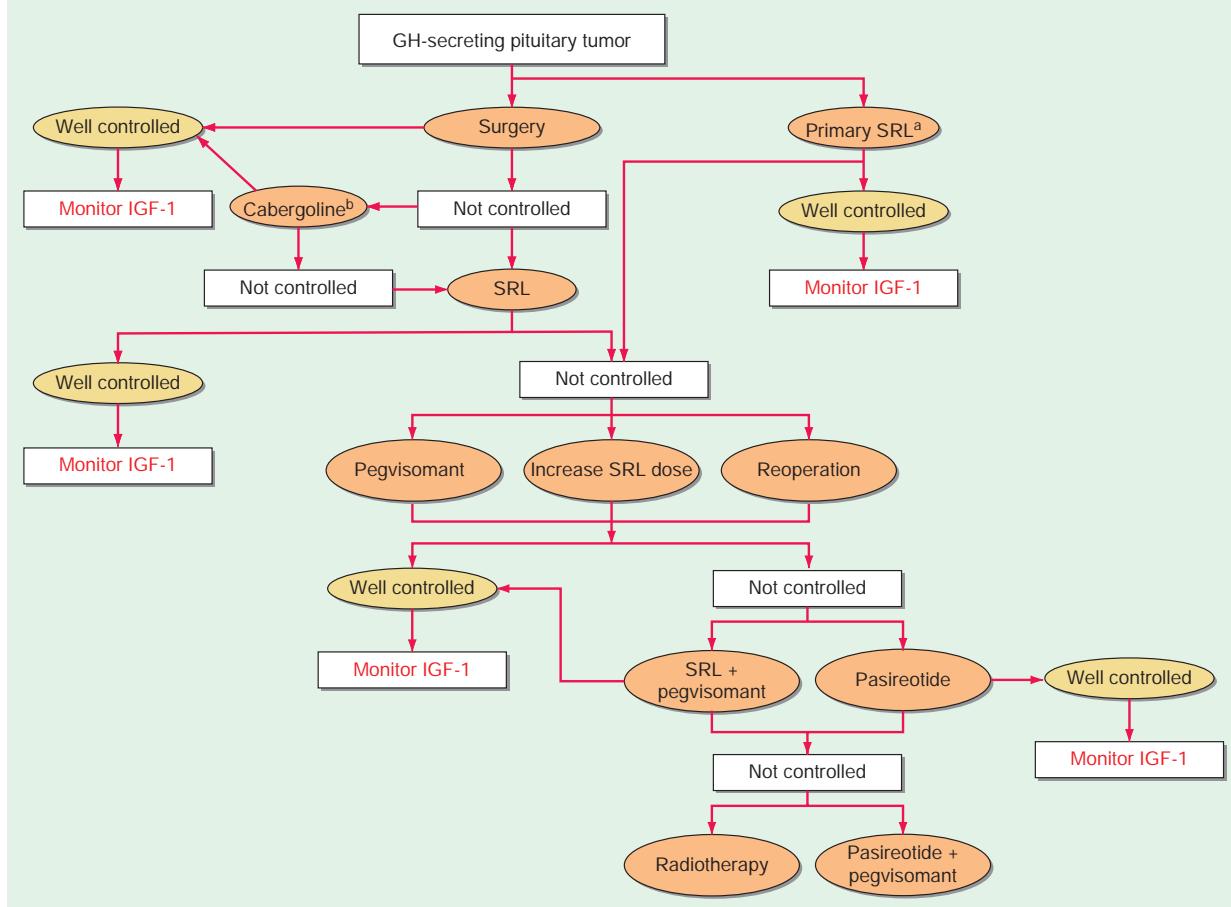
The preferred medical treatments for patients with acromegaly include long-acting injectable SRL depot formulations of octreotide and lanreotide as well as oral octreotide capsules. Although responses vary widely in individual patients, meta-analyses indicate that GH and IGF-1 levels are normalized in ~50% of patients. Octreotide acetate is an eight-amino-acid synthetic somatostatin analogue. In contrast to native somatostatin, the analogue is relatively resistant to plasma degradation. It has a 2-h serum half-life and possesses 40-fold greater potency than native somatostatin to suppress GH. *Octreotide LAR* is a sustained-release, long-acting formulation of octreotide incorporated into microspheres that sustain drug levels for several weeks after intramuscular injection. GH suppression occurs for as long as 6 weeks after a 30-mg intramuscular injection; long-term monthly treatment sustains GH and IGF-1 suppression and also reduces pituitary tumor size in

~50% of patients. *Lanreotide*, in a slow-release depot SRL preparation, is a cyclic somatostatin octapeptide analogue that suppresses GH and IGF-1 hypersecretion after a 60-mg subcutaneous injection. Long-term (every 4–6 weeks) administration controls GH hypersecretion in about two-thirds of treated patients and improves patient compliance because of the long interval required between drug injections. *Oral octreotide capsules* (40–80 mg daily) maintain biochemical control in patients previously maintained on injectable formulations. Rapid relief of headache and soft tissue swelling occurs in ~75% of patients within days to weeks of SRL initiation. Most patients report symptomatic improvement, including amelioration of headache, perspiration, obstructive apnea, and cardiac failure. *Pasireotide LAR*, a multireceptor ligand with preferential SST<sub>5</sub> binding (see below), has been shown to exhibit efficacy in achieving biochemical control in patients resistant to octreotide or lanreotide preparations.

**Side Effects** SRLs are well tolerated in most patients. Adverse effects are similar for injectable octreotide and lanreotide as well as for oral octreotide formulation. They are short-lived and mostly relate to drug-induced suppression of gastrointestinal motility and secretion. Transient nausea, abdominal discomfort, fat malabsorption, diarrhea, and flatulence occur in one-third of patients, and these symptoms usually remit within 2 weeks. Gallbladder contractility and emptying are attenuated; up to 30% of patients develop long-term echogenic sludge or asymptomatic cholesterol gallstones. Other side effects include mild glucose intolerance due to transient insulin suppression, asymptomatic bradycardia, hypothyroxinemia, and local injection site discomfort. Pasireotide is associated with similar gastrointestinal side effects but with a higher prevalence of glucose intolerance and new-onset diabetes mellitus.

### GH RECEPTOR ANTAGONIST

Pegvisomant antagonizes endogenous GH action by blocking peripheral GH binding to its receptor. Consequently, serum IGF-1 levels are suppressed, reducing the deleterious effects of excess endogenous GH. Pegvisomant is administered by daily



**FIGURE 380-8 Management of acromegaly.** <sup>a</sup>If curative surgery is not feasible. <sup>b</sup>Consider in cases of mild postoperative GH/IGF-1 elevations. GH, growth hormone; IGF, insulin-like growth factor; SRL, somatostatin receptor ligand (octreotide or lanreotide).

subcutaneous injection (10–30 mg) and normalizes IGF-1 in ~70% of patients. GH levels, however, remain elevated as the drug does not target the pituitary adenoma. Side effects include reversible liver enzyme elevation, lipodystrophy, and injection site pain. Tumor size should be monitored by MRI.

Combined treatment with monthly SRLs and weekly or biweekly pegvisomant injections has been used effectively in resistant patients.

#### DOPAMINE AGONISTS

Very high doses of cabergoline (0.5 mg/d) may achieve short-lived and modest GH therapeutic efficacy. Combined treatment with octreotide and cabergoline may induce additive biochemical control compared with either drug alone.

#### RADIATION THERAPY

External radiation therapy or high-energy stereotactic techniques are used as adjuvant therapy for acromegaly. An advantage of radiation is that patient compliance with long-term treatment is not required. Tumor mass is reduced, and GH levels are attenuated over time. However, 50% of patients require at least 8 years for GH levels to be suppressed to <5 µg/L; this level of GH reduction is achieved in ~90% of patients after 18 years but represents suboptimal GH suppression. Patients may require interim medical therapy for several years before attaining maximal radiation benefits. Most patients also experience hypothalamic-pituitary damage, leading to gonadotropin, ACTH, and/or TSH deficiency within 10 years of therapy.

#### SUMMARY

Surgery is the preferred primary treatment for GH-secreting microadenomas (Fig. 380-8). The high frequency of residual GH hypersecretion after macroadenoma resection usually necessitates adjuvant or primary medical therapy for these larger tumors. Patients unable to receive or respond to unimodal medical treatment may benefit from combined treatments, or they can be offered radiation. Very rarely, repeat surgery may be required.

#### CUSHING'S DISEASE (ACTH-PRODUCING ADENOMA) (See also Chap. 386)

**Etiology and Prevalence** Pituitary corticotrope adenomas (Cushing's disease) account for 70% of patients with endogenous causes of Cushing's syndrome. However, it should be emphasized that iatrogenic hypercortisolism is the most common cause of cushingoid features. Ectopic tumor ACTH production, cortisol-producing adrenal adenomas, adrenal carcinoma, and adrenal hyperplasia account for the other causes; rarely, ectopic tumor CRH production is encountered.

ACTH-producing adenomas account for ~10–15% of all pituitary tumors. Because the clinical features of Cushing's syndrome often lead to early diagnosis, most ACTH-producing pituitary tumors are relatively small microadenomas. However, macroadenomas also are seen and some ACTH-expressing adenomas are clinically silent. Cushing's disease is 5–10 times more common in women than in men. These pituitary adenomas exhibit unrestrained ACTH secretion, with

resultant hypercortisolemia. However, they retain partial suppressibility in the presence of high doses of administered glucocorticoids, providing the basis for dynamic testing to distinguish pituitary from nonpituitary causes of Cushing's syndrome.

**Presentation and Diagnosis** The diagnosis of Cushing's syndrome presents two great challenges: (1) to distinguish patients with pathologic cortisol excess from those with physiologic or other disturbances of cortisol production and (2) to determine the etiology of pathologic cortisol excess.

Typical features of chronic cortisol excess include thin skin, central obesity, hypertension, plethoric moon facies, purple striae and easy bruising, glucose intolerance or diabetes mellitus, gonadal dysfunction, osteoporosis, proximal muscle weakness, signs of hyperandrogenism (acne, hirsutism), and psychological disturbances (depression, mania, and psychoses) (Table 380-7). Hematopoietic features of hypercortisolism include leukocytosis, lymphopenia, and eosinopenia. Immune suppression includes delayed hypersensitivity and infection propensity. These protean yet commonly encountered manifestations of hypercortisolism make it challenging to decide which patients mandate formal laboratory evaluation. Certain features make pathologic causes of hypercortisolism more likely; they include characteristic central redistribution of fat, thin skin with striae and bruising, and proximal muscle weakness. In children and young females, early osteoporosis may be particularly prominent. The primary cause of death is cardiovascular disease, but life-threatening infections and risk of suicide are also increased.

Rapid development of features of hypercortisolism associated with skin hyperpigmentation and severe myopathy suggests an ectopic tumor source of ACTH. Hypertension, hypokalemic alkalosis, glucose intolerance, and edema are also more pronounced in these patients. Serum potassium levels <3.3 mmol/L are evident in ~70% of patients with ectopic ACTH secretion but are seen in <10% of patients with pituitary-dependent Cushing's syndrome.

**Laboratory Investigation** The diagnosis of Cushing's disease is based on laboratory documentation of endogenous hypercortisolism. Measurement of 24-h UFC is a precise and cost-effective screening test. Alternatively, the failure to suppress plasma cortisol after an overnight 1-mg dexamethasone suppression test can be used to identify patients with hypercortisolism. As nadir levels of cortisol occur

**TABLE 380-8 Differential Diagnosis of ACTH-Dependent Cushing's Syndrome<sup>a</sup>**

	ACTH-SECRETING PITUITARY TUMOR	ECTOPIC ACTH SECRETION
Etiology	Pituitary corticotrope adenoma Plurihormonal adenoma	Bronchial, abdominal carcinoid Small-cell lung cancer Thymoma, other sources
Sex	F > M	M > F
Clinical features	Slow onset	Rapid onset Pigmentation Severe myopathy
Serum potassium <3.3 µg/dL	<10%	75%
24-h UFC	High	High
Basal ACTH level	Inappropriately high	Very high
Dexamethasone suppression		
1 mg overnight		
Low-dose (0.5 mg q6h)	Cortisol >5 µg/dL	Cortisol >5 µg/dL
High-dose (2 mg q6h)	Cortisol <5 µg/dL	Cortisol >5 µg/dL
UFC >80% suppressed	Microadenomas: 90% Macroadenomas: 50%	10%
Inferior petrosal sinus sampling (IPSS)		
Basal		
IPSS: peripheral	>2	<2
CRH-induced		
IPSS: peripheral	>3	<3

<sup>a</sup>ACTH-independent causes of Cushing's syndrome are diagnosed by suppressed ACTH levels and an adrenal mass in the setting of hypercortisolism. Iatrogenic Cushing's syndrome is excluded by history.

Abbreviations: ACTH, adrenocorticotrophic hormone; CRH, corticotropin-releasing hormone; F, female; M, male; UFC, urinary free cortisol.

at night, elevated midnight serum or salivary samples of cortisol are suggestive of Cushing's disease. Basal plasma ACTH levels often distinguish patients with ACTH-independent (adrenal or exogenous glucocorticoid) from those with ACTH-dependent (pituitary, ectopic ACTH) Cushing's syndrome. Mean basal ACTH levels are about eightfold higher in patients with ectopic ACTH secretion than in those with pituitary ACTH-secreting adenomas. However, extensive overlap of ACTH levels in these two disorders precludes using ACTH measurements to make the distinction. Preferably, dynamic testing based on differential sensitivity to glucocorticoid feedback or ACTH stimulation in response to CRH or cortisol reduction is used to distinguish ectopic from pituitary sources of excess ACTH (Table 380-8). Very rarely, circulating CRH levels are elevated, reflecting ectopic tumor-derived secretion of CRH and often ACTH. **For further discussion of dynamic testing for Cushing's syndrome, see Chap. 386.**

Most ACTH-secreting pituitary tumors are <5 mm in diameter, and about half are undetectable by sensitive MRI. The high prevalence of incidental pituitary microadenomas diminishes the ability to distinguish ACTH-secreting pituitary tumors accurately from nonsecreting incidentalomas.

**Inferior Petrosal Venous Sampling** Because pituitary MRI with gadolinium enhancement is insufficiently sensitive to detect small (<2 mm) pituitary ACTH-secreting adenomas, bilateral inferior petrosal sinus ACTH sampling before and after CRH administration may be required to distinguish these lesions from ectopic ACTH-secreting tumors that may have similar clinical and biochemical characteristics. Simultaneous assessment of ACTH in each inferior petrosal vein and in the peripheral circulation provides a strategy for confirming and localizing pituitary ACTH production. Sampling is performed at baseline and 2, 5, and 10 min after intravenous bovine CRH (1 µg/kg) injection.

**TABLE 380-7 Clinical Features of Cushing's Syndrome (All Ages)**

SYMPTOMS/SIGNS	FREQUENCY, %
Obesity or weight gain (>115% ideal body weight)	80
Thin skin	80
Moon facies	75
Hypertension	75
Purple skin striae	65
Hirsutism	65
Menstrual disorders (usually amenorrhea)	60
Plthora	60
Abnormal glucose tolerance	55
Impotence	55
Proximal muscle weakness	50
Truncal obesity	50
Acne	45
Bruising	45
Mental changes	45
Osteoporosis	40
Edema of lower extremities	30
Hyperpigmentation	20
Hypokalemic alkalosis	15
Diabetes mellitus	15

Source: Adapted with permission from MA Magiokou et al, in Wierman ME: Diseases of the Pituitary. Totowa, NJ: Humana; 1997.

**2916** An increased ratio ( $>2$ ) of inferior petrosal:peripheral vein ACTH confirms pituitary Cushing's syndrome. After CRH injection, peak petrosal:peripheral ACTH ratios  $>3$  confirm the presence of a pituitary ACTH-secreting tumor. The sensitivity of this test is  $>95\%$ , with very rare false-positive results. False-negative results may be encountered in patients with aberrant venous drainage. Petrosal sinus catheterizations are technically difficult, and  $\sim 0.05\%$  of patients develop neurovascular complications. The procedure should not be performed in patients with hypertension, in patients with known cerebrovascular disease, or in the presence of a well-visualized pituitary adenoma on MRI.

## TREATMENT

### Cushing's Disease

Selective transsphenoidal resection is the treatment of choice for Cushing's disease (Fig. 380-9). The remission rate for this procedure is  $\sim 80\%$  for microadenomas but  $<50\%$  for macroadenomas. However, surgery is rarely successful when the adenoma is not visible on MRI. After successful tumor resection, most patients experience a postoperative period of symptomatic ACTH deficiency that may last up to 12 months. This usually requires low-dose cortisol replacement, as patients experience steroid withdrawal symptoms and have a suppressed hypothalamic-pituitary-adrenal axis. Biochemical recurrence occurs in  $\sim 5\%$  of patients in whom surgery was initially successful. As persistent hypercortisolism may cause blood clotting defects, prophylactic postoperative thromboembolic management has been advocated for vulnerable patients.

When initial surgery is unsuccessful, repeat surgery is sometimes indicated, particularly when a pituitary source for ACTH is well documented. In older patients, in whom issues of growth and fertility are less important, hemi- or total hypophysectomy may be necessary if a discrete pituitary adenoma is not recognized. Pituitary irradiation may be used after unsuccessful surgery, but it cures only  $\sim 15\%$  of patients. Because the effects of radiation are slow and only partially effective in adults, adrenal-targeted steroidogenic inhibitors are used in combination with pituitary irradiation to block adrenal responses to persistently high ACTH levels.

*Pasireotide LAR* 10–40 mg intramuscularly, an SRL with high affinity for SST<sub>5</sub> > SST<sub>2</sub> receptor subtypes, may control hypercortisolism in a subset of patients with ACTH-secreting pituitary tumors when surgery is not an option or has not been successful.

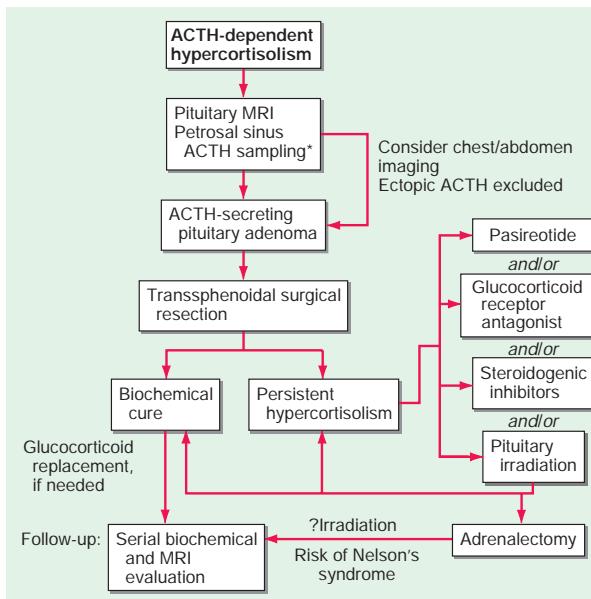


FIGURE 380-9 Management of Cushing's disease. ACTH, adrenocorticotropin hormone; MRI, magnetic resonance imaging; \*, Not usually required.

The drug lowers plasma ACTH levels and normalizes 24-h UFC levels in  $\sim 20\%$  of patients, and up to  $40\%$  of patients may experience pituitary tumor shrinkage. Side effects are similar to those encountered for other SRLs and include transient abdominal discomfort, diarrhea, nausea, and gallstones (20% of patients). Notably, hyperglycemia and new-onset diabetes develop in up to 70% of patients, likely due to suppressed pancreatic secretion of insulin and incretins. Because patients with hypercortisolism are insulin-resistant, hyperglycemia should be rigorously managed. The drug requires consistent long-term administration.

*Osilodrostat* (2 mg twice daily titrated up to 30 mg twice daily), an oral 11'-hydroxylase inhibitor that blocks adrenal gland cortisol biosynthesis, normalized 24-h UFC in 86% of patients. Mild, mostly transient gastrointestinal symptoms are common. Patients should be closely monitored for development of hypocortisolism and adrenal insufficiency. Elevated adrenal hormone precursors may lead to hypokalemia and hypertension. QTc prolongation and possibly increased tumor volume are also reported.

*Ketoconazole*, an imidazole derivative antimycotic agent, inhibits several P450 enzymes and effectively lowers cortisol in most patients with Cushing's disease when administered twice daily (600–1200 mg/d). Elevated hepatic transaminases, gynecomastia, impotence, gastrointestinal upset, and edema are common side effects.

*Mifepristone* (300–1200 mg/d), a glucocorticoid receptor antagonist, blocks peripheral cortisol action and is approved to treat hyperglycemia in Cushing's disease. Because the drug does not target the pituitary tumor, both ACTH and cortisol levels remain elevated, thus obviating a reliable circulating biomarker. Side effects are largely due to general antagonism of other steroid hormones and include hypokalemia, endometrial hyperplasia, hypoadrenalinism, and hypertension.

*Metyrapone* (2–4 g/d) inhibits 11'-hydroxylase activity and normalizes plasma cortisol in up to 75% of patients. Side effects include nausea and vomiting, rash, and exacerbation of acne or hirsutism. *Mitotane* (3–6 g/d orally in four divided doses) suppresses cortisol hypersecretion by inhibiting 11'-hydroxylase and cholesterol side-chain cleavage enzymes and by destroying adrenocortical cells. Side effects of mitotane include gastrointestinal symptoms, dizziness, gynecomastia, hyperlipidemia, skin rash, and hepatic enzyme elevation. It also may lead to hypoaldosteronism. Other agents include *aminoglutethimide* (250 mg tid), *trilostane* (200–1000 mg/d), *cypionate* (24 mg/d), and IV *etomidate* (0.3 mg/kg per h). Glucocorticoid insufficiency is a potential side effect of agents used to block steroidogenesis.

The use of steroidogenic inhibitors has decreased the need for bilateral adrenalectomy. Surgical removal of both adrenal glands corrects hypercortisolism but may be associated with significant morbidity rates and necessitates permanent glucocorticoid and mineralocorticoid replacement. Adrenalectomy in the setting of residual corticotropin adenoma tissue predisposes to the development of *Nelson's syndrome*, a disorder characterized by rapid pituitary tumor enlargement and increased pigmentation secondary to high ACTH levels. Prophylactic radiation therapy may be indicated to prevent the development of Nelson's syndrome after adrenalectomy.

## NONFUNCTIONING AND GONADOTROPIN-PRODUCING PITUITARY ADENOMAS

**Etiology and Prevalence** Nonfunctioning pituitary adenomas include those that secrete little or no pituitary hormones into the systemic circulation, as well as tumors that produce too little hormone to result in recognizable clinical features. They are the most common type of pituitary adenoma and are usually macroadenomas at the time of diagnosis because clinical features are not apparent until tumor mass effects occur. Based on immunohistochemistry, most clinically nonfunctioning adenomas can be shown to originate from gonadotrope cells or from pituitary null cells. These tumors typically produce small

amounts of intact gonadotropins (usually FSH) as well as uncombined LH and FSH subunits. Tumor secretion may lead to elevated and FSH subunits and, very rarely, to increased LH subunit levels. Some adenomas express subunits without FSH or LH. A TRH stimulation test often induces an atypical increase of tumor-derived gonadotropins or subunits.

**Presentation and Diagnosis** Clinically nonfunctioning tumors often present with optic chiasm pressure and other symptoms of local expansion or may be incidentally discovered on an MRI performed for another indication (incidentaloma). Rarely, menstrual disturbances or ovarian hyperstimulation occur in women with large tumors that produce FSH and LH. In these cases, ovaries may have features that resemble polycystic ovarian syndrome and may produce very high levels of estrogen. More commonly, adenoma compression of the pituitary stalk or surrounding pituitary tissue leads to attenuated LH and features of hypogonadism. PRL levels are usually slightly increased, also because of stalk compression. It is important to distinguish this circumstance from true prolactinomas, as nonfunctioning tumors do not shrink in response to treatment with dopamine agonists.

**Laboratory Investigation** The goal of laboratory testing in clinically nonfunctioning tumors is to classify the type of tumor, identify hormonal markers of tumor activity, and detect possible hypopituitarism. Free subunit levels may be elevated in 10–15% of patients with nonfunctioning tumors. In female patients, peri- or postmenopausal basal FSH concentrations are difficult to distinguish from tumor-derived FSH elevation. Premenopausal women have cycling FSH levels, also preventing clear-cut diagnostic distinction from tumor-derived FSH. In men, gonadotropin-secreting tumors may be diagnosed because of slightly increased gonadotropins ( $\text{FSH} > \text{LH}$ ) in the setting of a pituitary mass. Testosterone levels are usually low despite the normal or increased LH level, perhaps reflecting reduced LH bioactivity or the loss of normal LH pulsatility. Because this pattern of hormone test results is also seen in primary gonadal failure and, to some extent, with aging (Chap. 391), the finding of increased gonadotropins alone is insufficient for the diagnosis of a gonadotropin-secreting tumor. In the majority of patients with gonadotrope adenomas, TRH administration stimulates LH subunit secretion; this response is not seen in normal individuals. GnRH testing, however, is not helpful for making the diagnosis. For nonfunctioning and gonadotropin-secreting tumors, the diagnosis usually rests on immunohistochemical analyses of surgically resected tumor tissue, as the mass effects of these tumors usually necessitate resection.

Although acromegaly or Cushing's disease usually presents with unique clinical features, clinically inapparent (silent) somatotrope or corticotrope adenomas may only be diagnosed by immunostaining of resected tumor tissue. These silent tumors usually grow more aggressively and account for up to 20% of all nonfunctioning adenomas. If PRL levels are  $<100 \mu\text{g/L}$  in a patient harboring a pituitary mass, a nonfunctioning adenoma causing pituitary stalk compression should be considered.

## TREATMENT

### Nonfunctioning and Gonadotropin-Producing Pituitary Adenomas

Asymptomatic small nonfunctioning microadenomas with no threat to vision may be followed with regular MRI and visual field testing without immediate intervention. However, for macroadenomas, transsphenoidal surgery is indicated to reduce tumor size and relieve mass effects (Fig. 380-10). Although it is not usually possible to remove all adenoma tissue surgically, vision improves in 70% of patients with preoperative visual field defects. Preexisting hypopituitarism that results from tumor mass effects may improve or resolve completely. Beginning ~6 months postoperatively, MRI scans should be performed yearly to detect tumor regrowth. Within 5–6 years after successful surgical resection, ~15% of nonfunctioning tumors recur. When substantial tumor remains after transsphenoidal surgery, adjuvant radiotherapy may be indicated to prevent tumor regrowth. Radiotherapy may be deferred if no postoperative residual mass is evident. Nonfunctioning pituitary tumors respond poorly to dopamine agonist treatment, and SRLs are largely ineffective for shrinking these tumors. The selective GnRH antagonist Nal-Glu GnRH suppresses FSH hypersecretion but has no effect on adenoma size.

### TSH-SECRETING ADENOMAS

TSH-producing macroadenomas are very rare but are often large and locally invasive when they occur. Patients usually present with thyroid goiter and hyperthyroidism, reflecting chronic overproduction of TSH. Diagnosis is based on demonstrating elevated serum free  $\text{T}_4$  levels, inappropriately normal or high TSH secretion, and MRI evidence of a pituitary adenoma. Elevated free glycoprotein hormone subunits are seen in many patients.

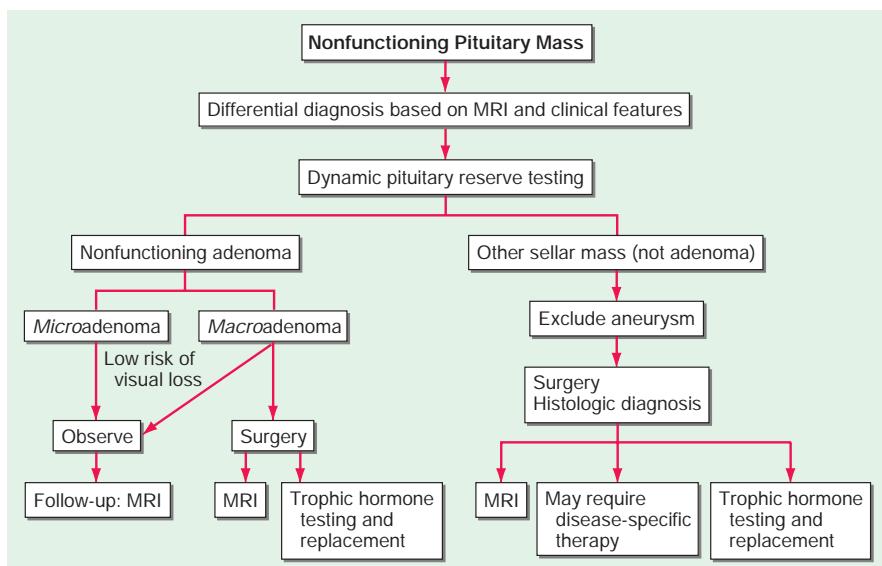


FIGURE 380-10 Management of a nonfunctioning pituitary mass. MRI, magnetic resonance imaging.

It is important to exclude other causes of inappropriate TSH secretion, such as resistance to thyroid hormone, an autosomal dominant disorder caused by mutations in the thyroid hormone receptor (**Chap. 382**). The presence of a pituitary mass and elevated subunit levels are suggestive of a TSH-secreting tumor. Dysalbuminemic hyperthyroxinemia syndromes, caused by mutations in serum thyroid hormone binding proteins, are also characterized by elevated thyroid hormone levels, but with normal rather than suppressed TSH levels. Moreover, free thyroid hormone levels are normal in these disorders, most of which are familial.

## TREATMENT

### TSH-Secreting Adenomas

The initial therapeutic approach is to remove or debulk the tumor mass surgically, usually using a transsphenoidal approach. Total resection is not often achieved as most of these adenomas are large and locally invasive. Normal circulating thyroid hormone levels are achieved in about two-thirds of patients after surgery. Thyroid ablation or antithyroid drugs (methimazole and propylthiouracil) can be used to reduce thyroid hormone levels. SRL treatment effectively normalizes TSH and subunit hypersecretion, shrinks the tumor mass in 50% of patients, and improves visual fields in 75% of patients; euthyroidism is restored in most patients. Because SRLs markedly suppress TSH, biochemical hypothyroidism often requires concomitant thyroid hormone replacement, which may also further control tumor growth.

### AGGRESSIVE ADENOMAS

Despite the rarity of malignant transformation and metastatic lesions, a subset of pituitary adenomas undergoes aggressive local growth and central nervous system invasion with high Ki67 levels (>4%). Silent corticotrope and somatotrope tumors, as well as prolactinomas occurring in middle-aged men, are particularly prone to aggressive growth and recurrence. Patients with these tumors usually require an integrated management approach including repeat surgeries and irradiation. Temozolomide has also been used with variable responses.

### FURTHER READING

- Coopmans EC et al: Multivariable prediction model for biochemical response to first-generation somatostatin receptor ligands in acromegaly. *J Clin Endocrinol Metab* 105:2964, 2020.
- Elbelt U et al: Efficacy of temozolomide therapy in patients with aggressive pituitary adenomas and carcinomas: A German survey. *J Clin Endocrinol Metab* 105:e660, 2020.
- Freida PU et al: Pituitary incidentaloma: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 96:894, 2011.
- Melmed S: Pituitary-tumor endocrinopathies. *N Engl J Med* 382:937, 2020.
- Melmed S et al: Diagnosis and treatment of hyperprolactinemia: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 96:273, 2011.
- Neou M et al: Pangenomic classification of pituitary neuroendocrine tumors. *Cancer Cell* 37:123, 2020.
- Nieman LK: Cushing's syndrome: Update on signs, symptoms and biochemical screening. *Eur J Endocrinol* 173:M33, 2015.
- Ntalai G et al: Clinical review: Functioning gonadotroph adenomas. *J Clin Endocrinol Metab* 99:4423, 2014.
- Pivonello R et al: The treatment of Cushing's disease. *Endocr Rev* 36:385, 2015.
- Samson SL et al: Maintenance of acromegaly control in patients switching from injectable somatostatin receptor ligands to oral octreotide. *J Clin Endocrinol Metab* 105:e3785, 2020.
- Theodoropoulou M et al: Tumor-directed therapeutic targets in Cushing disease. *J Clin Endocrinol Metab* 104:925, 2019.

Gary L. Robertson, Daniel G. Bichet



The neurohypophysis, or posterior pituitary, is composed of large neuronal axons that originate in cell bodies in the supraoptic and paraventricular nuclei of the hypothalamus, project through the diaphragm sella, and terminate as bulbous enlargements on a capillary plexus that drains into the superior vena cava. Some of these neurons produce arginine vasopressin (AVP), also known as antidiuretic hormone; others produce oxytocin. AVP acts on the renal tubules to reduce water loss by concentrating the urine. Oxytocin stimulates postpartum milk letdown in response to suckling. A deficiency of AVP secretion or action causes a syndrome characterized by the production of large amounts of dilute urine. Excessive or inappropriate AVP production impairs urinary water excretion and predisposes to hyponatremia if water intake is not reduced in parallel with urine output.

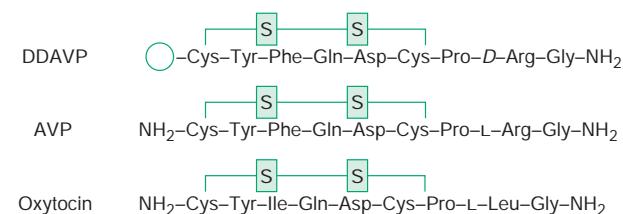
### VASOPRESSIN

#### SYNTHESIS AND SECRETION

AVP is a nonapeptide composed of a six-member disulfide ring and a tripeptide tail (**Fig. 381-1**). It is synthesized via a polypeptide precursor that includes AVP, neuropephsin, and copeptin, all encoded by a single gene on chromosome 20. After preliminary processing and folding, the precursor is packaged in neurosecretory vesicles, where it is transported down the axon; further processed to AVP, neuropephsin, and copeptin; and stored in neurosecretory vesicles until released by exocytosis into peripheral blood.

In healthy individuals, AVP secretion is regulated primarily by the "effective" osmotic pressure, which is determined largely by the plasma concentration of sodium and its anions. This regulation is mediated by specialized cells in the anteromedial hypothalamus, known as *osmoreceptors*. The osmoreceptors receive blood from small perforating branches of the anterior communicating artery. They are extremely sensitive to small changes in the plasma concentration of sodium and its anions but normally are insensitive to other naturally occurring plasma solutes such as urea and glucose. This osmoregulatory system includes inhibitory as well as stimulatory components that function in concert to form a threshold, or set point, control system. Below this threshold, plasma AVP is suppressed to levels that permit the development of a maximum water diuresis. Above it, plasma AVP rises steeply in direct proportion to plasma osmolarity, quickly reaching levels sufficient to produce maximum antidiuresis. The absolute levels of plasma osmolarity/sodium at which minimally and maximally effective levels of plasma AVP occur differ from person to person, apparently due to genetic influences on the set and sensitivity of the system. However, the average threshold, or set point, for AVP release corresponds to a plasma osmolarity and sodium of ~275 mosmol/L and 135 meq/L, respectively; levels only 2–4% higher normally result in maximum antidiuresis.

The set point of the osmostat decreases ~1% during the luteal phase of the menstrual cycle and ~3% during pregnancy. It is also reduced



**FIGURE 381-1** Primary structures of arginine vasopressin (AVP), oxytocin, and desmopressin (DDAVP).

by a decrease in blood pressure or by volume loss of >10–20%. These hemodynamic influences are mediated by neuronal afferents that originate in transmural pressure receptors of the heart and large arteries and project via the vagus and glossopharyngeal nerves to the brainstem, which sends postsynaptic projections to the hypothalamus. AVP secretion also can be stimulated by nausea, acute hypoglycemia, glucocorticoid deficiency, smoking, and, possibly, hyperangiotensinemia. Emetic stimuli are extremely potent since they typically elicit immediate, 50- to 100-fold increases in plasma AVP even when the nausea is transient and is not associated with vomiting or other symptoms. They appear to act via the emetic center in the medulla and can be blocked completely by treatment with antiemetics such as fluphenazine. There is no evidence that pain or other noxious stresses have any effect on AVP unless they elicit a vasovagal reaction with its associated nausea and hypotension.

### ACTION

The most important, if not the only, physiologic action of AVP is to reduce water excretion by promoting concentration of urine. This antidiuretic effect is achieved primarily by increasing the hydroosmotic permeability of principal cells that line the distal tubule and medullary collecting ducts of the kidney (Fig. 381-2). In the absence of AVP, these cells are impermeable to water and reabsorb little, if any, of the relatively large volume of dilute filtrate that enters from the proximal nephron. In this condition, the rate of urine output can be as high as 0.2 mL/kg per min and the specific gravity and osmolarity as low as ~1.000 and 50 mosmol/L, respectively. When AVP is secreted, it binds to V<sub>2</sub> receptors on the basal surface of principal cells causing water channels composed of aquaporin-2 to be inserted into the apical surface of the cell. These channels allow water to flow passively from the lumen through the cell down the osmotic gradient created by the hypertonicity of the renal medulla. The magnitude of this antidiuretic effect varies in direct proportion to plasma AVP, the rate of solute excretion, and the level of hypertonicity in the renal medulla. The maximum antidiuresis achievable in healthy humans occurs at plasma AVP levels in the range of 1 to 3 pg/mL and results in a urine osmolarity as high as 1200 mosmol/L and a rate of output as low as 0.35 mL/min. However, maximum concentrating capacity varies considerably depending on the level of hypertonicity in the renal medulla and that, in turn, is a function of the level and duration of AVP receptor 2 (AVPR2)-stimulated readorption of urea in the distal nephron. Hence, if basal AVP stimulation of AVPR2 is low for any reason (e.g., a high basal fluid intake), the rise in urine osmolarity that occurs immediately after an increase in hormone levels may be so blunted as to suggest a defect in antidiuretic function. This probably accounts for the shortcomings of the traditional indirect methods for the differential diagnosis of diabetes insipidus (see below).

At high concentrations, AVP also causes contraction of smooth muscle in blood vessels in the skin and gastrointestinal tract, induces glycogenolysis in the liver, and potentiates adrenocorticotrophic hormone (ACTH) release by corticotropin-releasing factor. These effects are mediated by V<sub>1a</sub> or V<sub>1b</sub> receptors that are coupled to phospholipase C. They may also affect the sensitivity of

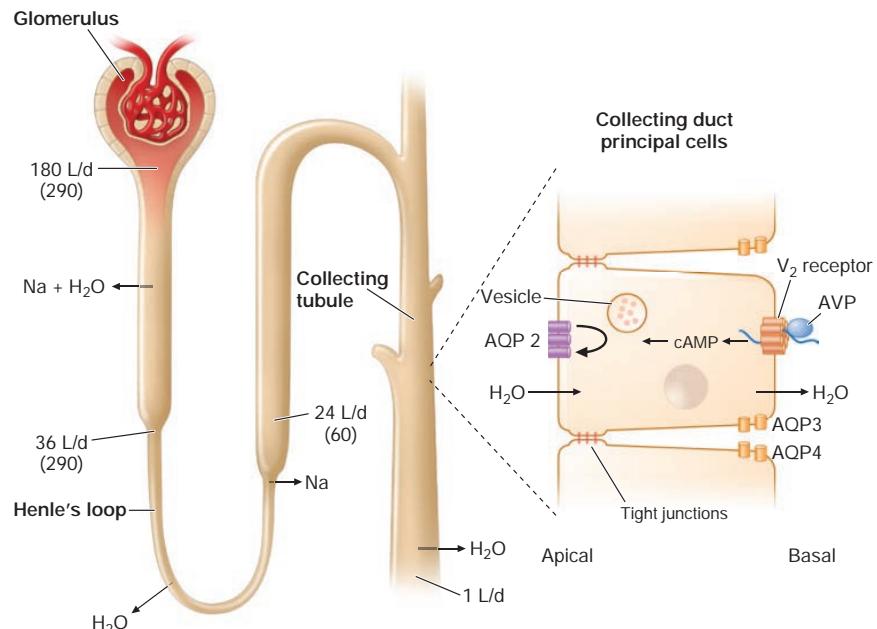
the baroreceptor and influence sympathetic and parasympathetic outflows to a variety of target organs, including the heart, the peripheral vasculature, and the kidneys. Their role, if any, in human physiology/pathophysiology remains to be determined.

### METABOLISM

AVP distributes rapidly into a space roughly equal to the extracellular fluid volume. It is cleared irreversibly with a half-life ( $t_{1/2}$ ) of 10–30 min. Most AVP clearance is due to degradation in the liver and kidneys. During pregnancy, the metabolic clearance of AVP is increased three- to fourfold due to placental production of an N-terminal peptidase.

### THIRST

Because AVP cannot reduce water loss below a certain minimum level obligated by urinary solute load and evaporation from skin and lungs, a mechanism for ensuring adequate intake is essential for preventing dehydration. This vital function is performed by the thirst mechanism. Like AVP, thirst and fluid intake are regulated primarily by an osmostat that is situated in the anteromedial hypothalamus and is able to detect very small changes in the plasma concentration of sodium and its anions. The thirst osmostat appears to be "set" about 3% higher than the AVP osmostat. This arrangement ensures that thirst, polydipsia, and dilution of body fluids do not occur until plasma osmolarity/sodium exceeds the defensive capacity of the antidiuretic mechanism. Defects in this mechanism result in hypodipsia, diverse abnormalities in AVP secretion, and a variety of clinical disorders of water balance (see below). The gastrointestinal tract also has a mechanism that detects fluid intake and inhibits thirst and AVP secretion before water is absorbed sufficiently to lower plasma osmolarity/sodium. However, the resultant inhibition of thirst and AVP is transient unless plasma



**FIGURE 381-2** Antidiuretic effect of arginine vasopressin (AVP) in the regulation of urine volume. In a typical 70-kg adult, the kidney filters ~180 L/d of plasma. Of this, ~144 L (80%) is reabsorbed isosmotically in the proximal tubule and another 8 L (4–5%) is reabsorbed without solute in the descending limb of Henle's loop. The remainder is diluted to an osmolarity of ~60 mmol/kg by selective reabsorption of sodium and chloride in the ascending limb. In the absence of AVP, the urine issuing from the loop passes largely unmodified through the distal tubules and collecting ducts, resulting in a maximum water diuresis. In the presence of AVP, solute-free water is reabsorbed osmotically through the principal cells of the collecting ducts, resulting in the excretion of a much smaller volume of concentrated urine. This antidiuretic effect is mediated via a G protein-coupled V<sub>2</sub> receptor that increases intracellular cyclic AMP, thereby inducing translocation of aquaporin 2 (AQP 2) water channels into the apical membrane. The resultant increase in permeability permits an influx of water that diffuses out of the cell through AQP 3 and AQP 4 water channels on the basal-lateral surface. The net rate of flux across the cell is determined by the number of AQP 2 water channels in the apical membrane and the strength of the osmotic gradient between tubular fluid and the renal medulla. Tight junctions on the lateral surface of the cells serve to prevent unregulated water flow. The V<sub>2</sub> receptors and AQP 2 are encoded by genes on chromosome Xq28 and 12q13, respectively.

## OXYTOCIN

Oxytocin is also a nonapeptide that differs from AVP only at positions 3 and 8 (Fig. 381-1). However, it has relatively little antidiuretic effect and seems to act mainly on mammary ducts to facilitate milk letdown during nursing. It also may help initiate or facilitate labor by stimulating contraction of uterine smooth muscle, but it is not clear if this action is physiologic or necessary for normal delivery.

## DEFICIENCIES OF AVP SECRETION AND ACTION

### DIABETES INSIPIDUS

**Clinical Characteristics** Diabetes insipidus (DI) is a syndrome characterized by the excretion of abnormally large volumes of dilute urine. The 24-h urine volume exceeds 40 mL/kg body weight, and the 24-h urine osmolarity is <280 mosm/L. The polyuria produces symptoms of urinary frequency, enuresis, and/or nocturia. It also results in a slight rise in plasma osmolarity/sodium that stimulates thirst and a commensurate increase in fluid intake (polydipsia). Hence, clinical symptoms and signs of dehydration are uncommon unless thirst and/or water intake are also impaired.

**Etiology** DI is divided into four different types based on the etiology. The most common is due to a primary deficiency of AVP secretion. It is referred to variously as *neurohypophyseal*, *neurogenic*, *pituitary*, *cranial*, or *central DI*. It can be caused by a variety of congenital, acquired, or genetic disorders but is often idiopathic (Table 381-1). Six genetic forms of pituitary DI are now known. By far, the most common is transmitted in an autosomal dominant mode and is caused by diverse mutations in the coding region of one allele of the AVP-neurophysin II (or *AVP-NPII*) gene. All the mutations alter one or more amino acids known to be critical for correct processing and/or folding of the prohormone, thus interfering with its trafficking through the endoplasmic reticulum. Presumably, the misfolded mutant precursor accumulates and interferes with the production of AVP by the normal allele. Eventually, it destroys the magnocellular neurons in which it is produced since histologic studies in a few patients show fibrosis and a lack of AVP-containing neurons in the posterior pituitary. The AVP deficiency usually is not present at birth but develops gradually over a period of months to years, progressing from partial to severe at different rates depending on the mutation and other unknown variables. Once established, the deficiency of AVP is permanent, but for unknown reasons, the DI occasionally improves or remits completely in late middle age. The parvocellular neurons that make AVP and the magnocellular neurons that make oxytocin appear to be unaffected. There are also rare autosomal recessive forms of pituitary DI. One is due to an inactivating mutation in the AVP portion of the gene that results in production of a biologically inactive form of AVP. Another is due to a large deletion that involves the majority of the AVP gene and regulatory sequences in the intergenic region. A third is caused by mutations of the *WFS1* gene responsible for Wolfram's syndrome (DI, diabetes mellitus, optic atrophy, and neural deafness [DIDMOAD]). An X-linked recessive form pituitary DI linked to Xq28 also has been reported, but the causative gene has not yet been identified. Finally, mutations in the *PCSK1* gene have been associated with severe early malabsorptive diarrhea and an undefined polyuria-polydipsia syndrome developing before 5 years of age.

The second type of DI is due to a suppression of AVP secretion by excessive intake of fluids. It is commonly referred to as *primary polydipsia* and can be subdivided into three subcategories. In one, called *dipsogenic DI*, the excessive fluid intake appears to be caused by inappropriate thirst. It can occur following head trauma or in association with multifocal diseases of the brain such as neurosarcoïd, tuberculous meningitis, and multiple sclerosis, but like pituitary DI, it is often idiopathic. The second subcategory, *psychogenic polydipsia*, is not associated with thirst, and the polydipsia seems to be a feature

**TABLE 381-1 Causes of Diabetes Insipidus**

Pituitary diabetes insipidus	Gestational diabetes insipidus
Acquired	Pregnancy (second and third trimesters)
Head trauma (closed and penetrating) including pituitary surgery	Nephrogenic diabetes insipidus
Neoplasms	Acquired
Primary	Drugs
Craniopharyngioma	Lithium
Pituitary adenoma (suprasellar)	Demeclocycline
Dysgerminoma	Methoxyflurane
Meningioma	Amphotericin B
Metastatic (lung, breast)	Aminoglycosides
Hematologic (lymphoma, leukemia)	Cisplatin
Granulomas	Rifampin
Sarcoidosis	Foscarnet
Histiocytosis	Metabolic
Xanthoma disseminatum	Hypercalcemia, hypercalciuria
Infectious	Hypokalemia
Chronic meningitis	Obstruction (ureter or urethra)
Viral encephalitis	Vascular
Toxoplasmosis	Sickle cell disease and trait
Inflammatory	Ischemia (acute tubular necrosis)
Lymphocytic infundibuloneurohypophysitis	Granulomas
Granulomatosis with polyangiitis (Wegener's)	Sarcoidosis
Lupus erythematosus	Neoplasms
Scleroderma	Sarcoma
Chemical toxins	Infiltration
Tetrodotoxin	Amyloidosis
Snake venom	Idiopathic
Vascular	Genetic
Sheehan's syndrome	X-linked recessive ( <i>AVP receptor-2</i> gene)
Aneurysm (internal carotid)	Autosomal recessive ( <i>AQP2</i> gene)
Aortocoronary bypass	Autosomal dominant ( <i>AQP2</i> gene)
Hypoxic encephalopathy	<b>Primary polydipsia</b>
Idiopathic	Acquired
Congenital malformations	Psychogenic
Septo-optic dysplasia	Schizophrenia
Midline craniofacial defects	Obsessive compulsive disorder
Holoprosencephaly	Dipsogenic (abnormal thirst)
Hypogenesis, ectopia of pituitary	Granulomas (sarcoidosis)
Genetic	Infectious (tuberculous meningitis)
Autosomal dominant ( <i>AVP-neurophysin</i> gene)	Head trauma (closed and penetrating)
Autosomal recessive	Demyelination (multiple sclerosis)
Type A ( <i>AVP-neurophysin</i> gene)	Drugs
Type B ( <i>AVP-neurophysin</i> gene)	Idiopathic
Type C ( <i>Wolfram's 4p-WFS1</i> gene)	Iatrogenic
X-linked recessive (Xq28)	

Abbreviation: AVP, arginine vasopressin.

of psychosis or obsessive-compulsive disorder. The third subcategory, *iatrogenic polydipsia*, is due to an increase in water intake motivated by a belief in its health benefits.

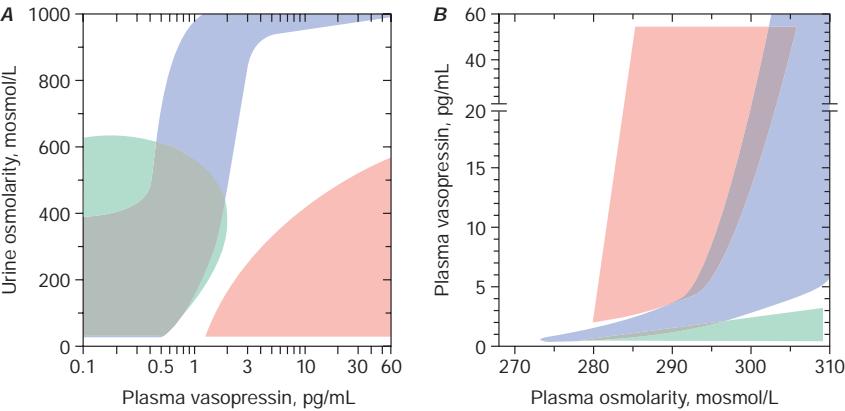
The third type of DI is also due to a deficiency of AVP caused by an increased rate of degradation by an N-terminal aminopeptidase produced in the placenta. It is referred to as *gestational DI* because the signs and symptoms manifest during pregnancy and usually remit several weeks after delivery. The fourth type of DI is caused by renal insensitivity to the antidiuretic action of AVP. It is called *nephrogenic DI* and can be caused by a drug such as lithium, a disorder such as

hypokalemia, or a genetic mutation (Table 381-1). The most common genetic form is transmitted in a semirecessive X-linked manner and is due to mutations in the gene on chromosome Xq28 that encodes the V<sub>2</sub> receptor. There are also autosomal recessive or dominant forms of nephrogenic DI. They are caused by mutations of the gene on chromosome 20 that encodes the aquaporin-2 water channels necessary for readorption of water from dilute urine in the renal collecting ducts.

**Pathophysiology** In pituitary and nephrogenic DI, the defect in urine concentration results in a rise in the rate of water excretion, a small (1–2%) decrease in body water, and a commensurate increase in plasma osmolarity/sodium, which stimulates thirst and a compensatory increase in water intake. The severity of the defect in antidiuretic function varies significantly from patient to patient. In some, it is nearly complete and cannot be overcome by even an intense stimulus such as nausea or severe dehydration. In others, the defect in AVP secretion or action is incomplete, and a modest stimulus such as a few hours of fluid deprivation, smoking, or a vasovagal reaction is sufficient to concentrate the urine. However, even in patients with a partial defect, the *maximum* level of urine osmolarity produced by these stimuli is usually less than normal partly because the prior deficiency in basal AVP stimulation temporarily diminishes renal concentrating capacity. Nevertheless, the underlying cause of the DI can be determined by analyzing the relationship of urine osmolarity to plasma AVP (Fig. 381-3A) and of plasma AVP to plasma osmolarity/sodium (Fig. 381-3B).

The pathophysiology of primary polydipsia is the reverse of that in pituitary and nephrogenic DI. The increase in fluid intake reduces plasma osmolarity/sodium and AVP secretion. The resultant urinary dilution produces a compensatory increase in urinary free-water excretion that usually offsets the increase in intake and stabilizes plasma osmolarity/sodium at a level below basal. Thus, hyponatremia is uncommon unless the polydipsia is very severe or the compensatory water diuresis is impaired. Fluid deprivation or hypertonic saline infusion produces a normal rise in plasma AVP, but the resultant increase in urine concentration is usually subnormal because the capacity of the kidney to concentrate the urine is temporarily diminished by the prior lack of AVP stimulation. Thus, the *maximum* level of urine osmolarity achieved is often indistinguishable from that produced by fluid deprivation and/or administration of antidiuretic hormone in partial pituitary or partial nephrogenic DI. However, unlike the other two types of DI, the *relationships* of the rise in plasma AVP to the rise in plasma and urine osmolarity are both normal in primary polydipsia (Fig. 381-3).

**Differential Diagnosis** If symptoms of urinary frequency, enuresis, nocturia, and/or persistent thirst are present in the absence of glucosuria, the possibility of DI should be evaluated by collecting a 24-h urine on unrestricted fluid intake. If the osmolarity is <280 mosm/L and the volume >50 mL/kg per day, the patient has DI and should be evaluated further to determine the type. Sometimes the type can be inferred from the clinical setting. Often, however, this information is lacking, ambiguous, or even misleading, and other approaches to differential diagnosis are needed. In the few patients in whom basal plasma osmolarity and/or sodium are above the normal range, a fluid deprivation test is unnecessary and potentially hazardous because primary polydipsia can be excluded. Therefore, an injection of AVP (0.5 IU) or desmopressin (2 µg) followed by a repeat measurement of urine osmolarity will suffice to determine if the DI is due to a severe defect in the secretion or action of AVP. However, if basal plasma osmolarity and sodium are within normal limits, as they usually are, some other method is needed to determine the type of DI. The traditional

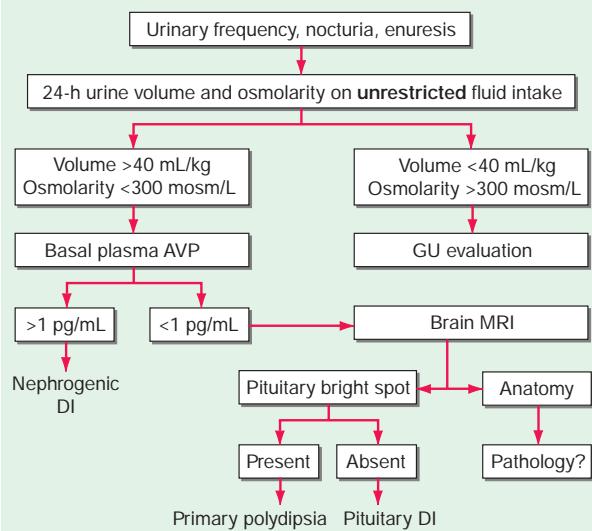


**FIGURE 381-3 Relationship of plasma arginine vasopressin (AVP) to urine osmolarity (A) and plasma osmolarity (B) before and during fluid deprivation-hypertonic saline infusion test in patients who are normal or have primary polydipsia (blue zones), pituitary diabetes insipidus (green zones), or nephrogenic diabetes insipidus (pink zones).**

approach is to stop all fluid intake for 4–6 h and closely monitor the changes in body weight, plasma osmolarity/sodium, and urine osmolarity. If plasma osmolarity and sodium rise above the normal range *without* concentrating the urine, primary polydipsia is excluded, and the effect on urine osmolarity of injecting AVP or desmopressin will determine if the patient has severe pituitary or severe nephrogenic DI. If, however, fluid deprivation results in concentration of the urine, the effect on urine osmolarity of injecting AVP or desmopressin does *not* distinguish reliably between *partial* pituitary DI, *partial* nephrogenic DI, and primary polydipsia because all three disorders temporarily diminish renal concentrating capacity to a variable extent depending on the severity of the basal polyuria.

The ambiguities inherent in the indirect method of differential diagnosis usually can be avoided by measuring plasma AVP before and during 4–6 h of complete fluid deprivation and relating these values to the concurrent level of plasma and urine osmolarity (Fig. 381-3). This approach distinguishes reliably between pituitary and nephrogenic DI even when the defect in AVP secretion or action is partial. It also differentiates partial pituitary DI from primary polydipsia if plasma osmolarity and sodium rise above the normal range. However, this level of dehydration is difficult to achieve by fluid deprivation alone when urinary concentration occurs. Therefore, it is usually necessary to add a short infusion of 3% saline (0.1 mL/kg body weight per minute for 60–90 min) and repeat the measurements of plasma AVP when plasma osmolarity and sodium rise above the normal range. This approach differentiates reliably between partial pituitary DI and primary polydipsia as evidenced by other findings, including MRI of the posterior pituitary and a properly dosed and closely monitored trial of antidiuretic therapy.

A simpler and less stressful but equally reliable way to differentiate among pituitary DI, nephrogenic DI, and primary polydipsia is to start by measuring basal plasma AVP and urine osmolarity under conditions of unrestricted fluid intake (Fig. 381-4). If AVP is normal or elevated (>1 pg/mL) and the concurrent urine osmolarity is low (<280 mosm/L), the patient has nephrogenic DI and the only additional evaluation required is to determine the cause. If, however, basal plasma AVP is low or undetectable (<1 pg/mL), nephrogenic DI is very unlikely, and a brain MRI can be performed to determine if the hyperintense signal normally emitted by the posterior pituitary on T1-weighted images is present. Because this “bright spot” is a function of pituitary stores of AVP, it is almost always present in primary polydipsia but is abnormally small or absent in pituitary DI even when the deficiency in AVP is partial or due to production of a biologically inactive form of the hormone. The bright spot is also faint or absent in nephrogenic DI presumably because the chronic stimulus to AVP secretion depletes pituitary stores of the hormone. Therefore, MRI does not differentiate between pituitary and nephrogenic DI unless it reveals other pathology involving the gland. The other caveat is that the pituitary bright spot is



**FIGURE 381-4 Simplified approach to the differential diagnosis of diabetes insipidus.** When symptoms suggest diabetes insipidus (DI), the syndrome should be differentiated from a genitourinary (GU) abnormality by measuring the 24-h urine volume and osmolarity on unrestricted fluid intake. If DI is confirmed, basal plasma arginine vasopressin (AVP) should be measured on unrestricted fluid intake. If AVP is normal or elevated ( $>1$  pg/mL), the patient probably has nephrogenic DI. However, if plasma AVP is low or undetectable, the patient has either pituitary DI or primary polydipsia. In that case, magnetic resonance imaging (MRI) of the brain can be performed to differentiate between these two conditions by determining whether or not the normal posterior pituitary bright spot is visible on T1-weighted midsagittal images. In addition, the MRI anatomy of the pituitary hypothalamic area can be examined to look for evidence of pathology that sometimes causes pituitary DI or the dipsogenic form of primary polydipsia. MRI is not reliable for differential diagnosis unless nephrogenic DI has been excluded because the bright spot is also absent, small, or faint in this condition.

also absent in patients with empty sella even in the absence of any type of DI and is sometimes present in infants in the early stages of familial pituitary DI.

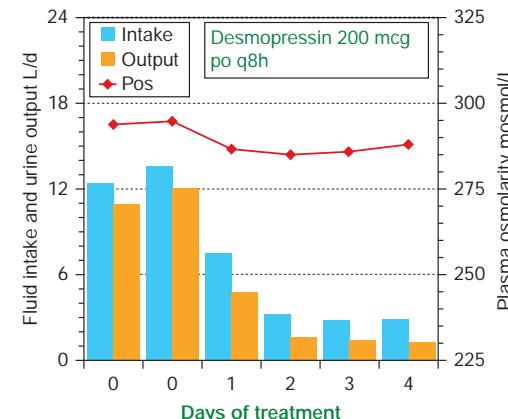
If MRI and/or AVP assays with the requisite sensitivity and specificity are unavailable and a fluid deprivation test is impractical or undesirable, a third way to differentiate among pituitary DI, nephrogenic DI, and primary polydipsia is a properly dosed and closely monitored trial of desmopressin therapy (see below). In nephrogenic DI, this treatment has no effect on urine output, fluid intake, or plasma osmolarity/sodium. In pituitary DI, it abolishes the polyuria and polydipsia and reduces plasma osmolarity/sodium by 1–2%. However, in primary polydipsia, antidiuretic therapy eliminates the polyuria but not the polydipsia and, as a consequence, produces moderate to severe hyponatremia within 8–24 h.

The measurement of plasma copeptin has also been advocated for the differential diagnosis of DI since it is synthesized and co-secreted with AVP. However, because copeptin is cleared from plasma more slowly than AVP, its relationship to plasma and urine osmolarity is not diagnostic, and another approach based on the change in plasma copeptin after an infusion of hypertonic saline has been reported. However, unlike the measurement of plasma AVP, the diagnoses obtained by the copeptin method do not correlate well with the MRI findings or the response to antidiuretic therapy. The reason for these disparities remains to be determined.

## TREATMENT

### Diabetes Insipidus

The signs and symptoms of uncomplicated pituitary DI can be eliminated by treatment with desmopressin (DDAVP), a synthetic analogue of AVP (Fig. 381-1). DDAVP acts selectively at V<sub>2</sub>



**FIGURE 381-5 Effect of desmopressin therapy on fluid intake (blue bars), urine output (orange bars), and plasma osmolarity (red line) in a patient with uncomplicated pituitary diabetes insipidus. Note that treatment rapidly reduces fluid intake and urine output to normal, with only a slight increase in body water as evidenced by the slight decrease in plasma osmolarity.**

receptors to increase urine concentration and decrease urine flow in a dose-dependent manner. It is also more resistant to degradation than is AVP and has a three- to fourfold longer duration of action. DDAVP can be given by IV or SC injection, nasal inhalation, or orally by means of a tablet or melt. The doses required to control pituitary DI vary depending on the patient and the route of administration. However, among adults, they usually range from 1–2 µg qd or bid by injection, 10–20 µg bid or tid by nasal spray, or 100–400 µg bid or tid orally. The onset of antidiuresis is rapid, ranging from as little as 15 min after injection to 60 min after oral administration. When given in a dose that normalizes 24-h urinary osmolarity (400–800 mosmol/L) and volume (15–30 mL/kg body weight), DDAVP produces a slight increase in total body water and a (1–2%) decrease in plasma osmolarity/sodium that rapidly eliminates thirst and polydipsia (Fig. 381-5). Consequently, water balance is maintained within the normal range. Hyponatremia rarely develops unless urine volume is reduced to <10 mL/kg per day or fluid intake is excessive due to an associated abnormality in thirst or cognition. Fortunately, thirst abnormalities are rare, and if the patient learns to drink only when truly thirsty, DDAVP can be given safely in doses sufficient to normalize urine output without the need for allowing intermittent escape to prevent water intoxication.

Primary polydipsia cannot be treated safely with DDAVP or any other antidiuretic drug because eliminating the polyuria does not eliminate the urge to drink. Therefore, it invariably produces hyponatremia and/or other signs of water intoxication, usually within 8–24 h if urine output is normalized completely. There is no consistently effective way to correct dipsogenic or psychogenic polydipsia, but the iatrogenic form may respond to patient education. To minimize the risk of water intoxication, all patients should be warned about the use of other drugs such as thiazide diuretics or carbamazepine (Tegretol) that can impair urinary free-water excretion directly or indirectly.

The polyuria and polydipsia of nephrogenic DI are not affected by treatment with standard doses of DDAVP. If resistance is partial, it may be overcome by tenfold higher doses, but this treatment is too expensive and inconvenient for long-term use. However, treatment with conventional doses of a thiazide diuretic and/or amiloride in conjunction with a low-sodium diet and a prostaglandin synthesis inhibitor (e.g., indomethacin) usually reduces the polyuria and polydipsia by 30–70% and may eliminate them completely. Side effects such as hypokalemia and gastric irritation can be minimized by the use of amiloride or potassium supplements and by taking medications with meals.

## HYPODIPSIC HYPERNATREMIA

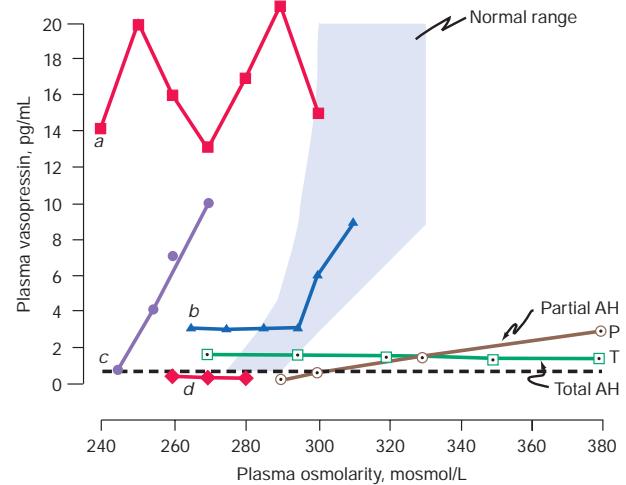
An increase in plasma osmolarity/sodium above the normal range (hypertonic hypernatremia) can be due to a decrease in total body water or an increase in total body sodium. The former results from a failure to drink enough water to replace normal or increased urinary and insensible loss due either to water deprivation or a lack of thirst (hypodipsia). This chapter focuses on hypodipsic hypernatremia, the form of hypernatremia due to a primary defect in the thirst mechanism. Hypernatremia caused by an increase in total body sodium is described elsewhere (Chap. 386).

**Clinical Characteristics** Hypodipsic hypernatremia is a syndrome characterized by chronic or recurrent hypertonic dehydration. The hypernatremia varies widely in severity and is often associated with signs of hypovolemia such as tachycardia, postural hypotension, azotemia, hyperuricemia, and hypokalemia due to secondary hyperaldosteronism. Muscle weakness, pain, rhabdomyolysis, hyperglycemia, hyperlipidemia, and acute renal failure may also occur. Obtundation or coma may be present. At presentation, plasma AVP is usually but not always subnormal relative to the concurrent hypernatremia/hyperosmolemia. DI is usually absent but may develop during rehydration as blood volume, blood pressure, and plasma osmolarity/sodium return toward normal.

**Etiology** Hypodipsia is usually due to hypogenesis or destruction of the osmoreceptors in the anterior hypothalamus that regulate thirst. The defect can result from various congenital malformations of midline brain structures or may be acquired due to diseases such as surgery or aneurysms of the anterior communicating artery, primary or metastatic tumors in the hypothalamus, head trauma, granulomatous diseases such as sarcoidosis and histiocytosis, AIDS, and cytomegalovirus encephalitis. Adipsic hypernatremia without demonstrable hypothalamic lesions has also been associated with autoantibodies directed against the subfornical organ. Episodes of transient hypodipsic hypernatremia have also been reported in association with depression or other psychological disorders, suggesting that reversible neurochemical defects in the osmoregulation of thirst can also occur.

**Pathophysiology** A deficiency in osmotically induced thirst results in a failure to drink enough water to replenish obligatory renal and extrarenal losses. Consequently, plasma osmolarity and sodium rise often to extremely high levels before the disorder is recognized. In most cases, plasma AVP is subnormal relative to the hyperosmolarity/hypernatremia, but it is still adequate to prevent DI. However, during rehydration, plasma AVP sometimes fails to levels that result in DI before the dehydration is fully corrected (Fig. 381-6). In other cases, plasma AVP does not decline even when the patient is overhydrated and a hyponatremic syndrome indistinguishable from inappropriate antidiuresis develops. This suggests that the AVP osmoreceptors normally provide inhibitory as well as stimulatory input to the neurohypophysis. In both situations, AVP secretion responds normally to nonosmotic stimuli such as nausea, hypovolemia, or hypotension, indicating that the neurohypophysis is intact. In a few patients, however, the neurohypophysis is also destroyed, resulting in a combination of chronic pituitary DI and hypodipsia, a life-threatening disorder that can be very difficult to manage. Rarely, the regulation of AVP secretion is completely normal, suggesting that the lack of thirst is due to a defect in post-osmoreceptor neural pathways to higher cognitive centers.

**Differential Diagnosis** Hypodipsic hypernatremia usually can be distinguished from other causes of inadequate fluid intake (e.g., coma, paralysis, restraints, absence of fresh water) by the clinical history and setting. Previous episodes and/or denial of thirst and failure to drink spontaneously when the patient is conscious, unrestrained, and hypernatremic are virtually diagnostic. The hypernatremia caused by excessive retention or intake of sodium can be distinguished by the presence of thirst as well as the physical and laboratory signs of hypovolemia rather than hypovolemia.



**FIGURE 381-6** Heterogeneity of osmoregulatory dysfunction in adipsic hypernatremia (AH) and the syndrome of inappropriate antidiuresis (SIAD). Each line depicts schematically the relationship of plasma arginine vasopressin (AVP) to plasma osmolarity during water loading and/or infusion of 3% saline in a patient with either AH (open symbols) or SIAD (closed symbols). The shaded area indicates the normal range of the relationship. The horizontal broken line indicates the plasma AVP level below which the hormone is undetectable and urinary concentration usually does not occur. Lines P and T represent patients with a selective deficiency in the osmoregulation of thirst and AVP that is either partial (○) or total (□). In the latter, plasma AVP does not change in response to increases or decreases in plasma osmolarity but remains within a range sufficient to concentrate the urine even if overhydration produces hypotonic hyponatremia. In contrast, if the osmoregulatory deficiency is partial (○), rehydration of the patient suppresses plasma AVP to levels that result in urinary dilution and polyuria before plasma osmolarity and sodium are reduced to normal. Lines a-d represent different defects in the osmoregulation of plasma AVP observed in patients with SIADH or SIAD. In a (■), plasma AVP is markedly elevated and fluctuates widely without relation to changes in plasma osmolarity, indicating complete loss of osmoregulation. In b (●), plasma AVP remains fixed at a slightly elevated level until plasma osmolarity reaches the normal range, at which point it begins to rise appropriately, indicating a selective defect in the inhibitory component of the osmoregulatory mechanism. In c (○), plasma AVP rises in close correlation with plasma osmolarity before the latter reaches the normal range, indicating downward resetting of the osmostat. In d (◆), plasma AVP appears to be osmoregulated normally, suggesting that the inappropriate antidiuresis is caused by some other abnormality.

## TREATMENT

### Hypodipsic Hypernatremia

Hypodipsic hypernatremia can be corrected by administering water orally if the patient is alert and cooperative or by infusing hypotonic fluids (0.45% saline or 5% dextrose and water) if the patient is not. The amount of free water in liters required to correct the deficit ( $FW$ ) can be estimated from body weight in kg ( $BW$ ) and the serum sodium concentration in mmol/L ( $S_{Na}$ ) by the formula  $FW = 0.5BW \times ([S_{Na} - 140]/140)$ . If serum glucose ( $S_{Glu}$ ) is elevated, the measured  $S_{Na}$  should be corrected ( $S_{Na}^*$ ) by the formula  $S_{Na}^* = S_{Na} + ([S_{Glu} - 90]/36)$ . This amount plus an allowance for continuing insensible and urinary losses should be given over a 24- to 48-h period. Close monitoring of serum sodium as well as fluid intake and urinary output is essential because, depending on the extent of osmoreceptor deficiency, some patients will develop AVP-deficient DI, requiring DDAVP therapy to complete rehydration; others will develop hyponatremia and a syndrome of inappropriate antidiuresis (SIAD)-like picture if overhydrated. If hyperglycemia and/or hypokalemia are present, insulin and/or potassium supplements should be given with the expectation that both can be discontinued when rehydration is complete. Plasma urea/creatinine should be monitored closely for signs of acute renal failure caused by rhabdomyolysis, hypovolemia, and hypotension.

Once the patient has been rehydrated, an MRI of the brain and tests of anterior pituitary function should be performed to look for

the cause and collateral defects in other hypothalamic functions. A long-term management plan to prevent or minimize recurrence of the fluid and electrolyte imbalance also should be developed. This should include a practical method to regulate fluid intake in accordance with variations in water balance as indicated by changes in body weight or serum sodium determined by home monitoring analyzers. Prescribing a fixed fluid intake is often problematic and potentially dangerous because insensible as well as urinary loss varies significantly over time due to changes in ambient temperature and physical activity as well as diet and antidiuresis.

## INAPPROPRIATE ANTIDIURESIS (SEE ALSO CHAP. 53)

**Clinical Characteristics** SIAD is characterized by hypoosmolar hyponatremia and impaired urinary dilution in the absence of hypovolemia, hypotension, or other nonosmotic stimuli to AVP secretion. If the hyponatremia develops gradually or exists for more than a few days, it may be largely asymptomatic. However, if the hyponatremia is severe or develops acutely, it can cause a variety of adverse symptoms and signs ranging from headache, confusion, and anorexia to nausea, vomiting, coma, and convulsions. SIAD occurs in many diverse clinical settings (Table 381-2).

**Etiology** The cause of SIAD is a failure to maximally dilute the urine and mount a water diuresis when total water intake exceeds urinary and insensible water loss. In most cases, the defect in urinary dilution is due to an abnormality in AVP secretion and is commonly referred to as the syndrome of inappropriate antidiuretic hormone (SIADH). The defect in the osmoregulation of AVP secretion can take any of several different forms (Fig. 381-6). The most common is one in which the AVP secretion responds normally to osmotic stimulation and suppression but the threshold or set point of the system is lower than normal. These patients are able to dilute their urine if plasma

osmolarity/sodium is reduced below the abnormal set point. In others, the secretion of AVP appears to be fixed or totally erratic. In ~10% of cases, there is no demonstrable defect in the osmoregulation of AVP secretion, and the failure to maximally dilute the urine may be due to an abnormality in the kidney. This form may be referred to as nephrogenic SIAD (NSIAD). In some of these patients, the inappropriate antidiuresis has been traced to a constitutive activating mutation of the V<sub>2</sub> receptor gene or the stimulatory G alpha protein GNAS. This form of inappropriate antidiuresis may be referred to as familial NSIAD.

**Pathophysiology** In SIADH and NSIAD, the failure to mount a water diuresis when intake exceeds urinary and insensible loss results in a slight expansion of total body water followed by a modest increase in urinary sodium excretion due at least in part to suppression of plasma renin activity and aldosterone secretion. As a result, expansion of extracellular volume is minimal, and clinically detectable edema does not develop. However, intracellular volume increases in proportion to the severity and rapidity of the change in plasma sodium. In the brain, this cellular swelling causes an increase in pressure that triggers a variety of symptoms. After several days, the swelling and symptoms may subside due to inactivation of some intracellular solutes and resultant decrease in cellular volume.

**Differential Diagnosis** SIADH and NSIAD must be differentiated from other types of hypo-osmolar hyponatremia associated with impaired urinary dilution. This is usually possible from the history, physical examination, and basic laboratory findings (Table 381-3). Hypervolemic hyponatremia (type I) typically occurs in patients with generalized edema due to severe congestive heart failure or cirrhosis. Plasma renin activity (PRA) and aldosterone are elevated. Hypovolemic hyponatremia (type II) occurs in patients with loss of sodium and water due to severe vomiting, diarrhea, or primary adrenal insufficiency. It is usually associated with hypotension in the recumbent or upright position and an elevation in PRA. Euvolemic hyponatremia (type III) is divisible into two groups, which need to be managed differently. In one group, the cause is a severe deficiency in cortisol or thyroxine and should be treated accordingly. The other group is composed of patients with SIADH and NSIAD. In both, edema and a history or signs of sodium and water loss are absent, urinary sodium may be slightly elevated, and PRA is often low. Measurement of plasma AVP is usually of little diagnostic value except to differentiate SIADH from NSIAD in children or families with two or more affected members. If it is undetectable, sequencing of the AVPR2 gene is indicated.

TABLE 381-2 Causes of Syndrome of Inappropriate Antidiuresis

Neoplasms	Neurologic
Carcinomas	Guillain-Barré syndrome
Lung	Multiple sclerosis
Duodenum	Delirium tremens
Pancreas	Amyotrophic lateral sclerosis
Ovary	Hydrocephalus
Bladder, ureter	Psychosis
Other neoplasms	Peripheral neuropathy
Thymoma	Congenital malformations
Mesothelioma	Agenesis corpus callosum
Bronchial adenoma	Cleft lip/palate
Carcinoid	Other midline defects
Gangliocytoma	
Ewing's sarcoma	
Genetic	Metabolic
AVP receptor-2	Acute intermittent porphyria
Head trauma (closed and penetrating)	Pulmonary
Infections	Asthma
Pneumonia, bacterial or viral	Pneumothorax
Abscess, lung or brain	Positive-pressure respiration
Cavitation (aspergillosis)	
Tuberculosis, lung or brain	Drugs
Meningitis, bacterial or viral	Vasopressin or desmopressin
Encephalitis	Serotonin reuptake inhibitors
AIDS	Oxytocin, high dose
Vascular	Vincristine
Cerebrovascular occlusions, hemorrhage	Carbamazepine
Cavernous sinus thrombosis	Nicotine
	Phenothiazines
	Cyclophosphamide
	Tricyclic antidepressants
	Monoamine oxidase inhibitors

Abbreviation: AVP, arginine vasopressin.

## TREATMENT

### Inappropriate Antidiuresis

The management of hyponatremia differs depending on the type as well as the severity and duration of symptoms. In hypervolemic hyponatremia, the objective should be to reduce total body sodium and water. If the hyponatremia is mild and symptomatic, restricting daily fluid intake to less than total urinary and insensible water loss usually suffices to prevent progression and gradually correct the defect. However, if basal urine output is very low or the hyponatremia is severe and/or symptomatic, an AVP antagonist such as tolvaptan or conivaptan can be given to increase the rate of urinary water excretion (see below). Their use should be limited to 30 days at a time because longer periods may cause or worsen abnormal liver chemistries. Infusion of hypertonic saline is absolutely contraindicated in hypervolemic hyponatremia because it further increases total body sodium and water, worsens the edema, and may precipitate cardiovascular decompensation.

In hypovolemic hyponatremia, fluid restriction and inhibitors of AVP action are absolutely contraindicated because they aggravate the hypovolemia and could precipitate hemodynamic collapse. Instead, an effort should be made to stop the loss of sodium and water and replace the deficits by mouth or infusion of isotonic or hypertonic saline. As in the treatment of other forms of

**TABLE 381-3 Differential Diagnosis of Hyponatremia Based on Clinical Assessment of Extracellular Fluid Volume (ECFV)**

Clinical Findings	Type I, Hypervolemic	Type II, Hypovolemic	Type III, Euvolemic	SIADH and SIAD Euvolemic
History				
CHF, cirrhosis, or nephrosis	Yes	No	No	No
Salt and water loss	No	Yes	No	No
ACTH-cortisol deficiency and/or nausea and vomiting	No	No	Yes	No
Physical examination				
Generalized edema, ascites	Yes	No	No	No
Postural hypotension	Maybe	Maybe	Maybe <sup>a</sup>	No
Laboratory				
BUN, creatinine	High-normal	High-normal	Low-normal	Low-normal
Uric acid	High-normal	High-normal	Low-normal	Low-normal
Serum potassium	Low-normal	Low-normal <sup>b</sup>	Normal <sup>c</sup>	Normal
Serum urate	High	High	Low	Low
Serum albumin	Low-normal	High-normal	Normal	Normal
Serum cortisol	Normal-high	Normal-high <sup>d</sup>	Low <sup>e</sup>	Normal
Plasma renin activity	High	High	Low <sup>f</sup>	Low
Urinary sodium (meq per unit of time) <sup>g</sup>	Low	Low <sup>h</sup>	High <sup>i</sup>	High <sup>i</sup>

<sup>a</sup>Postural hypotension may occur in secondary (ACTH-dependent) adrenal insufficiency even though extracellular fluid volume and aldosterone are usually normal. <sup>b</sup>Serum potassium may be high if hypovolemia is due to aldosterone deficiency. <sup>c</sup>Serum potassium may be low if vomiting causes alkalosis. <sup>d</sup>Serum cortisol is low if hypovolemia is due to primary adrenal insufficiency (Addison's disease). <sup>e</sup>Serum cortisol will be normal or high if the cause is nausea and vomiting rather than secondary (ACTH-dependent) adrenal insufficiency. <sup>f</sup>Plasma renin activity may be high if the cause is secondary (ACTH) adrenal insufficiency. <sup>g</sup>Urinary sodium should be expressed as the rate of excretion rather than the concentration. In a hyponatremic adult, an excretion rate >25 meq/d (or 25 μeq/mg of creatinine) could be considered high. <sup>h</sup>The rate of urinary sodium excretion may be high if the hypovolemia is due to diuretic abuse, primary adrenal insufficiency, or other causes of renal sodium wasting. <sup>i</sup>The rate of urinary sodium excretion may be low if intake is curtailed by symptoms or treatment.

**Abbreviations:** ACTH, adrenocorticotrophic hormone; BUN, blood urea nitrogen; CHF, congestive heart failure; SIAD, syndrome of inappropriate antidiuresis; SIADH, syndrome of inappropriate antidiuretic hormone.

hyponatremia, care must be taken to ensure that plasma sodium does not increase too rapidly or too far since doing so may produce osmotic demyelination in the brain.

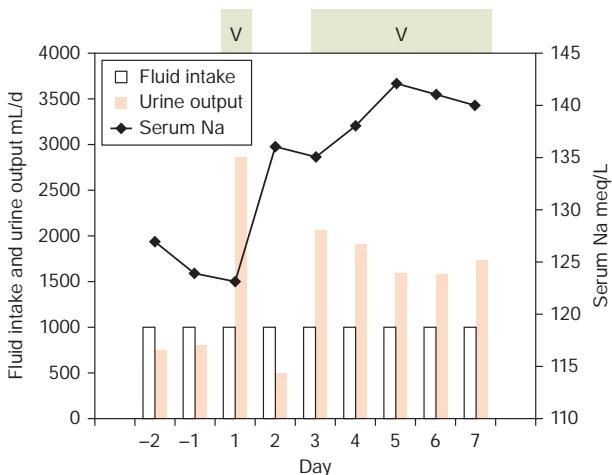
In euvolemic hyponatremia, the treatment of choice depends on the cause. If it is a deficiency in cortisol or thyroxine, gradual replacement usually suffices to eliminate all signs and symptoms, including the hyponatremia. In SIADH, the excess body water should be eliminated. If the hyponatremia is mild and largely asymptomatic, restricting total water intake to ~30 mL/kg per day less than urine output for several days or until the syndrome remits spontaneously will usually suffice. If, however, the hyponatremia is severe and symptomatic, the goal should be to partially correct it by intravenous infusion of hypertonic (3%) saline or administration of an AVP antagonist such as tolvaptan or conivaptan (Fig. 381-7). Infusion of 3% saline at a rate of ~0.05 mL/kg body weight per min raises serum sodium at a rate of ~1–2 meq/L per h, not only by replacing the slight sodium deficit but also by promoting a solute diuresis, which reduces total body water. Alternatively, a vaptan can be used to reduce body water by increasing urine output. Tolvaptan should be started at a dose of 15 mg PO qd and increased to 30 mg and 60 mg as needed to produce a brisk water diuresis. Conivaptan can be given intravenously starting with a loading dose of 20 mg over 30 min followed by another 20 mg IV over 24 h. With either vaptan, fluid intake should be monitored and restricted so as to underreplace urine output by ~5 mL/kg body weight per h. With hypertonic saline or vaptan therapy, plasma osmolarity and/or sodium should be checked every 1–2 h, and water intake or the treatment should be adjusted to keep the rate of rise at ~1% an hour until the values reach ~270 mosm/L or 130 meq/L, at which point the treatment should be discontinued. Raising the plasma sodium faster or farther may increase the risk of central pontine myelinolysis, an acute, potentially fatal neurologic syndrome characterized by quadriplegia, ataxia, and abnormal extraocular movements.

In SIAD due to an activating mutation of the V<sub>2</sub> receptor, the V<sub>2</sub> antagonists may not block the antidiuresis or raise plasma osmolarity/sodium. In that condition, use of an osmotic diuretic such as urea is reported to be effective in long-term prevention or

correction of hyponatremia. However, some vaptans may be effective in patients with a different type of activating mutation of the V<sub>2</sub> receptor, so the response to this therapy may be neither predictable nor diagnostic.

## GLOBAL PERSPECTIVES

The incidence, clinical characteristics, etiology, pathophysiology, differential diagnosis, and treatments of fluid and electrolyte disorders in tropical and nonindustrialized countries differ in some respects from those in the United States and other industrialized parts of the world.



**FIGURE 381-7** The effect of vaptan therapy on water balance in a patient with chronic syndrome of inappropriate antidiuretic hormone (SIADH). The periods of vaptan (V) therapy are indicated by the green shaded boxes at the top. Urine output is indicated by orange bars. Fluid intake is shown by the open bars. Intake was restricted to 1 L/d throughout. Serum sodium is indicated by the black line. Note that sodium increased progressively when vaptan increased urine output to levels that clearly exceeded fluid intake.

**2926** Hyponatremia, for example, appears to be more common and is more likely to be due to infectious diseases such as cholera, shigellosis, and other diarrheal disorders. In these circumstances, hyponatremia is probably due to gastrointestinal losses of salt and water (hypovolemia type II), but other abnormalities, including undefined infectious toxins, also may contribute. The causes of DI are similar worldwide except that malaria and venoms from snake or insect bites are much more common in some tropical climates.

### FURTHER READING

- Bichet DG: Regulation of thirst and vasopressin release. *Annu Rev Physiol* 81:359, 2019.
- Bichet DG et al: GNAS: A new nephrogenic cause of inappropriate antidiuresis. *J Am Soc Nephrol* 30:722, 2019.
- Fenske W et al: A copeptin based approach in the diagnosis of diabetes insipidus. *N Engl J Med* 379:428, 2018.
- Oiso Y et al: Clinical review: Treatment of neurohypophyseal diabetes insipidus. *J Clin Endocrinol Metab* 98:3958, 2013.
- Robertson GL: Vaptans for the treatment of hyponatremia. *Nat Rev Endocrinol* 7:151, 2011.
- Robertson GL: Diabetes insipidus: Differential diagnosis and management. *Best Pract Res Clin Endocrinol Metab* 30:205, 2016.

### ANATOMY AND DEVELOPMENT

The thyroid (Greek *thyreos*, shield, plus *eidos*, form) consists of two lobes connected by an isthmus. It is located anterior to the trachea between the cricoid cartilage and the suprasternal notch. The normal thyroid is 12–20 g in size, highly vascular, and soft in consistency. Four parathyroid glands, which produce parathyroid hormone (Chap. 410), are located posterior to each pole of the thyroid. The recurrent laryngeal nerves traverse the lateral borders of the thyroid gland and must be identified during thyroid surgery to avoid injury and vocal cord paralysis.

The thyroid gland develops from the floor of the primitive pharynx during the third week of gestation. The developing gland migrates along the thyroglossal duct to reach its final location in the neck. This feature accounts for the rare ectopic location of thyroid tissue at the base of the tongue (lingual thyroid) as well as the occurrence of thyroglossal duct cysts along this developmental tract. Thyroid hormone synthesis begins at about 11 weeks' gestation.

Neural crest derivatives from the ultimobranchial body give rise to thyroid medullary C cells that produce calcitonin, a calcium-lowering hormone. The C cells are interspersed throughout the thyroid gland, although their density is greatest in the juncture of the upper one-third and lower two-thirds of the gland. Calcitonin plays a minimal role in calcium homeostasis in humans, but the C cells are important because of their involvement in medullary thyroid cancer.

Thyroid gland development is orchestrated by the coordinated expression of several developmental transcription factors. Thyroid transcription factor (TTF)-1, TTF-2, NKX2-1, and paired homeobox-8 (PAX-8) are expressed selectively, but not exclusively, in the thyroid gland. In combination, they dictate thyroid cell development and the induction of thyroid-specific genes such as thyroglobulin (Tg), thyroid peroxidase (TPO), the sodium iodide symporter ( $\text{Na}^+/\text{I}^-$ , NIS), and the thyroid-stimulating hormone (TSH) receptor (TSH-R). Mutations in these developmental transcription factors or their downstream target genes are rare causes of thyroid agenesis or dyshormonogenesis, although the causes of most forms of congenital hypothyroidism remain unknown (see Chap. 383, Table 383-1). Because congenital hypothyroidism occurs in ~1 in 4000 newborns, neonatal screening is now performed in most industrialized countries. Transplacental passage of maternal thyroid hormone occurs before the fetal thyroid gland begins to function and provides significant hormone support to a fetus with congenital hypothyroidism. Early thyroid hormone replacement in newborns with congenital hypothyroidism prevents potentially severe developmental abnormalities.

The thyroid gland consists of numerous spherical follicles composed of thyroid follicular cells that surround secreted colloid, a proteinaceous fluid containing large amounts of thyroglobulin, the protein precursor of thyroid hormones (Fig. 382-2). The thyroid follicular cells are polarized—the basolateral surface is apposed to the bloodstream and an apical surface faces the follicular lumen. Increased demand for thyroid hormone is regulated by TSH, which binds to its receptor on the basolateral surface of the follicular cells. This binding leads to Tg reabsorption from the follicular lumen and proteolysis within the cytoplasm, yielding thyroid hormones for secretion into the bloodstream.

### REGULATION OF THE THYROID AXIS

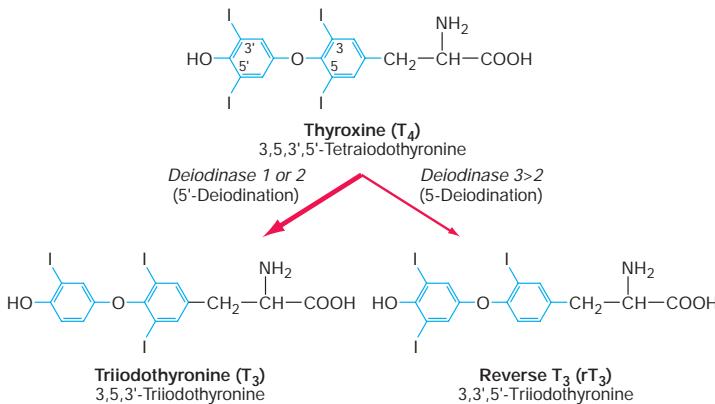
TSH, secreted by the thyrotrope cells of the anterior pituitary, plays a pivotal role in control of the thyroid axis and serves as the most useful physiologic marker of thyroid hormone action. TSH is a 31-kDa hormone composed of  $\alpha$  and  $\beta$  subunits; the  $\alpha$  subunit is common to the other glycoprotein hormones (luteinizing hormone, follicle-stimulating hormone, human chorionic gonadotropin [hCG]), whereas the  $\beta$  subunit is unique to TSH. The extent and nature of carbohydrate modification are modulated by thyrotropin-releasing hormone (TRH) and influence the biologic activity of the hormone.

The thyroid axis is a classic example of an endocrine feedback loop (Chap. 377). Hypothalamic TRH stimulates pituitary production of TSH, which, in turn, stimulates

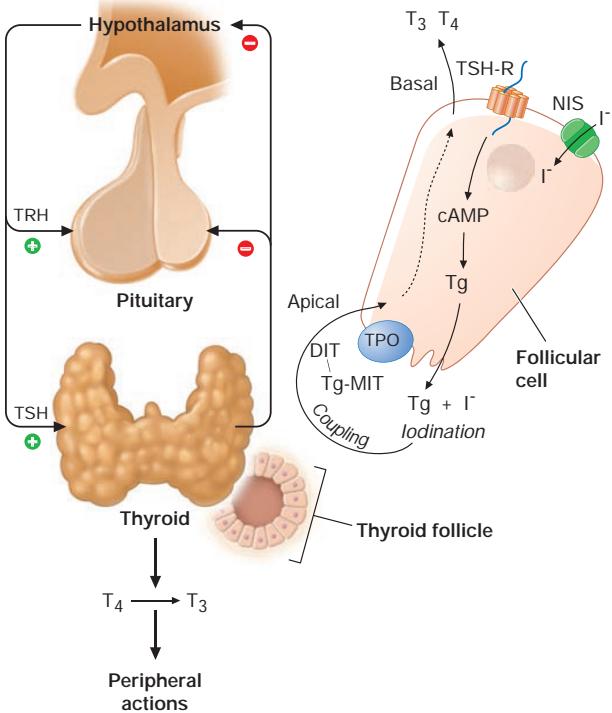
## 382 Thyroid Gland Physiology and Testing

J. Larry Jameson, Susan J. Mandel,  
Anthony P. Weetman

The thyroid gland produces two related hormones, thyroxine ( $\text{T}_4$ ) and triiodothyronine ( $\text{T}_3$ ) (Fig. 382-1). Acting through thyroid hormone receptors (TR)  $\alpha_1$  and  $\alpha_2$ , these hormones play a critical role in cell differentiation and organogenesis during development and help maintain thermogenic and metabolic homeostasis in the adult. Autoimmune disorders of the thyroid gland can stimulate overproduction of thyroid hormones (*thyrotoxicosis*) or cause glandular destruction and hormone deficiency (*hypothyroidism*). Benign nodules and various forms of thyroid cancer are relatively common and amenable to detection by physical examination, ultrasound, and other imaging techniques.



**FIGURE 382-1** Structures of thyroid hormones. Thyroxine ( $\text{T}_4$ ) contains four iodine atoms. Deiodination leads to production of the potent hormone triiodothyronine ( $\text{T}_3$ ) or the inactive hormone reverse  $\text{T}_3$ .



**FIGURE 382-2** Regulation of thyroid hormone synthesis. **Left.** Thyroid hormones  $T_4$  and  $T_3$  feed back to inhibit hypothalamic production of thyrotropin-releasing hormone (TRH) and pituitary production of thyroid-stimulating hormone (TSH). TSH stimulates thyroid gland production of  $T_4$  and  $T_3$ . **Right.** Thyroid follicles are formed by thyroid epithelial cells surrounding proteinaceous colloid, which contains thyroglobulin. Follicular cells, which are polarized, synthesize thyroglobulin and carry out thyroid hormone biosynthesis (see text for details). DIT, diiodotyrosine; MIT, monoiodotyrosine; NIS, sodium iodide symporter; Tg, thyroglobulin; TPO, thyroid peroxidase; TSH-R, thyroid-stimulating hormone receptor.

thyroid hormone synthesis and secretion. Thyroid hormones act via negative feedback predominantly through thyroid hormone receptor

2 (TR 2) to inhibit TRH and TSH production (Fig. 382-2). The “set point” in this axis is established by TSH. TRH is the major positive regulator of TSH synthesis and secretion. Peak TSH secretion occurs ~15 min after administration of exogenous TRH. Dopamine, glucocorticoids, and somatostatin suppress TSH but are not of major physiologic importance except when these agents are administered in pharmacologic doses. Reduced levels of thyroid hormone increase basal TSH production and enhance TRH-mediated stimulation of TSH. High thyroid hormone levels rapidly and directly suppress TSH gene expression and inhibit TRH stimulation of TSH secretion, indicating that thyroid hormones are the dominant regulator of TSH production. Like other pituitary hormones, TSH is released in a pulsatile manner and exhibits a diurnal rhythm; its highest levels occur at night. However, these TSH excursions are modest in comparison to those of other pituitary hormones, in part, because TSH has a relatively long plasma half-life (50 min). Consequently, single measurements of TSH are adequate for assessing its circulating level. TSH is measured using immunoradiometric assays that are highly sensitive and specific. These assays readily distinguish between normal and suppressed TSH values; thus, TSH can be used for the diagnosis of primary hyperthyroidism (low TSH) or primary hypothyroidism (high TSH).

## THYROID HORMONE SYNTHESIS, METABOLISM, AND ACTION

### THYROID HORMONE SYNTHESIS

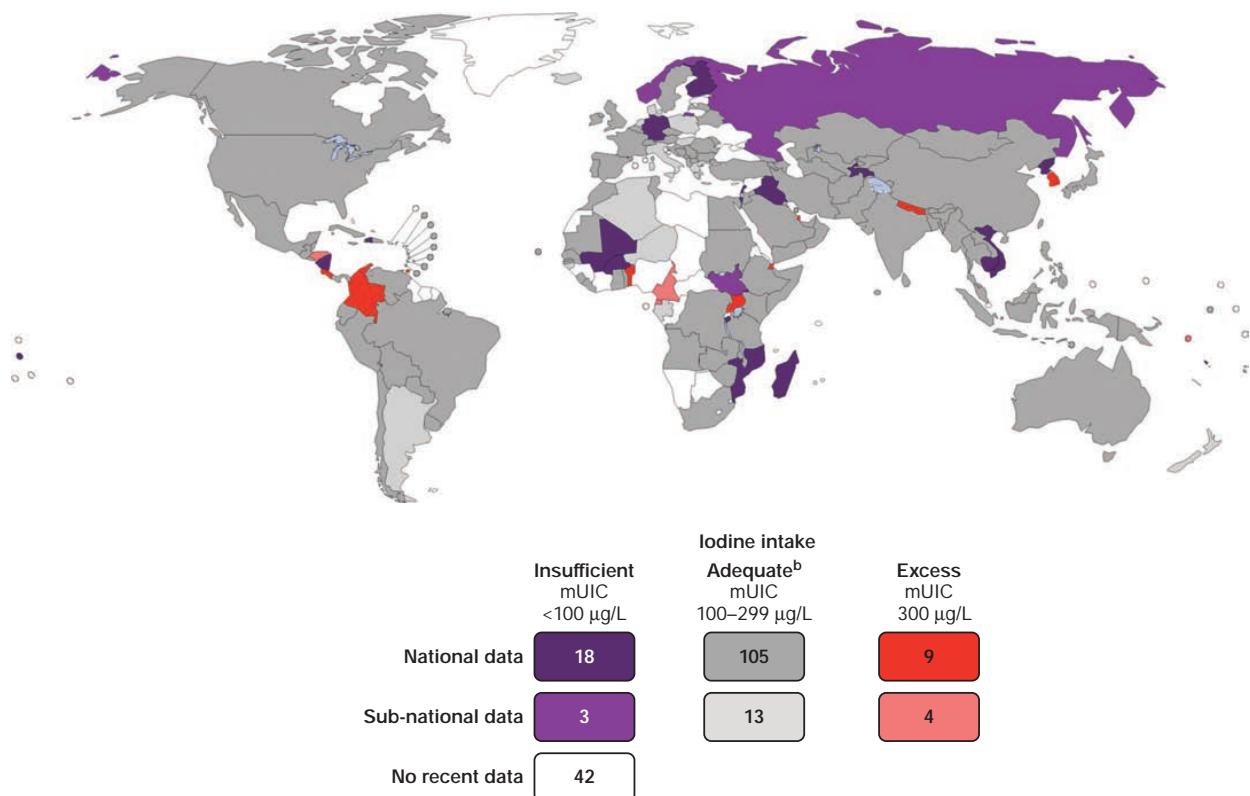
Thyroid hormones are derived from Tg, a large iodinated glycoprotein. After secretion into the thyroid follicle, Tg is iodinated on tyrosine residues that are subsequently coupled via an ether linkage. Reuptake

of Tg into the thyroid follicular cell allows proteolysis and the release of newly synthesized  $T_4$  and  $T_3$ .

**Iodine Metabolism and Transport** Iodide uptake is a critical first step in thyroid hormone synthesis. Ingested iodine is bound to serum proteins, particularly albumin. Unbound iodine is excreted in the urine. The thyroid gland extracts iodine from the circulation in a highly efficient manner. For example, 10–25% of radioactive tracer (e.g.,  $^{123}\text{I}$ ) is taken up by the normal thyroid gland over 24 h in an iodine-replete state; this value can rise to 70–90% in Graves’ disease. Iodide uptake is mediated by NIS, which is expressed at the basolateral membrane of thyroid follicular cells. NIS is most highly expressed in the thyroid gland, but low levels are present in the salivary glands, lactating breast, and placenta. The iodide transport mechanism is highly regulated, allowing adaptation to variations in dietary supply. Low iodine levels increase the amount of NIS and stimulate uptake, whereas high iodine levels suppress NIS expression and uptake. The selective expression of NIS in the thyroid allows isotopic scanning, treatment of hyperthyroidism, and ablation of thyroid cancer with radioisotopes of iodine, without significant effects on other organs. Mutation of the *NIS* gene is a rare cause of congenital hypothyroidism, underscoring its importance in thyroid hormone synthesis. Another iodine transporter, pendrin, is located on the apical surface of thyroid cells and mediates iodine efflux into the lumen. Mutation of the *Pendrin* gene causes *Pendred syndrome*, a disorder characterized by defective organification of iodine, goiter, and sensorineural deafness.

Iodine deficiency is prevalent in many mountainous regions and in central Africa, central South America, and northern Asia (Fig. 382-3). Europe remains mildly iodine-deficient, and health surveys indicate that iodine intake has been falling in the United States and Australia. The World Health Organization (WHO) estimates that about 2 billion people are iodine-deficient, based on urinary excretion data. In areas of relative iodine deficiency, there is an increased prevalence of goiter and, when deficiency is severe, hypothyroidism and cretinism. *Cretinism* is characterized by intellectual disability and growth retardation and occurs when children who live in iodine-deficient regions are not treated with iodine or thyroid hormone to restore normal thyroid hormone levels during early life. These children are often born to mothers with iodine deficiency, and it is likely that maternal thyroid hormone deficiency worsens the condition. Concomitant selenium deficiency may also contribute to the neurologic manifestations of cretinism. Iodine supplementation of salt, bread, and other food substances has markedly reduced the prevalence of cretinism. Unfortunately, however, iodine deficiency remains the most common cause of preventable intellectual disability, often because of societal resistance to food additives or the cost of supplementation. In addition to overt cretinism, mild iodine deficiency can lead to subtle reduction of IQ. Oversupply of iodine, through supplements or foods enriched in iodine (e.g., shellfish, kelp), is associated with an increased incidence of autoimmune thyroid disease. The Recommended Dietary Allowance (RDA) is 220  $\mu\text{g}$  iodine per day for pregnant women and 290  $\mu\text{g}$  iodine per day for breastfeeding women. Because the effects of iodine deficiency are most severe in pregnant women and their babies, the American Thyroid Association has recommended that all pregnant and breastfeeding women in the United States and Canada take a prenatal multivitamin containing 150  $\mu\text{g}$  iodine per day. Urinary iodine is >100  $\mu\text{g/L}$  in iodine-sufficient populations.

**Organification, Coupling, Storage, and Release** After iodide enters the thyroid, it is trapped and transported to the apical membrane of thyroid follicular cells, where it is oxidized in an organification reaction that involves TPO and hydrogen peroxide produced by dual oxidase (DUOX) and DUOX maturation factor (DUOXA). The reactive iodine atom is added to specific tyrosyl residues within Tg, a large (660 kDa) dimeric protein that consists of 2769 amino acids. The iodotyrosines in Tg are then coupled via an ether linkage in a reaction that is also catalyzed by TPO. Either  $T_4$  or  $T_3$  can be produced by this reaction, depending on the number of iodine atoms present in the iodotyrosines. After coupling, Tg is taken back into the thyroid cell,



**FIGURE 382-3 Worldwide iodine nutrition.** <sup>a</sup>In population monitoring of iodine status using urinary iodine concentration (UIC), school-age children (SAC) serve as a proxy for the general population; therefore, preference has been given to studies carried out in SAC. The UIC data have been selected for each country in the following order of priority: data from the most recent known nationally representative survey carried out between 2005 and 2020 in (i) SAC, (ii) SAC and adolescents, (iii) adolescents, (iv) women of reproductive age, (v) other adults (excluding pregnant or lactating women), and (vi) other eligible populations. In the absence of recent national surveys, subnational data were used in the same order of priority. Subnational UIC surveys are commonly carried out to provide a rapid assessment of population iodine status, but due to a lack of sampling rigor, they may over- or underestimate the iodine status at the national level and should be interpreted with caution. <sup>b</sup>Adequate iodine intake in school-age children corresponds to median UIC values in the range 100–299 µg/L, and includes categories previously referred to as "Adequate" (100–199 µg/L) and "More than adequate" (200–299 µg/L). (Reproduced with permission from The Iodine Global Network. Global scorecard of iodine nutrition in 2021 in the general population based on data in school-age children (SAC). IGN: Ottawa, Canada. 2021.)

where it is processed in lysosomes to release T<sub>4</sub> and T<sub>3</sub>. Uncoupled mono- and diiodotyrosines (MIT, DIT) can be deiodinated by the enzyme dehalogenase, thereby recycling any iodide that is not converted into thyroid hormones.

Disorders of thyroid hormone synthesis are rare causes of congenital hypothyroidism (Chap. 383). The vast majority of these disorders are due to recessive mutations in TPO or Tg, but defects have also been identified in the TSH-R, NIS, pendrin, hydrogen peroxide generation, and dehalogenase, as well as genes involved in thyroid gland development. In the case of biosynthetic defects, the gland is incapable of synthesizing adequate amounts of hormone, leading to increased TSH and a large goiter.

**TSH Action** TSH regulates thyroid gland function through the TSH-R, a seven-transmembrane G protein-coupled receptor (GPCR). The TSH-R is coupled to the  $\beta\gamma$  subunit of stimulatory G protein ( $G_s$ ), which activates adenylyl cyclase, leading to increased production of cyclic adenosine monophosphate (cAMP). TSH also stimulates phosphatidylinositol turnover by activating phospholipase C. Recessive loss-of-function TSH-R mutations cause thyroid hypoplasia and congenital hypothyroidism. Dominant gain-of-function mutations cause sporadic or familial hyperthyroidism that is characterized by goiter, thyroid cell hyperplasia, and autonomous function (Chap. 384). Most of these activating mutations occur in the transmembrane domain of

the receptor. They mimic the conformational changes induced by TSH binding or the interactions of thyroid-stimulating immunoglobulins (TSIs) in Graves' disease. Activating TSH-R mutations also occur as somatic events, leading to clonal selection and expansion of the affected thyroid follicular cell and autonomously functioning thyroid nodules.

#### Other Factors That Influence Hormone Synthesis and Release

Although TSH is the dominant hormonal regulator of thyroid gland growth and function, a variety of growth factors, most produced locally in the thyroid gland, also influence thyroid hormone synthesis. These include insulin-like growth factor 1 (IGF-1), epidermal growth factor, transforming growth factor (TGF-), endothelins, and various cytokines. The quantitative roles of these factors are not well understood, but they are important in selected disease states. In acromegaly, for example, increased levels of growth hormone and IGF-1 are associated with goiter and predisposition to multinodular goiter (MNG). Certain cytokines and interleukins (ILs) produced in association with autoimmune thyroid disease induce thyroid growth, whereas others lead to apoptosis. Iodine deficiency increases thyroid blood flow and upregulates the NIS, stimulating more efficient iodine uptake. Excess iodide transiently inhibits thyroid iodide organification, a phenomenon known as the *Wolff-Chaikoff effect*. In individuals with a normal thyroid, the gland escapes from this inhibitory effect and iodide organification resumes; the suppressive action of high iodide

may persist, however, in patients with underlying autoimmune thyroid disease.

## THYROID FUNCTION IN PREGNANCY

Five factors alter thyroid function in pregnancy: (1) the transient increase in hCG during the first trimester, which weakly stimulates the TSH-R; (2) the estrogen-induced rise in TBG during the first trimester, which is sustained during pregnancy; (3) alterations in the immune system, leading to the onset, exacerbation, or amelioration of an underlying autoimmune thyroid disease; (4) increased thyroid hormone metabolism by the placental type III deiodinase; and (5) increased urinary iodide excretion, which can cause impaired thyroid hormone production in areas of marginal iodine sufficiency. Women with a precarious iodine intake ( $<50 \mu\text{g}/\text{d}$ ) are most at risk of developing a goiter during pregnancy or giving birth to an infant with a goiter and hypothyroidism. The World Health Organization recommends a daily iodine intake of 250  $\mu\text{g}$  during pregnancy and lactation, and prenatal vitamins should contain 150  $\mu\text{g}$  per tablet.

The rise in circulating hCG levels during the first trimester is accompanied by a reciprocal fall in TSH that persists into the middle of pregnancy. This reflects the weak binding of hCG, which is present at very high levels, to the TSH-R. Rare individuals have variant TSH-R sequences that enhance hCG binding and TSH-R activation. hCG-induced changes in thyroid function can result in transient gestational hyperthyroidism that may be associated with *hyperemesis gravidarum*, a condition characterized by severe nausea and vomiting and risk of volume depletion. However, since the hyperthyroidism is not causal, antithyroid drugs are not indicated unless concomitant Graves' disease is suspected. Parenteral fluid replacement usually suffices until the condition resolves.

Normative values for most thyroid function tests differ during pregnancy and, if available, trimester-specific reference ranges should be used when diagnosing thyroid dysfunction during pregnancy. TSH levels decrease at the end of the first trimester and then rise as gestation progresses so that the nonpregnant reference ranges can be used from mid-gestation to delivery. Total  $T_4$  and  $T_3$  levels are  $\sim 1.5\times$  higher throughout pregnancy, but the free  $T_4$ , which is the same or slightly higher at the end of the 1st trimester, then progressively decreases so that third-trimester values in healthy pregnancies are often below the nonpregnant lower reference cutoff.

During pregnancy, subclinical hypothyroidism occurs in 2% of women, but overt hypothyroidism is present in only 1 in 500. Prospective randomized controlled trials have not shown a benefit for universal thyroid disease screening in pregnancy. Targeted TSH testing for hypothyroidism is recommended for women planning a pregnancy if they have a strong family history of autoimmune thyroid disease, other autoimmune disorders (e.g., type 1 diabetes), infertility, prior preterm delivery or recurrent miscarriage, or signs or symptoms of thyroid disease, or are older than 30 years. Thyroid hormone requirements are increased by up to 45% during pregnancy in levothyroxine-treated hypothyroid women.

## THYROID HORMONE TRANSPORT AND METABOLISM

**Serum-Binding Proteins**  $T_4$  is secreted from the thyroid gland in about twentyfold excess over  $T_3$  (Table 382-1). Both hormones are bound to plasma proteins, including thyroxine-binding globulin (TBG), transthyretin (TTR, formerly known as thyroxine-binding prealbumin [TBPA]), and albumin. The plasma-binding proteins increase the pool of circulating hormone, delay hormone clearance, and may modulate hormone delivery to selected tissue sites. The concentration of TBG is relatively low (1–2 mg/dL), but because of its high affinity for thyroid hormones ( $T_4 > T_3$ ), it carries  $\sim 80\%$  of the bound hormones. Albumin has relatively low affinity for thyroid hormones but has a high plasma concentration (~3.5 g/dL), and it binds up to 10% of  $T_4$  and 30% of  $T_3$ . TTR carries  $\sim 10\%$  of  $T_4$  but little  $T_3$ .

When the effects of the various binding proteins are combined, ~99.98% of  $T_4$  and 99.7% of  $T_3$  are protein-bound. Because  $T_3$  is less

TABLE 382-1 Characteristics of Circulating  $T_4$  and  $T_3$

HORMONE PROPERTY	$T_4$	$T_3$
Serum concentrations		
Total hormone	8 $\mu\text{g}/\text{dL}$	0.14 $\mu\text{g}/\text{dL}$
Fraction of total hormone in the unbound form	0.02%	0.3%
Unbound (free) hormone	$21 \times 10^{-12} M$	$6 \times 10^{-12} M$
Serum half-life	7 d	2 d
Fraction directly from the thyroid	100%	20%
Production rate, including peripheral conversion	90 $\mu\text{g}/\text{d}$	32 $\mu\text{g}/\text{d}$
Intracellular hormone fraction	~20%	~70%
Relative metabolic potency	0.3	1
Receptor binding	$10^{-10} M$	$10^{-11} M$

tightly bound than  $T_4$ , the fraction of unbound  $T_3$  is greater than unbound  $T_4$ , but there is less unbound  $T_3$  in the circulation because it is produced in smaller amounts and cleared more rapidly than  $T_4$ . The unbound or "free" concentrations of the hormones are  $\sim 2 \times 10^{-11} M$  for  $T_4$  and  $\sim 6 \times 10^{-12} M$  for  $T_3$ , which roughly correspond to the thyroid hormone receptor-binding constants for these hormones (see below). The unbound hormone is thought to be biologically available to tissues. The homeostatic mechanisms that regulate the thyroid axis are directed toward maintenance of normal concentrations of unbound hormones.

### Abnormalities of Thyroid Hormone-Binding Proteins

A number of inherited and acquired abnormalities affect thyroid hormone-binding proteins. X-linked TBG deficiency is associated with very low levels of total  $T_4$  and  $T_3$ . However, because unbound hormone levels are normal, patients are euthyroid and TSH levels are normal. It is important to recognize this disorder to avoid efforts to normalize total  $T_4$  levels, because this leads to thyrotoxicosis and is futile because of rapid hormone clearance in the absence of TBG. TBG levels are elevated by estrogen, which increases sialylation and delays TBG clearance. Consequently, in women who are pregnant or taking estrogen-containing contraceptives, elevated TBG increases total  $T_4$  and  $T_3$  levels; however, unbound  $T_4$  and  $T_3$  levels are normal. These features are part of the explanation for why women with hypothyroidism require increased amounts of l-thyroxine replacement as TBG levels are increased by pregnancy or estrogen treatment. Mutations in TBG, TTR, and albumin may increase the binding affinity for  $T_4$  and/or  $T_3$  and cause disorders known as *euthyroid hyperthyroxinemia* or *familial dysalbuminemic hyperthyroxinemia* (FDH) (Table 382-2). These disorders result in increased total  $T_4$  and/or  $T_3$ , but unbound hormone levels are normal. The familial nature of the disorders, and the fact that TSH levels are normal rather than suppressed, should suggest this diagnosis. Unbound hormone levels (ideally measured by dialysis) are normal in FDH. The diagnosis can be confirmed by using tests that measure the affinities of radiolabeled hormone binding to specific transport proteins or by performing DNA sequence analyses of the abnormal transport protein genes.

Certain medications, such as salicylates and salsalate, can displace thyroid hormones from circulating binding proteins. Although these drugs transiently perturb the thyroid axis by increasing free thyroid hormone levels, TSH is suppressed until a new steady state is reached, thereby restoring euthyroidism. Circulating factors associated with acute illness may also displace thyroid hormone from binding proteins (Chap. 384).

**Deiodinases**  $T_4$  may be thought of as a precursor for the more potent  $T_3$ .  $T_4$  is converted to  $T_3$  by the deiodinase enzymes (Fig. 382-1). Type I deiodinase, which is located primarily in thyroid, liver, and kidneys, has a relatively low affinity for  $T_4$ . Type II deiodinase has a higher affinity for  $T_4$  and is found primarily in the pituitary gland, brain, brown fat, and thyroid gland. Expression of type II deiodinase allows it to regulate  $T_3$  concentrations locally, a property that may be important in the context of levothyroxine ( $T_4$ ) replacement. Type II deiodinase

**TABLE 382-2** Conditions Associated with Euthyroid Hyperthyroxinemia

DISORDER	CAUSE	TRANSMISSION	CHARACTERISTICS
Familial dysalbuminemic hyperthyroxinemia (FDH)	Albumin mutations, usually R218H	AD	Increased T <sub>4</sub> Normal unbound T <sub>4</sub> Rarely increased T <sub>3</sub>
TBG Familial excess	Increased TBG production	XL	Increased total T <sub>4</sub> , T <sub>3</sub> Normal unbound T <sub>4</sub> , T <sub>3</sub>
Acquired excess	Medications (estrogen), pregnancy, cirrhosis, hepatitis		Increased total T <sub>4</sub> , T <sub>3</sub> Normal unbound T <sub>4</sub> , T <sub>3</sub>
Transthyretin <sup>a</sup> Excess Mutations	Islet tumors	Acquired AD	Usually normal T <sub>4</sub> , T <sub>3</sub>
	Increased affinity for T <sub>4</sub> or T <sub>3</sub>		Increased total T <sub>4</sub> , T <sub>3</sub> Normal unbound T <sub>4</sub> , T <sub>3</sub>
Medications: propranolol, ipodate, iopanoic acid, amiodarone	Decreased T <sub>4</sub> → T <sub>3</sub> conversion	Acquired	Increased T <sub>4</sub> Decreased T <sub>3</sub> Normal or increased TSH
Resistance to thyroid hormone (RTH)	Thyroid hormone receptor β mutations	AD	Increased unbound T <sub>4</sub> , T <sub>3</sub> Normal or increased TSH Some patients clinically thyrotoxic

<sup>a</sup>Also known as thyroxine-binding prealbumin (TBPA).

Abbreviations: AD, autosomal dominant; TBG, thyroxine-binding globulin; TSH, thyroid-stimulating hormone; XL, X-linked.

is also regulated by thyroid hormone; hypothyroidism induces the enzyme, resulting in enhanced T<sub>4</sub> → T<sub>3</sub> conversion in tissues such as brain and pituitary. T<sub>4</sub> → T<sub>3</sub> conversion is impaired by fasting, systemic illness or acute trauma, oral contrast agents, and a variety of medications (e.g., propylthiouracil, propranolol, amiodarone, glucocorticoids). Type III deiodinase inactivates T<sub>4</sub> and T<sub>3</sub> and is the most important source of reverse T<sub>3</sub> (rT<sub>3</sub>), including in the sick euthyroid syndrome. This enzyme is expressed in the human placenta but is not active in healthy individuals. In the sick euthyroid syndrome, especially with hypoperfusion, the type III deiodinase is activated in muscle and liver. Massive hemangiomas and other tumors that express type III deiodinase are a rare cause of consumptive hypothyroidism.

## THYROID HORMONE ACTION

**Thyroid Hormone Transport** Circulating thyroid hormones enter cells by passive diffusion and via specific transporters such as the monocarboxylate 8 transporter (MCT8), MCT10, and organic anion-transporting polypeptide 1C1. Mutations in the *MCT8* gene have been identified in patients with X-linked psychomotor retardation and thyroid function abnormalities (low T<sub>4</sub>, high T<sub>3</sub>, and high TSH). After entering cells, thyroid hormones act primarily through nuclear receptors, although they also have nongenomic actions through stimulating mitochondrial enzymatic responses and may act directly on blood vessels and the heart through integrin receptors.

**Nuclear Thyroid Hormone Receptors** Thyroid hormones bind with high affinity to nuclear TRs and RXRs. Both TR and RXR are expressed in most tissues, but their relative expression levels vary among organs; TR is particularly abundant in brain, kidneys, gonads, muscle, and heart, whereas RXR expression is relatively high in the pituitary and liver. Both receptors are variably spliced to form unique isoforms. The TR 2 isoform, which has a unique amino terminus, is selectively expressed in the hypothalamus and pituitary, where it plays a role in feedback control of the thyroid axis (see above). The TR 2 isoform contains a unique carboxy terminus that precludes thyroid hormone binding; it may function to inhibit the action of other TR isoforms.

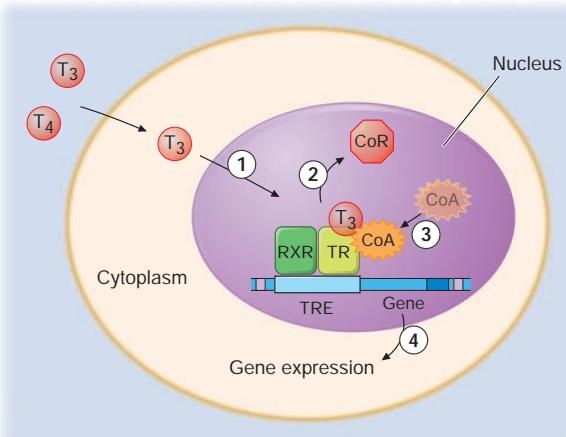
The TRs contain a central DNA-binding domain and a C-terminal ligand-binding domain. They bind to specific DNA sequences, termed *thyroid response elements* (TREs), in target genes (Fig. 382-4). The receptors bind as homodimers or, more commonly, as heterodimers

with retinoic acid X receptors (RXRs) (Chap. 377). The activated receptor can either stimulate gene transcription (e.g., myosin heavy chain ) or inhibit transcription (e.g., TSH -subunit gene), depending on the nature of the regulatory elements in the target gene.

Thyroid hormones (T<sub>3</sub> and T<sub>4</sub>) bind with similar affinities to TR and RXR. However, structural differences in the ligand-binding domains provide the potential for developing receptor-selective agonists or antagonists, and these are under investigation. T<sub>3</sub> is bound with 10–15 times greater affinity than T<sub>4</sub>, which explains its increased potency. Although T<sub>4</sub> is produced in excess of T<sub>3</sub>, receptors are occupied mainly by T<sub>3</sub>, reflecting T<sub>4</sub> conversion by peripheral tissues, T<sub>3</sub> bioavailability in the plasma, and the greater affinity of receptors for T<sub>3</sub>. After binding to TRs, thyroid hormone induces conformational changes in the receptors that modify its interactions with accessory transcription factors. Importantly, in the absence of thyroid hormone binding, the aporeceptors

bind to co-repressor proteins that inhibit gene transcription. Hormone binding dissociates the co-repressors and allows the recruitment of co-activators that enhance transcription. The discovery of TR interactions with co-repressors explains the fact that TR silences gene expression in the absence of hormone binding. Consequently, hormone deficiency has a profound effect on gene expression because it causes gene repression as well as loss of hormone-induced stimulation. This concept has been corroborated by the finding that targeted deletion of the TR genes in mice has a less pronounced phenotypic effect than hormone deficiency.

**Thyroid Hormone Resistance** Resistance to thyroid hormone (RTH) is an autosomal dominant disorder characterized by elevated thyroid hormone levels and inappropriately normal or elevated TSH.



**FIGURE 382-4** Mechanism of thyroid hormone receptor action. The thyroid hormone receptor (TR) and retinoid X receptor (RXR) form heterodimers that bind specifically to thyroid hormone response elements (TRE) in the promoter regions of target genes. In the absence of hormone, TR binds co-repressor (CoR) proteins that silence gene expression. The numbers refer to a series of ordered reactions that occur in response to thyroid hormone: (1) T<sub>4</sub> or T<sub>3</sub> enters the nucleus; (2) T<sub>3</sub> binding dissociates CoR from TR; (3) co-activators (CoA) are recruited to the T<sub>3</sub>-bound receptor; and (4) gene expression is altered.

Individuals with RTH do not, in general, exhibit signs and symptoms that are typical of hypothyroidism because hormone resistance is partial and is compensated by increased levels of thyroid hormone. The clinical features of RTH can include goiter, attention deficit disorder, mild reduction in IQ, delayed skeletal maturation, tachycardia, and impaired metabolic responses to thyroid hormone.

Classical forms of RTH are caused by mutations in the TR gene. These mutations, located in restricted regions of the ligand-binding domain, cause loss of receptor function. However, because the mutant receptors retain the capacity to dimerize with RXRs, bind to DNA, and recruit co-repressor proteins, they function as antagonists of the remaining normal TR and TR receptors. This property, referred to as "dominant negative" activity, explains the autosomal dominant mode of transmission. The diagnosis is suspected when unbound thyroid hormone levels are increased without suppression of TSH. Similar hormonal abnormalities are found in other affected family members, although the TR mutation arises de novo in ~20% of patients. DNA sequence analysis of the TR gene provides a definitive diagnosis. RTH must be distinguished from other causes of euthyroid hyperthyroxinemia (e.g., FDH) and inappropriate secretion of TSH by TSH-secreting pituitary adenomas (Chap. 380). In most patients, no treatment is indicated; the importance of making the diagnosis is to avoid inappropriate treatment of mistaken hyperthyroidism and to provide genetic counseling.

A distinct form of RTH is caused by mutations in the TR gene. Affected patients have many clinical features of congenital hypothyroidism including growth retardation, skeletal dysplasia, and severe constipation. In contrast to RTH caused by mutations in TR, thyroid function tests include normal TSH, low or normal T<sub>4</sub>, and normal or elevated T<sub>3</sub> levels. These distinct clinical and laboratory features underscore the different tissue distribution and functional roles of TR and TR. Thyroxine treatment appears to alleviate some of the clinical manifestations of patients with RTH caused by TR mutations.

## PHYSICAL EXAMINATION

In addition to the examination of the thyroid itself, the physical examination should include a search for signs of abnormal thyroid function and the extrathyroidal features of ophthalmopathy and dermopathy (Chap. 384). Examination of the neck begins by inspecting the seated patient from the front and side and noting any surgical scars, obvious masses, or distended veins. The thyroid can be palpated with both hands from behind or while facing the patient, using the thumbs to palpate each lobe. It is best to use a combination of these methods, especially when nodules are small. The patient's neck should be slightly flexed to relax the neck muscles. After locating the cricoid cartilage, the isthmus, which is attached to the lower one-third of the thyroid lobes, can be identified and then followed laterally to locate either lobe (normally, the right lobe is slightly larger than the left). By asking the patient to swallow sips of water, thyroid consistency can be better appreciated as the gland moves beneath the examiner's fingers.

Features to be noted include thyroid size, consistency, nodularity, and any tenderness or fixation. An estimate of thyroid size (normally 12–20 g) should be made, and a drawing is often the best way to record findings. Ultrasound imaging provides the most accurate measurement of thyroid volume and nodularity and is useful for assessment of goiter prevalence in iodine-deficient regions. However, ultrasound is not indicated if the thyroid physical examination is normal. The size, location, and consistency of any nodules should also be defined. A bruit or thrill over the gland, located over the insertion of the superior and inferior thyroid arteries (supero- or inferolaterally), indicates increased vascularity, associated with turbulent rather than laminar blood flow, as occurs in hyperthyroidism. If the lower borders of the thyroid lobes are not clearly felt, a goiter may be retrosternal. Large retrosternal goiters can cause venous distention over the neck and difficulty breathing, especially when the arms are raised (Pemberton's sign). With any central mass above the thyroid, the tongue should be extended, as thyroglossal cysts then move upward. The thyroid examination is not complete without assessment for lymphadenopathy in the suprACLAVicular and cervical regions of the neck.

## LABORATORY EVALUATION

**Measurement of Thyroid Hormones** The enhanced sensitivity and specificity of TSH assays have greatly improved laboratory assessment of thyroid function. Because TSH levels change dynamically in response to alterations of T<sub>4</sub> and T<sub>3</sub>, a logical approach to thyroid testing is to first determine whether TSH is suppressed, normal, or elevated. With rare exceptions (see below), a normal TSH level excludes a primary abnormality of thyroid function. This strategy depends on the use of immunochemiluminometric assays (ICMAs) for TSH that are sensitive enough to discriminate between the lower limit of the reference interval and the suppressed values that occur with thyrotoxicosis. Extremely sensitive assays can detect TSH levels 0.004 mIU/L, but, for practical purposes, assays sensitive to 0.1 mIU/L are sufficient. The widespread availability of the TSH ICMA has rendered the TRH stimulation test obsolete, because the failure of TSH to rise after an intravenous bolus of 200–400 µg TRH has the same implications as a suppressed basal TSH measured by ICMA. Because the antibodies used in the ICMA are biotinylated, biotin supplements, including biotin in multivitamins, can interfere with TSH measurement, resulting in falsely low TSH values and falsely high T<sub>4</sub> or T<sub>3</sub> levels. Therefore, patients should be advised to stop taking biotin for at least 2 days prior to thyroid function testing.

The finding of an abnormal TSH level must be followed by measurements of circulating thyroid hormone levels to confirm the diagnosis of hyperthyroidism (suppressed TSH) or hypothyroidism (elevated TSH). Automated immunoassays are widely available for serum total T<sub>4</sub> and total T<sub>3</sub>. T<sub>4</sub> and T<sub>3</sub> are highly protein-bound, and numerous factors (illness, medications, genetic factors) can influence protein binding. It is useful, therefore, to measure the free, or unbound, hormone levels, which correspond to the biologically available hormone pool. Two direct methods are used to measure *unbound thyroid hormones*: (1) unbound thyroid hormone competition with radiolabeled T<sub>4</sub> (or an analogue) for binding to a solid-phase antibody, and (2) physical separation of the unbound hormone fraction by ultracentrifugation or equilibrium dialysis. Although early unbound hormone immunoassays suffered from artifacts, newer assays correlate well with the results of the more technically demanding and expensive physical separation methods. An indirect method that is now less commonly used to estimate unbound thyroid hormone levels is to calculate the free T<sub>3</sub> or free T<sub>4</sub> index from the total T<sub>3</sub> or T<sub>4</sub> concentration and the *thyroid hormone binding ratio* (THBR). The latter is derived from the T<sub>3</sub>-resin uptake test, which determines the distribution of radiolabeled T<sub>3</sub> between an absorbent resin and the unoccupied thyroid hormone-binding proteins in the sample. The binding of the labeled T<sub>3</sub> to the resin is increased when there is reduced unoccupied protein binding sites (e.g., TBG deficiency) or increased total thyroid hormone in the sample; it is decreased under the opposite circumstances. The product of THBR and total T<sub>3</sub> or T<sub>4</sub> provides the *free T<sub>3</sub> or T<sub>4</sub> index*. In effect, the index corrects for anomalous total hormone values caused by variations in hormone-protein binding.

Total thyroid hormone levels are *elevated* when TBG is increased due to estrogens (pregnancy, oral contraceptives, hormone therapy, tamoxifen, selective estrogen receptor modulators, inflammatory liver disease) and *decreased* when TBG binding is reduced (androgens, nephrotic syndrome). Genetic disorders and acute illness can also cause abnormalities in thyroid hormone-binding proteins, and various drugs (phenytoin, carbamazepine, salicylates, and nonsteroidal anti-inflammatory drugs [NSAIDs]) can interfere with thyroid hormone binding. Because unbound thyroid hormone levels are normal and the patient is euthyroid in all of these circumstances, assays that measure unbound hormone are preferable to those for total thyroid hormones.

For most purposes, the unbound T<sub>4</sub> level is sufficient to confirm thyrotoxicosis, but 2–5% of patients have only an elevated T<sub>3</sub> level (T<sub>3</sub> toxicosis). Thus, unbound T<sub>3</sub> levels should be measured in patients with a suppressed TSH but normal unbound T<sub>4</sub> levels.

There are several clinical conditions in which the use of TSH as a screening test may be misleading, particularly without simultaneous unbound T<sub>4</sub> determinations. Any severe nonthyroidal illness can cause abnormal TSH levels. Although hypothyroidism is the most common cause of an elevated TSH level, rare causes include a TSH-secreting

pituitary tumor (**Chap. 380**), thyroid hormone resistance, and assay artifact. Conversely, a suppressed TSH level, particularly  $<0.01$  mIU/L, usually indicates thyrotoxicosis. However, subnormal TSH levels between 0.01 and 0.1 mIU/L may be seen during the first trimester of pregnancy (due to hCG secretion), after treatment of hyperthyroidism (because TSH can remain suppressed for several months), and in response to certain medications (e.g., high doses of glucocorticoids or dopamine). TSH levels measured by immunoassay may also be suppressed in patients ingesting biotin supplements  $<18$  h prior to a blood draw because the TSH capture antibodies are biotinylated and the exogenous biotin can interfere with the subsequent streptavidin capture. Importantly, secondary hypothyroidism, caused by hypothalamic-pituitary disease, is associated with a variable (low to high-normal) TSH level, which is inappropriate for the low  $T_4$  level. Thus, *TSH should not be used as an isolated laboratory test to assess thyroid function in patients with suspected or known hypothalamic or pituitary disease.*

Tests for the end-organ effects of thyroid hormone excess or depletion, such as estimation of basal metabolic rate, tendon reflex relaxation rates, or serum cholesterol, are relatively insensitive and are not useful as clinical determinants of thyroid function.

### Tests to Determine the Etiology of Thyroid Dysfunction

Autoimmune thyroid disease is detected most easily by measuring circulating antibodies against TPO and Tg. Because antibodies to Tg alone are less common, it is reasonable to measure only TPO antibodies. About 5–15% of euthyroid women and up to 2% of euthyroid men have thyroid antibodies; such individuals are at increased risk of developing thyroid dysfunction. Almost all patients with autoimmune hypothyroidism, and up to 80% of those with Graves' disease, have TPO antibodies, usually at high levels.

TSIs are antibodies that stimulate the TSH-R in Graves' disease. They are most commonly measured by commercially available tracer displacement assays called TRAb (TSH receptor antibody) with the assumption that elevated levels in the setting of clinical hyperthyroidism reflect stimulatory effects on the TSH receptor. A bioassay is less commonly used. Remission rates in patients with Graves' disease after antithyroid drug cessation are higher with disappearance rather than persistence of TRAb. Furthermore, the TRAb assay is used to predict both fetal and neonatal thyrotoxicosis caused by transplacental passage of high maternal levels of TRAb or TSI ( $>3\times$  upper limit of normal) in the last trimester of pregnancy.

Serum Tg levels are increased in all types of thyrotoxicosis except *thyrotoxicosis factitia* caused by self-administration of thyroid hormone. Tg levels are particularly increased in thyroiditis, reflecting thyroid tissue destruction and release of Tg. The main role for Tg measurement, however, is in the follow-up of thyroid cancer patients. After total thyroidectomy and radioablation, Tg levels should be  $<0.2$  ng/mL in the absence of anti-Tg antibodies; measurable levels indicate incomplete ablation or recurrent cancer.

### Radioiodine Uptake and Thyroid Scanning

The thyroid gland selectively transports radioisotopes of iodine ( $^{123}\text{I}$ ,  $^{125}\text{I}$ ,  $^{131}\text{I}$ ) and  $^{99m}\text{Tc}$  pertechnetate, allowing thyroid imaging and quantitation of radioactive tracer fractional uptake.

Nuclear imaging of Graves' disease is characterized by an enlarged gland and increased tracer uptake that is distributed homogeneously. Toxic adenomas appear as focal areas of increased uptake, with suppressed tracer uptake in the remainder of the gland (reflecting suppressed TSH). In toxic MNG, the gland is enlarged—often with distorted architecture—and there are multiple areas of relatively increased (functioning nodules) or decreased tracer uptake (suppressed thyroid parenchyma or nonfunctioning nodules). Subacute, viral, and postpartum thyroiditis are associated with very low uptake because of follicular cell damage and TSH suppression. Thyrotoxicosis factitia is also associated with low uptake because exogenous hormone suppresses TSH. In addition, if there is excessive circulating exogenous iodine (e.g., from dietary sources of iodinated contrast dye), the radionuclide uptake is low even in the presence of increased thyroid hormone production.

Thyroid scintigraphy is not used in the routine evaluation of patients with thyroid nodules but should be performed if the serum TSH level

is subnormal to determine if functioning thyroid nodules are present. Functioning or “hot” nodules are almost never malignant, and fine-needle aspiration (FNA) biopsy is not indicated. The vast majority of thyroid nodules do not produce thyroid hormone (“cold” nodules), and these are more likely to be malignant (~5–10%). Whole-body and thyroid scanning is also used in the treatment and, now less frequently, in the surveillance of thyroid cancer. After thyroidectomy for thyroid cancer, the TSH level is raised by either using a thyroid hormone withdrawal protocol or recombinant human TSH injection (**Chap. 385**). Administration of either  $^{131}\text{I}$  or  $^{123}\text{I}$  (in higher activities than used to image the thyroid gland alone) allows whole-body scanning (WBS) to detect the thyroid remnant. WBS imaging is also performed after therapeutic administration of  $^{131}\text{I}$ , which confirms remnant ablation and may reveal iodine-avid metastases.

**Thyroid Ultrasound** Ultrasonography is the most valuable tool for the diagnosis and evaluation of patients with nodular thyroid disease (**Chap. 385**). Evidence-based guidelines recommend thyroid ultrasonography for all patients suspected of having thyroid nodules by either physical examination or another imaging study. Using 10- to 12-MHz linear transducers, resolution and image quality are excellent, allowing the characterization of nodules and cysts  $>3$  mm. Sonographic patterns that combine suspicious sonographic features are highly suggestive of malignancy (e.g., hypoechoic solid nodules with infiltrative borders and microcalcifications,  $>90\%$  cancer risk), whereas other patterns correlate with a lower likelihood of cancer (isoechoic solid nodules, 5–10% cancer risk). Some patterns suggest benignity (e.g., spongiform nodules, defined as those with multiple small internal cystic areas, or simple cysts,  $<3\%$  cancer risk) (*see Chap. 385, Fig. 385-2*). These patterns have been incorporated into validated risk stratification systems (RSSs) for sonographic imaging of thyroid nodules (American College of Radiology [ACR] Thyroid Imaging Reporting and Data System [TI-RADS], American Thyroid Association, European Thyroid Association [EU-TIRADS] and others) (*see Chap. 385, Fig. 385-1*). These systems are relatively concordant in the classification of thyroid nodules; they differ in size cutoff recommendations for FNA. Not surprisingly, the RSSs with lower size cutoffs have higher sensitivity and lower specificity for thyroid cancer diagnosis than those with higher cutoffs. Nevertheless, all have been shown to reduce unnecessary FNAs by at least 45%, in part due to the recommendation not to perform FNA of spongiform nodules.

In addition to evaluating thyroid nodules, ultrasound is useful for monitoring nodule size and for the aspiration of nodules or cystic lesions. Ultrasound-guided FNA biopsy of thyroid lesions lowers the rate of inadequate sampling and decreases sample error, thereby reducing both the nondiagnostic and false-negative rates of FNA cytology. Ultrasonography of the central and lateral cervical lymph node compartments is indispensable in the evaluation of thyroid cancer patients, preoperatively and during follow-up. In addition, the ACR recommends a survey of the cervical lymph nodes as part of every diagnostic thyroid sonographic examination.

### FURTHER READING

- Alexander EK et al: 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *Thyroid* 27:315, 2017.
- Braun D, Schweizer U: Thyroid hormone transport and transporters. *Vitam Horm* 106:19, 2018.
- Ortiga-Cavalho TM et al: Thyroid hormone receptors and resistance to thyroid hormone disorders. *Nat Rev Endocrinol* 10:582, 2014.
- Rugge JB et al: Screening and treatment of thyroid dysfunction: An evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med* 162:35, 2015.
- Stoupa A et al: Update of thyroid developmental genes. *Endocrinol Metab Clin North Am* 45:243, 2016.
- Tessler FN et al: ACR Thyroid Imaging Reporting and Data System (TI-RADS): White paper of the ACR TI-RADS Committee. *J Am Coll Radiol* 14:587 2017.
- Zimmermann MB, Boelaert K: Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol* 3:286, 2015.

# 383

## Hypothyroidism

J. Larry Jameson, Susan J. Mandel,  
Anthony P. Weetman



### HYPOTHYROIDISM

Iodine deficiency remains a common cause of hypothyroidism worldwide. In areas of iodine sufficiency, autoimmune disease (Hashimoto's thyroiditis) and iatrogenic causes (treatment of hyperthyroidism) are most common (Table 383-1).

### CONGENITAL HYPOTHYROIDISM

**Prevalence** Hypothyroidism occurs in about 1 in 2000–4000 newborns, and neonatal screening is performed in most industrialized countries. It may be transient, especially if the mother has thyroid-stimulating hormone (TSH) receptor (TSH-R)-blocking antibodies or has received antithyroid drugs, but permanent hypothyroidism occurs in the majority. The causes of neonatal hypothyroidism include thyroid gland dysgenesis in 65%, inborn errors of thyroid hormone synthesis in 30%, and TSH-R antibody mediated in 5% of affected newborns. The developmental abnormalities are twice as common in girls. Mutations that cause congenital hypothyroidism are being increasingly identified, but most remain idiopathic. These can be broadly categorized as mutations causing (1) central hypothyroidism because of abnormal hypothalamic-pituitary development or the loss of specific components of the thyrotropin-releasing hormone (TRH)/TSH hormonal pathways; (2) abnormal thyroid gland development or dysgenesis; or (3) abnormal thyroid hormone synthesis and processing, or dys hormonogenesis (Table 383-2). Transplacental passage of maternal thyroid hormone occurs before the fetal thyroid gland begins to function and provides partial hormone support to a fetus with congenital hypothyroidism.

**TABLE 383-1 Causes of Hypothyroidism**

#### Primary

Autoimmune hypothyroidism: Hashimoto's thyroiditis, atrophic thyroiditis

Iatrogenic: <sup>131</sup>I treatment, subtotal or total thyroidectomy, external irradiation of neck for lymphoma or cancer

Drugs: iodine excess (including iodine-containing contrast media), amiodarone, lithium, antithyroid drugs, *p*-aminosalicylic acid, interferon  $\alpha$  and other cytokines, aminoglutethimide, tyrosine kinase inhibitors (e.g., sunitinib), immune checkpoint inhibitors (e.g., ipilimumab, nivolumab, pembrolizumab)

Congenital hypothyroidism: absent or ectopic thyroid gland, dysmorphogenesis, TSH-R mutation

Iodine deficiency

Infiltrative disorders: amyloidosis, sarcoidosis, hemochromatosis, scleroderma, cystinosis, Riedel's thyroiditis

Overexpression of type 3 deiodinase in infantile hemangioma and other tumors

#### Transient

Silent thyroiditis, including postpartum thyroiditis

Subacute thyroiditis

Withdrawal of supraphysiologic thyroxine treatment in individuals with an intact thyroid

After <sup>131</sup>I treatment or subtotal thyroidectomy for Graves' disease

#### Secondary

Hypopituitarism: tumors, pituitary surgery or irradiation, infiltrative disorders, Sheehan's syndrome, trauma, genetic forms of combined pituitary hormone deficiencies

Isolated TSH deficiency or inactivity

Bexarotene treatment

Hypothalamic disease: tumors, trauma, infiltrative disorders, idiopathic

*Abbreviations:* TSH, thyroid-stimulating hormone; TSH-R, TSH receptor.

**Clinical Manifestations** The majority of infants appear normal at birth, and with the use of biochemical screening, few cases are now diagnosed based on clinical features, which include prolonged jaundice, feeding problems, hypotonia, enlarged tongue, delayed bone maturation, and umbilical hernia. Importantly, permanent neurologic damage results if treatment is delayed. Typical features of adult hypothyroidism may also be present (Table 383-3). Other congenital malformations, especially cardiac, are four times more common in congenital hypothyroidism.

**Diagnosis and Treatment** Because of the severe neurologic consequences of untreated congenital hypothyroidism, neonatal screening programs have been established. These are generally based on measurement of TSH or T<sub>4</sub> levels in heel-prick blood specimens. When the diagnosis is confirmed, T<sub>4</sub> is instituted at a dose of 10–15 µg/kg per d, and the dose is adjusted by close monitoring of TSH levels. T<sub>4</sub> requirements are relatively great during the first year of life, and a high circulating T<sub>4</sub> level is usually needed to normalize TSH. Early treatment with T<sub>4</sub> results in normal IQ levels, but subtle neurodevelopmental abnormalities may occur in those with the most severe hypothyroidism at diagnosis or when treatment is delayed or suboptimal. If transient hypothyroidism is suspected, or the diagnosis is unclear, treatment can be stopped safely after the age of 3 years followed by further evaluation.

### AUTOIMMUNE HYPOTHYROIDISM

**Classification** Autoimmune hypothyroidism may be associated with a goiter (*Hashimoto's*, or *goitrous thyroiditis*) or minimal residual thyroid tissue (*atrophic thyroiditis*). Because the autoimmune process gradually reduces thyroid function, there is a phase of compensation when normal thyroid hormone levels are maintained by a rise in TSH. Although some patients may have minor symptoms, this state is called *subclinical hypothyroidism*. Later, unbound T<sub>4</sub> levels fall and TSH levels rise further; symptoms become more readily apparent at this stage (usually TSH >10 mIU/L), which is referred to as *clinical hypothyroidism* or *overt hypothyroidism*.

**Prevalence** The mean annual incidence rate of autoimmune hypothyroidism is up to 4 per 1000 women and 1 per 1000 men. It is more common in certain populations, such as the Japanese, probably because of genetic factors and chronic exposure to a high-iodine diet. The mean age at diagnosis is 60 years, and the prevalence of overt hypothyroidism increases with age. Subclinical hypothyroidism is found in 6–8% of women (10% over the age of 60) and 3% of men. The annual risk of developing clinical hypothyroidism is ~4% when subclinical hypothyroidism is associated with positive thyroid peroxidase (TPO) antibodies.

**Pathogenesis** In Hashimoto's thyroiditis, there is a marked lymphocytic infiltration of the thyroid with germinal center formation, atrophy of the thyroid follicles accompanied by oxyphil metaplasia, absence of colloid, and mild to moderate fibrosis. In atrophic thyroiditis, the fibrosis is much more extensive, lymphocyte infiltration is less pronounced, and thyroid follicles are almost completely absent. Atrophic thyroiditis usually represents the end stage of Hashimoto's thyroiditis rather than a separate disorder, although a distinct form of marked fibrosis occurs in which the gland is infiltrated with IgG4-positive plasma cells.

As with most autoimmune disorders, susceptibility to autoimmune hypothyroidism is determined by a combination of genetic and environmental factors, and the risk of either autoimmune hypothyroidism or Graves' disease is increased among siblings. HLA-DR polymorphisms are the best documented genetic risk factors for autoimmune hypothyroidism, especially HLA-DR3, DR4, and DR5 in Caucasians. A weak association also exists between polymorphisms in CTLA-4, a T cell-regulatory gene, and autoimmune hypothyroidism. Both of these genetic associations are shared by other autoimmune diseases, which may explain the relationship between autoimmune hypothyroidism and other autoimmune diseases, especially type 1 diabetes mellitus, Addison's disease, pernicious anemia, and vitiligo. HLA-DR and CTLA-4 polymorphisms account for approximately half of the

**TABLE 383-2 Examples of Genetic Causes of Congenital Hypothyroidism**

DEFECTIVE GENE PROTEIN	TYPE OF HYPOTHYROIDISM	INHERITANCE	CONSEQUENCES
PROP-1	Central, hypothyroidism	Homozygous recessive	Combined pituitary hormone deficiencies, including thyroid-stimulating hormone (TSH), with preservation of adrenocorticotrophic hormone
PIT-1	Central, hypothyroidism	Homozygous or Heterozygous loss of function	Combined deficiencies of growth hormone, prolactin, TSH
IGSF1	Central, hypothyroidism	X-linked loss of function	Loss of TSH receptor (TSH-R) expression, testicular enlargement
TSH $\beta$	Central, hypothyroidism	Heterozygous loss of function	TSH deficiency
TTF-1 (TITF-1)	Primary, thyroid dysgenesis	Heterozygous loss of function	Variable thyroid hypoplasia, choreoathetosis, pulmonary problems
TTF-2 (FOXE-1)	Primary, thyroid dysgenesis	Homozygous recessive	Thyroid agenesis, choanal atresia, spiky hair
PAX-8	Primary, thyroid dysgenesis	Heterozygous loss of function	Thyroid dysgenesis, kidney abnormalities
NKX2-1	Primary, thyroid dysgenesis	Heterozygous loss of function	Thyroid dysgenesis, brain, lung abnormalities
NKX2-5	Primary, thyroid dysgenesis	Heterozygous loss of function	Thyroid dysgenesis, heart abnormalities
GLIS3	Primary, thyroid dysgenesis	Homozygous recessive	Thyroid dysgenesis, neonatal diabetes, facial abnormalities
JAG-1	Primary, thyroid dysgenesis	Heterozygous loss of function	Thyroid dysgenesis, Alagille syndrome type 1, heart abnormalities
TSH receptor	Primary, thyroid dysgenesis and dyshormonogenesis	Homozygous recessive	Resistance to TSH
G $\alpha$ (Albright hereditary osteodystrophy)	Primary, thyroid dyshormonogenesis	Heterozygous loss of function, imprinting	Resistance to TSH
Na $^+$ /I $^-$ symporter (SLC5A5)	Primary, thyroid dyshormonogenesis	Homozygous recessive	Inability to transport iodide
DUOX2 (THOX2)	Primary, thyroid dyshormonogenesis	Heterozygous loss of function	Organification defect
DUOXA2	Primary, thyroid dyshormonogenesis	Homozygous recessive	Organification defect
Thyroid peroxidase	Primary, thyroid dyshormonogenesis	Homozygous recessive	Defective organification of iodide
Thyroglobulin	Primary, thyroid dyshormonogenesis	Homozygous recessive	Defective synthesis of thyroid hormone
Pendrin (SLC26A4)	Primary, thyroid dyshormonogenesis	Homozygous recessive	Pendred syndrome: sensorineural deafness and partial organification defect in thyroid
Dehalogenase 1 (IYD)	Primary, thyroid dyshormonogenesis	Homozygous recessive	Loss of iodide reutilization

genetic susceptibility to autoimmune hypothyroidism, and the role of other contributory loci remains to be clarified. A gene on chromosome 21 may be responsible for the association between autoimmune hypothyroidism and Down's syndrome. The female preponderance of thyroid autoimmunity is most likely due to sex steroid effects on the immune response, but an X chromosome–related genetic factor is also possible and may account for the high frequency of autoimmune hypothyroidism in Turner's syndrome. Environmental susceptibility factors are poorly defined at present. A high iodine or low selenium intake and decreased exposure to microorganisms in childhood increase the risk of autoimmune hypothyroidism. Smoking cessation transiently

increases incidence, whereas alcohol intake seems protective. These factors may account for the increase in prevalence over the past two to three decades.

The thyroid lymphocytic infiltrate in autoimmune hypothyroidism is composed of activated T cells as well as B cells. Thyroid cell destruction is primarily mediated by the CD8+ cytotoxic T cells, but local production of cytokines, such as tumor necrosis factor (TNF), interleukin-1 (IL-1), and interferon (IFN- $\gamma$ ), derived from the inflammatory infiltrate may render thyroid cells more susceptible to apoptosis mediated by death receptors, such as Fas, and by oxidative stress. These cytokines also impair thyroid cell function directly and induce the expression of other proinflammatory molecules by the thyroid cells themselves, such as cytokines, HLA class I and class II molecules, adhesion molecules, CD40, and nitric oxide. Administration of high concentrations of cytokines for therapeutic purposes (especially IFN- $\gamma$ ) is associated with increased autoimmune thyroid disease, possibly through mechanisms similar to those in sporadic disease. Novel anticancer and immunomodulatory treatments, such as tyrosine kinase inhibitors, immune checkpoint inhibitors, and alemtuzumab, can also induce thyroiditis via their effects on T-cell regulation.

Antibodies to TPO and thyroglobulin (Tg) are clinically useful markers of thyroid autoimmunity, but any pathogenic effect is restricted to a secondary role in amplifying an ongoing autoimmune response. TPO antibodies fix complement, and complement membrane-attack complexes are present in the thyroid in autoimmune hypothyroidism. However, transplacental passage of Tg or TPO antibodies has no effect on the fetal thyroid, which suggests that T cell-mediated injury is required to initiate autoimmune damage to the thyroid.

Up to 20% of patients with autoimmune hypothyroidism have antibodies against the TSH-R, which, in contrast to thyroid-stimulating immunoglobulin (TSI), do not stimulate the receptor but prevent the

**TABLE 383-3 Signs and Symptoms of Hypothyroidism (Descending Order of Frequency)**

SYMPTOMS	SIGNS
Tiredness, weakness	Dry coarse skin; cool peripheral extremities
Dry skin	
Feeling cold	Puffy face, hands, and feet (myxedema)
Hair loss	Diffuse alopecia
Difficulty concentrating and poor memory	Bradycardia
Constipation	Peripheral edema
Weight gain with poor appetite	Delayed tendon reflex relaxation
Dyspnea	Carpal tunnel syndrome
Hoarse voice	Serous cavity effusions
Menorrhagia (later oligomenorrhea or amenorrhea)	
Paresthesia	
Impaired hearing	

binding of TSH. These TSH-R-blocking antibodies, therefore, cause hypothyroidism and, especially in Asian patients, thyroid atrophy. Their transplacental passage may induce transient neonatal hypothyroidism. Rarely, patients have a mixture of TSI and TSH-R-blocking antibodies, and thyroid function can oscillate between hyperthyroidism and hypothyroidism as one or the other antibody becomes dominant. Predicting the course of disease in such individuals is difficult, and they require close monitoring of thyroid function. Bioassays can be used to document that TSH-R-blocking antibodies reduce the cyclic AMP-inducing effect of TSH on cultured TSH-R-expressing cells, but these assays are difficult to perform. Thyrotropin-binding inhibitory immunoglobulin (TBII) assays that measure the binding of antibodies to the receptor by competition with labeled TSH do not distinguish between TSI and TSH-R-blocking antibodies, but a positive result in a patient with spontaneous hypothyroidism is strong evidence for the presence of blocking antibodies. The use of these assays does not generally alter clinical management, although it may be useful to confirm the cause of transient neonatal hypothyroidism.

**Clinical Manifestations** The main clinical features of hypothyroidism are summarized in Table 383-3. The onset is usually insidious, and the patient may become aware of symptoms only when euthyroidism is restored. Patients with Hashimoto's thyroiditis may present because of goiter rather than symptoms of hypothyroidism. The goiter may not be large, but it is usually irregular and firm in consistency. Rarely, uncomplicated Hashimoto's thyroiditis is associated with pain.

Patients with atrophic thyroiditis or the later stage of Hashimoto's thyroiditis present with symptoms and signs of hypothyroidism. The skin is dry, and there is decreased sweating, thinning of the epidermis, and hyperkeratosis of the stratum corneum. Increased dermal glycosaminoglycan content traps water, giving rise to skin thickening without pitting (*myxedema*). Typical features include a puffy face with edematous eyelids and nonpitting pretibial edema (Fig. 383-1). There is pallor, often with a yellow tinge to the skin due to carotene accumulation. Nail growth is retarded, and hair is dry, brittle, difficult to manage, and falls out easily. In addition to diffuse alopecia, there is thinning of the outer third of the eyebrows, although this is not a specific sign of hypothyroidism.



FIGURE 383-1 Facial appearance in hypothyroidism. Note puffy eyes and thickened skin.

Other common features include constipation and weight gain (despite a poor appetite). In contrast to popular perception, the weight gain is usually modest and due mainly to fluid retention in the myxedematous tissues. Libido is decreased in both sexes, and there may be oligomenorrhea or amenorrhea in long-standing disease, but menorrhagia may occur at an early stage. Fertility is reduced, and the incidence of miscarriage is increased. Prolactin levels are often modestly increased (Chap. 380) and may contribute to alterations in libido and fertility and cause galactorrhea.

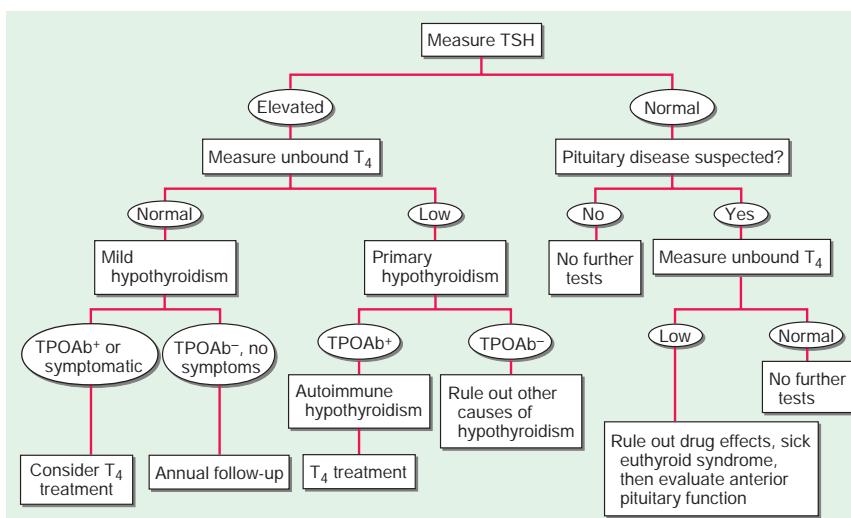
Myocardial contractility and pulse rate are reduced, leading to a reduced stroke volume and bradycardia. Increased peripheral resistance may be accompanied by hypertension, particularly diastolic. Blood flow is diverted from the skin, producing cool extremities. Pericardial effusions occur in up to 30% of patients but rarely compromise cardiac function. Although alterations in myosin heavy chain isoform expression have been documented, cardiomyopathy is rare. Fluid may also accumulate in other serous cavities and in the middle ear, giving rise to conductive deafness; sensorineural deafness may also occur. Pulmonary function is generally normal, but dyspnea may be caused by pleural effusion, impaired respiratory muscle function, diminished ventilatory drive, or sleep apnea.

Carpal tunnel and other entrapment syndromes are common, as is impairment of muscle function with stiffness, cramps, and pain. On examination, there may be slow relaxation of tendon reflexes and pseudomyotonia. Memory and concentration are impaired. Experimentally, positron emission tomography (PET) scans examining glucose metabolism in hypothyroid subjects show lower regional activity in the amygdala, hippocampus, and perigenual anterior cingulate cortex, among other regions, and this activity corrects after thyroxine replacement. Rare neurologic problems include reversible cerebellar ataxia, dementia, psychosis, and myxedema coma. *Hashimoto's encephalopathy* has been defined as a steroid-responsive syndrome associated with TPO antibodies, myoclonus, and slow-wave activity on electroencephalography, but the relationship with thyroid autoimmunity or hypothyroidism is not established, and if a patient is euthyroid, levothyroxine (LT4) therapy has not been shown to be efficacious in treatment. The hoarse voice and occasionally clumsy speech of hypothyroidism reflect fluid accumulation in the vocal cords and tongue.

The features described above are the consequence of thyroid hormone deficiency. However, autoimmune hypothyroidism may be associated with signs or symptoms of other autoimmune diseases, particularly vitiligo, pernicious anemia, Addison's disease (Schmidt's syndrome), alopecia areata, and type 1 diabetes mellitus (T1DM). In the polygenic disorder autoimmune polyendocrine syndrome type 2, autoimmune thyroid disease is present in 70–75%, T1DM in 40–60%, and Addison's disease in 40–50%. Less common associations include celiac disease, dermatitis herpetiformis, chronic active hepatitis, rheumatoid arthritis, systemic lupus erythematosus (SLE), myasthenia gravis, autoimmune hypoparathyroidism, primary hypogonadism, and Sjögren's syndrome. Thyroid-associated ophthalmopathy usually occurs in Graves' disease (see below), but in ~5% of patients, it is associated with autoimmune hypothyroidism.

Autoimmune hypothyroidism is uncommon in children and usually presents with slow growth and delayed facial and dental maturation. The pituitary may be enlarged due to thyrotroph hyperplasia. Myopathy, with muscle swelling, is more common in children than in adults. In most cases, puberty is delayed, but precocious puberty sometimes occurs. There may be intellectual impairment if the onset is before 3 years and the hormone deficiency is severe.

**Laboratory Evaluation** A summary of the investigations used to determine the existence and cause of hypothyroidism is provided in Fig. 383-2. A normal TSH level excludes primary (but not secondary) hypothyroidism. If the TSH is elevated, a free or unbound  $T_4$  level ( $FT_4$ ) is needed to confirm the presence of clinical hypothyroidism, but  $T_4$  is inferior to TSH when used as a screening test because it will not detect subclinical hypothyroidism. Circulating unbound  $T_3$  levels are normal in ~25% of patients, reflecting adaptive deiodinase responses to hypothyroidism.  $T_3$  measurements are, therefore, not indicated.



**FIGURE 383-2** Evaluation of hypothyroidism. TPOAb<sup>+</sup>, thyroid peroxidase antibodies present; TPOAb<sup>-</sup>, thyroid peroxidase antibodies not present; TSH, thyroid-stimulating hormone.

Once clinical or subclinical hypothyroidism is confirmed, the etiology is usually easily established by demonstrating the presence of TPO and Tg antibodies, which are present in >95% of patients with autoimmune hypothyroidism. TBII can be found in 10–20% of patients, but measurement is not needed routinely. Other abnormal laboratory findings in hypothyroidism may include increased creatine phosphokinase, elevated cholesterol and triglycerides, and anemia (usually normocytic or macrocytic). Except when accompanied by iron deficiency, the anemia and other abnormalities gradually resolve with thyroxine replacement.

**Differential Diagnosis** An asymmetric goiter in Hashimoto's thyroiditis may be confused with a multinodular goiter (MNG) or thyroid carcinoma, in which thyroid antibodies may also be present. Ultrasound can be used to show the presence of a solitary lesion or an MNG rather than the thyroid enlargement with heterogeneous echogenicity typical of Hashimoto's thyroiditis. Fine-needle aspiration biopsy is useful in the investigation of focal nodules. Other causes of hypothyroidism are discussed below and in Table 383-1 but rarely cause diagnostic confusion.

### OTHER CAUSES OF HYPOTHYROIDISM

*Iatrogenic hypothyroidism* is a common cause of hypothyroidism and can often be detected by screening before symptoms develop. In the first 3–4 months after radioiodine treatment for Graves' disease, transient hypothyroidism may occur due to reversible radiation damage. Low-dose thyroxine treatment can be withdrawn if recovery occurs. Because TSH levels are suppressed by hyperthyroidism, unbound T<sub>4</sub> levels are a better measure of thyroid function than TSH in the months following radioiodine treatment. Mild hypothyroidism after subtotal thyroidectomy may also resolve after several months, as the gland remnant is stimulated by increased TSH levels.

Iodine deficiency is responsible for endemic goiter and cretinism but is an uncommon cause of adult hypothyroidism unless the iodine intake is very low or there are complicating factors, such as the consumption of thiocyanates in cassava or selenium deficiency. Although hypothyroidism due to iodine deficiency can be treated with thyroxine, public health measures to improve iodine intake should be advocated to eliminate this problem. Iodized salt or bread or a single bolus of oral or intramuscular iodized oil have all been used successfully.

Paradoxically, chronic iodine excess can also induce goiter and hypothyroidism. The intracellular events that account for this effect are unclear, but individuals with autoimmune thyroiditis are especially

susceptible. Iodine excess is responsible for the hypothyroidism that occurs in patients treated with amiodarone (Chap. 384). Other drugs, particularly lithium, may also cause hypothyroidism. Transient hypothyroidism caused by thyroiditis is discussed below.

*Secondary or central hypothyroidism* is usually diagnosed in the context of other anterior pituitary hormone deficiencies; isolated TSH deficiency is very rare (Chap. 379). TSH levels may be low, normal, or even slightly increased in secondary hypothyroidism; the latter is due to secretion of immunoactive but bioinactive forms of TSH. The diagnosis is confirmed by detecting a low unbound T<sub>4</sub> level. The goal of treatment is to maintain T<sub>4</sub> levels in the upper half of the reference interval because TSH levels cannot be used to monitor therapy.

## TREATMENT

### Hypothyroidism

#### CLINICAL HYPOTHYROIDISM

If there is no residual thyroid function, the daily replacement dose of LT<sub>4</sub> is usually 1.6 µg/kg body weight (typically 100–150 µg), ideally taken at least 30 min before breakfast. In many patients, however, lower doses suffice until residual thyroid tissue is destroyed. In patients who develop hypothyroidism after the treatment of Graves' disease, there is often underlying autonomous function, necessitating lower replacement doses (typically 75–125 µg/d).

Adult patients under 60 years old without evidence of heart disease may be started on 50–100 µg of LT<sub>4</sub> daily. The dose is adjusted on the basis of TSH levels, with the goal of treatment being a normal TSH, ideally in the lower half of the reference range. TSH responses are gradual and should be measured about 2 months after instituting treatment or after any subsequent change in LT<sub>4</sub> dosage. The clinical effects of LT<sub>4</sub> replacement are slow to appear. Patients may not experience full relief from symptoms until 3–6 months after normal TSH levels are restored. Adjustment of LT<sub>4</sub> dosage is made in 12.5- or 25-µg increments if the TSH is high; decrements of the same magnitude should be made if the TSH is suppressed. Patients with a suppressed TSH of any cause, including LT<sub>4</sub> overtreatment, have an increased risk of atrial fibrillation and reduced bone density.

About 10–15% of patients may have persistent symptoms despite restoration of euthyroidism with LT<sub>4</sub> for reasons that remain unclear. Although desiccated animal thyroid preparations (thyroid extract USP) are available, they are not recommended because the

ratio of T<sub>3</sub> to T<sub>4</sub> is nonphysiologic. The use of LT4 combined with liothyronine (triiodothyronine, T<sub>3</sub>) has been investigated, but benefit has not been confirmed in prospective studies. There is no place for liothyronine alone as long-term replacement, because the short half-life necessitates three or four daily doses and is associated with fluctuating T<sub>3</sub> levels.

Once full replacement is achieved and TSH levels are stable, follow-up measurement of TSH is recommended at annual intervals. It is important to ensure ongoing adherence as patients do not feel any symptomatic difference after missing a few doses of LT4, and this sometimes leads to self-discontinuation.

In patients of normal body weight who are taking 200 µg of LT4 per d, an elevated TSH level is often a sign of poor adherence to treatment. This is also the likely explanation for fluctuating TSH levels, despite a constant LT4 dosage. Such patients often have normal or high unbound T<sub>4</sub> levels, despite an elevated TSH, because they remember to take medication for a few days before testing; this is sufficient to normalize T<sub>4</sub>, but not TSH levels. It is important to consider variable adherence, because this pattern of thyroid function tests is otherwise suggestive of disorders associated with inappropriate TSH secretion (Chap. 382). Because T<sub>4</sub> has a long half-life (7 days), patients who miss a dose can be advised to take two doses of the skipped tablets at once. Other causes of increased LT4 requirements must be excluded, particularly malabsorption (e.g., celiac disease, small-bowel surgery, atrophic or *Helicobacter pylori*-related gastritis), oral estrogen-containing medications or selective estrogen receptor modulator therapy, ingestion with a meal, and drugs that interfere with T<sub>4</sub> absorption or metabolism such as bile acid sequestrants, ferrous sulfate, calcium supplements, sevelamer, sucralfate, proton pump inhibitors, lovastatin, aluminum hydroxide, rifampicin, amiodarone, carbamazepine, phenytoin, and tyrosine kinase inhibitors.

### SUBCLINICAL HYPOTHYROIDISM

By definition, subclinical hypothyroidism refers to biochemical evidence of thyroid hormone deficiency in patients who have few or no apparent clinical features of hypothyroidism. There are no universally accepted recommendations for the management of subclinical hypothyroidism, but LT4 is recommended if the patient is a woman who wishes to conceive or is pregnant or when TSH levels are >10 mIU/L. Most other patients can simply be monitored annually. A trial of treatment may be considered when young or middle-aged patients have symptoms of hypothyroidism or risk of heart disease. It is important to confirm that any elevation of TSH is sustained over a 3-month period before treatment is given. Treatment is administered by starting with a low dose of LT4 (25–50 µg/d) with the goal of normalizing TSH.

### SPECIAL TREATMENT CONSIDERATIONS

Rarely, LT4 replacement is associated with pseudotumor cerebri in children. Presentation appears to be idiosyncratic and occurs months after treatment has begun.

Because maternal hypothyroidism may both adversely affect fetal neural development and be associated with adverse gestational outcomes (miscarriage, preterm delivery), thyroid function should be monitored to preserve euthyroidism in women with a history or high risk of hypothyroidism. Although epidemiologic studies have demonstrated the association of miscarriage and preterm delivery with the presence of thyroid autoantibodies detected either during or prior to gestation, randomized controlled trials evaluating LT4 therapy in this population have not demonstrated benefit. Because of the known increase in thyroid hormone requirements during pregnancy in hypothyroid women, LT4 therapy should be targeted to maintain a serum TSH in the normal range but <2.5 mIU/L prior to conception. Subsequently, thyroid function should be evaluated immediately after pregnancy is confirmed and every 4 weeks during the first half of the pregnancy, with less frequent testing after 20 weeks' gestation (every 6–8 weeks depending on whether LT4 dose adjustment is ongoing). The increment of

LT4 dosage increase depends upon the etiology of hypothyroidism, with athyreotic women requiring more (~45%) than those with Hashimoto's who may have some residual thyroid function. Women should increase LT4 from once-daily dosing to nine doses per week as soon as pregnancy is confirmed to anticipate this change. Thereafter dosage should be closely monitored with a goal TSH in the lower half of the trimester-specific normative range, if available, or <2.5 mIU/L which allows for reserve if additional LT4 dosage increases are required as pregnancy progresses. However, it is important to recognize that the normal TSH range in pregnancy for the second and third trimesters is not significantly different from the nonpregnancy reference range. However, serum TSH decreases in the late first trimester, and if trimester-specific ranges are not available, an appropriate range for 7–12 weeks' gestation can be approximated by decreasing the upper limit of the nonpregnant reference range by 0.5 mIU/L (~4.0 mIU/L) and the lower limit by 0.4 mIU/L (~0.1 mIU/L).

After delivery, LT4 doses typically return to prepregnancy levels. Pregnant women should be counseled to separate ingestion of prenatal vitamins and iron supplements from LT4.

Elderly patients may require 20% less thyroxine than younger patients. In the elderly, especially patients with known coronary artery disease, the starting dose of LT4 is 12.5–25 µg/d with similar increments every 2–3 months until TSH is normalized. In some patients, it may be impossible to achieve full replacement despite optimal antianginal treatment. *Emergency surgery* is generally safe in patients with untreated hypothyroidism, although routine surgery in a hypothyroid patient should be deferred until euthyroidism is achieved.

*Myxedema coma* still has a 20–40% mortality rate, despite intensive treatment, and outcomes are independent of the T<sub>4</sub> and TSH levels. Clinical manifestations include reduced level of consciousness, sometimes associated with seizures, as well as the other features of hypothyroidism (Table 383-3). Hypothermia can reach 23°C (74°F). There may be a history of treated hypothyroidism with poor compliance, or the patient may be previously undiagnosed. Myxedema coma almost always occurs in the elderly and is usually precipitated by factors that impair respiration, such as drugs (especially sedatives, anesthetics, and antidepressants), pneumonia, congestive heart failure, myocardial infarction, gastrointestinal bleeding, or cerebrovascular accidents. Sepsis should also be suspected. Exposure to cold may also be a risk factor. Hypoventilation, leading to hypoxia and hypercapnia, plays a major role in pathogenesis; hypoglycemia and dilutional hyponatremia also contribute to the development of myxedema coma.

LT4 can initially be administered as a single IV bolus of 200–400 µg, which serves as a loading dose, followed by a daily oral dose of 1.6 µg/kg per d, reduced by 25% if administered IV. If suitable IV preparation is not available, the same initial dose of LT4 can be given by nasogastric tube (although absorption may be impaired in myxedema). Because T<sub>4</sub> → T<sub>3</sub> conversion is impaired in myxedema coma, there is a rationale for adding liothyronine (T<sub>3</sub>) intravenously or via nasogastric tube to LT4 treatment, although excess liothyronine has the potential to provoke arrhythmias. An initial loading dose of 5–20 µg liothyronine should be followed by 2.5–10 µg every 8 h, with lower doses chosen for smaller or older patients and those at cardiovascular risk.

Supportive therapy should be provided to correct any associated metabolic disturbances. External warming is indicated only if the temperature is <30°C, as it can result in cardiovascular collapse (Chap. 464). Space blankets should be used to prevent further heat loss. Parenteral hydrocortisone (50 mg every 6 h) should be administered because there is impaired adrenal reserve in profound hypothyroidism. Any precipitating factors should be treated, including the early use of broad-spectrum antibiotics, pending the exclusion of infection. Ventilatory support with regular blood gas analysis is usually needed during the first 48 h. Hypertonic saline or IV glucose may be needed if there is severe hyponatremia or hypoglycemia; hypotonic IV fluids should be avoided because they

may exacerbate water retention secondary to reduced renal perfusion and inappropriate vasopressin secretion. The metabolism of most medications is impaired, and sedatives should be avoided if possible or used in reduced doses. Medication blood levels should be monitored, when available, to guide dosage.

## FURTHER READING

- Alexander EK et al: 2017 Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *Thyroid* 27:315, 2017.
- Biondi B et al: Subclinical hypothyroidism: A review. *JAMA* 322:153, 2019.
- Chaker L et al: Hypothyroidism. *Lancet* 390:1550, 2017.
- Ettleson MD et al: Individualized therapy for hypothyroidism: Is T<sub>4</sub> enough for everyone? *J Clin Endocrinol Metab* 105:e3090, 2020.
- Jonklaas J et al: Guidelines for the treatment of hypothyroidism: Prepared by the American Thyroid Association Task Force on thyroid hormone replacement. *Thyroid* 24:1670, 2014.
- Wassner AJ: Congenital hypothyroidism. *Clin Perinatol* 45:1, 2018.

# 384

## Hyperthyroidism and Other Causes of Thyrotoxicosis

J. Larry Jameson, Susan J. Mandel,  
Anthony P. Weetman



### THYROTOXICOSIS

**Thyrotoxicosis** is defined as the state of thyroid hormone excess and is not synonymous with **hyperthyroidism**, which is the result of excessive thyroid function. However, the major etiologies of thyrotoxicosis are hyperthyroidism caused by Graves' disease, toxic multinodular goiter (MNG), and toxic adenomas. Other causes are listed in **Table 384-1**.

### GRAVES' DISEASE

**Epidemiology** Graves' disease accounts for 60–80% of thyrotoxicosis. The prevalence varies among populations, reflecting genetic factors and iodine intake (high iodine intake is associated with an increased prevalence of Graves' disease). Graves' disease occurs in up to 2% of women but is one-tenth as frequent in men. The disorder rarely begins before adolescence and typically occurs between 20 and 50 years of age; it also occurs in the elderly.

**Pathogenesis** As in autoimmune hypothyroidism, a combination of environmental and genetic factors, including polymorphisms in HLA-DR, the immunoregulatory genes *CTLA-4*, *CD25*, *PTPN22*, *FCRL3*, and *CD226*, as well as the gene encoding the thyroid-stimulating hormone (TSH) receptor (TSH-R), contributes to Graves' disease susceptibility. The concordance for Graves' disease in monozygotic twins is 20–30%, compared to <5% in dizygotic twins. Indirect evidence suggests that stress is an important environmental factor, presumably operating through neuroendocrine effects on the immune system. Smoking is a minor risk factor for Graves' disease and a major risk factor for the development of ophthalmopathy. Sudden increases in iodine intake may precipitate Graves' disease, and there is a threefold increase in the occurrence of Graves' disease in the postpartum period. Graves' disease may occur during the immune reconstitution phase after highly active antiretroviral therapy (HAART) or alemtuzumab treatment and following treatment with immune checkpoint inhibitors (e.g., nivolumab, pembrolizumab).

**TABLE 384-1 Causes of Thyrotoxicosis**

#### Primary Hyperthyroidism

- Graves' disease
- Toxic multinodular goiter
- Toxic adenoma
- Functioning thyroid carcinoma metastases
- Activating mutation of the TSH receptor
- Activating mutation of G<sub>s</sub>α (McCune-Albright syndrome)
- Struma ovarii
- Drugs: iodine excess (Jod-Basedow phenomenon)

#### Thyrotoxicosis without Hyperthyroidism

- Subacute thyroiditis
- Silent thyroiditis
- Other causes of thyroid destruction: amiodarone, radiation, infarction of adenoma
- Ingestion of excess thyroid hormone (thyrotoxicosis factitia) or thyroid tissue

#### Secondary Hyperthyroidism

- TSH-secreting pituitary adenoma
- Thyroid hormone resistance syndrome: occasional patients may have features of thyrotoxicosis
- Chorionic gonadotropin-secreting tumors<sup>a</sup>
- Gestational thyrotoxicosis<sup>a</sup>

<sup>a</sup>Circulating TSH levels are low in these forms of secondary hyperthyroidism.

Abbreviation: TSH, thyroid-stimulating hormone.

The hyperthyroidism of Graves' disease is caused by thyroid-stimulating immunoglobulins (TSIs) that are synthesized by lymphocytes in the thyroid gland as well as in bone marrow and lymph nodes. Such antibodies can be detected by bioassays or by using the more widely available immunoassays (TSH receptor antibody [TRAb]) that measure whether the patient's serum contains an antibody that can displace either labeled TSH or a monoclonal TSH receptor antibody from the TSH receptor. The presence of TRAb in a patient with thyrotoxicosis implies the existence of TSI, and these assays are useful in monitoring pregnant Graves' patients in whom high levels of TSI can cross the placenta and cause neonatal thyrotoxicosis. Other thyroid autoimmune responses, similar to those in autoimmune hypothyroidism (see above), occur concurrently in patients with Graves' disease. In particular, thyroid peroxidase (TPO) and thyroglobulin (Tg) antibodies occur in up to 80% of cases. Because the coexisting thyroiditis can also affect thyroid function, there is no direct correlation between the level of TSI and thyroid hormone levels in Graves' disease.

Cytokines appear to play a major role in thyroid-associated ophthalmopathy. There is infiltration of the extraocular muscles by activated T cells; the release of cytokines such as interferon (IFN- $\gamma$ ), tumor necrosis factor (TNF), and interleukin 1 (IL-1) results in fibroblast activation and increased synthesis of glycosaminoglycans that trap water, thereby leading to characteristic muscle swelling. Late in the disease, there is irreversible fibrosis of the muscles. Increased fat is an additional cause of retrobulbar tissue expansion. The increase in intraorbital pressure can lead to proptosis, diplopia, and optic neuropathy. Although the pathogenesis of thyroid-associated ophthalmopathy is incompletely understood, the TSH-R is a thyroid autoantigen and is expressed in orbital tissues. In addition, aberrant signaling via insulin-like growth factor 1 receptors (IGF-1R) on orbital fibroblasts has also been implicated. These mechanisms are the basis for new monoclonal antibody treatments (e.g., tepratuzumab) that reduce the levels of TSH-R/IGF-1R complexes and attenuate signaling.

**Clinical Manifestations** Signs and symptoms include features that are common to any cause of thyrotoxicosis (**Table 384-2**) as well as those specific for Graves' disease. The clinical presentation depends on the severity of thyrotoxicosis, the duration of disease, individual susceptibility to excess thyroid hormone, and the patient's age. In the

**TABLE 384-2 Signs and Symptoms of Thyrotoxicosis (Descending Order of Frequency)**

SYMPTOMS	SIGNS <sup>a</sup>
Hyperactivity, irritability, dysphoria	Tachycardia; atrial fibrillation in the elderly
Heat intolerance and sweating	
Palpitations	Tremor
Fatigue and weakness	Goiter
Weight loss with increased appetite	Warm, moist skin
Diarrhea	Muscle weakness, proximal myopathy
Polyuria	Lid retraction or lag
Oligomenorrhea, loss of libido	Gynecomastia

<sup>a</sup>Excludes the signs of ophthalmopathy and dermopathy specific for Graves' disease.

elderly, features of thyrotoxicosis may be subtle or masked, and patients may present mainly with fatigue and weight loss, a condition known as *apathetic thyrotoxicosis*.

Thyrotoxicosis may cause unexplained weight loss, despite an enhanced appetite, due to the increased metabolic rate. Weight gain occurs in 5% of patients, however, because of increased food intake. Other prominent features include hyperactivity, nervousness, and irritability, ultimately leading to a sense of easy fatigability in some patients. Insomnia and impaired concentration are common; apathetic thyrotoxicosis may be mistaken for depression in the elderly. Fine tremor is a frequent finding, best elicited by having patients stretch out their fingers while feeling the fingertips with the palm. Common neurologic manifestations include hyperreflexia, muscle wasting, and proximal myopathy without fasciculation. Chorea is rare. Thyrotoxicosis is sometimes associated with a form of hypokalemic periodic paralysis; this disorder is particularly common in Asian males with thyrotoxicosis, but it occurs in other ethnic groups as well.

The most common cardiovascular manifestation is sinus tachycardia, often associated with palpitations, occasionally caused by supraventricular tachycardia. The high cardiac output produces a bounding pulse, widened pulse pressure, and an aortic systolic murmur and can lead to worsening of angina or heart failure in the elderly or those with pre-existing heart disease. Atrial fibrillation is more common in patients >50 years of age. Treatment of the thyrotoxic state alone converts atrial fibrillation to normal sinus rhythm in about half of patients, suggesting the existence of an underlying cardiac problem in the remainder.

The skin is usually warm and moist, and the patient may complain of sweating and heat intolerance, particularly during warm weather. Palmar erythema, onycholysis, and, less commonly, pruritus, urticaria, and diffuse hyperpigmentation may be evident. Hair texture may become fine, and a diffuse alopecia occurs in up to 40% of patients, persisting for months after restoration of euthyroidism. Gastrointestinal transit time is decreased, leading to increased stool frequency, often with diarrhea and occasionally mild steatorrhea. Women frequently experience oligomenorrhea or amenorrhea; in men, there may be impaired sexual function and, rarely, gynecomastia. The direct effect of thyroid hormones on bone resorption leads to osteopenia in long-standing thyrotoxicosis; mild hypercalcemia occurs in up to 20% of patients, but hypercalcuria is more common. There is a small increase in fracture rate in patients with a previous history of thyrotoxicosis.

In Graves' disease, the thyroid is usually diffusely enlarged to two to three times its normal size. The consistency is firm, but not nodular. There may be a thrill or bruit, best detected at the inferolateral margins of the thyroid lobes, due to the increased vascularity of the gland and the hyperdynamic circulation.

Lid retraction, causing a staring appearance, can occur in any form of thyrotoxicosis and is the result of sympathetic overactivity. However, Graves' disease is associated with specific eye signs that comprise *Graves' ophthalmopathy* (Fig. 384-1A). This condition is also called *thyroid-associated ophthalmopathy*, because it occurs in the absence of hyperthyroidism in 10% of patients. Most of these individuals have autoimmune hypothyroidism or thyroid antibodies. The onset



**FIGURE 384-1** Features of Graves' disease. **A.** Ophthalmopathy in Graves' disease; lid retraction, periorbital edema, conjunctival injection, and proptosis are marked. **B.** Thyroid dermopathy over the lateral aspects of the shins. **C.** Thyroid acropachy.

of Graves' ophthalmopathy occurs within the year before or after the diagnosis of thyrotoxicosis in 75% of patients but can sometimes precede or follow thyrotoxicosis by several years, accounting for some cases of euthyroid ophthalmopathy.

About one-third of patients with Graves' disease have clinical evidence of ophthalmopathy. However, the enlarged extraocular muscles typical of the disease, and other subtle features, can be detected in most patients when investigated by ultrasound or computed tomography (CT) imaging of the orbits. Unilateral signs are found in up to 10% of ophthalmopathy patients. The earliest manifestations of ophthalmopathy are usually a sensation of grittiness, eye discomfort, and excess tearing. About one-third of patients have proptosis, best detected by visualization of the sclera between the lower border of the iris and the lower eyelid, with the eyes in the primary position. Proptosis can be measured using an exophthalmometer. In severe cases, proptosis may cause corneal exposure and damage, especially if the lids fail to close during sleep. Periorbital edema, scleral injection, and chemosis are also frequent. In 5–10% of patients, the muscle swelling is so severe that diplopia results, typically, but not exclusively, when the patient looks up and laterally. The most serious manifestation is compression of the optic nerve at the apex of the orbit, leading to papilledema; peripheral field defects; and, if left untreated, permanent loss of vision.

The “NO SPECS” scoring system to evaluate ophthalmopathy is an acronym derived from the following changes:

- 0 = No signs or symptoms
- 1 = Only signs (lid retraction or lag), no symptoms
- 2 = Soft tissue involvement (periorbital edema)
- 3 = Proptosis (>22 mm)
- 4 = Extraocular muscle involvement (diplopia)
- 5 = Corneal involvement
- 6 = Sight loss

Although useful as a mnemonic, the NO SPECS scheme is inadequate to describe the eye disease fully, and patients do not necessarily progress from one class to another; alternative scoring systems (e.g., the EUGOGO system developed by the European Group on Graves' Orbitopathy) that assess disease activity are preferable for monitoring and treatment purposes. When Graves' eye disease is active and severe, referral to an ophthalmologist is indicated and objective measurements are needed, such as lid-fissure width; corneal staining with fluorescein;

**Thyroid dermopathy** occurs in <5% of patients with Graves' disease (Fig. 384-1B), almost always in the presence of moderate or severe ophthalmopathy. Although most frequent over the anterior and lateral aspects of the lower leg (hence the term *pretibial myxedema*), skin changes can occur at other sites, particularly after trauma. The typical lesion is a noninflamed, indurated plaque with a deep pink or purple color and an "orange skin" appearance. Nodular involvement can occur, and the condition can rarely extend over the whole lower leg and foot, mimicking elephantiasis. **Thyroid acropachy** refers to a form of clubbing found in <1% of patients with Graves' disease (Fig. 384-1C). It is so strongly associated with thyroid dermopathy that an alternative cause of clubbing should be sought in a Graves' patient without coincident skin and orbital involvement. Ophthalmopathy, dermopathy, and acropachy have declined in incidence, probably due to better recognition and prompt treatment of the underlying thyroid disease.

**Laboratory Evaluation** Investigations used to determine the existence and cause of thyrotoxicosis are summarized in Fig. 384-2. In Graves' disease, the TSH level is suppressed, and total and unbound thyroid hormone levels are increased. In 2–5% of patients (and more in areas of borderline iodine intake), only  $T_3$  is increased ( $T_3$  toxicosis). The converse state of  $T_4$  toxicosis, with elevated total and unbound  $T_4$  and normal  $T_3$  levels, is occasionally seen when hyperthyroidism is induced by excess iodine, providing surplus substrate for thyroid hormone synthesis. Measurement of TPO antibodies or TRAb may be useful if the diagnosis is unclear clinically but is not needed routinely. Associated abnormalities that may cause diagnostic confusion in thyrotoxicosis include elevation of bilirubin, liver enzymes, and ferritin. Microcytic anemia and thrombocytopenia may occur.

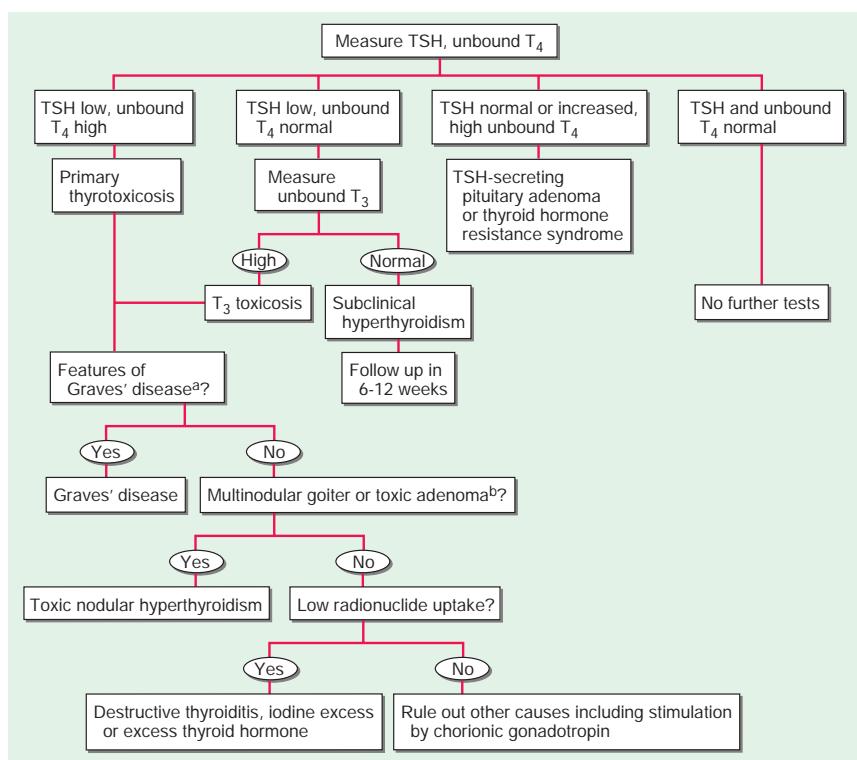
**Differential Diagnosis** Diagnosis of Graves' disease is straightforward in a patient with biochemically confirmed thyrotoxicosis, diffuse

goiter on palpation, ophthalmopathy, and often a personal or family history of autoimmune disorders. For patients with thyrotoxicosis who lack these features, the diagnosis can be established by a radionuclide ( $^{99m}\text{Tc}$ ,  $^{123}\text{I}$ , or  $^{131}\text{I}$ ) scan and uptake of the thyroid, which will distinguish the diffuse, high uptake of Graves' disease from destructive thyroiditis, ectopic thyroid tissue, and factitious thyrotoxicosis, as well as diagnose a toxic adenoma or toxic MNG. Increasingly, because of the rapidity of laboratory test results, TRAb measurement is used instead of radionuclide scanning to confirm the diagnosis of Graves' disease. Color-flow Doppler ultrasonography may distinguish between hyperthyroidism (with increased blood flow) and destructive thyroiditis. In secondary hyperthyroidism due to a TSH-secreting pituitary tumor, there is also a diffuse goiter. The presence of a nonsuppressed TSH level and the finding of a pituitary tumor on CT or magnetic resonance imaging (MRI) scan suggest this diagnosis.

Clinical features of thyrotoxicosis can mimic certain aspects of other disorders, including panic attacks, mania, pheochromocytoma, and weight loss associated with malignancy. The diagnosis of thyrotoxicosis can be easily excluded if the TSH and unbound  $T_4$  and  $T_3$  levels are normal. A normal TSH also excludes Graves' disease as a cause of diffuse goiter.

**Clinical Course** Clinical features generally worsen without treatment; mortality was 10–30% before the introduction of satisfactory therapy. Some patients with mild Graves' disease experience spontaneous relapses and remissions. Rarely, there may be fluctuation between hypo- and hyperthyroidism due to changes in the functional activity of TSH-R antibodies. About 15% of patients who enter remission after treatment develop hypothyroidism 10–15 years later as a result of the destructive autoimmune process.

The clinical course of ophthalmopathy does not follow that of the thyroid disease, although thyroid dysfunction can worsen eye signs. Ophthalmopathy typically worsens over the initial 3–6 months, followed by a plateau phase over the next 12–18 months,



**FIGURE 384-2** Evaluation of thyrotoxicosis. <sup>a</sup>Diffuse goiter, positive thyroid peroxidase (TPO) antibodies or thyroid-stimulating hormone (TSH) receptor antibody (TRAb), ophthalmopathy, dermopathy. <sup>b</sup>Can be confirmed by radionuclide scan.

and then some spontaneous improvement, particularly in the soft tissue changes. However, the course is more fulminant in up to 5% of patients, requiring intervention in the acute phase if there is optic nerve compression or corneal ulceration. Diplopia may appear late in the disease due to fibrosis of the extraocular muscles. Radioiodine treatment for hyperthyroidism worsens the eye disease in a small proportion of patients (especially smokers). Antithyroid drugs and surgery have no adverse effects on the clinical course of ophthalmopathy. Thyroid dermopathy, when it occurs, usually appears 1–2 years after the development of Graves' hyperthyroidism; it may improve spontaneously.

## TREATMENT

### Graves' Disease

The *hyperthyroidism* of Graves' disease is treated by reducing thyroid hormone synthesis, using an antithyroid drug, or reducing the amount of thyroid tissue with radioiodine ( $^{131}\text{I}$ ) treatment or by thyroidectomy. Antithyroid drugs are the predominant therapy in many centers in Europe, Latin America, and Japan, whereas radioiodine is more often the first line of treatment in North America. These differences reflect the fact that no single approach is optimal and that patients may require multiple treatments to achieve remission.

The main *antithyroid drugs* are thionamides: propylthiouracil, carbimazole (not available in the United States), and the active metabolite of the latter, methimazole. All inhibit the function of TPO, reducing oxidation and organification of iodide. These drugs also reduce thyroid antibody levels by mechanisms that remain unclear, and they appear to enhance spontaneous rates of remission. Propylthiouracil inhibits deiodination of  $\text{T}_4 \rightarrow \text{T}_3$ . However, this effect is of minor benefit, except in the most severe thyrotoxicosis, and is offset by the much shorter half-life of this drug (90 min) compared to methimazole (6 h). Due to the hepatotoxicity of propylthiouracil, the U.S. Food and Drug Administration (FDA) has limited indications for its use to the first trimester of pregnancy, the treatment of thyroid storm, and patients with minor adverse reactions to methimazole. If propylthiouracil is used, monitoring of liver function tests is recommended.

There are many variations of antithyroid drug regimens. The initial dose of carbimazole or methimazole is usually 10–20 mg every 8 or 12 h, but once-daily dosing is possible after euthyroidism is restored. Propylthiouracil is given at a dose of 100–200 mg every 6–8 h, and divided doses are usually given throughout the course. Lower doses of each drug may suffice in areas of low iodine intake. The starting dose of an antithyroid drug can be gradually reduced (titration regimen) as thyrotoxicosis improves. Less commonly, high doses may be given combined with levothyroxine (LT4) supplementation (block-replace regimen) to avoid drug-induced hypothyroidism. The titration regimen is preferred to minimize the dose of antithyroid drug and provide an index of treatment response.

Thyroid function tests and clinical manifestations are reviewed 4–6 weeks after starting treatment, and the dose is titrated based on unbound  $\text{T}_4$  levels. Most patients do not achieve euthyroidism until 6–8 weeks after treatment is initiated. TSH levels often remain suppressed for several months and therefore do not provide a sensitive index of treatment response. The usual daily maintenance doses of antithyroid drugs in the titration regimen are 2.5–10 mg of carbimazole or methimazole and 50–100 mg of propylthiouracil. In the block-replace regimen, the initial dose of antithyroid drug is held constant, and the dose of LT4 is adjusted to maintain normal unbound  $\text{T}_4$  levels. When TSH suppression is alleviated, TSH levels can also be used to monitor therapy.

Maximum remission rates (up to 30–60% in some populations) are achieved by 12–18 months for the titration regimen compared to 6 months for the block-replace regimen and are higher in patients in whom TRAb levels are no longer detected than in those with TRAb persistence. For unclear reasons, remission rates appear

to vary in different geographic regions. Younger patients, males, smokers, and patients with a history of allergy, severe hyperthyroidism, or large goiters are most likely to relapse when treatment stops, but outcomes are difficult to predict. All patients should be followed closely for relapse during the first year after treatment and at least annually thereafter.

The common minor side effects of antithyroid drugs are rash, urticaria, fever, and arthralgia (1–5% of patients). These may resolve spontaneously or after substituting an alternative antithyroid drug; rashes may respond to an antihistamine. Rare but major side effects include hepatitis (especially with propylthiouracil; avoid use in children) and cholestasis (methimazole and carbimazole); vasculitis; and, most important, agranulocytosis (<1%). It is essential that antithyroid drugs are stopped and not restarted if a patient develops major side effects. Written instructions should be provided regarding the symptoms of possible agranulocytosis (e.g., sore throat, fever, mouth ulcers) and the need to stop treatment pending an urgent complete blood count to confirm that agranulocytosis is not present. Management of agranulocytosis is described in **Chap. 102**. It is not useful to monitor blood counts prospectively, because the onset of agranulocytosis is idiosyncratic and abrupt.

*Propranolol* (20–40 mg every 6 h) or longer-acting selective  $\beta_1$  receptor blockers such as atenolol may be helpful to control adrenergic symptoms, especially in the early stages before antithyroid drugs take effect. Beta blockers are also useful in patients with thyrotoxic periodic paralysis, pending correction of thyrotoxicosis. In consultation with a cardiologist, anticoagulation should be considered in all patients with atrial fibrillation; there is often spontaneous reversion to sinus rhythm with control of hyperthyroidism, and long-term anticoagulation is not usually needed. Decreased warfarin doses are required when patients are thyrotoxic. If digoxin is used, increased doses are often needed in the thyrotoxic state.

*Radioiodine* causes progressive destruction of thyroid cells and can be used as initial treatment or for relapses after a trial of antithyroid drugs. There is a small risk of thyrotoxic crisis (see below) after radioiodine, which can be minimized by pretreatment with antithyroid drugs for at least a month before treatment. Antecedent treatment with an antithyroid drug and a beta blocker should be considered for all elderly patients or for those with cardiac problems. Carbimazole or methimazole must be stopped 2–3 days before radioiodine administration to achieve optimum iodine uptake and can be restarted 3–7 days after radioiodine in those at risk of complications from worsening thyrotoxicosis. Propylthiouracil appears to have a prolonged radioprotective effect and should be stopped for a longer period before radioiodine is given, or a larger dose of radioiodine will be necessary.

Efforts to calculate an optimal dose of radioiodine that achieves euthyroidism without a high incidence of relapse or progression to hypothyroidism have not been successful. Some patients inevitably relapse after a single dose because the biologic effects of radiation vary between individuals, and hypothyroidism cannot be uniformly avoided even using accurate dosimetry. A practical strategy is to give a fixed dose based on clinical features, such as the severity of thyrotoxicosis, the size of the goiter (increases the dose needed), and the level of radioiodine uptake (decreases the dose needed).  $^{131}\text{I}$  dosage generally ranges between 370 MBq (10 mCi) and 555 MBq (15 mCi). Most authorities favor an approach aimed at thyroid ablation (as opposed to euthyroidism), given that LT4 replacement is straightforward and most patients ultimately progress to hypothyroidism over 5–10 years, frequently with some delay in the diagnosis of hypothyroidism.

Certain radiation safety precautions are necessary in the first few days after radioiodine treatment, but the exact guidelines vary depending on local protocols. In general, patients need to avoid close, prolonged contact with children and pregnant women for 5–7 days because of possible transmission of residual isotope and exposure to radiation emanating from the gland. Rarely, there may be mild pain due to radiation thyroiditis 1–2 weeks after treatment. Hyperthyroidism can persist for 2–3 months before radioiodine

takes full effect. For this reason,  $\beta$ -adrenergic blockers or antithyroid drugs can be used to control symptoms during this interval. Persistent hyperthyroidism can be treated with a second dose of radioiodine, usually 6 months after the first dose. The risk of hypothyroidism after radioiodine depends on the dosage but is at least 10–20% in the first year and 5% per year thereafter. Patients should be informed of this possibility before treatment and require close follow-up during the first year followed by annual thyroid function testing.

Pregnancy and breast-feeding are absolute contraindications to radioiodine treatment, but patients can conceive safely 6 months after treatment. The presence of ophthalmopathy, especially in smokers, requires caution. Prednisone, 0.2–0.5 mg/kg per d (depending on ophthalmopathy severity), at the time of radioiodine treatment, tapered over 6–12 weeks, may prevent exacerbation of ophthalmopathy, but radioiodine should generally be avoided in patients with active moderate to severe eye disease. Although many physicians avoid radioiodine in children and adolescents because of the potential risks of malignancy, others have advocated radioiodine use in older children. A recent long-term follow-up study of adults found a modest increased lifetime risk of solid cancers after radioiodine treatment, contrary to previous findings. It is unclear how this will alter management in the future.

*Total or near-total thyroidectomy* is an option for patients who relapse after antithyroid drugs and prefer this treatment to radioiodine. Some experts recommend surgery in young individuals, particularly when the goiter is very large. Careful control of thyrotoxicosis with antithyroid drugs, followed by potassium iodide (SSKI; 1–2 drops orally tid for 10 days), is needed prior to surgery to avoid thyrotoxic crisis and to reduce the vascularity of the gland. The major complications of surgery—bleeding, laryngeal edema, hypoparathyroidism, and damage to the recurrent laryngeal nerves—are unusual when the procedure is performed by highly experienced surgeons. Recurrence rates in the best series are <2%, but the rate of hypothyroidism is similar to that following radioiodine treatment, especially with the current trend away from subtotal thyroidectomy.

Antithyroid drugs should be used to manage Graves' disease in *pregnancy*. Because transplacental passage of these drugs may produce fetal hypothyroidism and goiter if the maternal dose is excessive, maternal antithyroid dose titration should target serum free or total T<sub>4</sub> levels at or just above the pregnancy reference range. If available, propylthiouracil should be used until 14–16 weeks' gestation because of the association of rare cases of methimazole/carbimazole embryopathy, including *aplasia cutis* and other defects, such as choanal atresia and tracheoesophageal fistulae. Because of the potential for teratogenic effects, recent recommendations suggest discontinuation of antithyroid medication in a newly pregnant woman with Graves' disease who is euthyroid on a low dose of methimazole (<5–10 mg/d) or propylthiouracil (<100–200 mg/d), after evaluating recent thyroid function tests, disease history, goiter size, duration of therapy, and TRAb measurement. Following cessation, careful monitoring of maternal thyroid function tests is essential. On the other hand, for women at high risk of developing thyrotoxicosis if antithyroid drugs are discontinued (large goiter, requirement for higher antithyroid drug dosage), continued therapy is necessary, with propylthiouracil (if available) administration in the first trimester. However, because of its rare association with hepatotoxicity, propylthiouracil should be limited to the first trimester and then maternal therapy should be converted to methimazole (or carbimazole) at a ratio of 15–20 mg of propylthiouracil to 1 mg of methimazole. It is often possible to stop treatment in the last trimester because TSIs tend to decline in pregnancy. Nonetheless, the transplacental transfer of these antibodies if present at levels three times higher than the normative range rarely causes *fetal* or *neonatal thyrotoxicosis*. Poor intrauterine growth, a fetal heart rate of >160 beats/min, advanced bone age, fetal goiter, and high levels of maternal TSI after 26 weeks' gestation may herald this complication. Antithyroid drugs given to the mother can be used

to treat the fetus and may be needed for 1–3 months after delivery, until the maternal antibodies disappear from the baby's circulation. The postpartum period is a time of major risk for relapse of Graves' disease. Breast-feeding is safe with low doses of antithyroid drugs. Graves' disease in *children* is usually managed initially with methimazole or carbimazole (avoid propylthiouracil), often given as a prolonged course of the titration regimen. Surgery or radioiodine may be indicated for severe or relapsing disease.

*Thyrotoxic crisis*, or *thyroid storm*, is rare and presents as a life-threatening exacerbation of hyperthyroidism, accompanied by fever, delirium, seizures, coma, vomiting, diarrhea, and jaundice. The mortality rate due to cardiac failure, arrhythmia, or hyperthermia is as high as 30%, even with treatment. Thyrotoxic crisis is usually precipitated by acute illness (e.g., stroke, infection, trauma, diabetic ketoacidosis), surgery (especially on the thyroid), or radioiodine treatment of a patient with partially treated or untreated hyperthyroidism. Management requires intensive monitoring and supportive care, identification and treatment of the precipitating cause, and measures that reduce thyroid hormone synthesis. Large doses of propylthiouracil (500–1000 mg loading dose and 250 mg every 4 h) should be given orally or by nasogastric tube or per rectum; the drug's inhibitory action on T<sub>4</sub>  $\rightarrow$  T<sub>3</sub> conversion makes it the antithyroid drug of choice. If not available, methimazole can be used in doses of 20 mg every 6 h. One hour after the first dose of propylthiouracil, stable iodide (5 drops SSKI every 6 h) is given to block thyroid hormone synthesis via the Wolff-Chaikoff effect (the delay allows the antithyroid drug to prevent the excess iodine from being incorporated into new hormone). Propranolol should also be given to reduce tachycardia and other adrenergic manifestations (60–80 mg PO every 4 h, or 2 mg IV every 4 h). Although other  $\beta$ -adrenergic blockers can be used, high doses of propranolol decrease T<sub>4</sub>  $\rightarrow$  T<sub>3</sub> conversion, and the doses can be easily adjusted. Caution is needed to avoid acute negative inotropic effects, but controlling the heart rate is important, as some patients develop a form of high-output heart failure. Short-acting IV esmolol can be used to decrease heart rate while monitoring for signs of heart failure. Additional therapeutic measures include glucocorticoids (e.g., hydrocortisone 300 mg IV bolus, then 100 mg every 8 h), antibiotics if infection is present, cholestyramine to sequester thyroid hormones, cooling, oxygen, and IV fluids.

*Ophthalmopathy* requires no active treatment when it is mild or moderate, because there is usually spontaneous improvement. General measures include meticulous control of thyroid hormone levels, cessation of smoking, and an explanation of the natural history of ophthalmopathy. Discomfort can be relieved with artificial tears (e.g., hyromellose 0.3% or carbomer 0.2% ophthalmic gel), paraffin-based eye ointment, and the use of dark glasses with side frames. Periorbital edema may respond to a more upright sleeping position or a diuretic. Corneal exposure during sleep can be avoided by using patches or taping the eyelids shut. Minor degrees of diplopia improve with prisms fitted to spectacles. Some authorities also advocate selenium 100 µg bid. Severe ophthalmopathy, with optic nerve involvement or chemosis resulting in corneal damage, is an emergency requiring joint management with an ophthalmologist. Pulse therapy with IV methylprednisolone (e.g., 500 mg of methylprednisolone once weekly for 6 weeks, then 250 mg once weekly for 6 weeks) is preferable to oral glucocorticoids, which are used for moderately active disease. When glucocorticoids are ineffective, orbital decompression can be achieved by removing bone from any wall of the orbit, thereby allowing displacement of fat and swollen extraocular muscles. The transantral route is used most often because it requires no external incision. Proptosis recedes an average of 5 mm, but there may be residual or even worsened diplopia. Once the eye disease has stabilized, surgery may be indicated for relief of diplopia and correction of the appearance. External beam radiotherapy of the orbits has been used for many years, but the efficacy of this therapy remains unclear, and it is best reserved for those with moderately active disease who have failed or are not candidates for glucocorticoid therapy.

Teprotumumab, a human monoclonal antibody, received breakthrough designation and was approved by the FDA in 2020. Randomized clinical trials of patients with active thyroid eye disease demonstrated rapid effects on proptosis, diplopia, clinical activity score, and quality of life. Responses appear comparable to surgery. Teprotumumab is administered at an initial dose of 10 mg/kg IV and thereafter at 20 mg/kg IV every 3 weeks for 21 weeks.

*Thyroid dermopathy* does not usually require treatment, but it can cause cosmetic problems or interfere with the fit of shoes. Surgical removal is not indicated. If necessary, treatment consists of topical, high-potency glucocorticoid ointment under an occlusive dressing. Octreotide may be beneficial in some cases.

### OTHER CAUSES OF THYROTOXICOSIS

Destructive thyroiditis (subacute or silent thyroiditis) typically presents with a short thyrotoxic phase due to the release of preformed thyroid hormones and catabolism of Tg (see “Subacute Thyroiditis,” below). True hyperthyroidism is absent, as demonstrated by a low radionuclide uptake. Circulating Tg levels are typically increased. Other causes of thyrotoxicosis with low or absent thyroid radionuclide uptake include *thyrotoxicosis factitia*, iodine excess, and, rarely, ectopic thyroid tissue, particularly teratomas of the ovary (*struma ovarii*) and functional metastatic follicular carcinoma. Whole-body radionuclide studies can demonstrate ectopic thyroid tissue, and thyrotoxicosis factitia can be distinguished from destructive thyroiditis by the clinical features and low levels of Tg. Amiodarone treatment is associated with thyrotoxicosis in up to 10% of patients, particularly in areas of low iodine intake (see below).

*TSH-secreting pituitary adenoma* is a rare cause of thyrotoxicosis. It is characterized by the presence of an inappropriately normal or increased TSH level in a patient with hyperthyroidism, diffuse goiter, and elevated T<sub>4</sub> and T<sub>3</sub> levels (Chap. 380). Elevated levels of the subunit of TSH, released by the TSH-secreting adenoma, support this diagnosis, which can be confirmed by demonstrating the pituitary tumor on MRI or CT scan. A combination of transsphenoidal surgery, sella irradiation, and octreotide may be required to normalize TSH, because many of these tumors are large and locally invasive at the time of diagnosis. Radioiodine or antithyroid drugs can be used to control thyrotoxicosis.

Thyrotoxicosis caused by *toxic MNG* and *hyperfunctioning solitary nodules* is discussed below.

### THYROIDITIS

There are several classification systems to describe the clinical syndromes of thyroiditis. One is based on the onset and duration of disease (Table 384-3); others are based on the absence or presence of pain.

**TABLE 384-3 Causes of Thyroiditis**

Acute
Bacterial infection: especially <i>Staphylococcus</i> , <i>Streptococcus</i> , and <i>Enterobacter</i>
Fungal infection: <i>Aspergillus</i> , <i>Candida</i> , <i>Coccidioides</i> , <i>Histoplasma</i> , and <i>Pneumocystis</i>
Radiation thyroiditis after <sup>131</sup> I treatment
Amiodarone (may also be subacute or chronic)
Subacute
Viral (or granulomatous) thyroiditis
Silent thyroiditis (including postpartum thyroiditis)
Mycobacterial infection
Drug induced (interferon, amiodarone)
Chronic
Autoimmunity: focal thyroiditis, Hashimoto's thyroiditis, atrophic thyroiditis
Riedel's thyroiditis
Parasitic thyroiditis: echinococcosis, strongyloidiasis, cysticercosis
Traumatic: after palpation

### ACUTE THYROIDITIS

Acute thyroiditis is rare and due to suppurative infection of the thyroid. In children and young adults, the most common cause is the presence of a piriform sinus, a remnant of the fourth branchial pouch that connects the oropharynx with the thyroid. Such sinuses are predominantly left-sided. A long-standing goiter and degeneration in a thyroid malignancy are risk factors in the elderly. The patient presents with thyroid pain, often referred to the throat or ears, and a small, tender goiter that may be asymmetric. Fever, dysphagia, and erythema over the thyroid are common, as are systemic symptoms of a febrile illness and lymphadenopathy.

The differential diagnosis of *thyroid pain* includes subacute or, rarely, chronic thyroiditis; hemorrhage into a cyst; malignancy including lymphoma; and, rarely, amiodarone-induced thyroiditis or amyloidosis. However, the abrupt presentation and clinical features of acute thyroiditis rarely cause confusion. The erythrocyte sedimentation rate (ESR) and white cell count are usually increased, but thyroid function is normal. Fine-needle aspiration (FNA) biopsy shows infiltration by polymorphonuclear leukocytes; culture of the sample can identify the organism. Caution is needed in immunocompromised patients as fungal, mycobacterial, or *Pneumocystis* thyroiditis can occur in this setting. Antibiotic treatment is guided initially by Gram stain and, subsequently, by cultures of the FNA biopsy. Surgery may be needed to drain an abscess, which can be localized by CT scan or ultrasound. Tracheal obstruction, septicemia, retropharyngeal abscess, mediastinitis, and jugular venous thrombosis may complicate acute thyroiditis but are uncommon with prompt use of antibiotics.

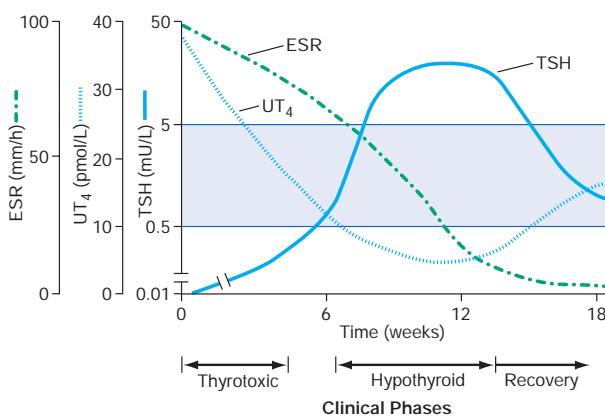
### SUBACUTE THYROIDITIS

This is also termed *de Quervain's thyroiditis*, *granulomatous thyroiditis*, or *viral thyroiditis*. Many viruses have been implicated, including mumps, coxsackie, influenza, adenoviruses, and echoviruses, but attempts to identify the virus in an individual patient are often unsuccessful and do not influence management. Recently, subacute thyroiditis associated with the SARS-CoV-2 was described. The diagnosis of subacute thyroiditis is often overlooked because the symptoms can mimic pharyngitis. The peak incidence occurs at 30–50 years, and women are affected three times more frequently than men.

**Pathophysiology** The thyroid shows a characteristic patchy inflammatory infiltrate with disruption of the thyroid follicles and multinucleated giant cells within some follicles. The follicular changes progress to granulomas accompanied by fibrosis. Finally, the thyroid returns to normal, usually several months after onset. During the initial phase of follicular destruction, there is release of Tg and thyroid hormones, leading to increased circulating T<sub>4</sub> and T<sub>3</sub> and suppression of TSH (Fig. 384-3). During this destructive phase, radioactive iodine uptake is low or undetectable. After several weeks, the thyroid is depleted of stored thyroid hormone and a phase of hypothyroidism typically occurs, with low unbound T<sub>4</sub> (and sometimes T<sub>3</sub>) and moderately increased TSH levels. Radioactive iodine uptake returns to normal or is even increased as a result of the rise in TSH. Finally, thyroid hormone and TSH levels return to normal as the disease subsides.

**Clinical Manifestations** The patient usually presents with a painful and enlarged thyroid, sometimes accompanied by fever. There may be features of thyrotoxicosis or hypothyroidism, depending on the phase of the illness. Malaise and symptoms of an upper respiratory tract infection may precede the thyroid-related features by several weeks. In other patients, the onset is acute, severe, and without obvious antecedent. The patient typically complains of a sore throat, and examination reveals a small goiter that is exquisitely tender. Pain is often referred to the jaw or ear. Complete resolution is the usual outcome, but late-onset permanent hypothyroidism occurs in 15% of cases, particularly in those with coincidental thyroid autoimmunity. A prolonged course over many months, with one or more relapses, occurs in a small percentage of patients.

**Laboratory Evaluation** As depicted in Fig. 384-3, thyroid function tests characteristically evolve through three distinct phases over



**FIGURE 384-3** Clinical course of subacute thyroiditis. The release of thyroid hormones is initially associated with a thyrotoxic phase and suppressed thyroid-stimulating hormone (TSH). A hypothyroid phase then ensues, with low  $T_4$  and TSH levels that are initially low but gradually increase. During the recovery phase, increased TSH levels combined with resolution of thyroid follicular injury lead to normalization of thyroid function, often several months after the beginning of the illness. ESR, erythrocyte sedimentation rate;  $UT_{4}$ , free or unbound  $T_4$ .

about 6 months: (1) thyrotoxic phase, (2) hypothyroid phase, and (3) recovery phase. In the thyrotoxic phase,  $T_4$  and  $T_3$  levels are increased, reflecting their discharge from the damaged thyroid cells, and TSH is suppressed. The  $T_4/T_3$  ratio is greater than in Graves' disease or thyroid autonomy, in which  $T_3$  is often disproportionately increased. The diagnosis is confirmed by a high ESR and low uptake of radioiodine (<5%) or  $^{99m}\text{Tc}$  pertechnetate (as compared to salivary gland pertechnetate concentration). The white blood cell count may be increased, and thyroid antibodies are negative. If the diagnosis is in doubt, FNA biopsy may be useful, particularly to distinguish unilateral involvement from bleeding into a cyst or neoplasm.

## TREATMENT

### Subacute Thyroiditis

Relatively large doses of aspirin (e.g., 600 mg every 4–6 h) or nonsteroidal anti-inflammatory drugs (NSAIDs) are sufficient to control symptoms in many cases. If this treatment is inadequate, or if the patient has marked local or systemic symptoms, glucocorticoids should be given. The usual starting dose is 15–40 mg of prednisone, depending on severity. The dose is gradually tapered over 6–8 weeks, in response to improvement in symptoms and the ESR. If a relapse occurs during glucocorticoid withdrawal, the dosage should be increased and then withdrawn more gradually. Thyroid function should be monitored every 2–4 weeks using TSH and free  $T_4$  levels. Symptoms of thyrotoxicosis improve spontaneously but may be ameliorated by  $\beta$ -adrenergic blockers; antithyroid drugs play no role in treatment of the thyrotoxic phase. LT<sub>4</sub> replacement may be needed if the hypothyroid phase is prolonged, but doses should be low enough (50–100  $\mu\text{g}$  daily) to allow TSH-mediated recovery.

### SILENT THYROIDITIS

*Painless thyroiditis*, or “silent” thyroiditis, occurs in patients with underlying autoimmune thyroid disease and has a clinical course similar to that of subacute thyroiditis. The condition occurs in up to 5% of women 3–6 months after pregnancy and is then termed *postpartum thyroiditis*. Typically, patients have a brief phase of thyrotoxicosis lasting 2–4 weeks, followed by hypothyroidism for 4–12 weeks, and then resolution; often, however, only one phase is apparent. The condition is associated with the presence of TPO antibodies antepartum, and it is three times more common in women with type 1 diabetes mellitus. As in subacute thyroiditis, the uptake of  $^{99m}\text{Tc}$  pertechnetate or radioactive

iodine is initially suppressed. In addition to the painless goiter, silent thyroiditis can be distinguished from subacute thyroiditis by a normal ESR and the presence of TPO antibodies. Glucocorticoid treatment is not indicated for silent thyroiditis. Severe thyrotoxic symptoms can be managed with a brief course of propranolol, 20–40 mg three or four times daily. Thyroxine replacement may be needed for the hypothyroid phase but should be withdrawn after 6–9 months, as recovery is the rule. Annual follow-up thereafter is recommended, because a proportion of these individuals develop permanent hypothyroidism. The condition may recur in subsequent pregnancies.

### DRUG-INDUCED THYROIDITIS

Patients receiving cytokines, such as IFN- $\alpha$ , tyrosine kinase inhibitors, and immune checkpoint inhibitors may develop painless thyroiditis. IFN- $\alpha$ , which is used to treat chronic hepatitis B or C and hematologic and skin malignancies, causes thyroid dysfunction in up to 5% of treated patients. It has been associated with painless thyroiditis, hypothyroidism, and Graves' disease and is most common in women with TPO antibodies prior to treatment. For discussion of amiodarone, see “Amiodarone Effects on Thyroid Function,” below.

### CHRONIC THYROIDITIS

Focal thyroiditis is present in 20–40% of euthyroid autopsy cases and is associated with serologic evidence of autoimmunity, particularly the presence of TPO antibodies. The most common clinically apparent cause of chronic thyroiditis is *Hashimoto's thyroiditis*, an autoimmune disorder that often presents as a firm or hard goiter of variable size (Chap. 383). *Riedel's thyroiditis* is a rare disorder that typically occurs in middle-aged women. It presents with an insidious, painless goiter with local symptoms due to compression of the esophagus, trachea, neck veins, or recurrent laryngeal nerves. Dense fibrosis disrupts normal gland architecture and can extend outside the thyroid capsule. Despite these extensive histologic changes, thyroid dysfunction is uncommon. The goiter is hard, nontender, often asymmetric, and fixed, leading to suspicion of a malignancy. Diagnosis requires open biopsy as FNA biopsy is usually inadequate. Treatment is directed to surgical relief of compressive symptoms. Tamoxifen may also be beneficial. There is an association between Riedel's thyroiditis and IgG4-related disease causing idiopathic fibrosis at other sites (retroperitoneum, mediastinum, biliary tree, lung, and orbit).

### SICK EUTHYROID SYNDROME (NONTHYROIDAL ILLNESS)

Any acute, severe illness can cause abnormalities of circulating TSH or thyroid hormone levels in the absence of underlying thyroid disease, making these measurements potentially misleading. The major cause of these hormonal changes is the release of cytokines such as IL-6. Unless a thyroid disorder is strongly suspected, the routine testing of thyroid function should be avoided in acutely ill patients.

The most common hormone pattern in sick euthyroid syndrome (SES), also called nonthyroidal illness (NTI), is a decrease in total and unbound  $T_3$  levels (low  $T_3$  syndrome) with normal levels of  $T_4$  and TSH. The magnitude of the fall in  $T_3$  correlates with the severity of the illness.  $T_4$  conversion to  $T_3$  via peripheral 5' (outer ring) deiodination is impaired, leading to increased reverse  $T_3$  ( $rT_3$ ). Since  $rT_3$  is metabolized by 5' deiodination, its clearance is also reduced. Thus, decreased clearance rather than increased production is the major basis for increased  $rT_3$ . Also,  $T_4$  is alternately metabolized to the hormonally inactive  $T_3$  sulfate. It is generally assumed that this low  $T_3$  state is adaptive, because it can be induced in normal individuals by fasting. Teleologically, the fall in  $T_3$  may limit catabolism in starved or ill patients.

Very sick patients may exhibit a dramatic fall in total  $T_4$  and  $T_3$  levels (low  $T_4$  syndrome). With decreased tissue perfusion, muscle and liver expression of the type 3 deiodinase leads to accelerated  $T_4$  and  $T_3$  metabolism. This state has a poor prognosis. Another key factor in the fall in  $T_4$  levels is altered binding to thyroxine-binding globulin (TBG). The commonly used free  $T_4$  assays are subject to artifact when

serum binding proteins are low and underestimate the true free T<sub>4</sub> level. Fluctuation in TSH levels also creates challenges in the interpretation of thyroid function in sick patients. TSH levels may range from <0.1 mIU/L in very ill patients, especially with dopamine or glucocorticoid therapy, to >20 mIU/L during the recovery phase of SES. The exact mechanisms underlying the subnormal TSH seen in 10% of sick patients and the increased TSH seen in 5% remain unclear but may be mediated by cytokines including IL-12 and IL-18.

Any severe illness can induce changes in thyroid hormone levels, but certain disorders exhibit a distinctive pattern of abnormalities. Acute liver disease is associated with an initial rise in total (but not unbound) T<sub>3</sub> and T<sub>4</sub> levels due to TBG release; these levels become subnormal with progression to liver failure. A transient increase in total and unbound T<sub>4</sub> levels, usually with a normal T<sub>3</sub> level, is seen in 5–30% of acutely ill psychiatric patients. TSH values may be transiently low, normal, or high in these patients. In the early stage of HIV infection, T<sub>3</sub> and T<sub>4</sub> levels rise, even if there is weight loss. T<sub>3</sub> levels fall with progression to AIDS, but TSH usually remains normal. Renal disease is often accompanied by low T<sub>3</sub> concentrations, but with normal rather than increased rT<sub>3</sub> levels, due to an unknown factor that increases uptake of rT<sub>3</sub> into the liver.

The diagnosis of NTI is challenging. Historic information may be limited, and patients often have multiple metabolic derangements. Useful features to consider include previous history of thyroid disease and thyroid function tests, evaluation of the severity and time course of the patient's acute illness, documentation of medications that may affect thyroid function or thyroid hormone levels, and measurements of rT<sub>3</sub> together with unbound thyroid hormones and TSH. The diagnosis of NTI is frequently presumptive, given the clinical context and pattern of laboratory values; only resolution of the test results with clinical recovery can clearly establish this disorder. Treatment of NTI with thyroid hormone (T<sub>4</sub> and/or T<sub>3</sub>) is controversial, but most authorities recommend monitoring the patient's thyroid function tests during recovery, without administering thyroid hormone, unless there is historic or clinical evidence suggestive of hypothyroidism. Sufficiently large randomized controlled trials using thyroid hormone are unlikely to resolve this therapeutic controversy in the near future, because clinical presentations and outcomes are highly variable.

## AMIODARONE EFFECTS ON THYROID FUNCTION

Amiodarone is a commonly used type III antiarrhythmic agent (*Chap. 252*). It is structurally related to thyroid hormone and contains 39% iodine by weight. Thus, typical doses of amiodarone (200 mg/d) are associated with very high iodine intake, leading to greater than fortyfold increases in plasma and urinary iodine levels. Moreover, because amiodarone is stored in adipose tissue, high iodine levels persist for >6 months after discontinuation of the drug. Amiodarone inhibits deiodinase activity, and its metabolites function as weak antagonists of thyroid hormone action. Amiodarone has the following effects on thyroid function: (1) acute, transient suppression of thyroid function; (2) inhibition of T<sub>4</sub> to T<sub>3</sub> conversion causing either euthyroid hyperthyroxinemia or increased dosage requirement in LT4-treated hypothyroid patients; (3) hypothyroidism in patients susceptible to the inhibitory effects of a high iodine load; and (4) thyrotoxicosis that may be caused by either a Jod-Basedow effect from the iodine load, in the setting of MNG or incipient Graves' disease, or a thyroiditis-like condition due to a toxic effect on thyroid follicular cells.

The initiation of amiodarone treatment is associated with a transient decrease of T<sub>4</sub> levels, reflecting the inhibitory effect of iodine on T<sub>4</sub> release. Soon thereafter, most individuals escape from iodide-dependent suppression of the thyroid (Wolff-Chaikoff effect), and the inhibitory effects on deiodinase activity and thyroid hormone receptor action become predominant. These events lead to the following pattern of thyroid function tests: increased T<sub>4</sub>, decreased T<sub>3</sub>, increased rT<sub>3</sub>, and

a transient TSH increase (up to 20 mIU/L). TSH levels normalize or are slightly suppressed within 1–3 months.

The incidence of hypothyroidism from amiodarone varies geographically, apparently correlating with iodine intake. Hypothyroidism occurs in up to 13% of amiodarone-treated patients in iodine-replete countries, such as the United States, but is less common (<6% incidence) in areas of lower iodine intake, such as Italy or Spain. The pathogenesis appears to involve an inability of the thyroid gland to escape from the Wolff-Chaikoff effect in autoimmune thyroiditis. Consequently, amiodarone-associated hypothyroidism is more common in women and individuals with positive TPO antibodies. It is usually unnecessary to discontinue amiodarone for this side effect, because LT4 can be used to normalize thyroid function. TSH levels should be monitored, because T<sub>4</sub> levels are often increased for the reasons described above. In addition, TSH levels need to be monitored in LT4-replaced hypothyroid patients because a dosage increase is often required.

The management of amiodarone-induced thyrotoxicosis (AIT) is complicated by the fact that there are different causes of thyrotoxicosis and because the increased thyroid hormone levels exacerbate underlying arrhythmias and coronary artery disease. Amiodarone treatment causes thyrotoxicosis in 10% of patients living in areas of low iodine intake and in 2% of patients in regions of high iodine intake. There are two major forms of AIT, although some patients have features of both. Type 1 AIT is associated with an underlying thyroid abnormality (preclinical Graves' disease or nodular goiter). Thyroid hormone synthesis becomes excessive as a result of increased iodine exposure (Jod-Basedow phenomenon). Type 2 AIT occurs in individuals with no intrinsic thyroid abnormalities and is the result of drug-induced lysosomal activation leading to destructive thyroiditis with histiocyte accumulation in the thyroid; the incidence rises as cumulative amiodarone dosage increases. Mild forms of type 2 AIT can resolve spontaneously or can occasionally lead to hypothyroidism. Color-flow Doppler ultrasonography shows increased vascularity in type 1 AIT but decreased vascularity in type 2 AIT. Thyroid scintiscans are difficult to interpret in this setting because the high endogenous iodine levels diminish tracer uptake. However, the presence of normal or rarely increased uptake favors type 1 AIT.

In AIT, the drug should be stopped, if possible, although this is often impractical because of the underlying cardiac disorder. Discontinuation of amiodarone will not have an acute effect because of its storage and prolonged half-life. High doses of antithyroid drugs can be used in type 1 AIT but are often ineffective. Potassium perchlorate, 200 mg every 6 h, has been used to reduce thyroidal iodide content. Perchlorate treatment has been associated with agranulocytosis, although the risk appears relatively low with short-term use. Glucocorticoids, as administered for subacute thyroiditis, have modest benefit in type 2 AIT and are generally initiated as prednisone 40 mg PO daily. Lithium blocks thyroid hormone release and can also provide some benefit. Near-total thyroidectomy rapidly decreases thyroid hormone levels and may be the most effective long-term solution if the patient can undergo the procedure safely.

## FURTHER READING

- Biondi B, Cooper DS: Subclinical hyperthyroidism. *N Engl J Med* 378:2411, 2018.
- De Leo S et al: Hyperthyroidism. *Lancet* 388:906, 2016.
- Kitahara CM et al: Association of radioactive iodine treatment with cancer mortality in patients with hyperthyroidism. *JAMA Intern Med* 179:1034, 2019.
- Ross DS et al: 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. *Thyroid* 26:1343, 2016.
- Smith TJ et al: Teprotumumab for the treatment of active thyroid eye Disease. *N Engl J Med* 382:341, 2020.

J. Larry Jameson, Susan J. Mandel,  
Anthony P. Weetman

sympporter [NIS]), thyroglobulin (Tg) synthesis, organification and coupling (thyroid peroxidase [TPO]), and the regeneration of iodide (dehalogenase), have been described.

## CLINICAL MANIFESTATIONS AND DIAGNOSIS

If thyroid function is preserved, most goiters are asymptomatic. Examination of a diffuse goiter reveals a symmetrically enlarged, nontender, generally soft gland without palpable nodules. Goiter is defined, somewhat arbitrarily, as a lateral lobe with a volume greater than the thumb of the individual being examined. On ultrasound, total thyroid volume exceeding 30 mL is considered abnormal. If the thyroid is markedly enlarged, it can cause tracheal or esophageal compression. These features are unusual, however, in the absence of nodular disease and fibrosis. *Substernal goiter* may obstruct the thoracic inlet. *Pemberton's sign* refers to facial and neck congestion due to jugular venous obstruction when the arms are raised above the head, a maneuver that draws the thyroid into the thoracic inlet. Respiratory flow measurements and CT or MRI should be used to evaluate substernal goiter in patients with obstructive signs or symptoms.

Thyroid function tests should be performed in all patients with goiter to exclude thyrotoxicosis or hypothyroidism. It is not unusual, particularly in iodine deficiency, to find a low total T<sub>4</sub> with normal T<sub>3</sub> and TSH, reflecting enhanced T<sub>4</sub> → T<sub>3</sub> conversion. A low TSH with a normal free T<sub>3</sub> and free T<sub>4</sub>, particularly in older patients, suggests the possibility of thyroid autonomy or undiagnosed Graves' disease, and is termed *subclinical thyrotoxicosis*. The benefit of treatment (typically with radioiodine) in subclinical thyrotoxicosis, versus follow-up and implementing treatment if free T<sub>3</sub> or free T<sub>4</sub> levels become abnormal, is unclear, but treatment is increasingly recommended in the elderly to reduce the risk of atrial fibrillation and bone loss. Low urinary iodine levels (<50 µg/L) support a diagnosis of iodine deficiency. Thyroid scanning is not generally necessary but will reveal increased uptake in iodine deficiency and most cases of dyshormonogenesis.

## TREATMENT

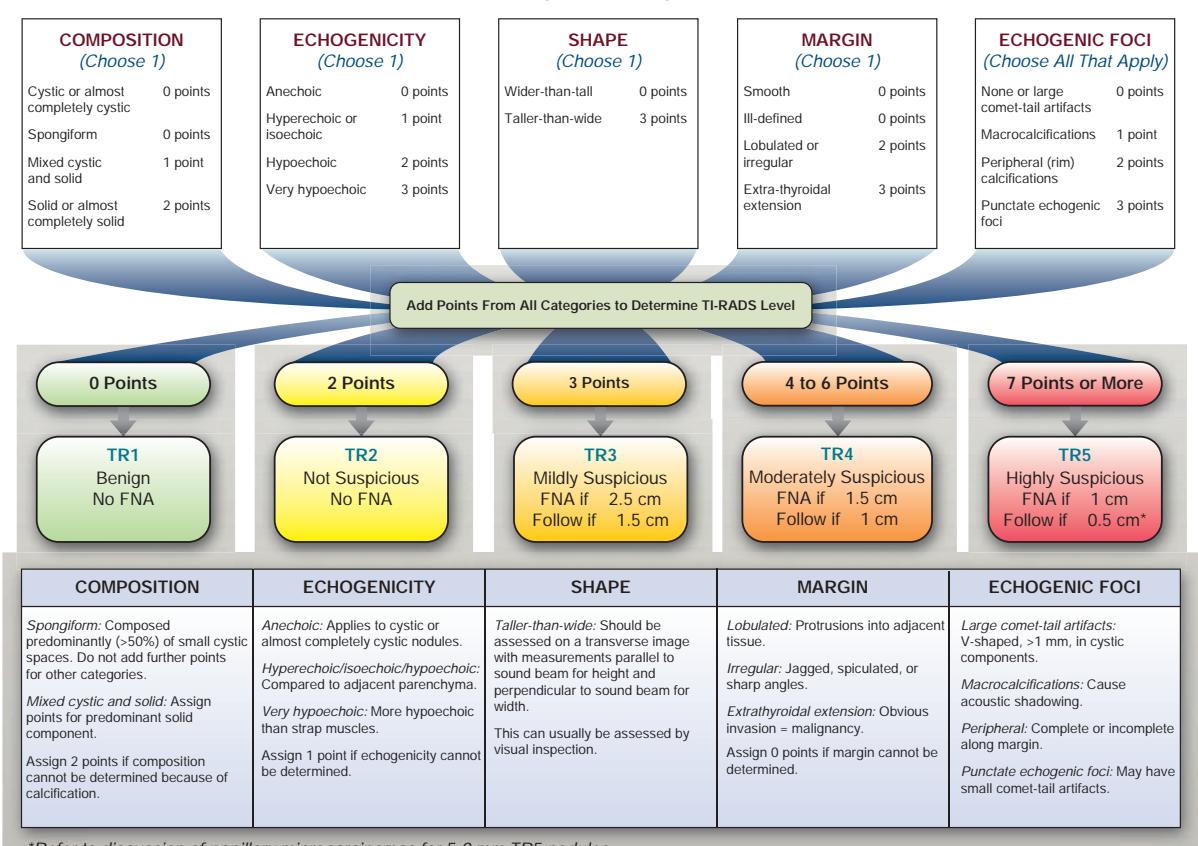
### Diffuse Nontoxic (Simple) Goiter

Iodine replacement induces variable regression of goiter in iodine deficiency, depending on duration and the degree of hyperplasia, with accompanying fibrosis, and autonomous function that may have developed. Surgery is rarely indicated for diffuse goiter. Exceptions include documented evidence of tracheal compression or obstruction of the thoracic inlet, which are more likely to be associated with substernal MNGs (see below). Subtotal or near-total thyroidectomy for these or cosmetic reasons should be performed by an experienced surgeon to minimize complication rates. Surgery should be followed by replacement with levothyroxine (LT4).

### NONTOXIC MULTINODULAR GOITER

**Etiology and Pathogenesis** Depending on the population studied, MNG or the presence of nodules in a thyroid of normal size occurs in up to 12% of adults. MNG should be distinguished from the presence of nodules in a normal-size thyroid gland (see "Approach to the Patient with Thyroid Nodules"). MNG is more common in women than men and increases in prevalence with age. It is more common in iodine-deficient regions but also occurs in regions of iodine sufficiency, reflecting multiple genetic, autoimmune, and environmental influences on the pathogenesis.

There is typically wide variation in nodule size. Histology reveals a spectrum of morphologies ranging from hypercellular, hyperplastic regions to cystic areas filled with colloid. Fibrosis is often extensive, and areas of hemorrhage or lymphocytic infiltration may be seen. Using molecular techniques, most nodules within an MNG are polyclonal in origin, suggesting a hyperplastic response to locally produced growth factors and cytokines. TSH, which is usually not elevated, may play a permissive or contributory role. Monoclonal neoplastic lesions may also occur, reflecting mutations in genes that confer a selective growth advantage to the progenitor cell.



\*Refer to discussion of papillary microcarcinomas for 5-9 mm TR5 nodules.

**FIGURE 385-1** American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS). TI-RADS is a five-tiered system categorizing the sonographic appearance of thyroid nodules based on increased risk for malignancy. For each level (TR1–5), there are recommendations for both fine-needle aspiration (FNA) minimum size cutoffs and follow-up. (Reproduced with permission from FN Tessler et al: ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. J Am Coll Radiol 14:587, 2017.)

**Clinical Manifestations** Most patients with nontoxic MNG are asymptomatic and euthyroid. MNG typically develops over many years and is detected on routine physical examination, when an individual notices an enlargement in the neck, or as an incidental finding on imaging. If the goiter is large enough, it can ultimately lead to compressive symptoms including difficulty swallowing, respiratory distress (tracheal compression), or plethora (venous congestion), but these symptoms are uncommon. Symptomatic MNGs are usually extraordinarily large and/or develop fibrotic areas that cause compression. Sudden pain in an MNG is usually caused by hemorrhage into a nodule. Hoarseness, reflecting laryngeal nerve involvement may suggest malignancy but more commonly is due to others causes such as gastroesophageal reflux.

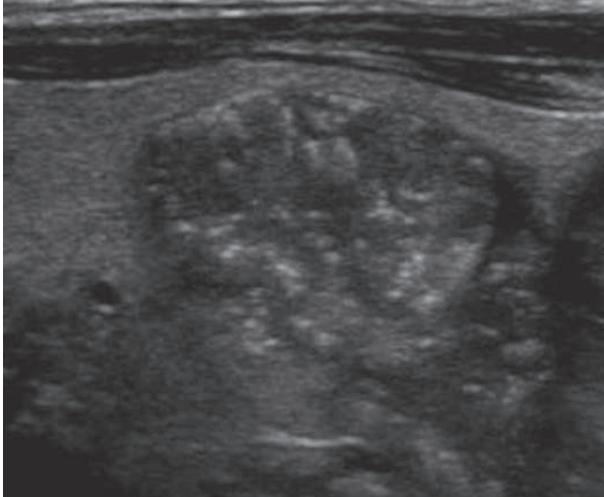
**Diagnosis** On examination, thyroid architecture is distorted, and multiple nodules of varying size can be appreciated. Because many nodules are deeply embedded in thyroid tissue or reside in posterior or substernal locations, it is not possible to palpate all nodules. Pemberton's sign, characterized by facial suffusion when the patient's arms are elevated above the head, suggests that the goiter has increased pressure in the thoracic inlet. A TSH level should be measured to exclude subclinical hyper- or hypothyroidism, but thyroid function is usually normal. Tracheal deviation is common, but compression must usually exceed 70% of the tracheal diameter before there is significant airway compromise. Pulmonary function testing can be used to assess the functional effects of compression, which characteristically causes inspiratory stridor. CT or MRI can be used to evaluate the anatomy of

the goiter and the extent of substernal extension or tracheal narrowing. A barium swallow may reveal the extent of esophageal compression. The risk of malignancy in MNG is similar to that in solitary nodules. Ultrasonography should be used to identify which nodules should be biopsied based on a combination of size and sonographic pattern (Fig. 385-1) (Chap. 382). For nodules with more suspicious sonographic patterns (e.g., hypoechoic solid nodules with infiltrative borders), biopsy is recommended at a lower size cutoff than those with less suspicious imaging features (Figs. 385-1 and 385-2).

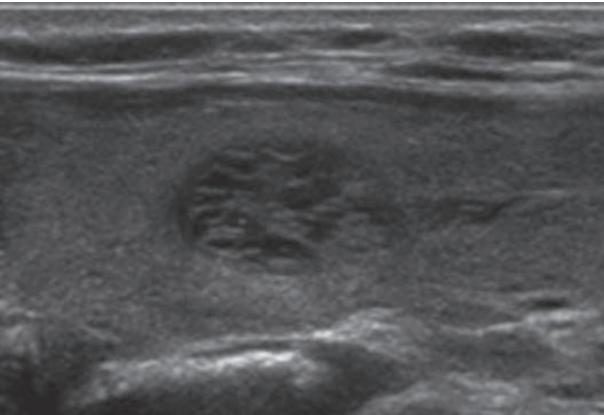
## TREATMENT

### Nontoxic Multinodular Goiter

Most nontoxic MNGs can be managed conservatively.  $T_4$  suppression is rarely effective for reducing goiter size and introduces the risk of subclinical or overt thyrotoxicosis, particularly if there is underlying autonomy or if it develops during treatment. Contrast agents and other iodine-containing substances should be avoided because of the risk of inducing the *Jod-Basedow effect*, characterized by enhanced thyroid hormone production by autonomous nodules. Radioiodine has been used when surgery is contraindicated in areas where large nodular goiters are more prevalent (e.g., some areas of Europe and Brazil) because it can decrease MNG volume and may selectively ablate regions of autonomy. Dosage of  $^{131}\text{I}$  depends on the size of the goiter and radioiodine uptake but is usually about 3.7 MBq (0.1 mCi) per gram of tissue, corrected for uptake



A



B

**FIGURE 385-2 Sonographic patterns of thyroid nodules.** A. High suspicion ultrasound pattern for thyroid malignancy ACR TI-RADS TR5 (hypoechoic solid nodule with irregular borders and punctate echogenic foci). B. Very low suspicion ultrasound pattern for thyroid malignancy ACR TI-RADS TR1 (spongiform nodule with microcystic areas comprising >50% of nodule volume). ACR TI-RADS, American College of Radiology Thyroid Imaging Reporting and Data System.

(typical dose 370–1070 MBq [10–29 mCi]). Repeat treatment may be needed, and effectiveness may be increased by concurrent administration of low-dose recombinant TSH (0.1 mg IM). It is possible to achieve a 40–50% reduction in goiter size in most patients. Earlier concerns about radiation-induced thyroid swelling and tracheal compression have diminished, as studies have shown this complication to be rare. When acute compression occurs, glucocorticoid treatment or surgery may be needed. Radiation-induced hypothyroidism is less common than after treatment for Graves' disease. However, posttreatment autoimmune thyrotoxicosis may occur in up to 5% of patients treated for nontoxic MNG. Surgery remains highly effective but is not without risk, particularly in older patients with underlying cardiopulmonary disease.

### TOXIC MULTINODULAR GOITER

The pathogenesis of toxic MNG appears to be similar to that of non-toxic MNG; the major difference is the presence of functional autonomy in toxic MNG. The molecular basis for autonomy in toxic MNG remains unknown. As in nontoxic goiters, many nodules are polyclonal, whereas others are monoclonal and vary in their clonal origins. Genetic abnormalities known to confer functional autonomy, such as

activating TSH-R or G<sub>s</sub> mutations (see below), are not usually found in the autonomous regions of toxic MNG goiter.

In addition to features of goiter, the clinical presentation of toxic MNG includes subclinical or mild overt hyperthyroidism. The patient is usually elderly and may present with atrial fibrillation or palpitations, tachycardia, nervousness, tremor, or weight loss. Recent exposure to iodine, from contrast dyes or other sources, may precipitate or exacerbate thyrotoxicosis. The TSH level is low. The uncombined T<sub>4</sub> level may be normal or minimally increased; T<sub>3</sub> is often elevated to a greater degree than T<sub>4</sub>. Thyroid scan shows heterogeneous uptake with multiple regions of increased and decreased uptake; 24-h uptake of radioiodine may not be increased but is usually in the upper normal range.

Prior to definitive treatment of the hyperthyroidism, ultrasound imaging should be performed to assess the presence of discrete nodules corresponding to areas of decreased uptake ("cold" nodules). If present, fine-needle aspiration (FNA) may be indicated based on sonographic patterns and size cutoffs (see "Approach to the Patient with Thyroid Nodules"). The cytology results, if indeterminate or suspicious, may direct the therapy to surgery.

### TREATMENT

#### Toxic Multinodular Goiter

Antithyroid drugs normalize thyroid function and are particularly useful in the elderly or ill patients with limited life span. In contrast to Graves' disease, spontaneous remission does not occur and so treatment is long term. Radioiodine is generally the treatment of choice; it treats areas of autonomy as well as decreasing the mass of the goiter by ablating the functioning nodules. Sometimes, however, a degree of autonomy may persist, presumably because multiple autonomous regions may emerge after others are treated, and further radioiodine treatment may be necessary. Surgery provides definitive treatment of underlying thyrotoxicosis as well as goiter. Patients should be rendered euthyroid using an antithyroid drug before operation.

### HYPERFUNCTIONING SOLITARY NODULE

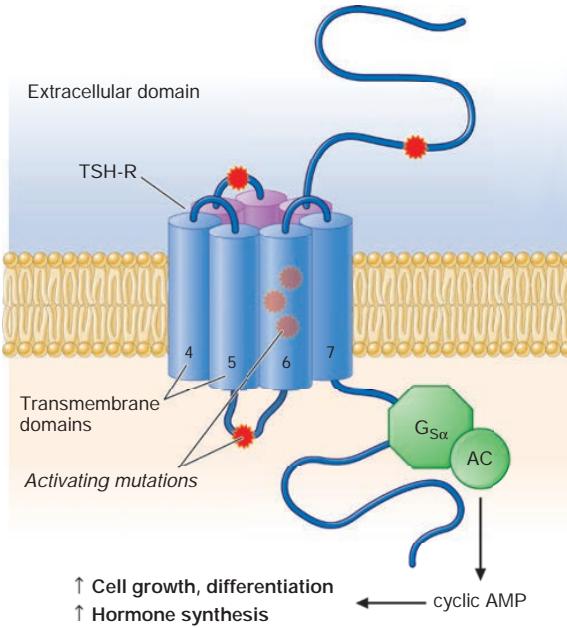
A solitary, autonomously functioning thyroid nodule is referred to as *toxic adenoma*. The pathogenesis of this disorder has been unraveled by demonstrating the functional effects of mutations that stimulate the TSH-R signaling pathway. Most patients with solitary hyperfunctioning nodules have acquired somatic, activating mutations in the TSH-R (Fig. 385-3). These mutations, located primarily in the receptor transmembrane domain, induce constitutive receptor coupling to G<sub>s</sub>, increasing cyclic adenosine monophosphate (AMP) levels and leading to enhanced thyroid follicular cell proliferation and function. Less commonly, somatic mutations are identified in G<sub>s</sub>. These mutations, which are similar to those seen in McCune-Albright syndrome (Chap. 412) or in a subset of somatotrope adenomas (Chap. 380), impair guanosine triphosphate (GTP) hydrolysis, causing constitutive activation of the cyclic AMP signaling pathway. In most series, activating mutations in either the TSH-R or the G<sub>s</sub> subunit genes are identified in >90% of patients with solitary hyperfunctioning nodules.

Thyrotoxicosis is usually mild and is generally only detected when a nodule is >3 cm. The disorder is suggested by a subnormal TSH level; the presence of the thyroid nodule, often large enough to be palpable; and the absence of clinical features suggestive of Graves' disease or other causes of thyrotoxicosis. A thyroid scan provides a definitive diagnostic test, demonstrating focal uptake in the hyperfunctioning nodule and diminished uptake in the remainder of the gland, as activity of the normal thyroid is suppressed.

### TREATMENT

#### Hyperfunctioning Solitary Nodule

Radioiodine ablation is usually the treatment of choice. Because normal thyroid function is suppressed, <sup>131</sup>I is concentrated in the hyperfunctioning nodule with minimal uptake and damage to



**FIGURE 385-3** Activating mutations of the thyroid-stimulating hormone receptor (TSH-R). Mutations (\*) that activate TSH-R reside mainly in transmembrane 5 and intracellular loop 3, although mutations have occurred in a variety of different locations. The effect of these mutations is to induce conformational changes that mimic TSH binding, thereby leading to coupling to stimulatory G protein ( $G_{s\alpha}$ ) and activation of adenylate cyclase (AC), an enzyme that generates cyclic AMP.

normal thyroid tissue. Relatively large radioiodine doses (e.g., 370–1110 MBq [10–29.9 mCi]  $^{131}\text{I}$ ) have been shown to correct thyrotoxicosis in ~75% of patients within 3 months. Hypothyroidism occurs in <10% of those patients over the next 5 years. Surgical resection is also effective and is usually limited to lobectomy, thereby preserving thyroid function and minimizing risk of hypoparathyroidism or damage to the recurrent laryngeal nerves. Medical therapy using antithyroid drugs and beta blockers can normalize thyroid function but is not an optimal long-term treatment. Using ultrasound guidance, percutaneous radiofrequency ablation has been used successfully in some centers to ablate hyperfunctioning nodules, and this technique has also been used to reduce the size of nonfunctioning thyroid nodules.

## BENIGN LESIONS

The various types of benign thyroid nodules are listed in **Table 385-1**. Benign nodules may be hyperplastic and reflect a combination of both macro- and microfollicular architecture or they may be neoplastic, encapsulated adenomas that generally have a more monotonous microfollicular pattern. If the adenoma is composed of oncocytic follicular cells arranged in a follicular pattern, this is termed Hürthle cell adenoma. Hyperplastic nodules generally appear as mixed cystic/solid or spongiform lesions on ultrasound. The definition of spongiform requires the presence of microcystic areas comprising >50% of the nodule volume, with the concept that this microcystic sonographic pattern recapitulates the histology of macrofollicles containing colloid (**Fig. 385-2B**). However, the majority of solid nodules (whether hypo-, iso-, or hyperechoic) are also benign. FNA, usually performed with ultrasound guidance, is the diagnostic procedure of choice to evaluate thyroid nodules (see the “Approach to the Patient with Thyroid Nodules” section). Pure thyroid cysts, <1% of all thyroid growths, consist of colloid and are benign as well. Cysts frequently recur, even after repeated aspiration, and may require surgical excision if they are large. Ethanol ablation to sclerose the cyst has been used successfully for patients who are symptomatic.

TSH suppression with LT4 therapy does not decrease thyroid nodule size in iodine-sufficient populations. However, if there is relative iodine

**TABLE 385-1** Classification of Thyroid Growths

Benign	
Hyperplasia	
Colloid nodule	
Follicular epithelial cell adenomas	
Conventional	
Oncocytic (Hürthle cell)	
Malignant	Approximate Prevalence, %
Follicular epithelial cell	
Papillary carcinomas	80–85
Classic variant	
Follicular variant	
Diffuse sclerosing variant	
Tall cell, columnar cell variants	
Follicular carcinomas	5–12
Conventional	
Oncocytic (Hürthle cell)	
Poorly differentiated carcinomas	3–5
Anaplastic (undifferentiated) carcinomas	1
C-cell origin (calcitonin-producing)	
Medullary thyroid cancer	<10
Sporadic	
Familial	
MEN 2	
Other malignancies	
Lymphomas	1
Metastases	
Breast, melanoma, lung, kidney	
Others	

Abbreviation: MEN, multiple endocrine neoplasia.

deficiency, both iodine and LT4 therapy have been demonstrated to decrease nodule volume. If LT4 is administered in this situation, the TSH should be maintained at or just below the lower limit of normal, but not frankly suppressed. If the nodule has not decreased in size after 6–12 months of therapy, treatment should be discontinued because little benefit is likely to accrue from long-term treatment; the risk of iatrogenic subclinical thyrotoxicosis should also be considered.

## THYROID CANCER

Thyroid carcinoma is the most common malignancy of the endocrine system. Malignant tumors derived from the follicular epithelium are classified according to histologic features. Differentiated tumors, such as papillary thyroid cancer (PTC) or follicular thyroid cancer (FTC), are often curable, and the prognosis is good for patients identified with early-stage disease. In contrast, anaplastic thyroid cancer (ATC) is aggressive, responds poorly to treatment, and is associated with a bleak prognosis.

Over the past 30 years, the incidence of thyroid cancer has increased from 4.9 to >15 cases per 100,000 individuals in the United States. However, disease-specific mortality has only minimally increased. The increased incidence is predominantly attributable to small T1 papillary cancer tumors (<2 cm) and has led experts to consider that thyroid cancer is being overdiagnosed, suggesting that cancers are being detected that would otherwise be unlikely to harm a patient. The concept of cancer overdiagnosis is predicated upon the presence of a disease reservoir (the autopsy prevalence of PTC is ~25%), activities leading to disease detection (increased diagnostic imaging with incidental detection of nodules), and a mismatch in the directional rate between diagnosis and mortality (thyroid cancer disease-specific mortality not changed in 40 years). Similar trends have been observed worldwide, especially in countries with a higher proportion of privately financed health care, leading to increased resource utilization including imaging. The 20-year disease-specific mortality for low-risk

**TABLE 385-2 Risk Factors for Thyroid Carcinoma in Patients with Thyroid Nodule from History and Physical Examination**

History of head and neck irradiation before the age of 18, including mantle radiation for Hodgkin's disease and brain radiation for childhood leukemia or other cranial malignancies	Family history of papillary thyroid cancer in two or more first-degree relatives, MEN 2, or other genetic syndromes associated with thyroid malignancy (e.g., Cowden's syndrome, familial adenomatous polyposis, Carney complex, PTEN [phosphatase and tensin homolog] hamartoma tumor)
Exposure to ionizing radiation from fallout in childhood or adolescence	Vocal cord paralysis, hoarse voice
Age <20 or >65 years	Nodule fixed to adjacent structures
Rapidly enlarging neck mass	Lateral cervical lymphadenopathy (ipsilateral to the nodule)
Male gender	

Abbreviation: MEN, multiple endocrine neoplasia.

thyroid cancer is 1%. Fortunately, epidemiologic data in the United States document a decrease in new thyroid cancer diagnoses (62,450 cases in 2015 and 52,070 cases in 2019), and this trend correlates with the implementation of evidence-based guidelines that recommend higher size thresholds for nodule FNA.

Current trends in thyroid cancer care focus on (1) avoiding overdiagnosis by limiting FNA by sonographic risk stratification with size cutoffs; (2) limiting surgery, radioiodine, and subsequent surveillance for low-risk tumors; and (3) identifying patients at higher recurrence risk for more aggressive treatment and monitoring. Prognosis is worse in older persons (>65 years). Thyroid cancer is twice as common in women as men, but male gender is associated with a worse prognosis. Additional important risk factors include a history of childhood (before age 18) head or neck irradiation, evidence for local tumor fixation or gross metastatic involvement of lymph nodes, and the presence of distant metastases (**Table 385-2**).

Several unique features of thyroid cancer facilitate its management: (1) thyroid nodules are amenable to biopsy by FNA; (2) iodine radioisotopes can be used to diagnose ( $^{123}\text{I}$  and  $^{131}\text{I}$ ) and potentially treat ( $^{131}\text{I}$ ) differentiated thyroid cancer, reflecting the unique uptake of this anion by the thyroid gland; and (3) serum markers allow the detection of residual or recurrent disease, including the use of Tg levels for PTC and calcitonin for medullary thyroid cancer (MTC).

## CLASSIFICATION

Thyroid neoplasms can arise in each of the cell types that populate the gland, including thyroid follicular cells, calcitonin-producing C cells, lymphocytes, and stromal and vascular elements, as well as metastases from other sites (Table 385-1). The American Joint Committee on Cancer (AJCC) staging system using the tumor, node, metastasis (TNM) classification is most commonly used (**Table 385-3**).

## PATHOGENESIS AND GENETIC BASIS

**Radiation** Early studies of the pathogenesis of thyroid cancer focused on the role of external radiation, which predisposes to chromosomal breaks, leading to genetic rearrangements and loss of tumor-suppressor genes. External radiation of the mediastinum, face, head, and neck region was administered in the past to treat an array of conditions, including acne and enlargement of the thymus, tonsils, and adenoids. Radiation exposure increases the risk of benign and malignant thyroid nodules, is associated with multicentric cancers, and shifts the incidence of thyroid cancer to an earlier age group. Radiation from nuclear fallout also increases the risk of thyroid cancer. Children seem more predisposed to the effects of radiation than adults.

**TSH and Growth Factors** Many differentiated thyroid cancers express TSH receptors and, therefore, remain responsive to TSH. Higher serum TSH levels, even within normal range, are associated with increased thyroid cancer risk in patients with thyroid nodules. These observations provide the rationale for  $\text{T}_4$  suppression of TSH in patients with thyroid cancer. Residual expression of TSH receptors also allows TSH-stimulated uptake of  $^{131}\text{I}$  therapy (see below).

**Oncogenes and Tumor-Suppressor Genes** Thyroid cancers are monoclonal in origin, consistent with the idea that they originate as a

**TABLE 385-3 Definitions of AJCC TNM**

### Definition of Primary Tumor (T)

Papillary, Follicular, Poorly Differentiated, Hürthle Cell and Anaplastic Thyroid Carcinoma	
T Category	T Criteria
TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
T1	Tumor 2 cm in greatest dimension limited to the thyroid
T1a	Tumor 1 cm in greatest dimension limited to the thyroid
T1b	Tumor >1 cm but 2 cm in greatest dimension limited to the thyroid
T2	Tumor >2 cm but 4 cm in greatest dimension limited to the thyroid
T3	Tumor >4 cm limited to the thyroid, or gross extrathyroidal extension invading only strap muscles
T3a	Tumor >4 cm limited to the thyroid
T3b	Gross extrathyroidal extension invading only strap muscles (sternohyoid, sternothyroid, thyrohyoid, or omohyoid muscles) from a tumor of any size
T4	Includes gross extrathyroidal extension beyond the strap muscles
T4a	Gross extrathyroidal extension invading subcutaneous soft tissues, larynx, trachea, esophagus, or recurrent laryngeal nerve from a tumor of any size
T4b	Gross extrathyroidal extension invading prevertebral fascia or encasing the carotid artery or mediastinal vessels from a tumor of any size

Note: All categories may be subdivided: (s) solitary tumor and (m) multifocal tumor (the largest tumor determines the classification).

### Definition of Regional Lymph Node (N)

N Category	N Criteria
NX	Regional lymph nodes cannot be assessed
N0	No evidence of locoregional lymph node metastasis
N0a	One or more cytologically or histologically confirmed benign lymph nodes
N0b	No radiologic or clinical evidence of locoregional lymph node metastasis
N1	Metastasis to regional nodes
N1a	Metastasis to level VI or VII (pretracheal, paratracheal, or prelaryngeal/Delphian, or upper mediastinal) lymph nodes. This can be unilateral or bilateral disease.
N1b	Metastasis to unilateral, bilateral, or contralateral lateral neck lymph nodes (levels I, II, III, IV, or V) or retropharyngeal lymph nodes

### Definition of Distant Metastasis (M)

M Category	M Criteria
M0	No distant metastasis
M1	Distant metastasis
MX	Distant metastasis cannot be assessed

Source: Used with permission of the American College of Surgeons, Chicago, Illinois. The original source for this information is the AJCC Cancer Staging System (2020).

consequence of mutations that confer a growth advantage to a single cell. In addition to increased rates of proliferation, some thyroid cancers exhibit impaired apoptosis and features that enhance invasion, angiogenesis, and metastasis. Thyroid neoplasms have been analyzed for a variety of genetic alterations, but without clear evidence of an ordered acquisition of somatic mutations as they progress from the benign to the malignant state. On the other hand, certain mutations, such as *RET/PTC* and *PAX-8-PPAR 1* rearrangements, are relatively specific for thyroid neoplasia.

As described above, activating mutations of the TSH-R and the  $G_s$  subunit are associated with autonomously functioning nodules. Although these mutations induce thyroid cell growth, this type of nodule is almost always benign, likely because they drive differentiation pathways.

Activation of the RET-RAS-BRAF signaling pathway is seen in up to 70% of PTCs, although the types of mutations are heterogeneous. A variety of rearrangements involving the *RET* gene on chromosome 10 bring this receptor tyrosine kinase under the control of other promoters, leading to receptor overexpression. *RET* rearrangements occur in 20–40% of PTCs in different series and were observed with increased

frequency in tumors developing after the Chernobyl radiation accident. Rearrangements in PTC have also occurred for another tyrosine kinase gene, *TRK1*, which is located on chromosome 1. To date, the identification of PTC with *RET* or *TRK1* rearrangements has not proven useful for predicting prognosis or treatment responses. *BRAFV600E* mutations appear to be the most common genetic alteration in PTC. These mutations activate the kinase, which stimulates the mitogen-activated protein kinase (MAPK) cascade. *RAS* mutations, which also stimulate the MAPK cascade, are found in ~20–30% of thyroid neoplasms (*NRAS* > *HRAS* > *KRAS*), including both PTC follicular variant and FTC. Of note, simultaneous *RET*, *BRAF*, and *RAS* mutations rarely occur in the same tumor, suggesting that activation of the MAPK cascade is critical for tumor development, independent of the step that initiates the cascade.

*RAS* mutations also occur in FTCs. In addition, a rearrangement of the thyroid developmental transcription factor *PAX8* with the nuclear receptor *PPAR* is identified in a significant fraction of FTCs. Overall, ~70% of follicular cancers have mutations or genetic rearrangements. Loss of heterozygosity of 3p or 11q, consistent with deletions of tumor-suppressor genes, is also common in FTCs.

Most of the mutations seen in differentiated thyroid cancers have also been detected in ATCs. *TERT* promoter mutations occur in <10% of differentiated PTCs but are more common in ATC. *BRAF* mutations are seen in up to 50% of ATCs. Mutations in *CTNNB1*, which encodes -catenin, occur in about two-thirds of ATCs, but not in PTC or FTC. Mutations of the tumor-suppressor *P53* also play an important role in the development of ATC. Because *P53* plays a role in cell-cycle surveillance, DNA repair, and apoptosis, its loss may contribute to the rapid acquisition of genetic instability as well as poor treatment responses (Chap. 72).

The role of molecular diagnostics in the clinical management of thyroid cancer is under investigation. In principle, analyses of specific mutations might aid in classification, prognosis, or choice of treatment. Although *BRAF* V600E mutations are associated with loss of iodine uptake by tumor cells. As discussed below, trials of multikinase pathway inhibitors are ongoing as a means to restore iodine uptake and enhance sensitivity to radioiodine treatment. Higher recurrence rates have been variably reported in patients with *BRAF*-positive PTC, but the impact on survival rates is unclear.

MTC, when associated with multiple endocrine neoplasia (MEN) type 2, harbors an inherited mutation of the *RET* gene. Unlike the rearrangements of *RET* seen in PTC, the mutations in MEN 2 are point mutations that induce constitutive activity of the tyrosine kinase (Chap. 388). MTC is preceded by hyperplasia of the C cells, raising the likelihood that as-yet-unidentified “second hits” lead to cellular transformation. A subset of sporadic MTC contains somatic mutations that activate *RET*.

## WELL-DIFFERENTIATED THYROID CANCER

**Papillary** PTC is the most common type of thyroid cancer, accounting for 80–85% of well-differentiated thyroid malignancies. Microscopic PTC is present in up to 25% of thyroid glands at autopsy, but most of these lesions are very small (several millimeters) and are not clinically significant. Characteristic cytologic features of PTC help make the diagnosis by FNA or after surgical resection; these include, large, clear nuclei with powdery chromatin (described as an “orphan Annie eye” appearance) with nuclear grooves and prominent nucleoli. The histologic finding of these cells arranged in either papillary structures or follicles distinguishes the classic and follicular variants of PTC, respectively. There are several subtypes of papillary thyroid cancer. The more differentiated classic and follicular variants are likely to have an indolent course in the absence of angioinvasion or metastatic adenopathy. The aggressive variants (tall cell, columnar cell, hobnail, poorly differentiated) require more intensive therapy and closer follow-up. Recently, a subtype previously known as the encapsulated PTC follicular variant, without capsular or angioinvasion, is no longer considered malignant and has been renamed noninvasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP).

PTC may be multifocal and invade locally within the thyroid gland as well as through the thyroid capsule and into adjacent structures in the neck. It has a propensity to spread via the lymphatic system but can metastasize hematogenously as well, particularly to bone and lung. Because of the relatively slow growth of the tumor, a significant

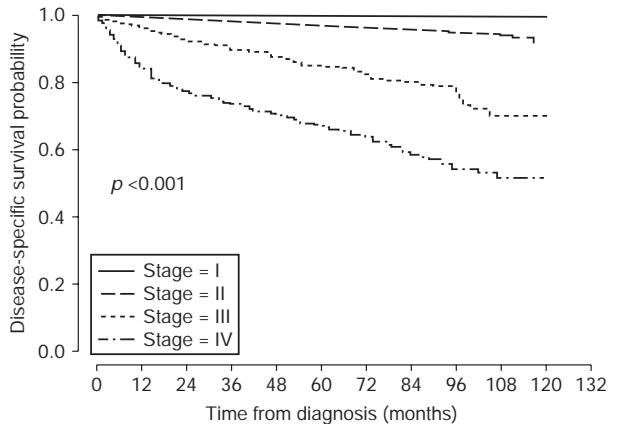


FIGURE 385-4 Unadjusted disease-specific survival curves for patients with papillary thyroid cancer in the AJCC/UICC 8th edition TNM staging system. (Reproduced with permission from LN Pontius et al: Projecting survival in papillary thyroid cancer: A comparison of the seventh and eighth editions of the American Joint Commission on Cancer/Union for International Cancer Control staging systems in two contemporary national patient cohorts. *Thyroid* 27:1408, 2017.)

burden of pulmonary metastases may accumulate, sometimes with remarkably few symptoms. The prognostic implication of lymph node spread depends on the volume of metastatic disease. Micrometastases, defined as <2 mm of cancer in a lymph node, do not affect prognosis. However, gross metastatic involvement of multiple 2- to 3-cm lymph nodes indicates a 25–30% chance of recurrence and may increase mortality in older patients. The staging of PTC by the TNM system is outlined in Table 385-3. Most papillary cancers are identified in the early stages (>95% stages I or II) and have an excellent prognosis, with survival curves similar to expected survival (Fig. 385-4). Mortality is markedly increased in stage IV disease, especially in the presence of distant metastases (stage IVB), but this group comprises only about 1% of patients. The treatment of PTC is described below.

**Follicular** The incidence of FTC varies widely in different parts of the world; it is more common in iodine-deficient regions. Currently, FTC accounts for only about 5% of all thyroid cancers diagnosed in the United States. FTC is difficult to diagnose by FNA because the distinction between benign and malignant follicular neoplasms requires histology because the nuclear features of follicular adenomas and carcinomas do not differ. Rather, follicular carcinoma is diagnosed by the presence of capsular and/or vascular invasion. Follicular carcinomas with only capsular invasion have a very low risk of metastasis, and lobectomy alone suffices. Angioinvasive FTC is more aggressive and may metastasize to bone, lung, and the central nervous system. Mortality rates associated with angioinvasive FTC are less favorable than for PTC, in part because a larger proportion of patients present with stage IV disease. Poor prognostic features include distant metastases, age >55 years, primary tumor size >4 cm, and the presence of marked vascular invasion.

## TREATMENT

### Surgery for Well-Differentiated Thyroid Cancer

All well-differentiated thyroid cancers >1 cm (T1b or larger) should be surgically excised, although active surveillance is an option for small intrathyroidal micropapillary thyroid cancers (T1a) without metastases. In addition to removing the primary lesion, surgery allows accurate histologic diagnosis and staging. Because there is no compelling evidence that bilateral thyroid surgery improves survival, the initial surgical procedure may be either a unilateral (lobectomy) or bilateral (near-total thyroidectomy) procedure for patients with intrathyroidal cancers >1 cm and <4 cm (T1b and T2 tumors) in the absence of metastatic disease and after a careful sonographic evaluation for metastatic cervical adenopathy. For patients at high risk for recurrence, bilateral surgery allows administration of radioiodine for remnant ablation and potential treatment

of iodine-avid metastases, if indicated, as well as for monitoring of serum Tg levels. Therefore, near-total thyroidectomy is appropriate for tumors >4 cm or in the presence of metastases or clinical evidence of extrathyroidal invasion. In addition, for patients found to have a high-risk tumor after lobectomy based upon aggressive pathology features (e.g., vascular invasion or a less differentiated subtype), completion surgery should be performed. Surgical complication rates are acceptably low if the surgeon is highly experienced in the procedure. Preoperative sonography should be performed in all patients to assess the central and lateral cervical lymph node compartments for suspicious adenopathy, which if present, should undergo FNA and be removed, as indicated, at surgery.

### TSH SUPPRESSION THERAPY

Because most tumors are still TSH-responsive, LT4 suppression of TSH is a mainstay of thyroid cancer treatment. Although TSH suppression clearly provides therapeutic benefit, there are no prospective studies that define the optimal level of TSH suppression. The degree of TSH suppression should be individualized based on a patient's risk of recurrence. It should be adjusted over time as surveillance blood tests and imaging confirm absence of disease or, alternatively, indicate possible residual/recurrent cancer. For patients at low risk of recurrence, TSH should be maintained in the lower normal limit (0.5–2.0 mIU/L). For patients either at intermediate or high risk of recurrence, TSH levels should be kept to 0.1–0.5 mIU/L and <0.1 mIU/L, respectively, if there are no strong contraindications to mild thyrotoxicosis. TSH should be <0.1 mIU/L for those with known metastatic disease.

### RADIOIODINE TREATMENT

After near-total thyroidectomy, <1 g of thyroid tissue remains in the thyroid bed. Postsurgical radioablation of the remnant thyroid eliminates residual normal thyroid, facilitating the use of Tg determinations. In addition, well-differentiated thyroid cancer often incorporates radioiodine, although less efficiently than normal thyroid follicular cells. Radioiodine uptake is determined primarily by expression of the NIS and is stimulated by TSH, requiring expression of the TSH-R. The retention time for radioactivity is influenced by the extent to which the tumor retains differentiated functions such as iodide trapping and organification. Consequently, for patients at higher risk of recurrence and for those with known distant metastatic disease, <sup>131</sup>I therapy may provide an adjuvant role and potentially treat residual tumor cells.

**Indications** Not all patients benefit from radioiodine therapy. Neither recurrence nor survival rates are improved in stage I patients with T1 tumors (<2 cm) confined to the thyroid. No benefit has been demonstrated for larger (>2 cm but <4 cm) low-risk tumors. However, in higher risk patients (larger tumors, more aggressive variants of papillary cancer, tumor vascular invasion, extrathyroidal invasion, presence of large-volume lymph node metastases), radioiodine reduces recurrence and may increase survival for older patients.

**<sup>131</sup>I Thyroid Ablation and Treatment** As noted above, the decision to use <sup>131</sup>I for thyroid ablation should be coordinated with the surgical approach, because radioablation is much more effective when there is minimal remaining normal thyroid tissue. Radioiodine is administered after iodine depletion (patient follows a low-iodine diet for 1–2 weeks) and in the presence of elevated serum TSH levels to stimulate uptake of the isotope into both the remnant and potentially any residual tumor. To achieve high serum TSH levels, there are two approaches. A patient may be withdrawn from thyroid hormone so that endogenous TSH is secreted and, ideally, the serum TSH level is >25 mIU/L at the time of <sup>131</sup>I therapy. A typical strategy is to treat the patient for several weeks postoperatively with liothyronine (25 µg qd or bid), followed by thyroid hormone withdrawal for 2 weeks. Alternatively, recombinant human TSH (rhTSH) is administered as two daily consecutive injections (0.9 mg) with administration of <sup>131</sup>I 24 h after the second injection. The patient can continue to take LT4 and remains euthyroid. Both approaches have equal success in achieving remnant ablation.

A pretreatment scanning dose of <sup>131</sup>I (usually 111 MBq [3 mCi]) or <sup>123</sup>I (74 MBq [2 mCi]) can reveal the amount of residual

tissue and provides guidance about the dose needed to accomplish ablation. However, because of concerns about radioactive "stunning" that impairs subsequent treatment, there is a trend to avoid pretreatment scanning with <sup>131</sup>I and use either <sup>123</sup>I or proceed directly to ablation, unless there is suspicion that the amount of residual tissue will alter therapy or that there is distant metastatic disease. In the United States, outpatient doses of up to 6475 MBq (175 mCi) can be given at most centers. The administered dose depends on the indication for therapy, with lower doses of 1100 MBq (30 mCi) given for remnant ablation but higher doses of up to 5500 MBq (150 mCi) reserved for use as adjuvant therapy when residual disease is suspected or present. Whole-body scanning (WBS) following radioiodine treatment is used to confirm the <sup>131</sup>I uptake in the remnant and to identify possible metastatic disease.

**Surveillance Testing** Serum thyroglobulin is a sensitive marker of residual/recurrent thyroid cancer after ablation of the residual postsurgical thyroid tissue. Current Tg assays have functional sensitivities as low as 0.1 ng/mL, as opposed to older assays with functional sensitivities of 1–2 ng/mL, reducing the number of patients with truly undetectable serum Tg levels. Because the vast majority of PTC recurrences are in cervical lymph nodes, a neck ultrasound should be performed about 6 months after thyroid ablation; ultrasound has been shown to be more sensitive than WBS in this scenario.

In low-risk patients who have no clinical evidence of residual disease after ablation, negative cervical sonography, and a basal Tg <0.2 ng/mL on LT4, the risk of structural recurrence is <3% at 5 years, and the frequency of follow-up testing can be decreased to annual TSH and Tg testing, with only periodic ultrasound examination.

The use of WBS is reserved for patients with known iodine-avid metastases or those with elevated serum thyroglobulin levels and negative imaging with ultrasound, chest CT, neck cross-sectional imaging, and positron emission tomography (PET) CT who may require additional <sup>131</sup>I therapy.

In addition to radioiodine, external beam radiotherapy is also used to treat gross residual neck disease or specific metastatic lesions, particularly when they cause bone pain or threaten neurologic injury (e.g., vertebral metastases).

**New Potential Therapies** Kinase inhibitors target pathways known to be active in thyroid cancer, including the RAS, BRAF, RET, EGFR, VEGFR, and angiogenesis pathways. Treatment has been shown to stabilize progressive metastatic disease that is refractory to radioiodine therapy, although only one study has demonstrated improved survival. Given the significant associated toxicities and the need for ongoing therapy, patient selection is critical to limit systemic therapy to those with significant morbidity risk. The American Thyroid Association guidelines recommend active surveillance for asymptomatic patients with metastatic tumors between 1 and 2 cm and then intervention as the rate of tumor growth increases. In addition, based on genetic analyses of metastases, mutation-selective kinase inhibitors are now being used. Ongoing trials are also exploring whether differentiation protocols, targeting the MAPK pathway, might enhance radioiodine uptake and efficacy.

## ANAPLASTIC AND OTHER FORMS OF THYROID CANCER

**Anaplastic Thyroid Cancer** As noted above, ATC is a poorly differentiated and aggressive cancer. The prognosis is poor, and most patients die within 6 months of diagnosis. Because of the undifferentiated state of these tumors, the uptake of radioiodine is usually negligible, but it can be used therapeutically if there is residual uptake. Chemotherapy has been attempted with multiple agents, including anthracyclines and paclitaxel, but it is usually ineffective. External beam radiation therapy can be attempted and continued if tumors are responsive. Both multitargeted and mutation-directed kinase inhibitors are in clinical trials and may prolong survival by a few months.

**Thyroid Lymphoma** Lymphoma in the thyroid gland often arises in the background of Hashimoto's thyroiditis. A rapidly expanding thyroid mass suggests the possibility of this diagnosis. Diffuse large-cell

lymphoma is the most common type in the thyroid. Biopsies reveal sheets of lymphoid cells that can be difficult to distinguish from small-cell lung cancer or ATC. These tumors are often highly sensitive to external radiation. Surgical resection should be avoided as initial therapy because it may spread disease that is otherwise localized to the thyroid. If staging indicates disease outside of the thyroid, treatment should follow guidelines used for other forms of lymphoma ([Chap. 108](#)).

### MEDULLARY THYROID CARCINOMA

MTC can be sporadic or familial and accounts for ~5% of thyroid cancers. There are three familial forms of MTC: MEN 2A, MEN 2B, and familial MTC without other features of MEN ([Chap. 388](#)). In general, MTC is more aggressive in MEN 2B than in MEN 2A, and familial MTC is more aggressive than sporadic MTC. Elevated serum calcitonin provides a marker of residual or recurrent disease. All patients with MTC should be tested for *RET* mutations, because genetic counseling and testing of family members can be offered to those individuals who test positive for mutations.

The management of MTC is primarily surgical. Prior to surgery, pheochromocytoma should be excluded in all patients with a *RET* mutation. Unlike tumors derived from thyroid follicular cells, these tumors do not take up radioiodine. External radiation treatment and targeted kinase inhibitors may provide palliation in patients with advanced disease ([Chap. 388](#)).

## APPROACH TO THE PATIENT

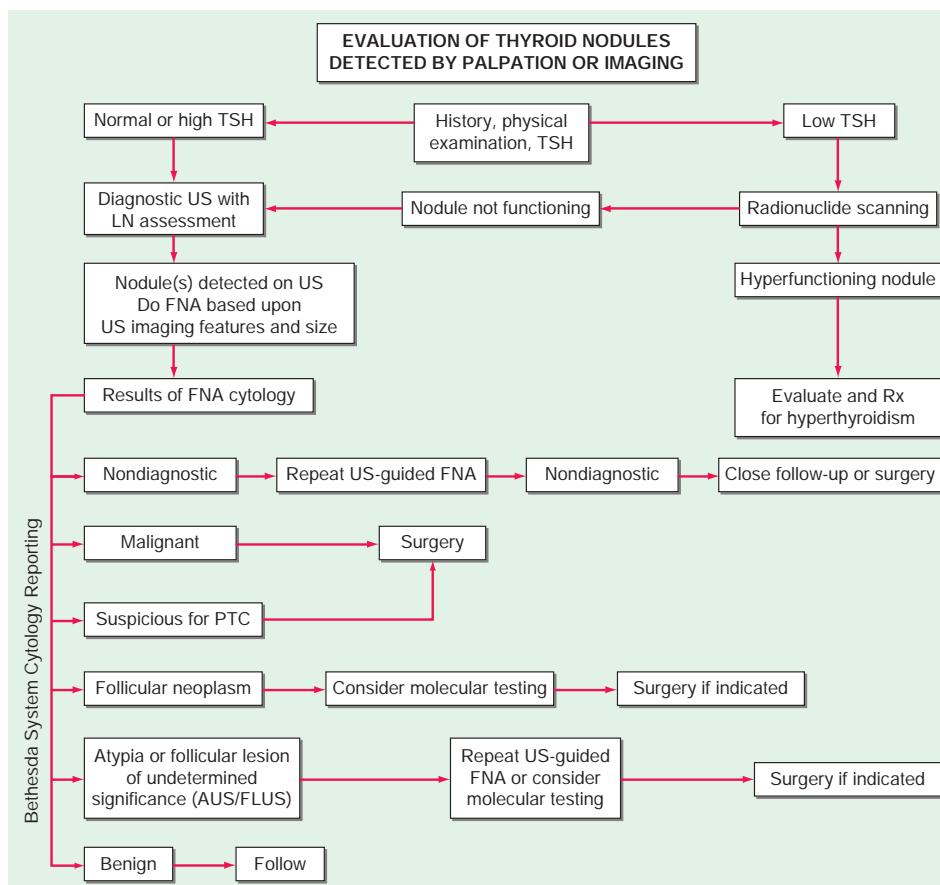
### Thyroid Nodules

Palpable thyroid nodules are found in ~5% of adults, but the prevalence varies considerably worldwide. Given this high prevalence

rate, practitioners may identify thyroid nodules on physical examination. However, the increased usage of diagnostic medical imaging (e.g., carotid ultrasound, cervical spine MRI) has led to an increased frequency of incidental nodule detection, accounting for the majority of patients currently presenting for nodule evaluation. The main goal of this evaluation is to identify, in a cost-effective manner, the small subgroup of individuals with malignant lesions that have the potential to be clinically significant.

Nodules are more common in iodine-deficient areas, in women, and with aging. Most palpable nodules are >1 cm in diameter, but the ability to feel a nodule is influenced by its location within the gland (superficial vs deeply embedded), the anatomy of the patient's neck, and the experience of the examiner. More sensitive methods of detection, such as CT, thyroid ultrasound, and pathologic studies, reveal thyroid nodules in up to 50% of glands in individuals aged >50 years. The presence of these thyroid incidentalomas has led to much debate about how to detect nodules and which nodules to investigate further.

An approach to the evaluation of thyroid nodules detected by either palpation or imaging is outlined in [Fig. 385-5](#). Most patients with thyroid nodules have normal thyroid function tests. Nonetheless, thyroid function should be assessed by measuring a TSH level, which may be suppressed by one or more autonomously functioning nodules. If the TSH is suppressed, a radionuclide scan is indicated to determine if the identified nodule is "hot," as lesions with increased uptake are almost never malignant and FNA is unnecessary. Otherwise, the next step in evaluation is performance of a thyroid ultrasound for three reasons: (1) For nodules detected on physical examination, ultrasound will confirm if the palpable nodule is indeed a nodule. About 15% of "palpable" nodules are



**FIGURE 385-5** Approach to the patient with a thyroid nodule. See text and references for details. FNA, fine-needle aspiration; LN, lymph node; PTC, papillary thyroid cancer; Rx, therapy; TSH, thyroid-stimulating hormone; US, ultrasound.

not confirmed on imaging, and therefore no further evaluation is required. (2) Ultrasound will assess if there are additional non-palpable nodules for which FNA may be recommended based on imaging features and size. (3) Ultrasound will characterize the imaging pattern of the nodule, which, combined with the nodule's size, facilitates decision-making about FNA. There are several validated risk stratification systems (RSS) for sonographic imaging of thyroid nodules (American College of Radiology [ACR] Thyroid Imaging Reporting and Data System [TI-RADS], American Thyroid Association, European Thyroid Association [EU-TIRADS], among others). These demonstrate consistent risk estimates for thyroid cancer based on certain sonographic patterns. All provide size cutoff recommendations for nodule FNA based on sonographic patterns, with lower size cutoffs for nodules with more suspicious ultrasound patterns, but the specific size cutoff criteria differ among the RSS. Not surprisingly, the RSSs with lower size cutoffs have higher sensitivity and lower specificity for thyroid cancer diagnosis than those with higher cutoffs. Nevertheless, all have been shown to reduce unnecessary FNAs by at least 45%, in part due to the recommendation not to perform FNA for spongiform nodules. ACR TI-RADS is currently the most widely used RSS in the United States, and nodules are classified from TR1 to TR5 (Fig. 385-1).

For example, a spongiform nodule (TR1) has a <3% chance of cancer, and observation rather than FNA is generally recommended by all RSSs, whereas 10–20% of solid hypoechoic nodules with smooth borders (TR4) are malignant and FNA is recommended at size cutoffs ranging from 1–1.5 cm. All the RSSs recommend FNA at 1 cm for the highest suspicion pattern nodule, TR5 (Figs. 385-1 and 385-2). Given what is known about the prevalence and generally indolent behavior of small thyroid cancers <1 cm, none of the RSSs recommend FNA for any nodule <1 cm unless metastatic cervical lymph nodes are present.

FNA biopsy, ideally performed with ultrasound guidance, is the best diagnostic test when performed by physicians familiar with the procedure and when the results are interpreted by experienced cytopathologists. The technique is particularly useful for detecting PTC. However, the distinction between benign and malignant follicular patterned lesions is often not possible using cytology alone because of the absence of characteristic nuclear features in follicular carcinoma. Using the current ultrasound RSS for FNA decision-making, FNA biopsies yield the following spectrum of cytology diagnoses: 55–60% benign, 5% malignant or suspicious for malignancy, 5–7% nondiagnostic or yielding insufficient material for diagnosis, and 25–30% indeterminate. The Bethesda System is now widely used to provide more uniform terminology for reporting thyroid nodule FNA cytology results. This six-tiered classification system with the respective estimated malignancy rates is shown in **Table 385-4**. Importantly, because NIFTP can only be diagnosed by surgical pathology, NIFTP is included in the malignancy estimates. Specifically, the Bethesda System subcategorized cytology specimens previously labeled as indeterminate into three categories: atypia or follicular lesion of undetermined significance (AUS/FLUS), follicular neoplasm, and suspicious for malignancy.

**TABLE 385-4** Bethesda Classification for Thyroid Cytology Version 2

DIAGNOSTIC CATEGORY	RISK OF MALIGNANCY (INCLUDING NIFTP)
I. Nondiagnostic or unsatisfactory	5–10%
II. Benign	0–3%
III. Atypia or follicular lesion of unknown significance (AUS/FLUS)	~10–30%
IV. Follicular neoplasm	25–40%
V. Suspicious for malignancy	50–75%
VI. Malignant	97–99%

Abbreviation: NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear features.

Cytology results indicative of malignancy generally mandate surgery, after performing preoperative sonography to evaluate the cervical lymph nodes. Nondiagnostic cytology specimens most often result from cystic lesions but may also occur in fibrous long-standing nodules or very vascular nodules where a longer needle dwell time may result in a hemorrhagic specimen. Ultrasound-guided FNA is indicated when a repeat FNA is necessary. Repeat FNA will yield a diagnostic cytology in ~50% of cases. Given the low false-negative rate of a benign cytology (<3%), benign nodules with a low suspicion sonographic pattern (TR2, TR3) can be followed. Those with more worrisome ultrasound features, especially TR5 nodules, should undergo repeat FNA because of a higher likelihood of a missed malignancy. The use of LT4 to suppress serum TSH is not effective in shrinking nodules in iodine-replete populations, and therefore, LT4 suppression should not be used. The three indeterminate cytology classifications introduced by the Bethesda System are associated with different risks of malignancy (Table 385-4). For nodules with suspicious for malignancy cytology, surgery is recommended after ultrasound assessment of cervical lymph nodes. Options to be discussed with the patient include lobectomy versus total thyroidectomy.

On the other hand, the majority of nodules with AUS/FLUS and follicular neoplasm cytology results are benign; the range of malignancy (ROM) varies from 10 to 40%. The traditional approach for these patients is diagnostic lobectomy for histopathologic diagnosis. Therefore, many patients undergo surgery for benign nodules. Over the past decade, the uncertainty about the ROM for indeterminate cytology nodules has been the driver for the development of molecular testing, which can better differentiate benign from malignant nodules. Based on results from next-generation sequencing, which includes point mutations, small insertions/deletions, and gene fusions, as well as results from microRNA analyses and gene expression, the current validated and commercially available molecular tests combine these techniques with the following two goals: (1) risk stratification of thyroid nodules based on a positive result; and (2) reduction in cancer risk to an acceptable level for nonsurgical surveillance based on a negative result. Assuming a 25–30% ROM for nodules with indeterminate cytology, the negative predictive values for the currently validated molecular tests are >95%.

The evaluation of a thyroid nodule is stressful for most patients. They are concerned about the possibility of thyroid cancer, whether verbalized or not. It is constructive, therefore, to review the diagnostic approach and to reassure patients when no malignancy is found. When a suspicious lesion or thyroid cancer is identified, the generally favorable prognosis and available treatment options can be reassuring.

## FURTHER READING

- Cibas ES et al: The 2017 Bethesda system for reporting thyroid cytopathology. *Thyroid* 27:1341, 2017.
- Davies L, Hoang J: Thyroid cancer in the USA: Current trends and outstanding questions. *Lancet Diab Endocrinol* 9:11, 2021.
- Dunn LA et al: Vemurafenib redifferentiation of BRAF mutant, RAI-refractory thyroid cancers. *J Clin Endocrinol Metab* 104:1417, 2019.
- Durante C et al: The diagnosis and management of thyroid nodules: A review. *JAMA* 319:914, 2018.
- Fagin JA, Wells SA: Biologic and clinical perspectives on thyroid cancer. *N Engl J Med* 375:1054, 2016.
- Haugen BR et al: 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer. *Thyroid* 26:1, 2016.
- Tessler FN et al: ACR Thyroid Imaging Reporting and Data System (TI-RADS): White paper of the ACR TI-RADS committee. *J Am Coll Radiol* 14:587, 2017.
- Tuttle RM et al: Updated American Joint Committee and Cancer/Tumor-Node-Metastasis Staging System for differentiated and anaplastic thyroid cancer (eighth edition): What changed and why? *Thyroid* 27:751, 2017.

# 386

## Disorders of the Adrenal Cortex

Wiebke Arlt

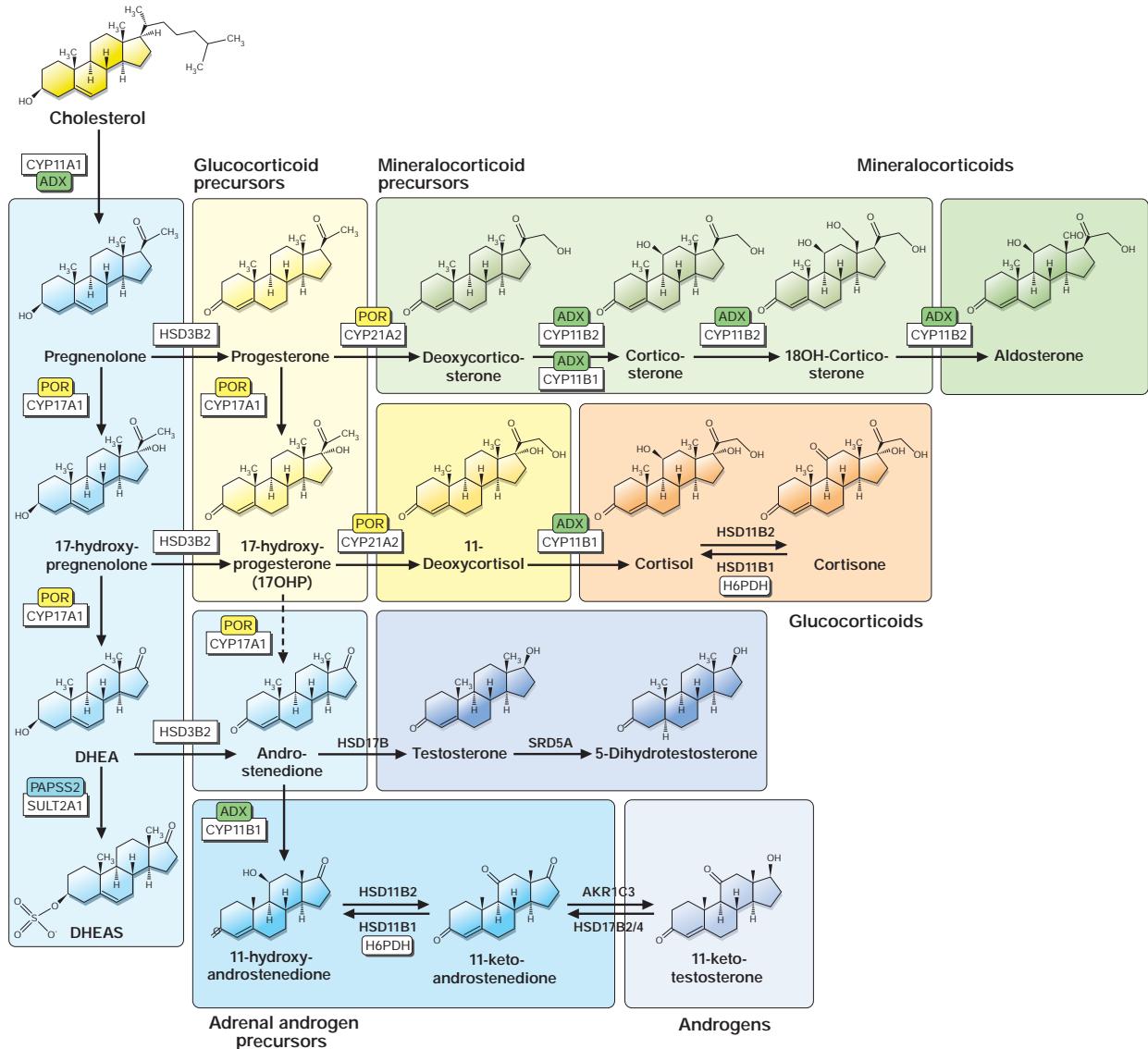


The adrenal cortex produces three classes of corticosteroid hormones: glucocorticoids (e.g., cortisol), mineralocorticoids (e.g., aldosterone), and adrenal androgen precursors (e.g., dehydroepiandrosterone [DHEA]) (Fig. 386-1). Glucocorticoids and mineralocorticoids act through specific nuclear receptors, regulating aspects of the physiologic stress response as well as blood pressure and electrolyte homeostasis. Adrenal androgen precursors are converted in the gonads and peripheral target cells to sex steroids that act via nuclear androgen and estrogen receptors.

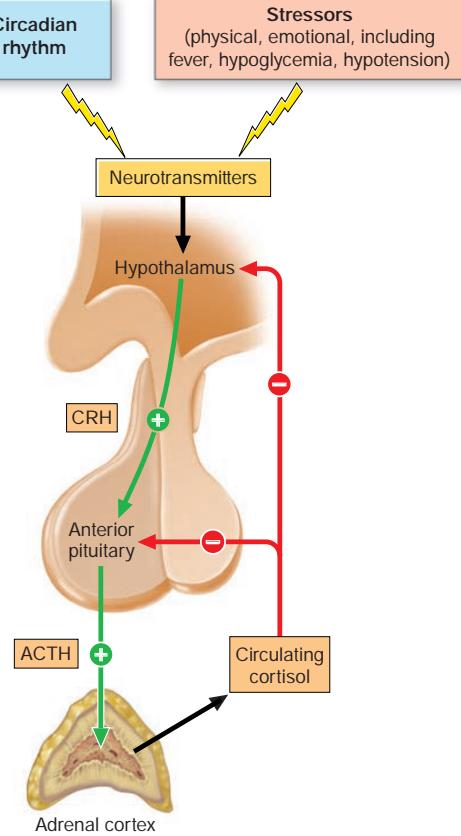
Disorders of the adrenal cortex are characterized by deficiency or excess of one or several of the three major corticosteroid classes. Hormone deficiency can be caused by inherited glandular or enzymatic disorders or by destruction of the pituitary or adrenal gland by autoimmune disorders, infection, infarction, or iatrogenic events such as surgery or hormonal suppression. Hormone excess is usually the result of neoplasia, leading to increased production of adrenocorticotrophic hormone (ACTH) by the pituitary or neuroendocrine ectopic ACTH-producing cells or increased production of glucocorticoids, mineralocorticoids, or adrenal androgen precursors by adrenal nodules. Adrenal nodules are increasingly identified incidentally during cross-sectional imaging of chest or abdomen performed for other reasons.

### ADRENAL ANATOMY AND DEVELOPMENT

The normal adrenal glands weigh 6–11 g each. They are located above the kidneys and have their own blood supply. Arterial blood flows



**FIGURE 386-1** Adrenal steroidogenesis. ADX, adrenodoxin; AKR1C3, aldo-keto reductase 1C3; CYP11A1, side chain cleavage enzyme; CYP11B1, 11 $\beta$ -hydroxylase; CYP11B2, aldosterone synthase; CYP17A1, 17 $\alpha$ -hydroxylase/17,20 lyase; CYP21A2, 21-hydroxylase; DHEA, dehydroepiandrosterone; DHEAS, dehydroepiandrosterone sulfate; H6PDH, hexose-6-phosphate dehydrogenase; HSD11B1, 11 $\beta$ -hydroxysteroid dehydrogenase type 1; HSD11B2, 11 $\beta$ -hydroxysteroid dehydrogenase type 2; HSD17B, 17 $\beta$ -hydroxysteroid dehydrogenase; HSD3B2, 3 $\beta$ -hydroxysteroid dehydrogenase type 2; PAPSS2, PAPS synthase type 2; POR, P450 oxidoreductase; SRD5A, 5 $\alpha$ -reductase; SULT2A1, DHEA sulfotransferase.



**FIGURE 386-2** Regulation of the hypothalamic-pituitary-adrenal (HPA) axis. ACTH, adrenocorticotrophic hormone; CRH, corticotropin-releasing hormone.

initially to the subcapsular region and then meanders from the outer cortical zona glomerulosa through the intermediate zona fasciculata to the inner zona reticularis and eventually to the adrenal medulla. The right suprarenal vein drains directly into the vena cava, while the left suprarenal vein drains into the left renal vein.

During early embryonic development, the adrenals originate from the urogenital ridge and then separate from gonads and kidneys at about the sixth week of gestation. Concordant with the time of sexual differentiation (seventh to ninth week of gestation, [Chap. 390](#)), the adrenal cortex starts to produce cortisol and the adrenal sex steroid precursor DHEA. The orphan nuclear receptors SF1 (steroidogenic factor 1; encoded by the gene *NR5A1*) and DAX1 (dosage-sensitive sex reversal gene 1; encoded by the gene *NR0B1*), among others, play a crucial role during this period of development, as they regulate a multitude of adrenal genes involved in steroidogenesis.

### REGULATORY CONTROL OF STEROIDOGENESIS

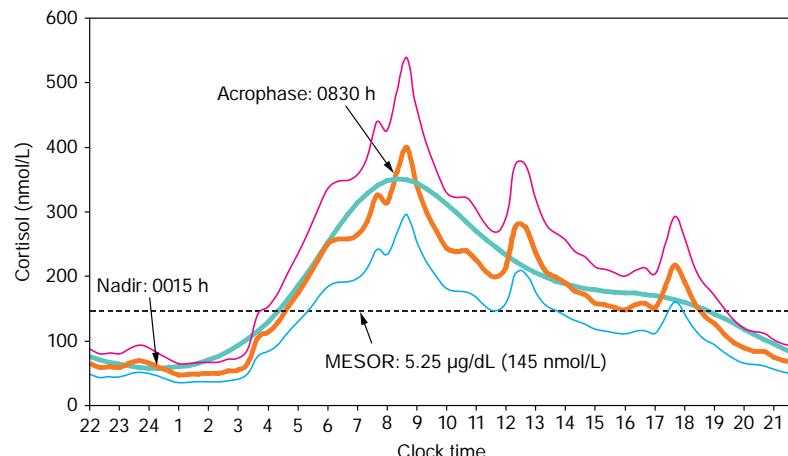
Production of glucocorticoids and adrenal androgens is under the control of the hypothalamic-pituitary-adrenal (HPA) axis, whereas mineralocorticoids are regulated by the renin-angiotensin-aldosterone (RAA) system.

Glucocorticoid synthesis is under inhibitory feedback control by the hypothalamus and the pituitary ([Fig. 386-2](#)). Hypothalamic release of corticotropin-releasing hormone (CRH) occurs

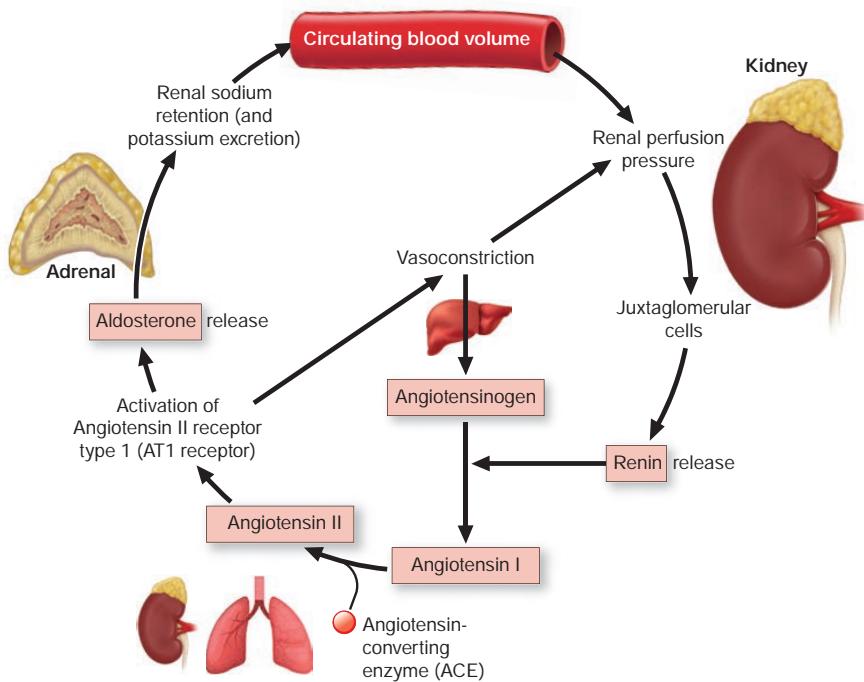
in response to endogenous or exogenous stress. CRH stimulates the cleavage of the 241-amino acid polypeptide proopiomelanocortin (POMC) by pituitary-specific prohormone convertase 1 (PC1), yielding the 39-amino acid peptide ACTH. ACTH is released by the corticotrope cells of the anterior pituitary and acts as the pivotal regulator of adrenal cortisol synthesis, with additional short-term effects on mineralocorticoid and adrenal androgen synthesis. The release of CRH, and subsequently ACTH, occurs in a pulsatile fashion that follows a circadian rhythm under the control of the hypothalamus, specifically its suprachiasmatic nucleus (SCN), with additional regulation by a complex network of cell-specific clock genes. Reflecting the pattern of ACTH secretion, adrenal cortisol secretion exhibits a distinct circadian rhythm, starting to rise in the early morning hours prior to awakening, with peak levels in the morning and low levels in the evening ([Fig. 386-3](#)).

Diagnostic tests assessing the HPA axis make use of the fact that it is regulated by negative feedback. Glucocorticoid excess is diagnosed by employing a dexamethasone suppression test. Dexamethasone, a potent synthetic glucocorticoid, suppresses CRH/ACTH by binding hypothalamic-pituitary glucocorticoid receptors (GRs) and, therefore, results in downregulation of endogenous cortisol synthesis. Various versions of the dexamethasone suppression test are described in detail in [Chap. 380](#). If cortisol production is autonomous (e.g., adrenal nodule), ACTH is already suppressed, and dexamethasone has little additional effect. If cortisol production is driven by an ACTH-producing pituitary adenoma, dexamethasone suppression is ineffective at low doses but usually induces suppression at high doses. If cortisol production is driven by an ectopic source of ACTH, the tumors are usually resistant to dexamethasone suppression. Thus, the dexamethasone suppression test is useful to establish the diagnosis of Cushing's syndrome and assist with the differential diagnosis of cortisol excess.

Conversely, to assess glucocorticoid deficiency, ACTH stimulation of cortisol production is used. The ACTH peptide contains 39 amino acids, but the first 24 are sufficient to elicit a physiologic response. The standard ACTH stimulation test involves administration of cosyntropin (ACTH 1-24), 0.25 mg IM or IV, and collection of blood samples at 0, 30, and 60 min for cortisol. A normal response is defined as a cortisol level >15–20 µg/dL (>400–550 nmol/L) 30–60 min after cosyntropin stimulation, with the precise cutoff dependent on the assay used. A low-dose (1 µg cosyntropin IV) version of this test has been advocated; however, it has no superior diagnostic value and is more cumbersome to carry out. Alternatively, an insulin tolerance test (ITT) can be used to assess adrenal function. It involves injection of insulin to induce hypoglycemia, which represents a strong stress signal that triggers



**FIGURE 386-3** Physiologic cortisol circadian rhythm. Circulating cortisol concentrations (geometrical mean ± standard deviation values and fitted cosinor) drop under the rhythm-adjusted mean (MESOR) in the early evening hours, with nadir levels around midnight and a rise in the early morning hours; peak levels are observed ~8:30 A.M. (acrophase). (Reproduced with permission from M Debono et al: Modified-release hydrocortisone to provide circadian cortisol profiles. *J Clin Endocrinol Metab* 94:1548, 2009.)



**FIGURE 386-4** Regulation of the renin-angiotensin-aldosterone (RAA) system.

hypothalamic CRH release and activation of the entire HPA axis. The ITT involves administration of regular insulin 0.1 U/kg IV (dose should be lower if hypopituitarism is likely) and collection of blood samples at 0, 30, 60, and 120 min for glucose, cortisol, and growth hormone (GH), if also assessing the GH axis. Oral or IV glucose is administered after the patient has achieved symptomatic hypoglycemia (usually plasma glucose <40 mg/dL). A normal response is defined as a cortisol >20 µg/dL and GH >5.1 µg/L, again with assay-specific cutoff variability. The ITT requires careful clinical monitoring and sequential measurements of glucose. It is contraindicated in patients with coronary disease, cerebrovascular disease, or seizure disorders, which has made the short cosyntropin test the commonly accepted first-line test.

Mineralocorticoid production is controlled by the RAA regulatory cycle, which is initiated by the release of renin from the juxtaglomerular cells in the kidney, resulting in cleavage of angiotensinogen to angiotensin I in the liver (Fig. 386-4). Angiotensin-converting enzyme (ACE) cleaves angiotensin I to angiotensin II, which binds and activates the angiotensin II receptor type 1 (AT1 receptor [AT1R]), resulting in increased adrenal aldosterone production and vasoconstriction. Aldosterone enhances sodium retention and potassium excretion and increases the arterial perfusion pressure, which in turn regulates renin release. Because mineralocorticoid synthesis is primarily under the control of the RAA system, hypothalamic-pituitary damage does not significantly impact the capacity of the adrenal to synthesize aldosterone.

Similar to the HPA axis, the assessment of the RAA system can be used for diagnostic purposes. If mineralocorticoid excess is present, there is a counter-regulatory downregulation of plasma renin (see below for testing). Conversely, in mineralocorticoid deficiency, plasma renin is markedly increased. Physiologically, oral or IV sodium loading results in suppression of aldosterone, a response that is attenuated or absent in patients with autonomous mineralocorticoid excess.

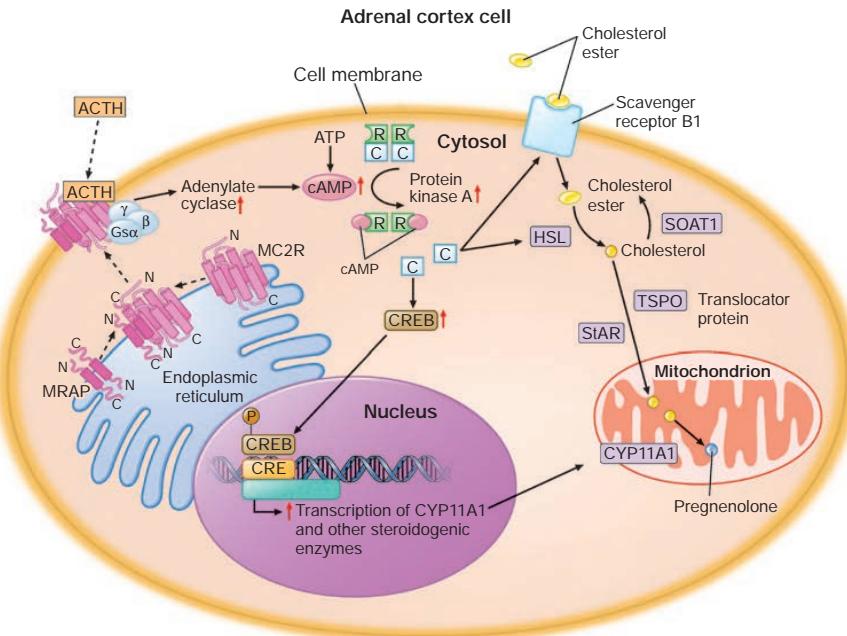
### STEROID HORMONE SYNTHESIS, METABOLISM, AND ACTION

ACTH stimulation is required for the initiation of steroidogenesis. The ACTH receptor MC2R (melanocortin 2 receptor) interacts with the

MC2R-accessory protein MRAP, and the complex is transported to the adrenocortical cell membrane, where it binds to ACTH (Fig. 386-5). ACTH stimulation generates cyclic AMP (cAMP), which upregulates the protein kinase A (PKA) signaling pathway. Inactive PKA is a tetramer of two regulatory and two catalytic subunits that is dissociated by cAMP into a dimer of two regulatory subunits bound to cAMP and two free and active catalytic subunits. PKA activation impacts steroidogenesis in three distinct ways: (1) increases the import of cholesterol esters; (2) increases the activity of hormone-sensitive lipase, which cleaves cholesterol esters to cholesterol for import into the mitochondrion; and (3) increases the availability and phosphorylation of CREB (cAMP response element binding), a transcription factor that enhances transcription of CYP11A1 and other enzymes required for glucocorticoid synthesis.

Adrenal steroidogenesis occurs in a zone-specific fashion, with mineralocorticoid synthesis occurring in the outer zona glomerulosa, glucocorticoid synthesis in the zona fasciculata, and adrenal androgen biosynthesis in the inner zona reticularis serving as precursors for both classic and 11-oxygenated androgens (Fig. 386-1). All steroidogenic pathways require cholesterol import into the mitochondrion, a process initiated by the action of the steroidogenic acute regulatory (STAR) protein, which shuttles cholesterol from the outer to the inner mitochondrial membrane. The majority of steroidogenic enzymes are cytochrome P450 (CYP) enzymes, which are either located in the mitochondrion (side chain cleavage enzyme, CYP11A1; 11-hydroxylase, CYP11B1; aldosterone synthase, CYP11B2) or in the endoplasmic reticulum membrane (17-hydroxylase, CYP17A1; 21-hydroxylase, CYP21A2; aromatase, CYP19A1). These enzymes require electron donation via specific redox cofactor enzymes, P450 oxidoreductase (POR), and adrenodoxin/adrenodoxin reductase (ADX/ADR) for the microsomal and mitochondrial CYP enzymes, respectively. In addition, the short-chain dehydrogenase 3-hydroxysteroid dehydrogenase type 2 (3-HSD2), also termed 4,5-isomerase, plays a major role in adrenal steroidogenesis.

The cholesterol side chain cleavage enzyme CYP11A1 generates pregnenolone. Glucocorticoid synthesis requires conversion of pregnenolone to progesterone by 3-HSD2, followed by conversion



**FIGURE 386-5** ACTH effects on adrenal steroidogenesis. ACTH, adrenocorticotrophic hormone; CREB, cAMP response element-binding protein; HSL, hormone-sensitive lipase; MRAP, MC2R-accessory protein; protein kinase A catalytic subunit (C; PRKACA); PKA regulatory subunit (R; PRKAR1A); SOAT1, sterol O-acyltransferase 1; STAR, steroidogenic acute regulatory (protein); TSPO, translocator protein.

to 17-hydroxyprogesterone (17OHP) by CYP17A1, further hydroxylation at carbon 21 by CYP21A2, and eventually, 11'-hydroxylation by CYP11B1 to generate active cortisol (Fig. 386-1). Mineralocorticoid synthesis also requires progesterone, which is first converted to deoxycorticosterone (DOC) by CYP21A2 and then converted via corticosterone and 18-hydroxycorticosterone to aldosterone in three steps catalyzed by CYP11B2. For adrenal androgen synthesis, pregnenolone undergoes conversion by CYP17A1, which uniquely catalyzes two enzymatic reactions. Via its 17'-hydroxylase activity, CYP17A1 converts pregnenolone to 17-hydroxypregnенolone, followed by generation of the universal sex steroid precursor DHEA via CYP17A1 17,20 lyase activity. The majority of DHEA is secreted by the adrenal in the form of its sulfate ester, DHEAS, generated by DHEA sulfotransferase (SULT2A1). DHEA is converted to androstenedione, which can be activated to testosterone or channeled into the 11'-oxygenated androgen pathway by 11'-hydroxylation (CYP11B1).

Following its release from the adrenal, cortisol circulates in the bloodstream mainly bound to cortisol-binding globulin (CBG) and, to a lesser extent, to albumin, with only a minor fraction circulating as free, unbound hormone. Free cortisol is thought to enter cells directly, not requiring active transport. In addition, in a multitude of peripheral target tissues of glucocorticoid action, including adipose, liver, muscle, and brain, cortisol is generated from inactive cortisone within the cell by the enzyme 11'-hydroxysteroid dehydrogenase type 1 (11'-HSD1) (Fig. 386-6). Thereby, 11'-HSD1 functions as a tissue-specific pre-receptor regulator of glucocorticoid action. For the conversion of inactive cortisone to active cortisol, 11'-HSD1 requires nicotinamide adenine dinucleotide phosphate (NADPH [reduced form]), which is provided by the enzyme hexose-6-phosphate dehydrogenase (H6PDH). Like the catalytic domain of 11'-HSD1, H6PDH is located in the lumen of the endoplasmic reticulum and converts glucose-6-phosphate (G6P) to 6-phosphogluconate (6PGL), thereby regenerating NADP<sup>+</sup> to NADPH, which drives the activation of cortisol from cortisone by 11'-HSD1.

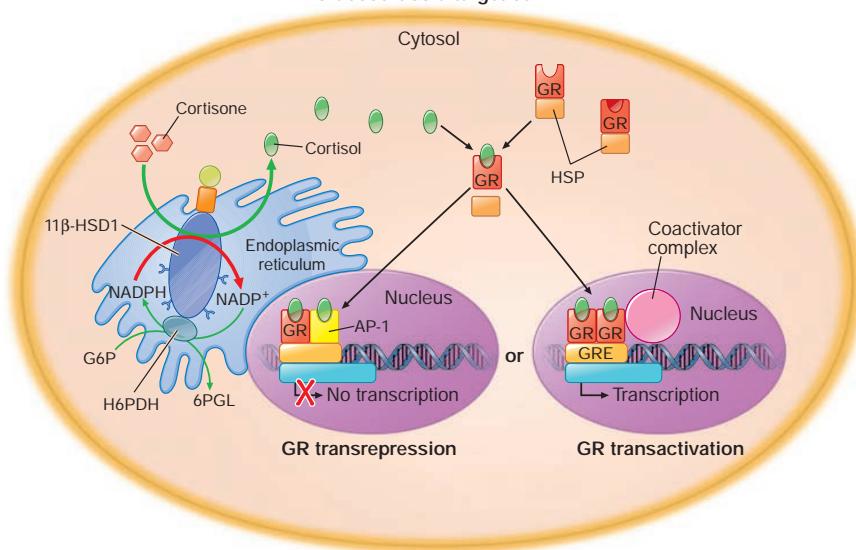
In the cytosol of target cells, cortisol binds and activates the GR, which results in dissociation of heat shock proteins (HSPs) from the receptor and subsequent dimerization (Fig. 386-6). Cortisol-bound GR dimers translocate to the nucleus and activate glucocorticoid response elements (GREs) in the DNA sequence, thereby enhancing

transcription of glucocorticoid-regulated genes (GR transactivation). However, cortisol-bound GR can also form heterodimers with transcription factors such as AP-1 or NF- $\kappa$ B, resulting in transrepression of proinflammatory genes, a mechanism of major importance for the anti-inflammatory action of glucocorticoids. It is important to note that corticosterone also exerts glucocorticoid activity, albeit much weaker than cortisol itself. However, in rodents, corticosterone is the major glucocorticoid, and in patients with 17-hydroxylase deficiency, lack of cortisol can be compensated for by higher concentrations of corticosterone that accumulates as a consequence of the enzymatic block.

Cortisol is inactivated to cortisone by the microsomal enzyme 11'-hydroxysteroid dehydrogenase type 2 (11'-HSD2) (Fig. 386-7), mainly in the kidney, but also in the colon, salivary glands, and other target tissues. Cortisol and aldosterone bind the mineralocorticoid receptor (MR) with equal affinity; however, cortisol circulates in the bloodstream at about a 1000-fold higher concentration. Thus, only rapid inactivation of cortisol to cortisone by 11'-HSD2 prevents MR activation by excess cortisol, thereby acting as a tissue-specific modulator of the MR pathway. In addition to cortisol and aldosterone, DOC (Fig. 386-1) also exerts mineralocorticoid activity. DOC accumulation due to 11'-hydroxylase deficiency or due to tumor-related excess production can result in mineralocorticoid excess.

Aldosterone synthesis in the adrenal zona glomerulosa cells is driven by the enzyme aldosterone synthase (CYP11B2). The binding of angiotensin II to the AT1 receptor causes glomerulosa cell membrane depolarization by increasing intracellular sodium through inhibition of sodium potassium ( $\text{Na}^+/\text{K}^+$ ) ATPase enzymes as well as potassium channels. This drives an increase in intracellular calcium by opening of voltage-dependent calcium channels or inhibition of calcium ( $\text{Ca}^{2+}$ ) ATPase enzymes. Consequently, the calcium signaling pathway is triggered, resulting in upregulation of CYP11B2 transcription (Fig. 386-8).

Analogous to cortisol action via the GR, aldosterone (or cortisol) binding to the MR in the kidney tubule cell dissociates the HSP-receptor complex, allowing homodimerization of the MR and translocation of the hormone-bound MR dimer to the nucleus (Fig. 386-7). The activated MR enhances transcription of the epithelial sodium channel (ENaC) and serum glucocorticoid-inducible kinase 1 (SGK-1). In the cytosol, interaction of ENaC with Nedd4 prevents cell surface



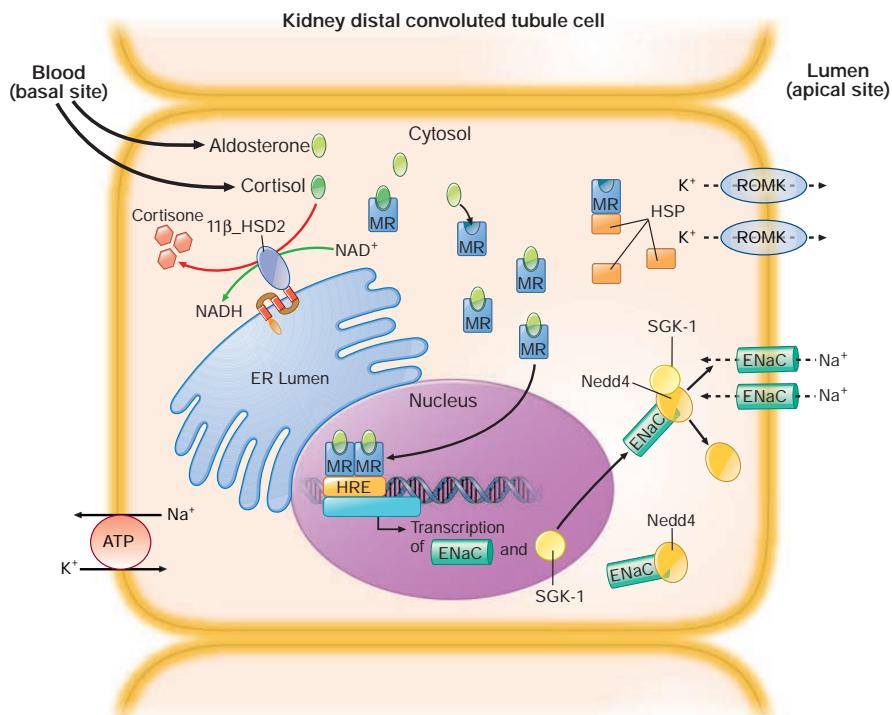
**FIGURE 386-6** Prereceptor activation of cortisol and glucocorticoid receptor (GR) action. AP-1, activator protein-1; G6P, glucose-6-phosphate; GREs, glucocorticoid response elements; HSPs, heat shock proteins; NADPH, nicotinamide adenine dinucleotide phosphate (reduced form); 6PGL, 6-phosphogluconate.

expression of ENaC. However, SGK-1 phosphorylates serine residues within the Nedd4 protein, reduces the interaction between Nedd4 and ENaC, and consequently, enhances the trafficking of ENaC to the cell surface, where it mediates sodium retention.

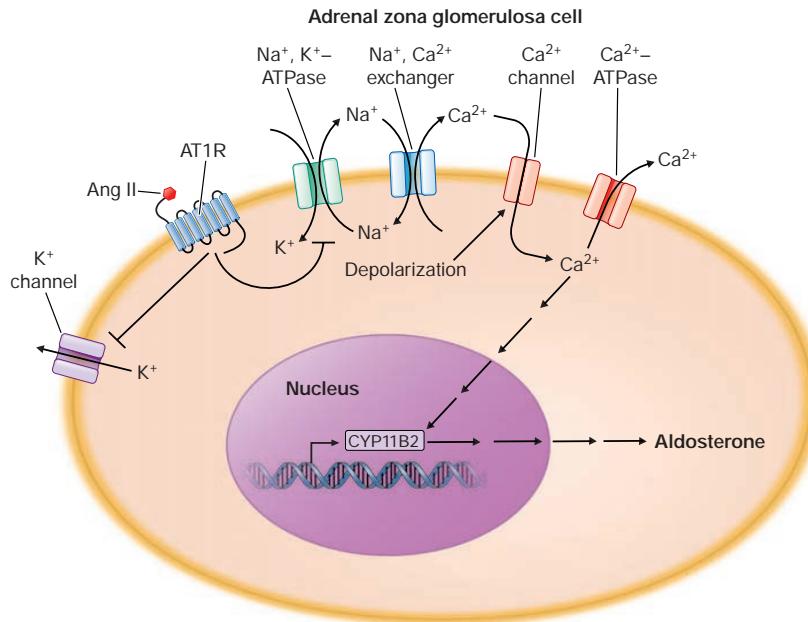
### CUSHING'S SYNDROME

(See also Chap. 380) Cushing's syndrome reflects a constellation of clinical features that result from chronic exposure to excess

glucocorticoids of any etiology. The disorder can be ACTH-dependent (e.g., pituitary corticotrope adenoma, ectopic secretion of ACTH by nonpituitary tumor) or ACTH-independent (e.g., adrenocortical adenoma, adrenocortical carcinoma [ACC], nodular adrenal hyperplasia), as well as iatrogenic (e.g., administration of exogenous glucocorticoids to treat various inflammatory conditions). The term *Cushing's disease* refers specifically to Cushing's syndrome caused by a pituitary corticotrope adenoma.



**FIGURE 386-7** Prereceptor inactivation of cortisol and mineralocorticoid receptor action. ENaC, epithelial sodium channel; HRE, hormone response element; Na<sup>+</sup>/K<sup>+</sup>-ATPase, sodium-potassium adenosine triphosphatase; NADH, nicotinamide adenine dinucleotide; ROMK, renal outer medullary potassium channel; SGK-1, serum glucocorticoid-inducible kinase-1.



**FIGURE 386-8** Regulation of adrenal aldosterone synthesis. Ang II, angiotensin II; AT1R, angiotensin II receptor type 1; CYP11B2, aldosterone synthase.

**Epidemiology** Cushing's syndrome is generally considered a rare disease. It occurs with an incidence of 1–2 per 100,000 population per year. However, it is debated whether mild cortisol excess may be more prevalent among patients with features of Cushing's such as centripetal obesity, type 2 diabetes, and osteoporotic vertebral fractures, recognizing that these are relatively nonspecific and common in the population.

In the overwhelming majority of patients with endogenous Cushing's syndrome, the underlying cause is an ACTH-producing corticotrope adenoma of the pituitary (Table 386-1), as initially described by Harvey Cushing in 1912. Cushing's disease more frequently affects women, with the exception of prepubertal cases, where it is more common in boys. By contrast, ectopic ACTH syndrome is more frequently identified in men. Only 10% of patients with Cushing's syndrome have a primary, adrenal cause of their disease (e.g., autonomous cortisol excess independent of ACTH), and most of these patients are women. However, overall, the medical use of glucocorticoids for immunosuppression or for the treatment of inflammatory disorders is the most common cause of Cushing's syndrome, also termed iatrogenic Cushing's.

**Etiology** In at least 90% of patients with Cushing's disease, ACTH excess is caused by a corticotrope pituitary microadenoma, often only a few millimeters in diameter. Pituitary macroadenomas (i.e., tumors >1 cm in size) are found in only 5–10% of patients. Pituitary corticotrope adenomas usually occur sporadically but very rarely can be found in the context of multiple endocrine neoplasia type 1 (MEN 1) (Chap. 388). Pituitary adenomas causative of Cushing's disease frequently harbor mutations in the deubiquitinase USP8, which leads to constitutive activation of epidermal growth factor (EGF) signaling and consequent upregulated expression of the ACTH precursor POMC. USP8 mutations are found more frequently in adults (41 vs 17% in children) and in women (43 vs 17% in men) with Cushing's disease.

Ectopic ACTH production is predominantly caused by occult carcinoid tumors, most frequently in the lung, but also in thymus or pancreas. Because of their small size, these tumors are often difficult to locate. Advanced small-cell lung cancer can cause ectopic ACTH production. In rare cases, ectopic CRH and/or ACTH production has been found to originate from medullary thyroid carcinoma or pheochromocytoma, the

latter co-secreting catecholamines and ACTH.

The majority of patients with endogenous ACTH-independent cortisol excess harbor a cortisol-producing adrenal adenoma, and somatic mutations in the PKA catalytic subunit *PRKAC*A have been identified as cause of disease in 40% of these tumors. ACCs may also cause ACTH-independent disease and are often large, with excess production of several corticosteroid classes.

A rare but notable cause of adrenal cortisol excess is primary bilateral macronodular adrenal hyperplasia (PBMAH) with low circulating ACTH but with evidence for autocrine stimulation of cortisol production via intraadrenal ACTH production. These hyperplastic nodules are often also characterized by ectopic expression of G protein-coupled receptors not usually found in the adrenal, including receptors for luteinizing hormone, vasopressin, serotonin, interleukin 1, catecholamines, or gastric inhibitory peptide (GIP), the cause of food-dependent Cushing's. Activation of these receptors results in upregulation of PKA signaling, as physiologically occurs with ACTH, with a subsequent increase in cortisol production. A combination of germline and somatic mutations in the tumor-suppressor gene *ARMC5* have been identified as a prevalent cause of Cushing's due to bilateral macronodular adrenal hyperplasia; these patients often present with biochemical evidence of Cushing's but lack specific clinical signs, which develop slowly over decades and accelerate cardiovascular risk. Constitutively activating mutations in the PKA catalytic subunit *PRKAC*A are found as somatic mutations in one-third of cortisol-producing adrenocortical adenomas and, as germline mutations, can also represent a rare cause of macronodular adrenal hyperplasia associated with cortisol excess.

Germline mutations in one of the regulatory subunits of PKA, *PRKAR1A*, are found in patients with primary pigmented nodular adrenal disease (PPNAD) as part of *Carney's complex*, an autosomal dominant multiple neoplasia condition associated with cardiac myxomas, hyperpigmentation, Sertoli cell tumors, and PPNAD. PPNAD can present as micronodular or macronodular hyperplasia, or both. Phosphodiesterases can influence intracellular cAMP and can thereby impact PKA activation. Mutations in *PDE11A* and *PDE8B* have been identified in patients with bilateral adrenal hyperplasia and Cushing's, with and without evidence of PPNAD.

Another rare cause of ACTH-independent Cushing's is *McCune-Albright syndrome*, also associated with polyostotic fibrous dysplasia, unilateral café-au-lait spots, and precocious puberty. McCune-Albright

**TABLE 386-1 Causes of Cushing's Syndrome**

CAUSES OF CUSHING'S SYNDROME	FEMALE:MALE RATIO	%
<b>ACTH-Dependent Cushing's</b>		90
Cushing's disease (= ACTH-producing pituitary adenoma)	4:1	75
Ectopic ACTH syndrome (due to ACTH secretion by bronchial or pancreatic carcinoid tumors, small-cell lung cancer, medullary thyroid carcinoma, pheochromocytoma, and others)	1:1	15
<b>ACTH-Independent Cushing's</b>	4:1	10
Adrenocortical adenoma		5–10
Adrenocortical carcinoma		1
Rare causes: macronodular adrenal hyperplasia; primary pigmented nodular adrenal disease (micro- and/or macronodular); McCune-Albright syndrome		<1

Abbreviation: ACTH, adrenocorticotrophic hormone.

**TABLE 386-2 Signs and Symptoms of Cushing's Syndrome**

BODY COMPARTMENT/ SYSTEM	SIGNS AND SYMPTOMS
Body fat	Weight gain, central obesity, rounded face, fat pad on back of neck ("buffalo hump")
Skin	Facial plethora, thin and brittle skin, easy bruising, broad and purple stretch marks, acne, hirsutism
Bone	Osteopenia, osteoporosis (vertebral fractures), decreased linear growth in children
Muscle	Weakness, proximal myopathy (prominent atrophy of gluteal and upper leg muscles with difficulty climbing stairs or getting up from a chair)
Cardiovascular system	Hypertension, hypokalemia, edema, atherosclerosis
Metabolism	Glucose intolerance/diabetes, dyslipidemia
Reproductive system	Decreased libido, in women amenorrhea (due to cortisol-mediated inhibition of gonadotropin release)
Central nervous system	Irritability, emotional lability, depression, sometimes cognitive defects; in severe cases, paranoid psychosis
Blood and immune system	Increased susceptibility to infections, increased white blood cell count, eosinopenia, hypercoagulation with increased risk of deep vein thrombosis and pulmonary embolism

syndrome is caused by activating mutations in the stimulatory G protein alpha subunit 1, GNAS-1 (guanine nucleotide-binding protein alpha stimulating activity polypeptide 1), and such mutations have also been found in bilateral macronodular hyperplasia without other McCune-Albright features and, in rare instances, also in isolated cortisol-producing adrenal adenomas (Table 386-1; **Chap. 412**).

**Clinical Manifestations** Glucocorticoids affect almost all cells of the body; thus, signs of cortisol excess impact multiple physiologic systems (Table 386-2), with upregulation of gluconeogenesis, lipolysis,

and protein catabolism causing the most prominent features. In addition, excess glucocorticoid secretion overcomes the ability of 11-HSD2 to rapidly inactivate cortisol to cortisone in the kidney, thereby exerting mineralocorticoid actions, manifest as diastolic hypertension, hypokalemia, and edema. Excess glucocorticoids also interfere with central regulatory systems, leading to suppression of gonadotropins with subsequent hypogonadism and amenorrhea and suppression of the hypothalamic-pituitary-thyroid axis, resulting in decreased thyroid-stimulating hormone (TSH) secretion.

The majority of clinical signs and symptoms observed in Cushing's syndrome are relatively nonspecific and include features such as obesity, diabetes, diastolic hypertension, hirsutism, and depression that are commonly found in patients who do not have Cushing's. Therefore, careful clinical assessment is an important aspect of evaluating suspected cases. A diagnosis of Cushing's should be considered when several clinical features are found in the same patient, in particular when more specific features are found or manifest at an unusual age, e.g., osteoporosis in a young patient. Distinct features include fragility of the skin, with easy bruising and broad (>1 cm), purplish striae (Fig. 386-9), and signs of proximal myopathy, which becomes most obvious when trying to stand up from a chair without the use of hands or when climbing stairs. Clinical manifestations of Cushing's do not differ substantially among the different causes of Cushing's. In ectopic ACTH syndrome, hyperpigmentation of the knuckles, scars, or skin areas exposed to increased friction can be observed (Fig. 386-9) and is caused by stimulatory effects of excess ACTH and other POMC cleavage products on melanocyte pigment production. Furthermore, patients with ectopic ACTH syndrome, and some with ACC as the cause of Cushing's, may have a more brisk onset and rapid progression of clinical signs and symptoms, namely of edema, hypokalemia, and hypertension.

Patients with Cushing's syndrome can be acutely endangered by deep vein thrombosis, with subsequent pulmonary embolism, due to a hypercoagulable state associated with Cushing's. The majority of patients also experience psychiatric symptoms, mostly in the form of



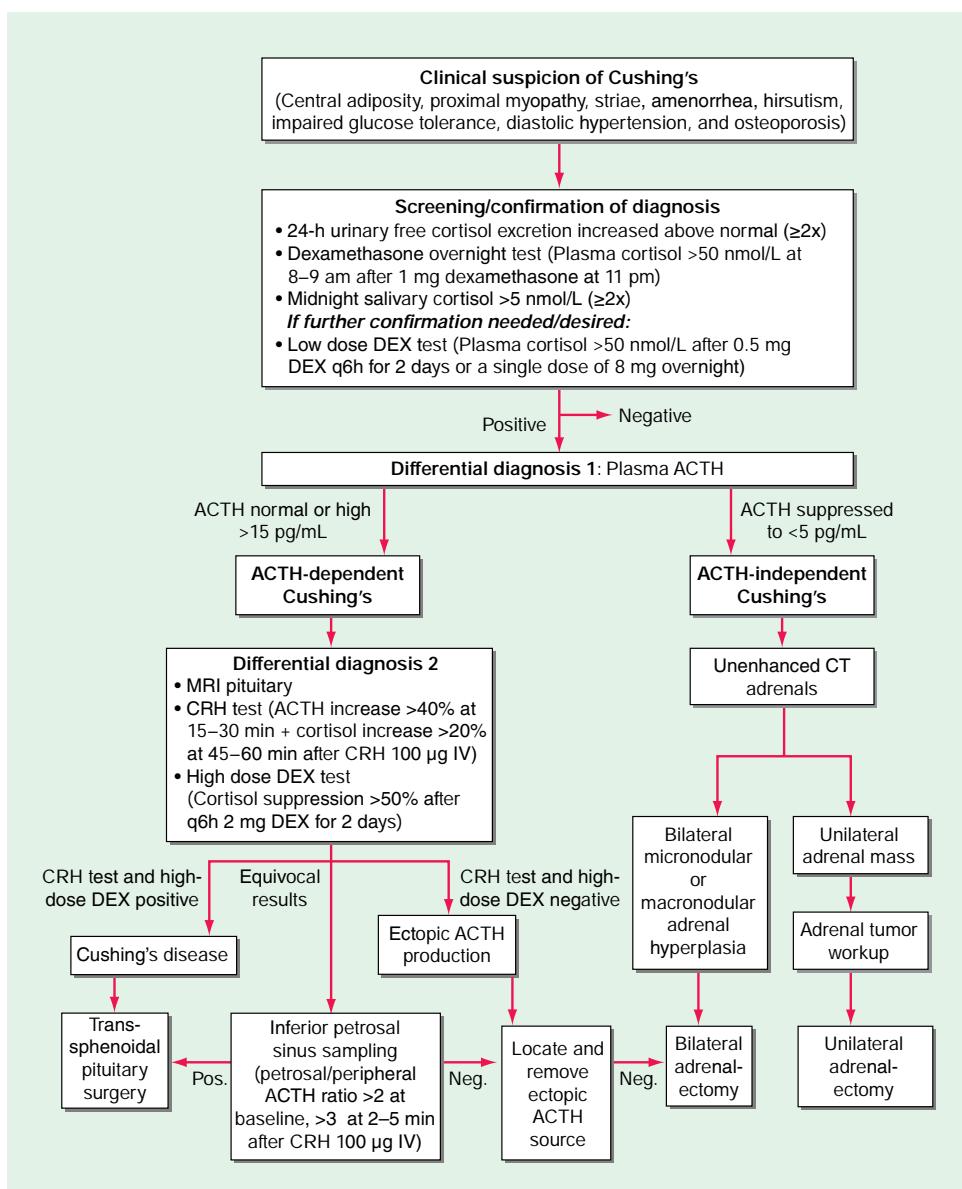
**FIGURE 386-9 Clinical features of Cushing's syndrome. A.** Note central obesity and broad, purple stretch marks (**B**, close-up). **C.** Note thin and brittle skin in an elderly patient with Cushing's syndrome. **D.** Hyperpigmentation of the knuckles in a patient with ectopic adrenocorticotrophic hormone (ACTH) excess.

**2962** anxiety or depression, but acute paranoid or depressive psychosis may occur. Even after cure, long-term health may be affected by persistently impaired health-related quality of life and increased risk of cardiovascular disease and osteoporosis with vertebral fractures, depending on the duration and degree of exposure to significant cortisol excess.

**Diagnosis** The most important first step in the management of patients with suspected Cushing's syndrome is to establish the correct diagnosis. Most mistakes in clinical management, leading to unnecessary imaging or surgery, are made because the diagnostic protocol is not followed (**Fig. 386-10**). This protocol requires establishing the diagnosis of Cushing's beyond doubt prior to employing any tests used for the differential diagnosis of the condition. In principle, after excluding exogenous glucocorticoid use as the cause of clinical signs and symptoms, suspected cases should be tested if there are multiple and progressive features of Cushing's, particularly features with a

potentially higher discriminatory value. Exclusion of Cushing's is also indicated in patients with incidentally discovered adrenal masses.

A diagnosis of Cushing's can be considered as established if the results of several tests are consistently suggestive of Cushing's. These tests may include increased 24-h urinary free cortisol excretion in three separate collections, failure to appropriately suppress morning cortisol after overnight exposure to dexamethasone, and evidence of loss of diurnal cortisol secretion with high levels at midnight, the time of the physiologically lowest secretion (Fig. 386-10). Factors potentially affecting the outcome of these diagnostic tests have to be excluded such as incomplete 24-h urine collection or rapid inactivation of dexamethasone due to concurrent intake of CYP3A4-inducing drugs (e.g., antiepileptics, rifampicin). Concurrent intake of oral contraceptives that raise CBG and thus total cortisol can cause failure to suppress after dexamethasone. If in doubt, testing should be repeated after 4–6 weeks off estrogens. Patients with pseudo-Cushing states, i.e., alcohol-related, and those with cyclic Cushing's may require further



**FIGURE 386-10** Management of the patient with suspected Cushing's syndrome. ACTH, adrenocorticotrophic hormone; CRH, corticotropin-releasing hormone; CT, computed tomography; DEX, dexamethasone; MRI, magnetic resonance imaging.

testing to safely confirm or exclude the diagnosis of Cushing's. In addition, the biochemical assays employed can affect the test results, with specificity representing a common problem with antibody-based assays for the measurement of urinary free cortisol. These assays have been greatly improved by the introduction of highly specific tandem mass spectrometry.

**Differential Diagnosis** The evaluation of patients with confirmed Cushing's should be carried out by an endocrinologist and begins with the differential diagnosis of ACTH-dependent and ACTH-independent cortisol excess (Fig. 386-10). Generally, plasma ACTH levels are suppressed in cases of autonomous adrenal cortisol excess, as a consequence of enhanced negative feedback to the hypothalamus and pituitary. By contrast, patients with ACTH-dependent Cushing's have normal or increased plasma ACTH, with very high levels being found in some patients with ectopic ACTH syndrome. Importantly, imaging should only be used after it is established whether the cortisol excess is ACTH-dependent or ACTH-independent because nodules in the pituitary or the adrenal are a common finding in the general population. In patients with confirmed ACTH-independent excess, adrenal imaging is indicated (Fig. 386-11), preferably using an unenhanced computed tomography (CT) scan. This allows assessment of adrenal morphology and determination of precontrast tumor density in Hounsfield units (HUs), which helps to distinguish between benign and malignant adrenal lesions.

For ACTH-dependent cortisol excess (Chap. 380), a magnetic resonance image (MRI) of the pituitary is the investigation of choice, but it may not show an abnormality in up to 40% of cases because of small tumors below the sensitivity of detection. Characteristically, pituitary corticotrope adenomas fail to enhance following gadolinium administration on T1-weighted MRI images. In all cases of confirmed ACTH-dependent Cushing's, further tests are required for the differential diagnosis of pituitary Cushing's disease and ectopic ACTH syndrome. These tests exploit the fact that most pituitary corticotrope adenomas still display regulatory features, including residual ACTH suppression by high-dose glucocorticoids and CRH responsiveness. In contrast, ectopic sources of ACTH are typically resistant to dexamethasone suppression and unresponsive to CRH (Fig. 386-10). However, it should be noted that a small minority of ectopic ACTH-producing tumors exhibit dynamic responses similar to pituitary corticotrope tumors. If the two tests show discordant results or if there is any other reason for doubt, the differential diagnosis can be further clarified by performing bilateral inferior petrosal sinus sampling (IPSS) with concurrent blood sampling for ACTH in the right and left inferior petrosal sinus and a peripheral vein. An increased central/Peripheral plasma ACTH ratio  $>2$  at baseline and  $>3$  at 2–5 min after CRH injection is indicative of Cushing's disease (Fig. 386-10), with very high sensitivity and specificity. Of note, the results of the IPSS cannot be reliably used for lateralization (i.e., prediction of the location of the tumor within the pituitary) because there is broad interindividual variability in the venous drainage of the pituitary region. Importantly, no cortisol-lowering agents should be used prior to IPSS.

If the differential diagnostic testing indicates ectopic ACTH syndrome, then further imaging should include high-resolution, fine-cut CT scanning of the chest and abdomen for scrutiny of the lung, thymus, and pancreas. If no lesions are identified, an MRI of the chest can be considered because carcinoid tumors usually show high signal intensity on T2-weighted images. Furthermore, octreotide scintigraphy can be helpful in some cases because ectopic ACTH-producing tumors often express somatostatin receptors. Depending on the suspected cause, patients with ectopic ACTH syndrome should also undergo blood sampling for fasting gut hormones, chromogranin A, calcitonin, and biochemical exclusion of pheochromocytoma.

## TREATMENT

### Cushing's Syndrome

Overt Cushing's is associated with a poor prognosis if left untreated. In ACTH-independent disease, treatment consists of surgical

removal of the adrenal tumor. For smaller tumors, a minimally invasive approach can be used, whereas for larger tumors and those suspected of malignancy, an open approach is preferred.

In Cushing's disease, the treatment of choice is selective removal of the pituitary corticotrope tumor, usually via an endoscopic trans-sphenoidal approach. This results in an initial cure rate of 70–80% when performed by a highly experienced surgeon. However, even after initial remission following surgery, long-term follow-up is important because late relapse occurs in a significant number of patients. If pituitary disease recurs, there are several options, including second surgery, radiotherapy, stereotactic radiosurgery, and bilateral adrenalectomy. These options need to be applied in a highly individualized fashion.

In some patients with very severe, overt Cushing's (e.g., difficult to control hypokalemic hypertension or acute psychosis), it may be necessary to introduce medical therapy to rapidly control the cortisol excess during the period leading up to surgery, which also can help to alleviate hypercoagulability and, thus, operative risk. Similarly, patients with metastasized, glucocorticoid-producing carcinomas may require long-term antiglucocorticoid drug treatment. In case of ectopic ACTH syndrome, in which the tumor cannot be located, one must carefully weigh whether drug treatment or bilateral adrenalectomy is the most appropriate choice, with the latter facilitating immediate cure but requiring life-long corticosteroid replacement. In this instance, it is paramount to ensure regular imaging follow-up for identification of the ectopic ACTH source.

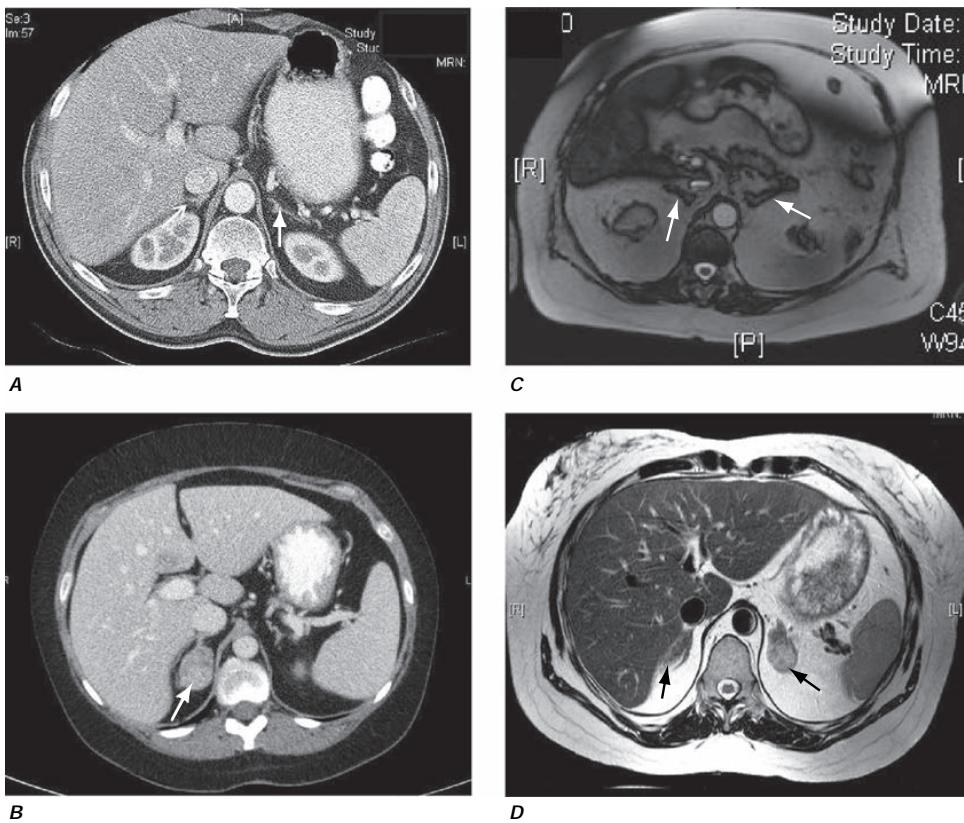
Oral agents with established efficacy in Cushing's syndrome are metyrapone and ketoconazole. Metyrapone inhibits cortisol synthesis at the level of 11'-hydroxylase (Fig. 386-1), whereas the antimycotic drug ketoconazole inhibits the early steps of steroidogenesis. Typical starting doses are 500 mg tid for metyrapone (maximum dose, 6 g) and 200 mg tid for ketoconazole (maximum dose, 1200 mg). Recently, the potent 11'-hydroxylase inhibitor osilodrostat has been introduced for the treatment of Cushing's, which also exerts strong inhibition of aldosterone synthase (CYP11B2). Mitotane, a derivative of the insecticide o,p'DDD, is an adrenolytic agent that is also effective for reducing cortisol. Because of its side effect profile, it is most commonly used in the context of ACC, but low-dose treatment (500–1000 mg/d) has also been used in benign Cushing's. In severe cases of cortisol excess, etomidate, an agent that potently blocks 11'-hydroxylase and aldosterone synthase, can be used to lower cortisol. It is administered by continuous IV infusion in low, nonanesthetic doses. For Cushing's disease, the subcutaneous administration of pasireotide, a somatostatin receptor agonist, represents another therapeutic option, if surgical cure cannot be achieved.

After the successful removal of an ACTH- or cortisol-producing tumor, the HPA axis will remain suppressed. Thus, hydrocortisone replacement needs to be initiated at the time of surgery and slowly tapered following recovery, to allow physiologic adaptation to normal cortisol levels. Depending on degree and duration of cortisol excess, the HPA axis may require many months or even years to resume normal function and sometimes does not recover. Generally, ectopic ACTH syndrome shows the best recovery rate (80%) and adrenal Cushing's has the lowest (40%), with Cushing's disease intermediate (60%).

## MINERALOCORTICOID EXCESS

**Epidemiology** Following the first description of a patient with an aldosterone-producing adrenal adenoma (*Conn's syndrome*), mineralocorticoid excess was thought to represent a rare cause of hypertension. However, in studies systematically screening all patients with hypertension, a much higher prevalence is now recognized, ranging from 5 to 12%. The prevalence is higher when patients are preselected for hypokalemic hypertension.

**Etiology** The most common cause of mineralocorticoid excess is primary aldosteronism, reflecting excess production of aldosterone by the adrenal zona glomerulosa. Bilateral micronodular hyperplasia is



**FIGURE 386-11** Adrenal imaging in Cushing's syndrome. **A.** Adrenal computed tomography (CT) showing normal bilateral adrenal morphology (arrows). **B.** CT scan depicting a right adrenocortical adenoma (arrow) causing Cushing's syndrome. **C.** Magnetic resonance imaging (MRI) showing bilateral adrenal hyperplasia due to excess adrenocorticotrophic hormone stimulation in Cushing's disease. **D.** MRI showing bilateral macronodular hyperplasia causing Cushing's syndrome.

somewhat more common than unilateral adrenal adenomas (**Table 386-3**). Somatic mutations in channels and enzymes responsible for increasing sodium and calcium influx in adrenal zona glomerulosa cells have been identified as prevalent causes of aldosterone-producing adrenal adenomas (**Table 386-3**) and, in the case of germline mutations, also of primary aldosteronism due to bilateral macronodular adrenal hyperplasia. However, bilateral adrenal hyperplasia as a cause of mineralocorticoid excess is usually micronodular but can also contain larger nodules that might be mistaken for a unilateral adenoma. In rare instances, primary aldosteronism is caused by an ACC. Carcinomas should be considered in younger patients and in those with larger tumors because benign aldosterone-producing adenomas usually measure <2 cm in diameter.

A rare cause of aldosterone excess is glucocorticoid-remediable aldosteronism (GRA), which is caused by a chimeric gene resulting from crossover of promoter sequences between the *CYP11B1* and *CYP11B2* genes that are involved in glucocorticoid and mineralocorticoid synthesis, respectively (**Fig. 386-1**). This rearrangement brings *CYP11B2* transcription under the control of ACTH receptor signaling; consequently, aldosterone production is regulated by ACTH rather than by renin. The family history can be helpful because there may be evidence for dominant transmission of hypertension. Recognition of the disorder is important because it can be associated with early-onset hypertension and strokes. In addition, glucocorticoid suppression can reduce aldosterone production.

Other rare causes of mineralocorticoid excess are listed in **Table 386-3**. An important cause is excess binding and activation of the MR by a steroid other than aldosterone. Cortisol acts as a potent mineralocorticoid if it escapes efficient inactivation to cortisone by 11-HSD2 in the kidney (**Fig. 386-7**). This can be caused by inactivating mutations in the *HSD11B2* gene resulting in the syndrome

of apparent mineralocorticoid excess (SAME) that characteristically manifests with severe hypokalemic hypertension in childhood. However, milder mutations may cause normokalemic hypertension manifesting in adulthood (type II SAME). Inhibition of 11-HSD2 by excess licorice ingestion also results in hypokalemic hypertension, as does overwhelming of 11-HSD2 conversion capacity by cortisol excess in Cushing's syndrome. DOC also binds and activates the MR and can cause hypertension if its circulating concentrations are increased. This can arise through autonomous DOC secretion by an ACC, but also when DOC accumulates as a consequence of an adrenal enzymatic block, as seen in congenital adrenal hyperplasia (CAH) due to *CYP11B1* (11-hydroxylase) or *CYP17A1* (17-hydroxylase) deficiency (**Fig. 386-1**). Progesterone can cause hypokalemic hypertension in rare individuals who harbor a MR mutation that enhances binding and activation by progesterone; physiologically, progesterone normally exerts antiminerlocorticoid activity. Finally, excess mineralocorticoid activity can be caused by mutations in the  $\alpha$  or  $\beta$  subunits of the ENaC, disrupting its interaction with Nedd4 (**Fig. 386-7**), and thereby decreasing receptor internalization and degradation. The constitutively active ENaC drives hypokalemic hypertension, resulting in an autosomal dominant disorder termed *Liddle's syndrome*.

**Clinical Manifestations** Excess activation of the MR leads to potassium depletion and increased sodium retention, with the latter causing an expansion of extracellular and plasma volume. Increased ENaC activity also results in hydrogen depletion that can cause metabolic alkalosis. Aldosterone also has direct effects on the vascular system, where it increases cardiac remodeling and decreases compliance. Aldosterone excess may cause direct damage to the myocardium and the kidney glomeruli, in addition to secondary damage due to systemic hypertension.

**TABLE 386-3 Causes of Mineralocorticoid Excess**

CAUSES OF MINERALOCORTICOID EXCESS	MECHANISM	%
<b>Primary Aldosteronism</b>		
Adrenal (Conn's) adenoma	Autonomous aldosterone excess can be caused by somatic (intratumor) mutations in the potassium channel GIRK4 (encoded by <i>KCNJ5</i> ; identified as cause of disease in 40% of aldosterone-producing adenomas; rare germline mutations can cause bilateral macronodular adrenal hyperplasia). Further causes include somatic mutations affecting the $\alpha$ -subunit of the $\text{Na}^+/\text{K}^+$ -ATPase (encoded by <i>ATP1A1</i> ), the plasma membrane calcium-transporting ATPase 3 (encoded by <i>ATP2B3</i> ), and somatic mutations in <i>CACNA1D</i> or <i>CACNA1H</i> encoding the voltage-gated calcium channel $\text{CaV}1.3$ and $\text{CaV}3.2$ , respectively. All mutations result in upregulation of <i>CYP11B2</i> and hence aldosterone synthesis.	40
Bilateral (micronodular) adrenal hyperplasia	Autonomous aldosterone excess, mostly micronodular and rarely macronodular, with germline <i>KCNJ5</i> mutations being a rare cause.	60
Glucocorticoid-remediable hyperaldosteronism (dexamethasone-suppressible hyperaldosteronism)	Crossover between the <i>CYP11B1</i> and <i>CYP11B2</i> genes results in ACTH-driven aldosterone production	<1
<b>Other Causes (Rare)</b>		<1
Syndrome of apparent mineralocorticoid excess (SAME)	Mutations in <i>HSD11B2</i> result in lack of renal inactivation of cortisol to cortisone, leading to excess activation of the MR by cortisol (inhibition of $11\beta$ -hydroxysteroid dehydrogenase type 2 by excess licorice ingestion can have similar effects)	
Cushing's syndrome	Cortisol excess overcomes the capacity of HSD11B2 to inactivate cortisol to cortisone, consequently flooding the MR	
Glucocorticoid resistance	Upregulation of cortisol production due to GR mutations results in flooding of the MR by cortisol	
Adrenocortical carcinoma	Autonomous aldosterone and/or DOC excess	
Congenital adrenal hyperplasia	Accumulation of DOC due to mutations in <i>CYP11B1</i> or <i>CYP17A1</i>	
Progesterone-induced hypertension	Progesterone acts as an abnormal ligand due to mutations in the MR gene	
Liddle's syndrome	Mutant ENaC $\beta$ or $\gamma$ subunits resulting in reduced degradation of ENaC keeping the membrane channel in open conformation for longer, enhancing mineralocorticoid action	

**Abbreviations:** ACTH, adrenocorticotrophic hormone; DOC, deoxycorticosterone; ENaC, epithelial sodium channel; GR, glucocorticoid receptor; HSD11B2,  $11\beta$ -hydroxysteroid dehydrogenase type 2; MR, mineralocorticoid receptor.

The clinical hallmark of mineralocorticoid excess is hypokalemic hypertension; however, only 50% of patients with primary aldosteronism exhibit hypokalemia. Serum sodium tends to be normal due to the concurrent fluid retention, which in some cases can lead to peripheral edema. Hypomagnesemia is also a common finding. Hypokalemia can be exacerbated by thiazide drug treatment, which leads to increased delivery of sodium to the distal renal tubule, thereby driving potassium excretion. Severe hypokalemia can be associated with muscle weakness, overt proximal myopathy, or even hypokalemic paralysis. Severe alkalosis contributes to muscle cramps and, in severe cases, can cause tetany.

Of note, patients with primary aldosteronism show increased rates of osteoporosis, type 2 diabetes, and cognitive dysfunction. A significant proportion of patients suffer from concurrent mild autonomous cortisol excess (MACE), termed *Connshing syndrome*.

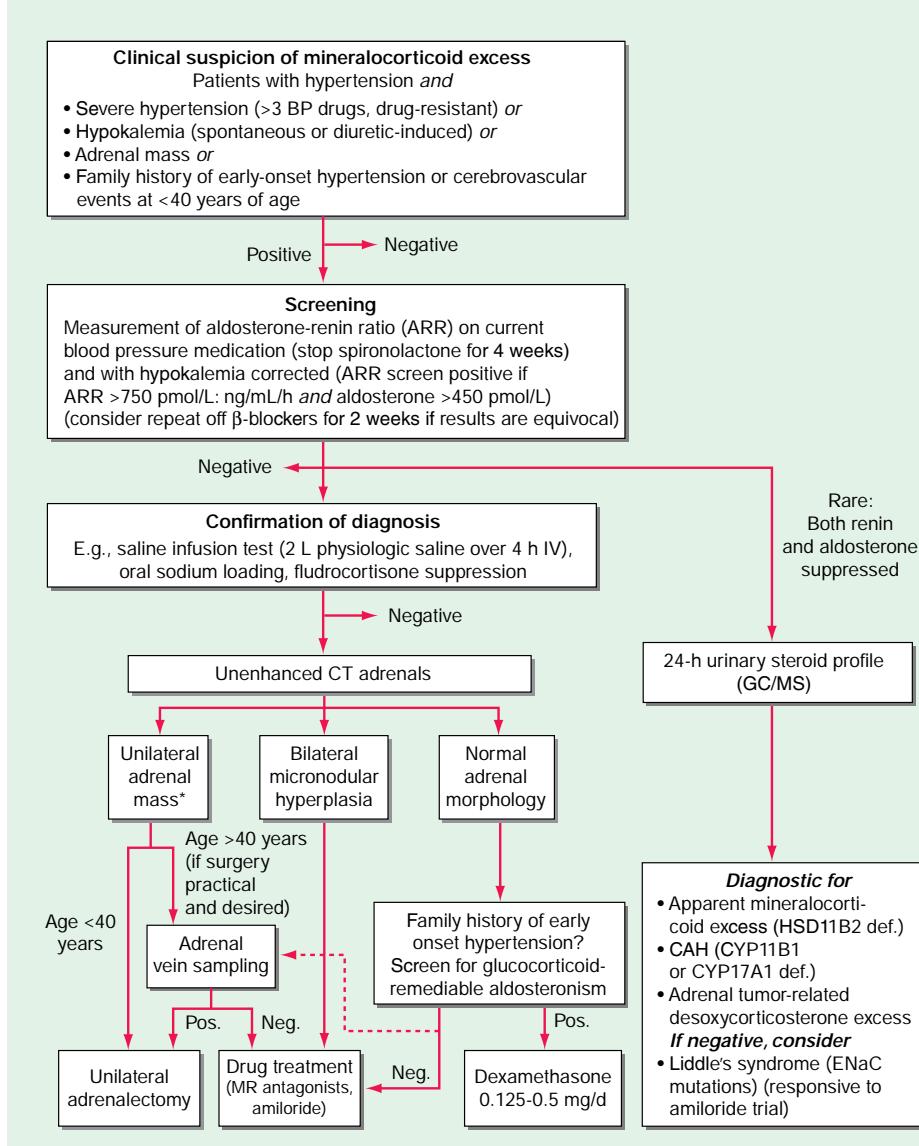
**Diagnosis** Diagnostic screening for mineralocorticoid excess is not currently recommended for all patients with hypertension but should be restricted to those who exhibit hypertension associated with drug resistance, hypokalemia, an adrenal mass, or onset of disease before the age of 40 years (Fig. 386-12). The accepted screening test is concurrent measurement of plasma renin and aldosterone with subsequent calculation of the aldosterone-renin ratio (ARR) (Fig. 386-12); serum potassium needs to be normalized prior to testing. Stopping antihypertensive medication can be cumbersome, particularly in patients with severe hypertension. Thus, for practical purposes, in the first instance, the patient can remain on the usual antihypertensive medications, with the exception that MR antagonists need to be ceased at least 4 weeks prior to ARR measurement. The remaining antihypertensive drugs usually do not affect the outcome of ARR testing, except that beta blocker treatment can cause false-positive results and ACE/AT1R inhibitors can cause false-negative results in milder cases (**Table 386-4**).

ARR screening is positive if the ratio is  $>750$  pmol/L per ng/mL per hour, with a concurrently high normal or increased aldosterone (Fig. 386-12). If one relies on the ARR only, the likelihood of a false-positive ARR becomes greater when renin levels are very low. The characteristics of the biochemical assays are also important. Some labs

measure plasma renin activity, whereas others measure plasma renin concentrations. Antibody-based assays for the measurement of serum aldosterone lack the reliability of tandem mass spectrometry assays, but these are not yet ubiquitously available.

Diagnostic confirmation of mineralocorticoid excess in a patient with a positive ARR screening result should be undertaken by an endocrinologist as the tests lack optimized validation. The most straightforward is the saline infusion test, which involves the IV administration of 2 L of physiologic saline over a 4-h period. Failure of aldosterone to suppress  $<140$  pmol/L (5 ng/dL) is indicative of autonomous mineralocorticoid excess. Alternative tests are the oral sodium loading test (300 mmol NaCl/d for 3 days) or the fludrocortisone suppression test (0.1 mg q6h with 30 mmol NaCl q8h for 4 days); the latter can be difficult because of the risk of profound hypokalemia and increased hypertension. In patients with overt hypokalemic hypertension, strongly positive ARR, and concurrently increased aldosterone levels, confirmatory testing is usually not necessary.

**Differential Diagnosis and Treatment** After the diagnosis of hyperaldosteronism is established, the next step is to use adrenal imaging to further assess the cause. Fine-cut CT scanning of the adrenal region is the method of choice because it provides excellent visualization of adrenal morphology and most aldosterone-producing adenomas are  $<1$  cm. CT will readily identify larger tumors suspicious of malignancy but may miss lesions  $<5$  mm. The differentiation between bilateral micronodular hyperplasia and a unilateral adenoma is only required if a surgical approach is feasible and desired. Consequently, selective adrenal vein sampling (AVS) should only be carried out in surgical candidates with either no obvious lesion on CT or evidence of a unilateral lesion but with age  $>40$  years because the latter patients have a high likelihood of harboring a coincidental, endocrine-inactive adrenal adenoma (Fig. 386-12). AVS is used to compare aldosterone levels in the inferior vena cava and between the right and left adrenal veins. AVS requires concurrent measurement of cortisol to document correct placement of the catheter in the adrenal veins and should demonstrate a cortisol gradient  $>3$  between the vena cava and each adrenal vein. Lateralization is confirmed by an aldosterone/cortisol ratio that is at least twofold higher on one side than the other. AVS is a



**FIGURE 386-12** Management of patients with suspected mineralocorticoid excess. \*Perform adrenal tumor workup (see Fig. 386-13). BP, blood pressure; CAH, congenital adrenal hyperplasia; CT, computed tomography; ENaC, epithelial sodium channel; GC/MS, gas chromatography/mass spectrometry; MR, mineralocorticoid receptor; PRA, plasma renin activity.

**TABLE 386-4 Effects of Antihypertensive Drugs on the Aldosterone-Renin Ratio (ARR)**

DRUG	EFFECT ON RENIN	EFFECT ON ALDOSTERONE	NET EFFECT ON ARR
β Blockers	↓	↑	↑
α <sub>1</sub> Blockers	→	→	→
α <sub>2</sub> Sympathomimetics	→	→	→
ACE inhibitors	↑	↓	↓
AT1R blockers	↑	↓	↓
Calcium antagonists	→	→	→
Diuretics	(↑)	(↑)	→/(↓)

Abbreviations: ACE, angiotensin-converting enzyme; AT1R, angiotensin II receptor type 1.

complex procedure that requires a highly skilled interventional radiologist. Even then, the right adrenal vein can be difficult to cannulate correctly, which, if not achieved, invalidates the procedure. There is also no agreement as to whether the two adrenal veins should be cannulated simultaneously or successively and whether ACTH stimulation enhances the diagnostic value of AVS.

Patients <40 years with confirmed mineralocorticoid excess and a unilateral lesion on CT can go straight to surgery, which is also indicated in patients with confirmed lateralization documented by a valid AVS procedure. Laparoscopic adrenalectomy is the preferred approach. Patients who are not surgical candidates, or with evidence of bilateral hyperplasia based on CT or AVS, should be treated medically (Fig. 386-12). Medical treatment, which can also be considered prior to surgery to avoid postsurgical hypoaldosteronism, consists primarily of the MR antagonist spironolactone. It can be started at 12.5–50 mg bid

and titrated up to a maximum of 400 mg/d to control blood pressure and normalize potassium. Side effects include menstrual irregularity, decreased libido, and gynecomastia. The more selective MR antagonist eplerenone can also be used. Doses start at 25 mg bid, and it can be titrated up to 200 mg/d. Another useful drug is the sodium channel blocker amiloride (5–10 mg bid).

In patients with normal adrenal morphology and family history of early-onset, severe hypertension, a diagnosis of GRA should be considered and can be evaluated using genetic testing. Treatment of GRA consists of administering dexamethasone, using the lowest dose possible to control blood pressure. Some patients also require additional MR antagonist treatment.

The diagnosis of non-aldosterone-related mineralocorticoid excess is based on documentation of suppressed renin and suppressed aldosterone in the presence of hypokalemic hypertension. This testing is best carried out by employing urinary steroid metabolite profiling by gas chromatography/mass spectrometry (GC/MS). An increased free cortisol over free cortisone ratio is suggestive of SAME and can be treated with dexamethasone. Steroid profiling by GC/MS also detects the steroids associated with CYP11B1 and CYP17A1 deficiency or the irregular steroid secretion pattern in a DOC-producing ACC (Fig. 386-12). If the GC/MS profile is normal, Liddle's syndrome should be considered. It is very sensitive to amiloride treatment but will not respond to MR antagonist treatment because the defect is due to a constitutively active ENaC.

### APPROACH TO THE PATIENT: INCIDENTALLY DISCOVERED ADRENAL MASS

**Epidemiology** Incidentally discovered adrenal masses, commonly termed adrenal “incidentalomas,” are common, with a prevalence of 2–5% in the general population as documented in CT and autopsy series. The prevalence increases with age, with 1% of 40-year-olds and 7% of 70-year-olds harboring an adrenal mass. The widespread use of cross-sectional imaging has also increased the recognized prevalence.

**Etiology** Most solitary adrenal tumors are monoclonal neoplasms. Several genetic syndromes, including MEN 1 (*MEN1*), MEN 2 (*RET*), Carney's complex (*PRKAR1A*), and McCune-Albright (*GNAS1*), can have adrenal tumors as one of their features. Somatic mutations in *MEN1*, *GNAS1*, and *PRKAR1A* have been identified in a small proportion of sporadic adrenocortical adenomas. Aberrant expression of membrane receptors (GIP, - and -adrenergic, luteinizing hormone, vasopressin V1, and interleukin 1 receptors) has been identified in some sporadic cases of macronodular adrenocortical hyperplasia.

The majority of adrenal nodules are endocrine-inactive adrenocortical adenomas. However, larger series suggest that up to 25% of adrenal nodules are hormonally active, due to a cortisol- or aldosterone-producing adrenocortical adenoma or a pheochromocytoma associated with catecholamine excess (Table 386-5). ACC is rare but is the cause of an adrenal mass in 5% of patients. However, metastases originating from another solid tissue tumor are an additional cause of adrenal incidentaloma and have a higher incidence in patients undergoing imaging for tumor staging or follow-up monitoring (Table 386-5).

**Differential Diagnosis and Treatment** Patients with an adrenal mass >1 cm require a diagnostic evaluation. Two key questions need to be addressed: (1) Does the tumor autonomously secrete hormones that could have a detrimental effect on health? (2) Is the adrenal mass benign or malignant?

Hormone secretion by an adrenal mass occurs along a continuum, with a gradual increase in clinical manifestations in parallel with hormone levels. Exclusion of catecholamine excess from a pheochromocytoma arising from the adrenal medulla is a mandatory part of the diagnostic workup (Fig. 386-13). Furthermore, autonomous cortisol resulting in Cushing's syndrome requires exclusion and, in patients with hypertension or low serum potassium, also primary aldosteronism. Adrenal incidentalomas can be associated with MACE, and patients usually lack overt clinical features of Cushing's syndrome. Nonetheless, they may exhibit one or more components of the

**TABLE 386-5 Classification of Unilateral Adrenal Masses**

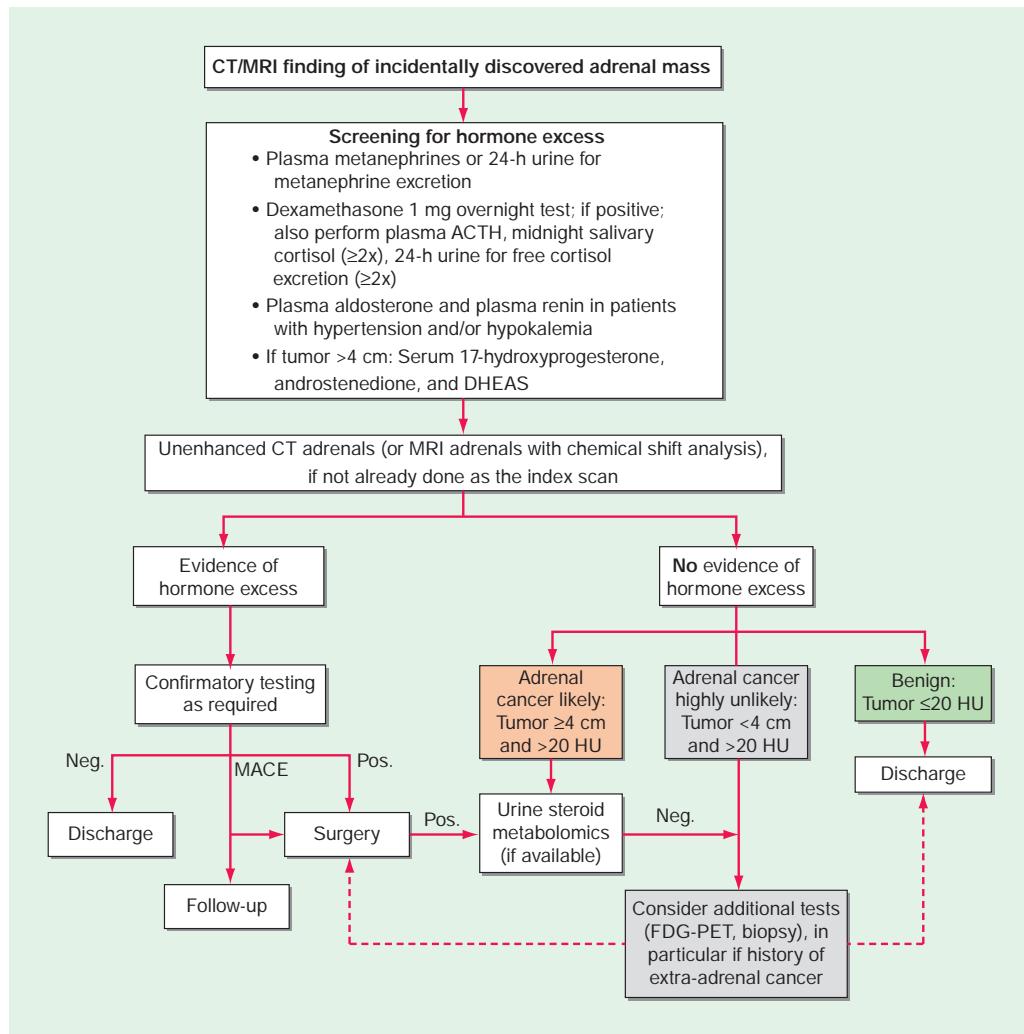
MASS	APPROXIMATE PREVALENCE (%)
<b>Benign</b>	
Adrenocortical adenoma	
Endocrine-inactive	60–85
Cortisol-producing	5–10
Aldosterone-producing	2–5
Pheochromocytoma	5–10
Adrenal myelolipoma	<1
Adrenal ganglioneuroma	<0.1
Adrenal hemangioma	<0.1
Adrenal cyst	<1
Adrenal hematoma/hemorrhagic infarction	<1
<b>Indeterminate</b>	
Adrenocortical oncocytoma	<1
<b>Malignant</b>	
Adrenocortical carcinoma	2–5
Malignant pheochromocytoma	<1
Adrenal neuroblastoma	<0.1
Lymphomas (including primary adrenal lymphoma)	<1
Metastases (most frequent: breast, lung)	1–2

*Note:* Bilateral adrenal enlargement/masses may be caused by congenital adrenal hyperplasia, bilateral macronodular hyperplasia, bilateral hemorrhage (due to antiphospholipid syndrome or sepsis-associated Waterhouse-Friderichsen syndrome), granuloma, amyloidosis, or infiltrative disease including tuberculosis.

metabolic syndrome (e.g., obesity, type 2 diabetes, or hypertension). There is ongoing debate about the optimal treatment for these patients. Overproduction of adrenal androgen precursors, DHEA and its sulfate, is rare and most frequently seen in the context of ACC, as are increased levels of steroid precursors such as 17OHP.

For the differentiation of benign from malignant adrenal masses, imaging is relatively sensitive, although specificity is suboptimal. Unenhanced CT is the procedure of choice for imaging the adrenal glands (Fig. 386-11). A diagnosis of ACC, pheochromocytoma, and benign adrenal myelolipoma becomes more likely with increasing diameter of the adrenal mass. However, size alone is of poor predictive value, with only 80% specificity for the differentiation of benign from malignant masses when using a 4-cm cutoff. Metastases are rare but are found with similar frequency in adrenal masses of all sizes. The tumor attenuation value on unenhanced CT is of high diagnostic value, as many adrenocortical adenomas are lipid rich and thus present with low attenuation values (i.e., densities of <20 Hounsfield units [HUs]). However, similar numbers of adrenocortical adenomas are lipid poor and present with higher HUs, making it difficult to differentiate them from ACCs, as well as also pheochromocytomas, both of which invariably have high attenuation values (i.e., densities >20 HU on precontrast scans). Generally, benign lesions are rounded and homogenous, whereas most malignant lesions appear lobulated and inhomogeneous. Pheochromocytoma and adrenomyelolipoma may also exhibit lobulated and inhomogeneous features. MRI also allows for the visualization of the adrenal glands with somewhat lower resolution than CT. However, because it does not involve exposure to ionizing radiation, it is preferred in children, young adults, and during pregnancy. MRI has a valuable role in the characterization of indeterminate adrenal lesions using chemical shift analysis, with malignant tumors rarely showing loss of signal on opposed-phase MRI; however, this may also be observed in a proportion of benign adrenocortical adenomas.

Fine-needle aspiration (FNA) or CT-guided biopsy of an adrenal mass is very rarely indicated. FNA of a pheochromocytoma can cause a life-threatening hypertensive crisis. FNA of an ACC violates the tumor capsule and can cause needle track metastasis. FNA should only be considered in a patient with a history of nonadrenal malignancy and a newly detected adrenal mass, after careful exclusion of



**FIGURE 386-13 Management of the patient with an incidentally discovered adrenal mass.** ACTH, adrenocorticotrophic hormone; CT, computed tomography; FDG-PET, fluorodeoxyglucose positron emission tomography; MACE, mild autonomous cortisol excess; MRI, magnetic resonance imaging.

pheochromocytoma, and if the outcome will influence therapeutic management. It is important to recognize that in 25% of patients with a previous history of nonadrenal malignancy, a newly detected mass on CT is not a metastasis. While FNA can diagnose extra-adrenal malignancies, it has very limited ability to differentiate between benign and malignant adrenocortical lesions and hence should not be used for diagnosis of ACC.

Adrenal masses associated with confirmed hormone excess or suspected malignancy are usually treated surgically (Fig. 386-13) or, if adrenalectomy is not feasible or desired, with medication. Preoperative exclusion of glucocorticoid excess is particularly important for the prediction of postoperative suppression of the contralateral adrenal gland, which requires glucocorticoid replacement peri- and postoperatively. Adrenal masses with normal endocrine biochemistry at diagnosis and a tumor radiodensity of  $<20$  HU on unenhanced CT can be considered benign and do not require further follow-up. In adrenal masses with suspicious imaging findings ( $>20$  HU), further tests and surgery are feasible options (Fig. 386-13); however, the latter will still result in unnecessary surgery for benign tumors. A recently introduced diagnostic test, urine steroid metabolomics, has a twofold higher positive predictive value than imaging in detecting adrenocortical carcinoma, based on a distinct "steroid fingerprint" with accumulating precursor steroids in 24-h urine.

## ADRENOCORTICAL CARCINOMA

ACC is a rare malignancy with an annual incidence of 1–2 per million population. ACC is generally considered a highly malignant tumor; however, it presents with broad interindividual variability with regard to biologic characteristics and clinical behavior. Somatic mutations in the tumor-suppressor gene *TP53* are found in 25% of apparently sporadic ACC. Germline *TP53* mutations are the cause of the Li-Fraumeni syndrome associated with multiple solid organ cancers including ACC and are found in 25% of pediatric ACC cases; the *TP53* mutation R337H is found in almost all pediatric ACC in Brazil. Other genetic changes identified in ACC include alterations in the Wnt/-catenin pathway and in the insulin-like growth factor 2 (IGF2) cluster; IGF2 overexpression is found in 90% of ACCs.

Patients with large adrenal tumors suspicious of malignancy should be managed by a multidisciplinary specialist team, including an endocrinologist, an oncologist, a surgeon, a radiologist, and a histopathologist. FNA is not indicated in suspected ACC: first, cytology and also histopathology of a core biopsy cannot differentiate between benign and malignant primary adrenal masses; second, FNA violates the tumor capsule and may even cause needle canal metastasis. Even when the entire tumor specimen is available, the histopathologic differentiation between benign and malignant adrenocortical lesions is a diagnostic challenge. The most common histopathologic classification

**TABLE 386-6 Classification System for Staging of Adrenocortical Carcinoma**

ENSAT STAGE	TNM STAGE	TNM DEFINITIONS
I	T1,N0,M0	T1, tumor <5 cm N0, no positive lymph node M0, no distant metastases
II	T2,N0,M0	T2, tumor >5 cm N0, no positive lymph node M0, no distant metastases
III	T1-T2,N1,M0	N1, positive lymph node(s)
	T3-T4,N0-N1,M0	M0, no distant metastases T3, tumor infiltration into surrounding tissue T4, tumor invasion into adjacent organs or venous tumor thrombus in vena cava or renal vein
IV	T1-T4,N0-N1,M1	M1, presence of distant metastases

Abbreviations: ENSAT, European Network for the Study of Adrenal Tumors; TNM, tumor, node, metastasis.

is the Weiss score, taking into account high nuclear grade; mitotic rate (>5/HPF); atypical mitosis; <25% clear cells; diffuse architecture; and presence of necrosis, venous invasion, and invasion of sinusoidal structures and tumor capsule. The presence of three or more elements suggests ACC. However, FNA is a feasible option if looking for metastases of an extra-adrenal primary or other adrenal tumor entities, such as ganglioneuroma.

Although 60–70% of ACCs show biochemical evidence of steroid overproduction, in many patients, this is not clinically apparent due to the relatively inefficient steroid production by the adrenocortical cancer cells. Excess production of glucocorticoids and adrenal androgen precursors are most common and indicative of malignancy.

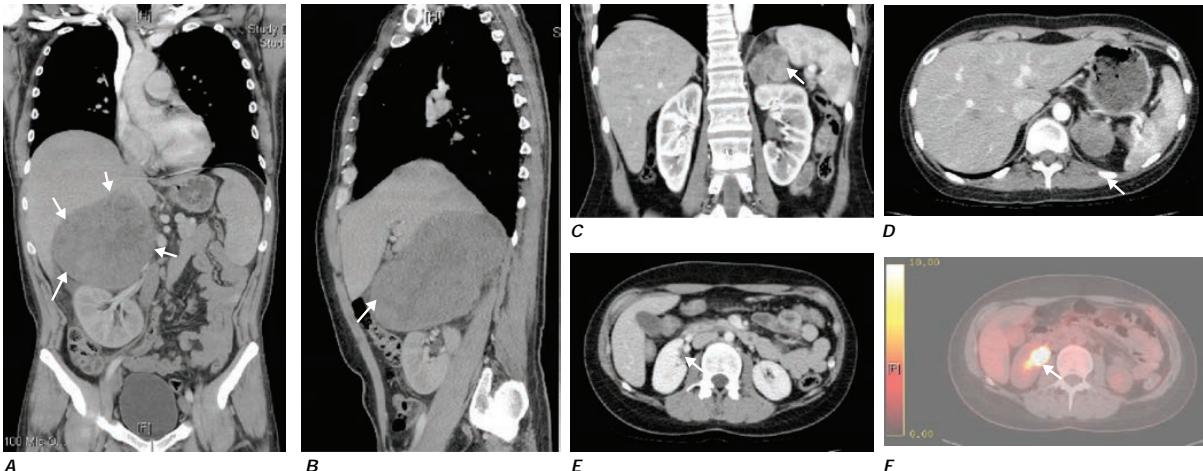
Tumor staging at ACC diagnosis (Table 386-6) has important prognostic implications and requires scanning of the chest and abdomen for local organ invasion, lymphadenopathy, and metastases. Intravenous contrast medium is necessary for maximum sensitivity for hepatic metastases. An adrenal origin may be difficult to determine on standard axial CT imaging if the tumors are large and invasive, but CT reconstructions and MRI are more informative (Fig. 386-14) using multiple planes and different sequences. Vascular and adjacent organ invasion is diagnostic of malignancy. 18-Fluoro-2-deoxy-D-glucose positron emission tomography (18-FDG-PET) is highly sensitive for

the detection of malignancy and can be used to detect small metastases or local recurrence that may not be obvious on CT (Fig. 386-14). However, FDG-PET has limited specificity and therefore cannot be used for differentiating benign from malignant adrenal lesions. Metastasis in ACC most frequently occurs to liver and lung.

There is no established grading system for ACC, and the Weiss score carries no prognostic value; the most important prognostic histopathologic parameter is the Ki67 proliferation index, with Ki67 <10% indicative of slow to moderate growth velocity, whereas a Ki67 >10% is associated with poor prognosis including high risk of recurrence and rapid progression.

Cure of ACC can only be achieved by early detection and complete surgical removal. Capsule violation during primary surgery, metastasis at diagnosis, and primary treatment in a nonspecialist center and by a nonspecialist surgeon are major determinants of poor survival. If the primary tumor invades adjacent organs, en bloc removal of kidney and spleen should be considered to reduce the risk of recurrence, and regional lymph node dissection may further reduce this risk. Surgery can also be considered in a patient with metastases if there is severe tumor-related hormone excess. This indication needs to be carefully weighed against surgical risk, including thromboembolic complications, and the resulting delay in the introduction of other therapeutic options. Patients with confirmed ACC and successful removal of the primary tumor should receive adjuvant treatment with mitotane (*o,p'*DDD), particularly in patients with a high risk of recurrence as determined by tumor size >8 cm, histopathologic signs of vascular invasion, capsule invasion or violation, and a Ki67 proliferation index

10%. Adjuvant mitotane should be continued for at least 2 years, if side effects are tolerated. Regular monitoring of plasma mitotane levels is mandatory (therapeutic range 14–20 mg/L; neurotoxic complications more frequent at >20 mg/L). Mitotane is usually started at 500 mg tid, with stepwise increases to a maximum dose of 2000 mg tid in days (high-dose saturation) or weeks (low-dose saturation) as tolerated. Once therapeutic range plasma mitotane levels are achieved, the dose can be tapered to maintenance doses mostly ranging from 1000–1500 mg tid. Mitotane treatment results in disruption of cortisol synthesis and thus requires glucocorticoid replacement; glucocorticoid replacement dose should be at least double of that usually used in adrenal insufficiency (i.e., 20 mg tid) because mitotane induces hepatic CYP3A4 activity, resulting in rapid inactivation of glucocorticoids. Mitotane also increases circulating CBG, thereby decreasing the available free cortisol fraction. Single metastases can be addressed surgically or with radiofrequency ablation as appropriate. If the tumor recurs or progresses during mitotane treatment, cytotoxic chemotherapy should be considered; the established first-line chemotherapy regimen is the



**FIGURE 386-14 Imaging in adrenocortical carcinoma (ACC).** Magnetic resonance imaging scan with (A) frontal and (B) lateral views of a right ACC that was detected incidentally. Computed tomography (CT) scan with (C) coronal and (D) transverse views depicting a right-sided ACC. Note the irregular border and inhomogeneous structure. CT scan (E) and positron emission tomography/CT (F) visualizing a peritoneal metastasis of an ACC in close proximity to the right kidney (arrow).

combination of cisplatin, etoposide, and doxorubicin plus continuing mitotane. Painful bone metastasis responds to irradiation. Overall survival in ACC is still poor, with 5-year survival rates of 30–40% and a median survival of 15 months in metastatic ACC.

## ADRENAL INSUFFICIENCY

**Epidemiology** The prevalence of well-documented, permanent adrenal insufficiency is 5 in 10,000 in the general population. Hypothalamic-pituitary origin of disease is most frequent, with a prevalence of 3 in 10,000, whereas primary adrenal insufficiency has a prevalence of 2 in 10,000. Approximately one-half of the latter cases are acquired, mostly caused by autoimmune destruction of the adrenal glands; the other one-half are genetic, most commonly caused by distinct enzymatic blocks in adrenal steroidogenesis affecting glucocorticoid synthesis (i.e., CAH).

Adrenal insufficiency arising from suppression of the HPA axis as a consequence of exogenous glucocorticoid treatment is much more common, occurring in 0.5–2% of the population in developed countries.

**Etiology** Primary adrenal insufficiency is most commonly caused by autoimmune adrenalitis. Isolated autoimmune adrenalitis accounts for 30–40%, whereas 60–70% develop adrenal insufficiency as part of autoimmune polyglandular syndromes (APSs) ([Chap. 388](#)) ([Table 386-7](#)).

APS1, also termed APECED (autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy), is the underlying cause in 10% of patients affected by APS. APS1 is transmitted in an autosomal recessive manner and is caused by mutations in the autoimmune regulator gene *AIRE*. Associated autoimmune conditions overlap with those seen in APS2 but may also include total alopecia, primary hypoparathyroidism, and, in rare cases, lymphoma. APS1 patients invariably develop chronic mucocutaneous candidiasis, usually manifested in childhood and preceding adrenal insufficiency by years or decades. The much more prevalent APS2 is of polygenic inheritance, with confirmed associations with the *HLA-DR3* gene region in the major histocompatibility complex and distinct gene regions involved in immune regulation (*CTLA-4*, *PTPN22*, *CLEC16A*). Coincident autoimmune disease most frequently includes thyroid autoimmune disease, vitiligo, and premature ovarian failure. Less commonly, additional features may include type 1 diabetes and pernicious anemia caused by vitamin  $B_{12}$  deficiency.

X-linked adrenoleukodystrophy has an incidence of 1:20,000 males and is caused by mutations in the *X-ALD* gene encoding the peroxisomal membrane transporter protein *ABCD1*; its disruption results in accumulation of very-long-chain (>24 carbon atoms) fatty acids. Approximately 50% of cases manifest in early childhood with rapidly progressive white matter disease (cerebral adrenoleukodystrophy); 35% present during adolescence or in early adulthood with neurologic features indicative of myelin and peripheral nervous system

**TABLE 386-7 Causes of Primary Adrenal Insufficiency**

DIAGNOSIS	GENE	ASSOCIATED FEATURES
Autoimmune polyglandular syndrome 1 (APS1)	<i>AIRE</i>	Hypoparathyroidism, chronic mucocutaneous candidiasis, other autoimmune disorders, rarely lymphomas
Autoimmune polyglandular syndrome 2 (APS2)	Associations with <i>HLA-DR3</i> , <i>CTLA-4</i>	Hypothyroidism, hyperthyroidism, premature ovarian failure, vitiligo, type 1 diabetes mellitus, pernicious anemia
Isolated autoimmune adrenalitis	Associations with <i>HLA-DR3</i> , <i>CTLA-4</i>	
Congenital adrenal hyperplasia (CAH)	<i>CYP21A2</i> , <i>CYP11B1</i> , <i>CYP17A1</i> , <i>HSD3B2</i> , <i>POR</i>	See Table 386-10 (see also <a href="#">Chap. 390</a> )
Congenital lipoid adrenal hyperplasia (CLAH)	<i>STAR</i> , <i>CYP11A1</i>	46,XY DSD, gonadal failure (see also <a href="#">Chap. 390</a> )
Adrenal hypoplasia congenita (AHC)	<i>NR0B1</i> ( <i>DAX-1</i> ), <i>NR5A1</i> ( <i>SF-1</i> )	46,XY DSD, gonadal failure (see also <a href="#">Chap. 390</a> )
Adrenoleukodystrophy (ALD), adrenomyeloneuropathy (AMN)	<i>ABCD1</i>	Demyelination of central nervous system (ALD) or spinal cord and peripheral nerves (AMN)
Familial glucocorticoid deficiency	<i>MC2R</i> <i>MRAP</i> <i>STAR</i> <i>NNT</i> <i>TXNRD2</i> <i>MCM4</i>	Tall stature None None None None Growth retardation, natural killer cell deficiency
Triple A syndrome	<i>AAAS</i>	Alacrima, achalasia, neurologic impairment
Smith-Lemli-Opitz syndrome	<i>SLOS</i>	Cholesterol synthesis disorder associated with mental retardation, craniofacial malformations, growth failure
Kearns-Sayre syndrome	Mitochondrial DNA deletions	Progressive external ophthalmoplegia, pigmentary retinal degeneration, cardiac conduction defects, gonadal failure, hypoparathyroidism, type 1 diabetes,
IMAGe syndrome	<i>CDKN1C</i>	Intrauterine growth retardation, metaphyseal dysplasia, genital anomalies
MIRAGE syndrome	<i>SAMD9</i>	Myelodysplasia, infection, restriction of growth, genital phenotypes, and enteropathy
Sphingosine-1-phosphate lyase deficiency	<i>SGPL1</i>	Steroid-resistant nephrotic syndrome, immunodeficiency, neurological defects, ichthyosis, primary hypothyroidism, cryptorchidism
Adrenal infections		Tuberculosis, HIV, CMV, cryptococcosis, histoplasmosis, coccidioidomycosis
Adrenal infiltration		Metastases, lymphomas, sarcoidosis, amyloidosis, hemochromatosis
Adrenal hemorrhage		Meningococcal sepsis (Waterhouse-Friderichsen syndrome), primary antiphospholipid syndrome
Drug-induced		Mitotane, aminoglutethimide, abiraterone, trilostane, etomidate, ketoconazole, osilodrostat, suramin, RU486, interferon-alpha, ribavirin, megestrol acetate, immune checkpoint inhibitors (rare)
Bilateral adrenalectomy		E.g., in the management of Cushing's syndrome or after bilateral nephrectomy

**Abbreviations:** AIRE, autoimmune regulator; CMV, cytomegalovirus; DSD, disordered sex development; MC2R, ACTH receptor; MCM4, mini chromosome maintenance-deficient 4 homologue; MRAP, MC2R-accessory protein; NNT, nicotinamide nucleotide transhydrogenase.

involvement (adrenomyeloneuropathy [AMN]). In the remaining 15%, adrenal insufficiency is the sole manifestation of disease. Of note, distinct mutations manifest with variable penetrance and phenotypes within affected families.

Rarer causes of adrenal insufficiency involve destruction of the adrenal glands as a consequence of infection, hemorrhage, or infiltration (Table 386-7); tuberculous adrenalitis is still a frequent cause of disease in developing countries. Adrenal metastases rarely cause adrenal insufficiency, and this occurs only with bilateral, bulky metastases.

Inborn causes of primary adrenal insufficiency other than CAH are rare, causing <1% of cases. However, their elucidation provides important insights into adrenal gland development and physiology. Mutations causing primary adrenal insufficiency (Table 386-7) include factors regulating adrenal development and steroidogenesis (DAX-1, SF-1), cholesterol synthesis, import and cleavage (DHCR7, StAR, CYP11A1), elements of the adrenal ACTH response pathway (MC2R, MRAP) (Fig. 386-5), and factors involved in redox regulation (NNT, TXNRD2) and DNA repair (MCM4, CDKN1C).

**Secondary (or central) adrenal insufficiency** is the consequence of dysfunction of the hypothalamic-pituitary component of the HPA axis (Table 386-8). Excluding iatrogenic suppression, the overwhelming majority of cases are caused by pituitary or hypothalamic tumors or their treatment by surgery or irradiation (Chap. 380). Rarer causes include pituitary apoplexy, either as a consequence of an infarcted pituitary adenoma or transient reduction in the blood supply of the pituitary during surgery or after rapid blood loss associated with parturition, also termed Sheehan's syndrome. Isolated ACTH deficiency is rarely caused by autoimmune disease or pituitary infiltration (Table 386-8). Mutations in the ACTH precursor POMC or in factors regulating pituitary development are genetic causes of ACTH deficiency (Table 386-8).

**Clinical Manifestations** In principle, the clinical features of primary adrenal insufficiency (Addison's disease) are characterized by the loss of both glucocorticoid and mineralocorticoid secretion (Table 386-9). In secondary adrenal insufficiency, only glucocorticoid deficiency is present, as the adrenal itself is intact and thus still amenable to regulation by the RAA system. Adrenal androgen secretion is disrupted in both primary and secondary adrenal insufficiency (Table 386-9). Hypothalamic-pituitary disease can lead to additional clinical manifestations due to involvement of other endocrine axes (thyroid, gonads, GH, prolactin) or visual impairment with bitemporal hemianopia caused by chiasmal compression. It is important to recognize that iatrogenic adrenal insufficiency caused by exogenous glucocorticoid suppression of the HPA axis may result in all symptoms associated with glucocorticoid deficiency (Table 386-9), if exogenous glucocorticoids are stopped abruptly. However, patients will appear clinically cushingoid as a result of the preceding overexposure to glucocorticoids.

**Chronic adrenal insufficiency** manifests with relatively nonspecific signs and symptoms, such as fatigue and loss of energy, often resulting in delayed or missed diagnoses (e.g., as depression or anorexia). A distinguishing feature of primary adrenal insufficiency is hyperpigmentation, which is caused by excess ACTH stimulation of melanocytes. Hyperpigmentation is most pronounced in skin areas exposed to increased friction or shear stress and is increased by sunlight (Fig. 386-15). Conversely, in secondary adrenal insufficiency, the skin has an alabaster-like palleness due to lack of ACTH secretion.

Hypotension is a characteristic biochemical feature in primary adrenal insufficiency and is found in 80% of patients at presentation. Hyperkalemia is present in 40% of patients at initial diagnosis. Hyponatremia is primarily caused by mineralocorticoid deficiency but can also occur in secondary adrenal insufficiency due to diminished inhibition of antidiuretic hormone (ADH) release by cortisol, resulting in mild syndrome of inappropriate secretion of antidiuretic hormone (SIADH). Glucocorticoid deficiency also results in slightly increased TSH concentrations that normalize within days to weeks after initiation of glucocorticoid replacement.

**Acute adrenal insufficiency**, also termed adrenal crisis, usually occurs after a prolonged period of nonspecific complaints and is more

**TABLE 386-8 Causes of Secondary Adrenal Insufficiency**

DIAGNOSIS	GENE	ASSOCIATED FEATURES
Pituitary tumors (endocrine active and inactive adenomas, very rare: carcinoma)		Depending on tumor size and location: visual field impairment (bilateral hemianopia), hyperprolactinemia, secondary hypothyroidism, hypogonadism, growth hormone deficiency
Other mass lesions affecting the hypothalamic-pituitary region		Craniopharyngioma, meningioma, ependymoma, metastases
Pituitary irradiation		Radiotherapy administered for pituitary tumors, brain tumors, or craniospinal irradiation in leukemia
Autoimmune hypophysitis		Often associated with pregnancy; may present with panhypopituitarism or isolated ACTH deficiency: can be associated with autoimmune thyroid disease, more rarely with vitiligo, premature ovarian failure, type 1 diabetes, pernicious anemia
Pituitary apoplexy/ hemorrhage		Hemorrhagic infarction of large pituitary adenomas or pituitary infarction consequent to traumatic major blood loss (e.g., surgery or pregnancy: Sheehan's syndrome)
Pituitary infiltration		Tuberculosis, actinomycosis, sarcoidosis, histiocytosis X, granulomatosis with polyangiitis (Wegener's), metastases
Drug-induced		Chronic glucocorticoid excess (endogenous or exogenous), immune check point inhibitors
Congenital isolated ACTH deficiency	<i>TBX19</i> (Tpit)	
Combined pituitary hormone deficiency (CPHD)	<i>PROP-1</i>  <i>HESX1</i> <i>LHX3</i>  <i>LHX4</i> <i>SOX3</i>	Progressive development of CPHD in the order GH, PRL, TSH, LH/FSH, ACTH  CPHD and septo-optic dysplasia CPHD and limited neck rotation, sensorineural deafness CPHD and cerebellar abnormalities CPHD and variable mental retardation
Proopiomelanocortin (POMC) deficiency	<i>POMC</i>	Early-onset obesity, red hair pigmentation

*Abbreviations:* ACTH, adrenocorticotrophic hormone; GH, growth hormone; LH/FSH, luteinizing hormone/follicle-stimulating hormone; PRL, prolactin; TSH, thyroid-stimulating hormone.

frequently observed in patients with primary adrenal insufficiency, due to the loss of both glucocorticoid and mineralocorticoid secretion. Postural hypotension may progress to hypovolemic shock. Adrenal insufficiency may mimic features of acute abdomen with abdominal tenderness, nausea, vomiting, and fever. In some cases, the primary presentation may resemble neurologic disease, with decreased responsiveness progressing to stupor and coma. An adrenal crisis can be triggered by an intercurrent illness, surgical or other stress, or increased glucocorticoid inactivation (e.g., hyperthyroidism). Prospective data indicate 8.3 adrenal crises and 0.5 adrenal crisis-related deaths per 100 patient-years.

**Diagnosis** The diagnosis of adrenal insufficiency is established by the short cosyntropin test, a safe and reliable tool with excellent predictive diagnostic value (Fig. 386-16). The cutoff for failure is usually defined at cortisol levels of <450–500 nmol/L (16–18 µg/dL) sampled 30–60 min after ACTH stimulation; the exact cutoff is dependent on the locally available assay, with generally lower cutoffs for mass spectrometry-based assays. During the early phase of HPA disruption (e.g., within 4 weeks of pituitary insufficiency), patients may still respond to exogenous ACTH stimulation. In this circumstance, the ITT is an

**TABLE 386-9 Signs and Symptoms of Adrenal Insufficiency**

Signs and Symptoms Caused by Glucocorticoid Deficiency	
Fatigue, lack of energy	
Weight loss, anorexia	
Myalgia, joint pain	
Fever	
Normochromic anemia, lymphocytosis, eosinophilia	
Slightly increased TSH (due to loss of feedback inhibition of TSH release)	
Hypoglycemia (more frequent in children)	
Low blood pressure, postural hypotension	
Hyponatremia (due to loss of feedback inhibition of AVP release)	
Signs and Symptoms Caused by Mineralocorticoid Deficiency (Primary Adrenal Insufficiency Only)	
Abdominal pain, nausea, vomiting	
Dizziness, postural hypotension	
Salt craving	
Low blood pressure, postural hypotension	
Increased serum creatinine (due to volume depletion)	
Hyponatremia	
Hyperkalemia	
Signs and Symptoms Caused by Adrenal Androgen Deficiency	
Lack of energy	
Dry and itchy skin (in women)	
Loss of libido (in women)	
Loss of axillary and pubic hair (in women)	
Other Signs and Symptoms	
Hyperpigmentation (primary adrenal insufficiency only) (due to excess of proopiomelanocortin [POMC]-derived peptides)	
Alabaster-colored pale skin (secondary adrenal insufficiency only) (due to deficiency of POMC-derived peptides)	

Abbreviations: AVP, arginine vasopressin; TSH, thyroid-stimulating hormone.

alternative choice but is more invasive and should be carried out only under a specialist's supervision (see above). Induction of hypoglycemia is contraindicated in individuals with diabetes mellitus, cardiovascular disease, or history of seizures. Random serum cortisol measurements are of limited diagnostic value because baseline cortisol levels may be coincidentally low due to the physiologic diurnal rhythm of cortisol secretion (Fig. 386-3). Similarly, many patients with secondary adrenal insufficiency have relatively normal baseline cortisol levels but fail to mount an appropriate cortisol response to ACTH, which can only be revealed by stimulation testing. Importantly, tests to establish the diagnosis of adrenal insufficiency should never delay treatment. Thus, in a patient with suspected adrenal crisis, it is reasonable to draw baseline cortisol levels, provide replacement therapy, and defer formal stimulation testing until a later time.

Once adrenal insufficiency is confirmed, measurement of plasma ACTH is the next step, with increased or inappropriately low levels defining primary and secondary origin of disease, respectively (Fig. 386-16). In primary adrenal insufficiency, increased plasma renin will confirm the presence of mineralocorticoid deficiency. At initial presentation, patients with primary adrenal insufficiency should undergo screening for steroid autoantibodies as a marker of autoimmune adrenalitis. If these tests are negative, adrenal imaging by CT is indicated to investigate possible hemorrhage, infiltration, or masses. In male patients with negative autoantibodies in the plasma, very-long-chain fatty acids should be measured to exclude X-ALD. Patients with inappropriately low ACTH, in the presence of confirmed cortisol deficiency, should undergo hypothalamic-pituitary imaging by MRI. Features suggestive of preceding pituitary apoplexy, such as sudden-onset severe headache or history of previous head trauma, should be carefully explored, particularly in patients with no obvious MRI lesion.

## TREATMENT

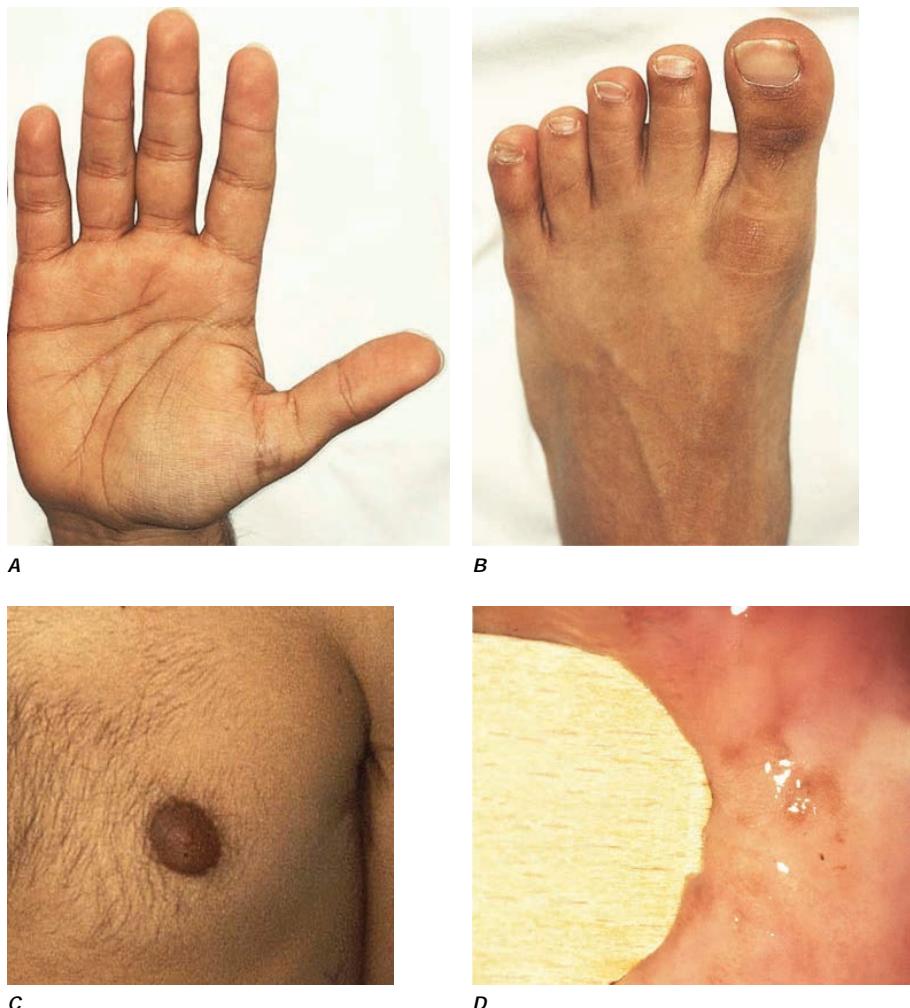
### Acute Adrenal Insufficiency

Acute adrenal insufficiency requires immediate initiation of rehydration, usually carried out by saline infusion at initial rates of 1 L/h with continuous cardiac monitoring. Glucocorticoid replacement should be initiated by bolus injection of 100 mg hydrocortisone, followed by the administration of 200 mg hydrocortisone over 24 h, preferably by continuous infusion or alternatively by bolus IV or IM injections. Mineralocorticoid replacement can be initiated once the daily hydrocortisone dose has been reduced to <50 mg because at higher doses hydrocortisone provides sufficient stimulation of MRs.

**Glucocorticoid replacement** for the treatment of chronic adrenal insufficiency should be administered at a dose that replaces the physiologic daily cortisol production, which is usually achieved by the oral administration of 15–25 mg hydrocortisone in two to three divided doses. Pregnancy may require an increase in hydrocortisone dose by 50% during the last trimester. In all patients, at least one-half of the daily dose should be administered in the morning. Currently available glucocorticoid preparations fail to mimic the physiologic cortisol secretion rhythm (Fig. 386-3). Long-acting glucocorticoids such as prednisolone or dexamethasone are not preferred because they result in increased glucocorticoid exposure due to extended GR activation at times of physiologically low cortisol secretion. There are no well-established dose equivalencies, but as a guide, equipotency can be assumed for 1 mg hydrocortisone, 1.6 mg cortisone acetate, 0.2 mg prednisolone, 0.25 mg prednisone, and 0.025 mg dexamethasone.

Monitoring of glucocorticoid replacement is mainly based on the history and examination for signs and symptoms suggestive of glucocorticoid over- or underreplacement, including assessment of body weight and blood pressure. Plasma ACTH, 24-h urinary free cortisol, or serum cortisol day curves reflect whether hydrocortisone has been taken or not but do not convey reliable information about replacement quality. In patients with isolated primary adrenal insufficiency, monitoring should include screening for autoimmune thyroid disease, and female patients should be made aware of the possibility of premature ovarian failure. Supraphysiologic glucocorticoid treatment with doses equivalent to 30 mg hydrocortisone or more will affect bone metabolism, and these patients should undergo regular bone mineral density evaluation. All patients with adrenal insufficiency need to be instructed about the requirement for stress-related glucocorticoid dose adjustments. These generally consist of doubling the routine oral glucocorticoid dose in the case of intercurrent illness with fever and bed rest and the need for immediate IV or IM injection of 100 mg hydrocortisone followed by intravenous infusion of 200 mg hydrocortisone/24 h in cases of prolonged vomiting, surgery, or trauma. All patients, but in particular those living or traveling in regions with delayed access to acute health care, should carry a hydrocortisone self-injection emergency kit, in addition to their usual steroid emergency cards and bracelets, and should receive training in its use.

**Mineralocorticoid replacement** in primary adrenal insufficiency should be initiated at a dose of 100–150 µg fludrocortisone. The adequacy of treatment can be evaluated by measuring blood pressure, sitting and standing, to detect a postural drop indicative of hypovolemia. In addition, serum sodium, potassium, and plasma renin should be measured regularly. Renin levels should be kept in the upper normal reference range. Changes in glucocorticoid dose may also impact on mineralocorticoid replacement as cortisol also binds the MR; 40 mg of hydrocortisone is equivalent to 100 µg of fludrocortisone. It is important to note that prednisone and prednisolone have reduced mineralocorticoid activity and dexamethasone has none. In patients living or traveling in areas with hot or tropical weather conditions, the fludrocortisone dose should be increased by 50–100 µg during the summer. Mineralocorticoid dose may also need to be adjusted during pregnancy due to the antiminerocorticoid activity of progesterone, but this is less often required



**FIGURE 386-15** Clinical features of Addison's disease. Note the hyperpigmentation in areas of increased friction including (A) palmar creases, (B) dorsal foot, (C) nipples and axillary region, and (D) patchy hyperpigmentation of the oral mucosa.

than hydrocortisone dose adjustment. Plasma renin cannot serve as a monitoring tool during pregnancy because renin rises physiologically during gestation.

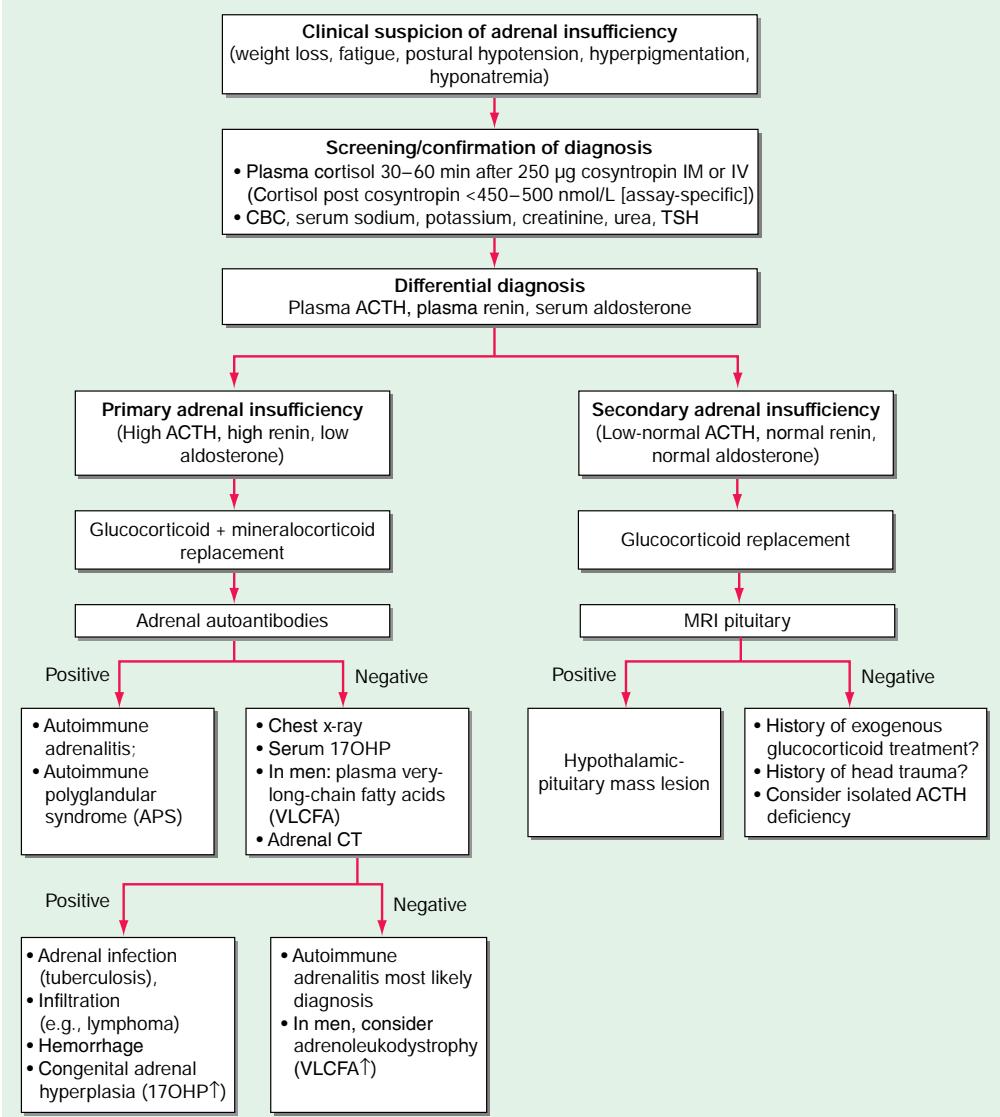
**Adrenal androgen replacement** is an option in patients with lack of energy, despite optimized glucocorticoid and mineralocorticoid replacement. It may also be indicated in women with features of androgen deficiency, including loss of libido. Adrenal androgen replacement can be achieved by once-daily administration of 25–50 mg DHEA. Treatment is monitored by measurement of DHEAS, androstenedione, testosterone, and sex hormone-binding globulin (SHBG) 24 h after the last DHEA dose.

### CONGENITAL ADRENAL HYPERPLASIA

(See also Chap. 390) CAH is caused by mutations in genes encoding steroidogenic enzymes involved in glucocorticoid synthesis (*CYP21A2*, *CYP17A1*, *HSD3B2*, *CYP11B1*) or in the cofactor enzyme P450 oxidoreductase that serves as an electron donor to *CYP21A2* and *CYP17A1* (Fig. 386-1). Invariably, patients affected by CAH exhibit glucocorticoid deficiency. Depending on the exact step of enzymatic block, they may also have excess production of mineralocorticoids or deficient production of sex steroids (Table 386-10). The diagnosis of CAH is readily established by measurement of the steroids accumulating before the distinct enzymatic block, either in serum or in urine, preferably by the use of mass spectrometry-based assays (Table 386-10).

Mutations in *CYP21A2* are the most prevalent cause of CAH, responsible for 90–95% of cases. 21-Hydroxylase deficiency disrupts glucocorticoid and mineralocorticoid synthesis (Fig. 386-1), resulting in diminished negative feedback via the HPA axis. This leads to increased pituitary ACTH release, which drives increased synthesis of adrenal androgen precursors and subsequent androgen excess. The degree of impairment of glucocorticoid and mineralocorticoid secretion depends on the severity of mutations. Major loss-of-function mutations result in combined glucocorticoid and mineralocorticoid deficiency (classic CAH, neonatal presentation), whereas less severe mutations affect glucocorticoid synthesis only (simple virilizing CAH, neonatal or early childhood presentation). The mildest mutations result in the least severe clinical phenotype, nonclassic CAH, usually presenting during adolescence and early adulthood and with preserved glucocorticoid production.

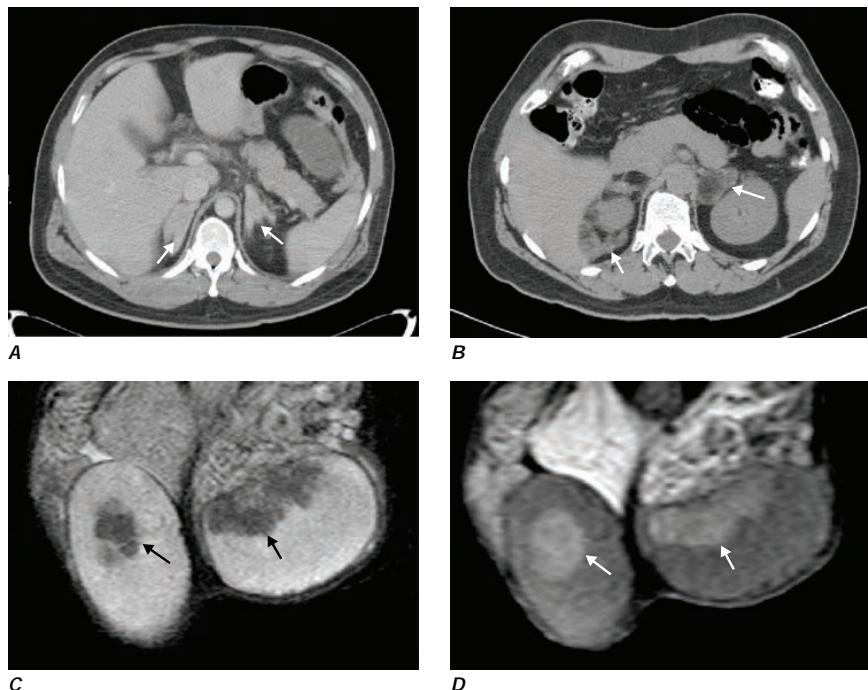
Androgen excess is present in all patients and manifests with broad phenotypic variability, ranging from severe virilization of the external genitalia in neonatal girls (e.g., 46,XX disordered sex development [DSD]) to hirsutism and oligomenorrhea resembling a polycystic ovary syndrome phenotype in young women with nonclassic CAH. In countries without neonatal screening for CAH, boys with classic CAH usually present with life-threatening adrenal crisis in the first few weeks of life (salt-wasting crisis); a simple-virilizing genotype manifests with precocious pseudopuberty and advanced bone age in early childhood, whereas men with nonclassic CAH are usually detected only through family screening.



**FIGURE 386-16** Management of the patient with suspected adrenal insufficiency. ACTH, adrenocorticotrophic hormone; CBC, complete blood count; MRI, magnetic resonance imaging; PRA, plasma renin activity; TSH, thyroid-stimulating hormone.

**TABLE 386-10** Variants of Congenital Adrenal Hyperplasia

VARIANT	GENE	IMPACT ON STEROID SYNTHESIS	DIAGNOSTIC MARKER STEROIDS IN SERUM (AND URINE)
21-Hydroxylase deficiency (21OHD)	CYP21A2	Glucocorticoid deficiency, mineralocorticoid deficiency, adrenal androgen excess	17-Hydroxyprogesterone, 21-deoxycortisol (pregnanetriol, 17-hydroxypregnanolone, pregnanetriolone)
11 $\beta$ -Hydroxylase deficiency (11OHD)	CYP11B1	Glucocorticoid deficiency, mineralocorticoid excess, adrenal androgen excess	11-Deoxycortisol, 11-deoxycorticosterone (tetrahydro-11-deoxycortisol, tetrahydro-11-deoxycorticosterone)
17 $\alpha$ -Hydroxylase deficiency (17OHD)	CYP17A1	(Glucocorticoid deficiency), mineralocorticoid excess, androgen deficiency	11-Deoxycorticosterone, corticosterone, pregnenolone, progesterone (tetrahydro-11-deoxycorticosterone, tetrahydrocorticosterone, pregnenediol, pregnanediol)
3 $\beta$ -Hydroxysteroid dehydrogenase deficiency (3 $\beta$ HSDD)	HSD3B2	Glucocorticoid deficiency, (mineralocorticoid deficiency), adrenal androgen excess (females and males), gonadal androgen deficiency (males)	17-Hydroxypregnanolone (pregnanetriol)
P450 oxidoreductase deficiency (PORD)	POR	Glucocorticoid deficiency, (mineralocorticoid excess), prenatal androgen excess and postnatal androgen deficiency, skeletal malformations	Pregnenolone, progesterone, 17-hydroxyprogesterone (pregnanediol, pregnanetriol)



**FIGURE 386-17** Imaging in congenital adrenal hyperplasia (CAH). Adrenal computed tomography scans showing homogenous bilateral hyperplasia in a young patient with classic CAH (**A**) and macronodular bilateral hyperplasia (**B**) in a middle-aged patient with classic CAH with longstanding poor disease control. Magnetic resonance imaging scan with T1-weighted (**C**) and T2-weighted (**D**) images showing bilateral testicular adrenal rest tumors (arrows) in a young patient with salt-wasting CAH. (Used with permission from N. Reisch.)

Glucocorticoid treatment is more complex than for other causes of primary adrenal insufficiency as it not only needed to replace missing glucocorticoids but also to control the increased ACTH drive and subsequent androgen excess. Current treatment is hampered by the lack of glucocorticoid preparations that mimic the diurnal cortisol secretion profile, resulting in a prolonged period of ACTH stimulation and subsequent androgen production during the early morning hours. In childhood, optimization of growth and pubertal development are important goals of glucocorticoid treatment, in addition to prevention of adrenal crisis and treatment of 46,XX DSD. In adults, the focus shifts to preserving fertility and preventing side effects of glucocorticoid overtreatment, namely, the metabolic syndrome and osteoporosis. Fertility can be compromised in women due to oligomenorrhea/amenorrhea with chronic anovulation as a consequence of androgen excess. Men may develop testicular adrenal rest tissue (TART) (Fig. 386-17) consisting of hyperplastic cells with shared adrenal and gonadal characteristics located in the rete testis, which should not be confused with testicular tumors. TART can compromise sperm production and induce testicular fibrosis that may be irreversible.

## TREATMENT

### Congenital Adrenal Hyperplasia

Hydrocortisone is a good treatment option for the prevention of adrenal crisis, but longer acting prednisolone may be needed to control androgen excess. In children, hydrocortisone is given in divided doses at 1–1.5 times the normal cortisol production rate (~10–13 mg/m<sup>2</sup> per day). In adults, if hydrocortisone does not suffice, intermediate-acting glucocorticoids (e.g., prednisone) may be given, using the lowest dose necessary to suppress excess androgen production. For achieving fertility, dexamethasone treatment may be required but should only be given for the shortest possible time period to limit adverse metabolic side effects. The recent introduction of modified and delayed-release hydrocortisone, which mimics the endogenous physiologic cortisol release pattern, is promising,

providing effective control of steroid precursor excess while the daily hydrocortisone dose is lower than required for immediate-release hydrocortisone.

Biochemical monitoring should include androstenedione and testosterone, aiming for the normal sex-specific reference range. 17OHP is a useful marker of overtreatment, indicated by 17OHP levels within the normal range of healthy controls. Glucocorticoid overtreatment may suppress the hypothalamic-pituitary-gonadal axis. Thus, treatment needs to be carefully titrated against clinical features of disease control. Stress-dose glucocorticoids should be given at double or triple the daily dose for surgery, acute illness, or severe trauma. Poorly controlled CAH can result in adrenocortical hyperplasia, which gave the disease its name, and may present as macronodular hyperplasia subsequent to long-standing ACTH excess (Fig. 386-17). The nodular areas can develop autonomous adrenal androgen production and may be unresponsive to glucocorticoid treatment. The prevalence of adrenomyelolipomas is increased in CAH; these are benign but can require surgical intervention due to lack of self-limiting growth.

Mineralocorticoid requirements change during life and are higher in children, explained by relative mineralocorticoid resistance that diminishes with ongoing maturation of the kidney. Children with CAH usually receive mineralocorticoid and salt replacement. However, young adults with CAH should undergo reassessment of their mineralocorticoid reserve. Plasma renin should be regularly monitored and kept within the upper half of the normal reference range.

## FURTHER READING

- Arlt W et al: Steroid metabolome analysis reveals prevalent glucocorticoid excess in primary aldosteronism. *JCI Insight* 2:e93136, 2017.
- Bancos I et al: Urine steroid metabolomics for the differential diagnosis of adrenal incidentalomas in the EURINE-ACT study: A prospective test validation study. *Lancet Diabetes Endocrinol* 8:773, 2020.
- Bornstein SR et al: Diagnosis and treatment of primary adrenal insufficiency: An Endocrine Society Clinical Practice guideline. *J Clin Endocrinol Metab* 101:364, 2016.

Claahsen-Van Der Grinten HL et al: Congenital adrenal hyperplasia: Current insights in pathophysiology, diagnostics and management. *Endocr Rev* 2021;7:bnab016.

Ebbehoj A et al: Epidemiology of adrenal tumours in Olmsted County, Minnesota, USA: A population-based cohort study. *Lancet Diabetes Endocrinol* 8:894, 2020.

Fassnacht M et al: Management of adrenal incidentalomas: European Society of Endocrinology Clinical Practice Guideline in collaboration with the European Network for the Study of Adrenal Tumors. *Eur J Endocrinol* 175:G1, 2016.

Feeiders RA et al: Advances in the medical treatment of Cushing's syndrome. *Lancet Diabetes Endocrinol* 7:300, 2019.

Funder JW et al: The management of primary aldosteronism: Case detection, diagnosis and treatment: An Endocrine Society Clinical Practice guideline. *J Clin Endocrinol Metab* 101:1889, 2016.

Hahner S et al: Adrenal insufficiency. *Nat Rev Dis Primers* 7:19, 2021.

Lodish M, Stratakis CA: A genetic and molecular update on adrenocortical causes of Cushing syndrome. *Nat Rev Endocrinol* 12:255, 2016.

Loriaux DL: Diagnosis and differential diagnosis of Cushing's syndrome. *N Engl J Med* 376:1451, 2017.

Merke DP, Auchus RJ: Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *N Engl J Med* 383:1248, 2020.

Merke DP et al: Modified-release hydrocortisone in congenital adrenal hyperplasia. *J Clin Endocrinol Metab* 106:e2063, 2021.

Mulatero P et al: Genetics, prevalence, screening and confirmation of primary aldosteronism: Position statement and consensus of the Working Group on Endocrine Hypertension of the European Society of Hypertension. *J Hypertens* 38:1919, 2020.

Nanba K, Rainey WE: Genetics in endocrinology: Impact of race and sex on genetic causes of aldosterone-producing adenomas. *Eur J Endocrinol* 185:R1, 2021.

Prete A et al: Prevention of adrenal crisis: Cortisol responses to major stress compared to stress dose hydrocortisone delivery. *J Clin Endocrinol Metab* 105:2262, 2020.

glomus tympanicum, glomus jugulare, glomus vagale) paraganglia (**Fig. 387-1**). The name *pheochromocytoma* reflects the formerly used black-colored staining caused by chromaffin oxidation of catecholamines; although a variety of terms have been used to describe these tumors, most clinicians use this designation to describe symptomatic catecholamine-producing tumors, including those in extra-adrenal retroperitoneal, pelvic, and thoracic sites. The term *paraganglioma* is used to describe catecholamine-producing tumors in the skull base and neck; these tumors may secrete little or no catecholamine. In contrast to common clinical parlance, the World Health Organization (WHO) restricts the term *pheochromocytoma* to adrenal tumors and applies the term *paraganglioma* to tumors at all other sites.

The etiology of sporadic pheochromocytomas and paragangliomas is unknown. However, 25–33% of patients have an inherited condition, including germline mutations in the classically recognized *RET* (rearranged during transfection), *VHL*, *NF1* (neurofibromatosis type 1), *SDHB*, *SDHC*, and *SDHD* (subunits of SDH) genes or in the more recently recognized *SDHA*, *SDHAF2*, *TMEM127* (transmembrane protein 127), *MAX* (myc-associated factor X), *FH* (fumarate hydratase), *PDH1*, *PDH2* (pyruvate dehydrogenase), *HIF1* and *HIF2* (hypoxia-inducible factor), *MDH2* (malate dehydrogenase), *KIF1B* (kinesin family member), *IDH1* (isocitrate dehydrogenase 1), *SLC25A11* (oxoglutarate/malate), *H-RAS* (transforming protein p21), and *DNMTA3* (DNA methyltransferase 3 alpha) genes. Biallelic gene inactivation, a characteristic of tumor-suppressor genes, has been demonstrated for the *VHL*, *NF1*, *SDHx*, *TMEM127*, *MAX*, *FH*, *PDH1*, *PDH2*, *MDH2*, and *KIF1B* genes. In contrast, *RET* is a protooncogene, and mutations activate receptor tyrosine kinase activity. Succinate dehydrogenase (SDH) is an enzyme of the Krebs cycle and the mitochondrial respiratory chain. The *VHL* protein is a component of a ubiquitin E3 ligase. *VHL* mutations reduce protein degradation, resulting in upregulation of components involved in cell-cycle progression, glucose metabolism, and oxygen sensing. In addition to germline mutations, somatic mutations have been observed in >20 genes, broadly grouped into three different clusters of pathogenetically relevant genes: cluster 1, the pseudohypoxia group comprising mainly the genes *SDHx* (subunits of SDH), *FH*, *VHL*, and *HIF2A*; cluster 2, the kinase signaling group (*RET*, *NF1*, *TMEM127*, *MAX*, *HRAS*, *KIF1B*, *PDH*); and cluster 3, the Wnt signaling group (*CSDE1*, *MAML3*).

## CLINICAL FEATURES

Its clinical presentation is so variable that pheochromocytoma has been termed “the great masquerader” (**Table 387-1**). Among the presenting manifestations, episodes of palpitation, headache, and profuse sweating are typical, and these manifestations constitute a classic triad. The presence of all three manifestations in association with hypertension makes pheochromocytoma a likely diagnosis. However, a pheochromocytoma can be asymptomatic for years, and some tumors grow to a considerable size before patients note symptoms.

The dominant sign is hypertension. Classically, patients have episodic hypertension, but sustained hypertension is also common. Catecholamine crises can lead to heart failure, pulmonary edema, arrhythmias, and intracranial hemorrhage. During episodes of hormone release, which can occur at widely divergent intervals, patients are anxious and pale, and they experience tachycardia and palpitations. These paroxysms generally last <1 h and may be precipitated by surgery, positional changes, exercise, pregnancy, urination (particularly with bladder pheochromocytomas), and various medications (e.g., tricyclic antidepressants, opiates, metoclopramide).

## DIAGNOSIS

The diagnosis is based on documentation of catecholamine excess by biochemical testing and localization of the tumor by imaging. These two criteria are of equal importance, although measurement of catecholamines or metanephrines (their methylated metabolites) is traditionally the first step in diagnosis.

**Biochemical Testing** Pheochromocytomas and paragangliomas synthesize and store catecholamines, which include norepinephrine (noradrenaline), epinephrine (adrenaline), and dopamine. Elevated

# 387

## Pheochromocytoma

Hartmut P. H. Neumann



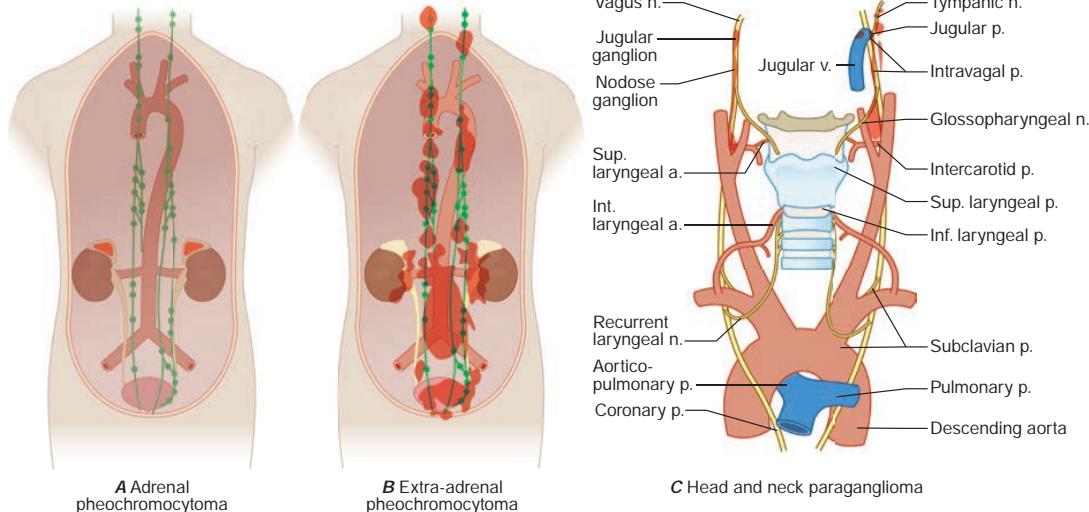
Pheochromocytomas and paragangliomas are catecholamine-producing tumors derived from the sympathetic or parasympathetic nervous system. These tumors may arise sporadically or be inherited as features of multiple endocrine neoplasia type 2 (MEN 2), von Hippel-Lindau (VHL) disease, or several other pheochromocytoma-associated syndromes. The diagnosis of pheochromocytomas identifies a potentially correctable cause of hypertension, and their removal can prevent hypertensive crises that can be lethal. The clinical presentation is variable, ranging from an adrenal incidentaloma to a hypertensive crisis with associated cerebrovascular or cardiac complications.

### EPIDEMIOLOGY

Pheochromocytoma is estimated to occur in 2–8 of 1 million persons per year, and ~0.1% of hypertensive patients harbor a pheochromocytoma. The mean age at diagnosis is ~40 years, although the tumors can occur from early childhood until late in life. The classic “rule of tens” for pheochromocytomas states that ~10% are bilateral, 10% are extra-adrenal, and 10% are metastatic.

### ETIOLOGY AND PATHOGENESIS

Pheochromocytomas and paragangliomas are well-vascularized tumors that arise from cells derived from the sympathetic (e.g., adrenal medulla or sympathetic trunk) or parasympathetic (e.g., carotid body,



**FIGURE 387-1** The paragangli system and topographic sites (in red) of pheochromocytomas and paragangliomas. (Figures A, B reproduced with permission from WM Manger, RW Gifford: *Clinical and experimental pheochromocytoma*. Cambridge: Blackwell Science; 1996.)

plasma and urinary levels of catecholamines and metanephrenes form the cornerstone of diagnosis. The characteristic fluctuations in the hormonal activity of tumors result in considerable variation in serial catecholamine measurements. However, most tumors continuously leak O-methylated metabolites, which are detected by measurement of metanephrenes.

Catecholamines and metanephrenes can be measured by different methods, including high-performance liquid chromatography, enzyme-linked immunosorbent assay, and liquid chromatography/mass spectrometry. When pheochromocytoma is suspected on clinical grounds (i.e., when values are three times the upper limit of normal), this diagnosis is highly likely regardless of the assay used. However, as summarized in **Table 387-2**, the sensitivity and specificity of available biochemical tests vary greatly, and these differences are important in assessing patients with borderline elevations of different compounds. Urinary tests for metanephrenes (total or fractionated) and catecholamines are widely available and are used commonly for initial evaluation. Among these tests, those for the fractionated metanephrenes and catecholamines are the most sensitive. Plasma tests are more convenient and include measurements of catecholamines and metanephrenes. Measurements of plasma metanephrenine are the most sensitive and are less susceptible to false-positive elevations from stress, including venipuncture. Although the incidence of false-positive test results has been reduced by the introduction of newer assays, physiologic stress responses and medications that increase catecholamine levels still can confound testing. Because the tumors are relatively rare, borderline elevations are likely to represent false-positive results. In this circumstance, it is important to exclude dietary or drug-related factors (withdrawal of levodopa or use of sympathomimetics, diuretics, tricyclic antidepressants, alpha and beta blockers) that might cause

false-positive results and then to repeat testing or perform a clonidine suppression test (i.e., the measurement of plasma normetanephrine 3 h after oral administration of 300 µg of clonidine). Other pharmacologic tests, such as the phentolamine test and the glucagon provocation test, are of relatively low sensitivity and are not recommended.

**Diagnostic Imaging** A variety of methods have been used to localize pheochromocytomas and paragangliomas (Table 387-2, **Figs. 387-2, 387-3, and 387-4**). CT and MRI are similar in sensitivity and should be performed with contrast. T2-weighted MRI with gadolinium contrast is optimal for detecting pheochromocytomas and is somewhat better than CT for imaging extra-adrenal pheochromocytomas and paragangliomas. About 5% of adrenal incidentalomas, which usually are detected by CT or MRI, prove to be pheochromocytomas upon endocrinologic evaluation, but the presence of pheochromocytomas is unlikely if unenhanced CT reveals an attenuation of <10 HU.

Tumors also can be localized by procedures using radioactive tracers, including <sup>131</sup>I- or <sup>123</sup>I-metaiodobenzylguanidine (MIBG) scintigraphy, <sup>111</sup>In-somatostatin analogue scintigraphy, <sup>18</sup>F-DOPA positron emission tomography (PET), <sup>68</sup>Ga-DOTATATE PET, or <sup>18</sup>F-fluorodeoxyglucose

**TABLE 387-2** Biochemical and Imaging Methods Used for Diagnosis of Pheochromocytoma and Paraganglioma

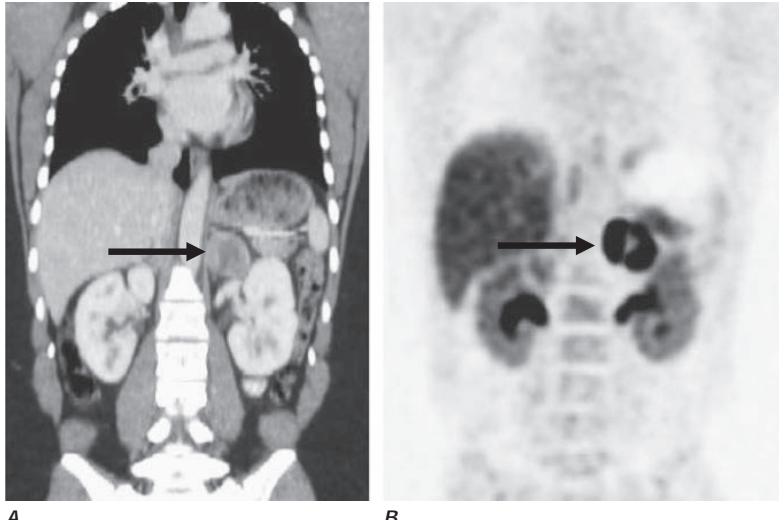
DIAGNOSTIC METHOD	SENSITIVITY	SPECIFICITY
24-h urinary tests		
Catecholamines	+++	+++
Fractionated metanephrenes	++++	++
Total metanephrenes	+++	++++
Plasma tests		
Catecholamines	+++	++
Free metanephrenes	++++	+++
Imaging		
CT	++++	+++
MRI	++++	+++
MIBG scintigraphy	++	++++
Somatostatin receptor scintigraphy <sup>a</sup>	++	++
<sup>18</sup> Fluoro-DOPA PET/CT	++++	++++
<sup>68</sup> Gallium-DOTATOC or DOTATATE PET/CT	++++	++++

<sup>a</sup>Values are particularly high in head and neck paragangliomas.

**Abbreviations:** MIBG, metaiodobenzylguanidine; PET/CT, positron emission tomography plus CT. For the biochemical tests, the ratings correspond globally to sensitivity and specificity rates as follows: ++, <85%; +++, 85–95%; and +++, >95%.

**TABLE 387-1** Clinical Features Associated with Pheochromocytoma, Listed by Frequency of Occurrence

- |  |  |
|--|--|
| 1. Headaches                             | 10. Weight loss                                    |
| 2. Profuse sweating                      | 11. Paradoxical response to antihypertensive drugs |
| 3. Palpitations and tachycardia          | 12. Polyuria and polydipsia                        |
| 4. Hypertension, sustained or paroxysmal | 13. Constipation                                   |
| 5. Anxiety and panic attacks             | 14. Orthostatic hypotension                        |
| 6. Pallor                                | 15. Dilated cardiomyopathy                         |
| 7. Nausea                                | 16. Erythrocytosis                                 |
| 8. Abdominal pain                        | 17. Elevated blood sugar                           |
| 9. Weakness                              | 18. Hypercalcemia                                  |



**FIGURE 387-2** Typical pheochromocytoma (adrenal unilateral). **A.** MRI. **B.**  $^{18}\text{F}$ -DOPA positron emission tomography (PET). Tumor marked by arrows. (Part A was provided courtesy of Dr. Tobias Krauss, Freiburg. Part B was provided courtesy of Dr. Juri Ruf, Freiburg.)

(FDG) PET (Fig. 387-2B and 387-4A and B). For PET-CT with both  $^{68}\text{Ga}$ -DOTATATE and  $^{18}\text{F}$ -DOPA, the sensitivity and specificity are very high (>95%). These agents are particularly useful in the documentation of hereditary syndromes but also in metastatic pheochromocytoma, because uptake is exhibited also in paragangliomas and metastases.

**Pathology** Pheochromocytomas and paragangliomas are found at the classical sites of the adrenal medulla (Fig. 387-2) and paraganglia (Fig. 387-3). Histologically, the tumors often show a characteristic “Zellballen” pattern, consisting of nests of neuroendocrine chief cells

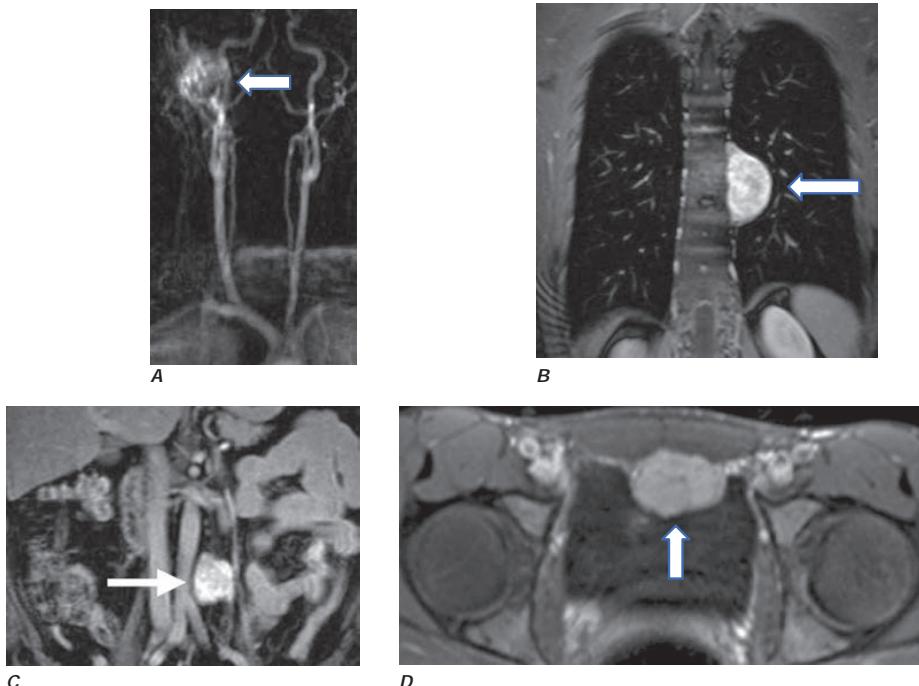
with peripheral glial-like sustentacular cells. However, a broad spectrum of architectural and cyto-logic features can be seen. Immunohistochemistry is positive for chromogranin and synaptophysin in the chief cells and S-100 in the sustentacular cells (Fig. 387-5A–D). Increasingly, staining with antibodies against the proteins encoded by susceptibility genes for hereditary pheochromocytomas, such as SDHB, is used to histologically demonstrate defects of these proteins, thereby making germline mutations more likely (Fig. 387-5E and F).

**Differential Diagnosis** When the possibility of a pheochromocytoma is being entertained, other disorders to consider include essential hypertension, anxiety attacks, use of cocaine or amphetamines, mastocytosis or carcinoid syndrome (usually without hypertension), intracranial lesions, clonidine withdrawal, autonomic epilepsy, and factitious crises (usually from use of sympathomimetic amines). When an asymptomatic adrenal mass is identified, likely diagnoses other than pheochromocytoma include a nonfunctioning adrenal adenoma, an aldosteronomia, and a cortisol-producing adenoma (Cushing’s syndrome).

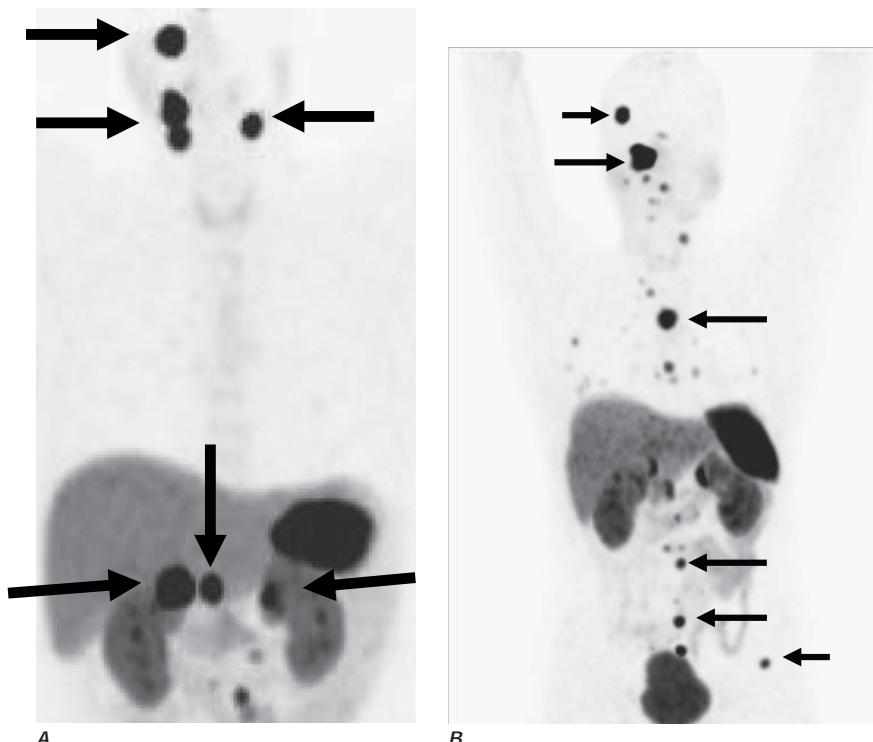
## TREATMENT

### Pheochromocytoma

Complete tumor removal, the ultimate therapeutic goal, can be achieved by partial or total adrenalectomy. It is important to preserve the normal adrenal cortex in order to prevent Addison’s disease, particularly in hereditary disorders in which bilateral pheochromocytomas are most likely. Preoperative preparation of the patient has to be considered, and blood pressure should be consistently <160/90 mmHg. Classically, blood pressure has been



**FIGURE 387-3** Paragangliomas (extra-adrenal pheochromocytomas). **A.** Carotid body tumor. **B.** Thoracic tumor. **C.** Paraaortal tumor. **D.** Pelvic tumor at the anterior wall of the urinary bladder. Tumors marked by arrows. (Part A was provided courtesy of Dr. Carsten Boedeker, Stralsund. Parts B and D were provided courtesy of Dr. Tobias Krauss, Freiburg. Part C was provided courtesy of Dr. Martin Walz, Essen.)



**FIGURE 387-4** Multiple and metastatic pheochromocytoma. **A.** Paraganglioma syndrome. A patient with the *SDHD* W5X mutation and PGL1  $^{68}\text{Ga}$ -DOTATATE positron emission tomography (PET) demonstrating tumor uptake in the right jugular glomus, the right and left carotid body, both adrenal glands, and an interaortocaval paraganglion (arrows). Note the physiologic accumulation of the radiopharmaceutical agent in the kidneys and the liver. **B.**  $^{18}\text{F}$ -DOPA PET of a patient with metastatic pheochromocytoma. Several metastases marked by arrows. (Parts A and B were provided courtesy of Dr. Juri Ruf, Freiburg.)

controlled by  $\alpha$ -adrenergic blockers (oral phenoxybenzamine, 0.5–4 mg/kg of body weight). Because patients are volume-constricted, liberal salt intake and hydration are necessary to avoid severe orthostasis. Oral prazosin or intravenous phentolamine can be used to manage paroxysms while adequate alpha blockade is awaited. Beta blockers (e.g., 10 mg of propranolol three or four times per day) can then be added. Other antihypertensives, such as calcium channel blockers or angiotensin-converting enzyme inhibitors, have also been used effectively.

Surgery should be performed by teams of surgeons and anesthesiologists with experience in the management of pheochromocytomas. Blood pressure can be labile during surgery, particularly at the outset of intubation or when the tumor is manipulated. Nitroprusside infusion is useful for intraoperative hypertensive crises, and hypotension usually responds to volume infusion. The latter side effect can, however, be avoided in normotensive pheochromocytoma patients by having only standby intraoperative nitroprusside, which has been shown to be safe and avoids postoperative hypotension often caused by alpha blockers; the long-lasting guideline for obligatory preoperative treatment with alpha blockers is under discussion.

Minimally invasive techniques (laparoscopy or retroperitoneoscopy) have become the standard approaches in pheochromocytoma surgery. They are associated with fewer complications, a faster recovery, and optimal cosmetic results. Extra-adrenal abdominal and most thoracic pheochromocytomas can also be removed endoscopically. Postoperatively, catecholamine normalization should be documented. An adrenocorticotrophic hormone (ACTH) test should be used to exclude cortisol deficiency when bilateral adrenal cortex-sparing surgery has been performed.

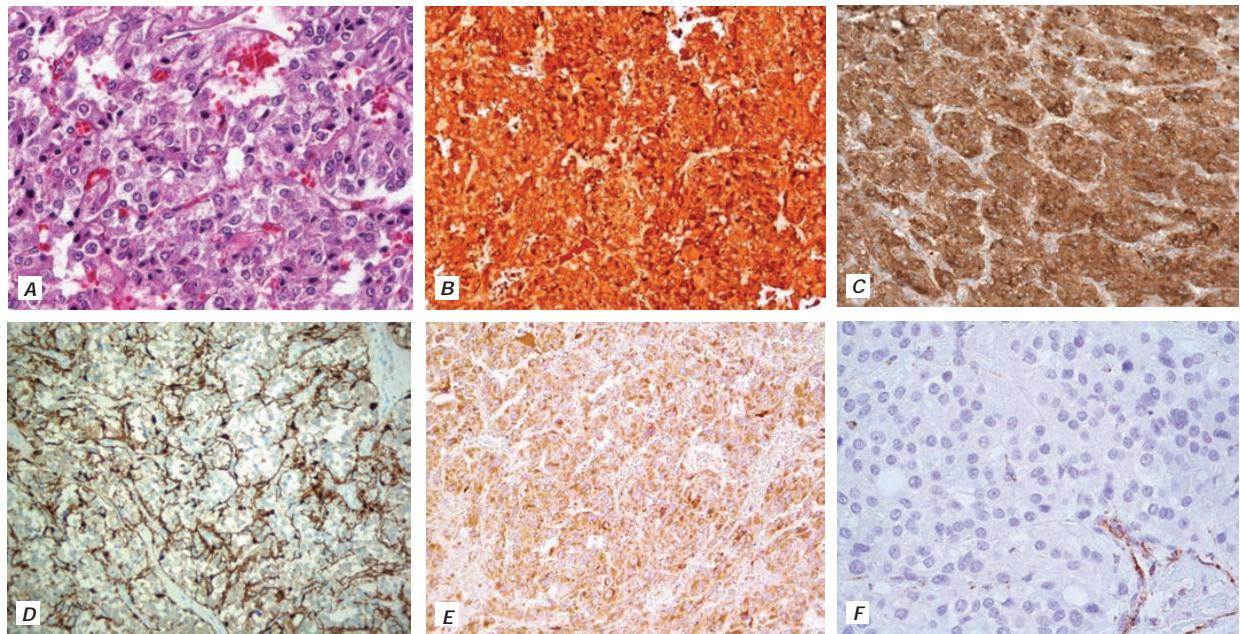
Head and neck paragangliomas are a challenge for surgeons, since damage of adjacent tissue, mainly vessels or cranial nerves, is a frequent permanent side effect. Careful consideration of best

management is important, and radiotherapy may be an alternative, especially for large head and neck paragangliomas.

### METASTATIC PHEOCHROMOCYTOMA

About 5–10% of pheochromocytomas and paragangliomas are metastatic. The diagnosis of malignant pheochromocytoma is problematic. The typical histologic criteria of cellular atypia, presence of mitoses, and invasion of vessels or adjacent tissues are insufficient for the diagnosis of malignancy in pheochromocytoma. Thus, the term *malignant pheochromocytoma* has been replaced by *metastatic pheochromocytoma* as suggested by the WHO and is restricted to tumors with lymph node or distant metastases, the latter most commonly found by nuclear medicine imaging in lungs, bone, or liver locations suggesting a vascular pathway of spread (Fig. 387-4B). Because hereditary syndromes are associated with multifocal tumor sites, these features should be anticipated in patients with germline mutations, especially of *RET*, *VHL*, *SDHD*, or *SDHB*. However, distant metastases also occur in these syndromes, especially in carriers of *SDHB* mutations.

Treatment of metastatic pheochromocytoma or paraganglioma is challenging. Options include tumor mass reduction, alpha blockers for symptoms, chemotherapy, nuclear medicine radiotherapy, and stereotactic radiation. The first-line choice is nuclear medicine therapy for scintigraphically documented metastases, preferably with  $^{131}\text{I}$ -MIBG in 100–300 mCi doses over 3–6 cycles. Other options for radionuclide treatment are somatostatin receptor ligands, e.g., DOTATOC labeled with yttrium-90 or lutetium-177, both for palliative outcomes. Averbuch's chemotherapy protocol includes dacarbazine (600 mg/m<sup>2</sup> on days 1 and 2), cyclophosphamide (750 mg/m<sup>2</sup> on day 1), and vincristine (1.4 mg/m<sup>2</sup> on day 1), all repeated every 21 days for 3–6 cycles. Palliation (stable disease to shrinkage) is achieved in about one-half of patients. Due to increasing insights in the genetics of pheochromocytoma and their molecular pathways, new targeted chemotherapeutic options such as



**FIGURE 387-5** Histology and immunohistochemistry of pheochromocytoma. **A.** Hematoxylin and eosin, **B.** chromogranin, **C.** synaptophysin, **C** and **B** stain chief cells; **D.** S-100 stains sustentacular cells. **E.**, **F.** Immunohistochemistry with SDHB antibody: positive staining (granular cytoplasmic staining) indicates intact SDHB (**E**), whereas negative staining (endothelial cells positive as internal control) (**F**) indicates structurally changed or absent SDHB due to a germline mutation in the *SDHB* gene, which was confirmed by molecular genetic analysis of a blood sample. (Parts A–D and F were used with permission from Dr Helena Leijon, Helsinki. Part E was provided courtesy of Dr. Kurt Werner Schmid, Essen.)

sunitinib and temozolomide/thalidomide are under development. The prognosis of metastatic pheochromocytoma or paraganglioma is variable, with 5-year survival rates of 30–60%.

### PHEOCHROMOCYTOMA IN PREGNANCY

Pheochromocytomas occasionally are diagnosed in pregnancy and can be very challenging to manage. Endoscopic removal, preferably in the fourth to sixth month of gestation, is possible and can be followed by uneventful childbirth. Regular screening in families with inherited pheochromocytomas provides an opportunity to identify and remove such tumors in women of reproductive age.

### PHEOCHROMOCYTOMA-ASSOCIATED SYNDROMES

About 25–33% of patients with a pheochromocytoma or paraganglioma have an inherited syndrome. At diagnosis, patients with inherited syndromes are a mean of ~15 years younger than patients with sporadic tumors.

The best-known pheochromocytoma-associated syndrome is the autosomal dominant disorder MEN 2 (Chap. 388). Both types of MEN 2 (2A and 2B) are caused by mutations in *RET*, which encodes a tyrosine kinase. The locations of *RET* mutations correlate with the severity of disease and the type of MEN 2 (Chap. 388). MEN 2A is characterized by medullary thyroid carcinoma (MTC), pheochromocytoma, and hyperparathyroidism; MEN 2B also includes MTC and pheochromocytoma as well as multiple mucosal neuromas, marfanoid habitus, and other developmental disorders, although it typically lacks hyperparathyroidism. MTC is found in virtually all patients with MEN 2, but pheochromocytoma occurs in only ~50% of these patients. Nearly all pheochromocytomas in MEN 2 are benign and located in the adrenals, often bilaterally. Pheochromocytoma may be symptomatic before MTC. Prophylactic thyroidectomy is being performed in many carriers of *RET* mutations; pheochromocytomas should be excluded before any surgery in these patients.

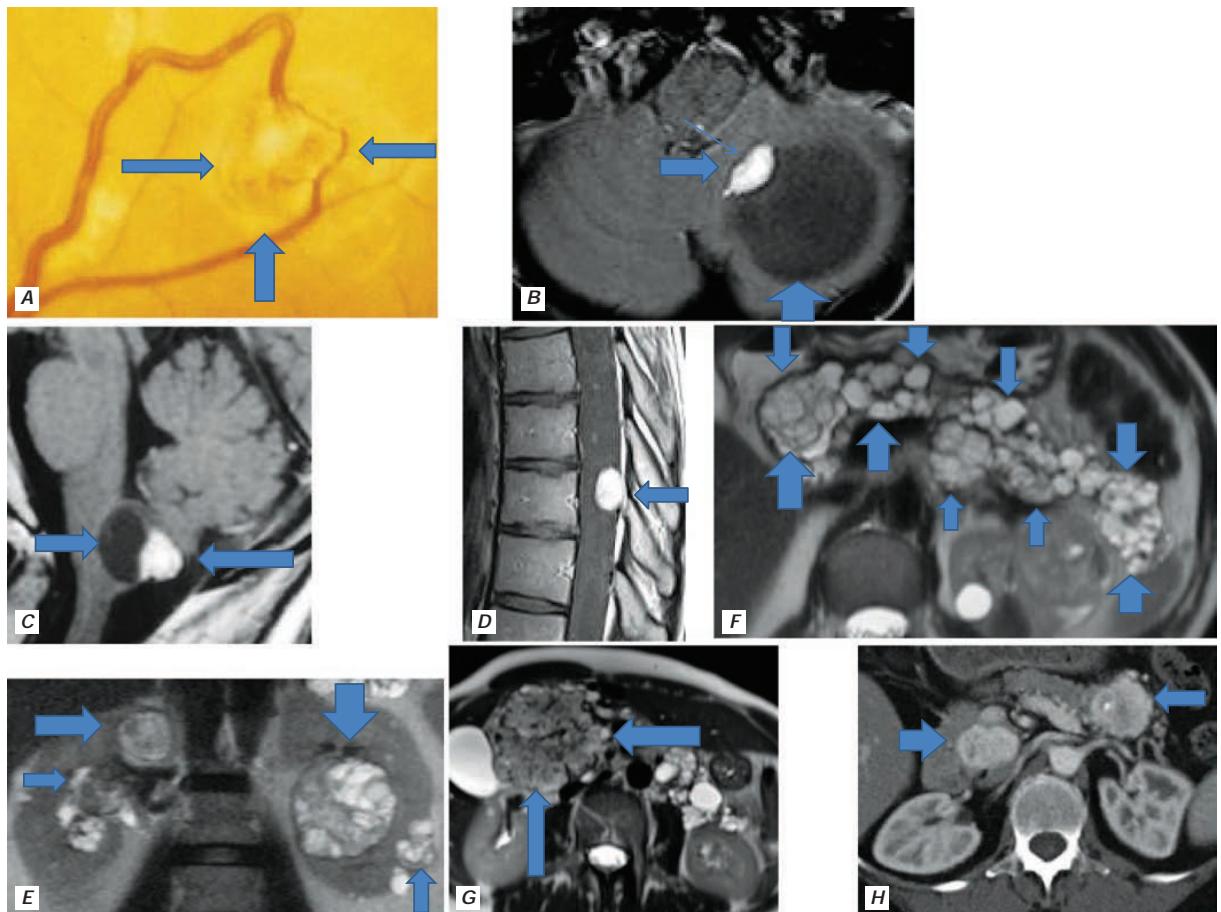
VHL is an autosomal dominant disorder that predisposes to retinal and cerebellar hemangioblastomas, which also occur in the brainstem and spinal cord (Fig. 387-6). Other important features of VHL are clear cell renal carcinomas, pancreatic neuroendocrine tumors,

endolymphatic sac tumors of the inner ear, cystadenomas of the epididymis and broad ligament, and multiple pancreatic or renal cysts. Although the *VHL* gene can be inactivated by all types of mutations, patients with pheochromocytoma predominantly have missense mutations. About 20–30% of patients with VHL have pheochromocytomas, but in some families, the incidence can reach 90%. The recognition of pheochromocytoma as a VHL-associated feature provides an opportunity to diagnose retinal, central nervous system, renal, and pancreatic tumors at a stage when effective treatment may still be possible.

NF1 was the first described pheochromocytoma-associated syndrome. The *NF1* gene functions as a tumor suppressor by regulating the Ras signaling cascade. Classic features of neurofibromatosis include multiple neurofibromas, café au lait spots, axillary freckling of the skin, and Lisch nodules of the iris. Pheochromocytomas occur in only ~1% of these patients and are located predominantly in the adrenals. Metastatic pheochromocytoma is not uncommon in NF1.

The *paraganglioma syndromes* (PGLs) have been classified by genetic analyses of families with head and neck paragangliomas. The susceptibility genes encode subunits of the enzyme SDH, a component in the Krebs cycle and the mitochondrial electron transport chain. SDH is formed by four subunits (A–D). Mutations of *SDHA* (PGL5), *SDHB* (PGL4), *SDHC* (PGL3), *SDHD* (PGL1), and *SDHAF2* (PGL2) predispose to the PGLs. The transmission of the disease in carriers of *SDHA*, *SDHB*, and *SDHC* germline mutations is autosomal dominant. In contrast, in virtually all *SDHD* and *SDHAF2* families, only the progeny of affected fathers develop tumors if they inherit the mutation. PGL1 is most common, followed by PGL4; PGL2, PGL3, and PGL5 are rare. Adrenal, extra-adrenal abdominal, and thoracic pheochromocytomas, which are components of PGL1, PGL4, and PGL5, are rare in PGL3 and absent in PGL2 (Fig. 387-4A). About one-third of patients with PGL4 develop metastases, which is the highest rate in pheochromocytoma-associated syndromes. Other syndromes with metastatic pheochromocytomas are mainly VHL, NF1, and PGL1.

Other familial pheochromocytoma has been attributed to hereditary, mainly adrenal tumors in patients with germline mutations in the genes *TMEM127* and *MAX*. Transmission is also autosomal dominant, and mutations of *MAX*, like those of *SDHD*, cause tumors only if inherited from the father.



**FIGURE 387-6 von Hippel-Lindau disease.** Tumors and cysts marked by arrows. **A.** Retinal angioma (arrows with a pair of feeding vessels). All subsequent panels show findings on MRI. **B–D.** Hemangioblastomas of the cerebellum (large cyst and a solid mural tumor) (**B**) in brainstem (in part cystic) (**C**) and spinal cord (thoracic) (**D**). **E.** Bilateral renal clear cell carcinomas with two tumors on each side. **F.** Multiple pancreatic cysts. **G.** Microcystic serous pancreatic cystadenoma (with multiple tiny spaces). **H.** Two pancreatic islet cell tumors. (Part A was provided courtesy of Dr. Dieter Schmidt. Part B was provided courtesy of Dr. Christian Taschner, Freiburg. Part C was provided courtesy of Dr. Sven Glaesker, Brussels. Part D was used with permission from Dr. Jan-Helge Klingler, Freiburg. Part E was provided courtesy of Dr. Cordula Jilg, Freiburg. Parts F–H were provided courtesy of Dr Tobias Krauss, Freiburg.)

### GENETIC SCREENING OF PATIENTS WITH PHEOCHROMOCYTOMA OR PARAGANGLIOMA

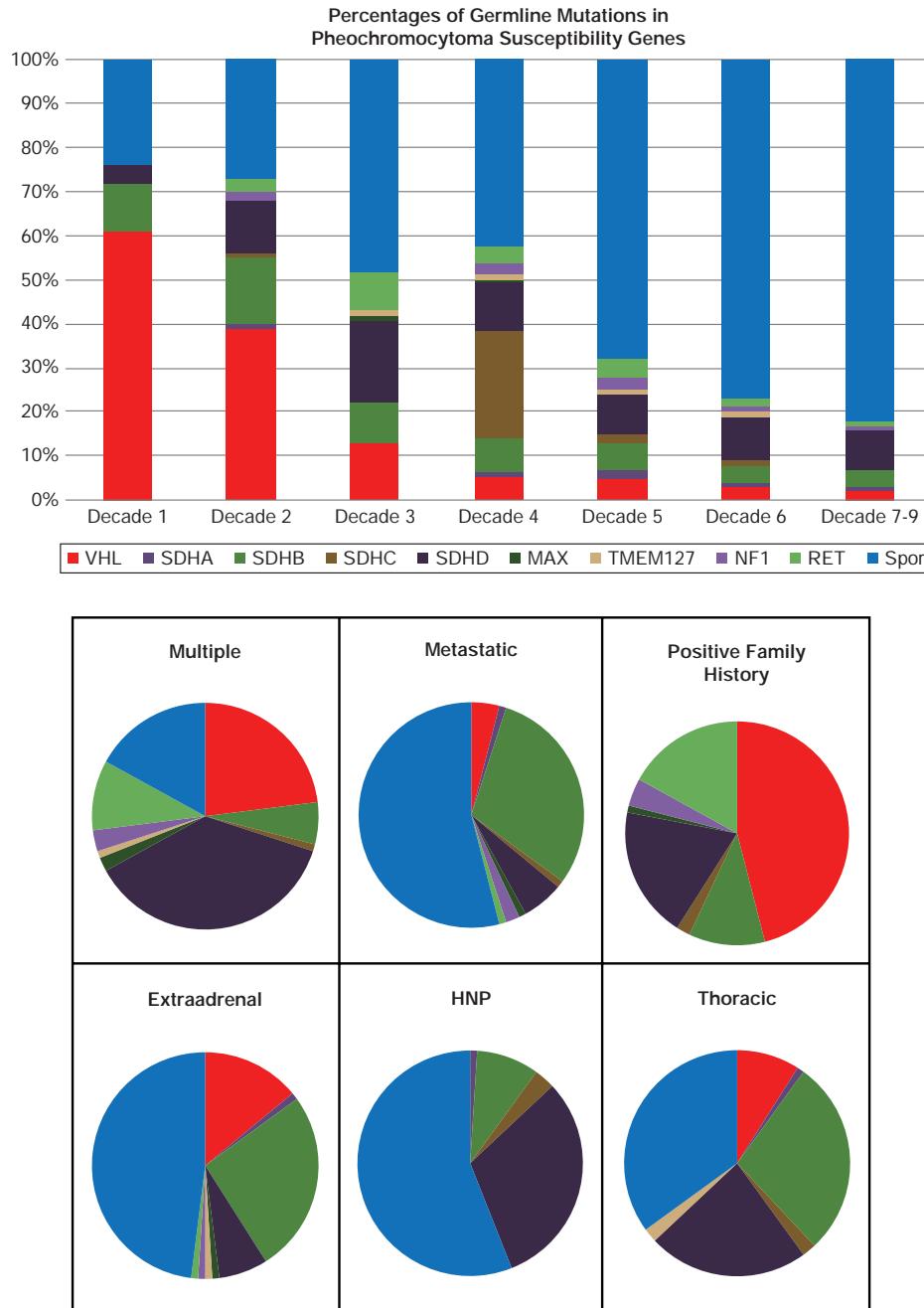
Effective preventive medicine for pheochromocytoma and pheochromocytoma-associated diseases requires management according to identified germline mutations in susceptibility genes. In addition to family history, general features suggesting an inherited syndrome

include young age, multifocal tumors, extra-adrenal tumors, and metastatic tumors (Table 387-3 and Fig. 387-7). Because of the relatively high prevalence of familial syndromes among patients who present with pheochromocytoma or paraganglioma, it is useful to identify germline mutations even in patients without a known family history. A first step is to search for clinical features of inherited syndromes and

**TABLE 387-3 Patterns of Occurrence in Inherited Pheochromocytoma and Paraganglioma-Associated Syndromes**

MUTATED GENE	ADRENAL TUMORS	HEAD AND NECK TUMORS	EXTRA-ADRENAL RETROPERITONEAL OR PELVIC TUMORS	THORACIC TUMORS	MULTIPLE TUMORS	BILATERAL ADRENAL TUMORS	METASTATIC TUMORS	FAMILY HISTORY IN PROBANDS FOR COMPONENTS OF THE GIVEN SYNDROME
MAX	++++	<X	+	<X	++++	+++	++	+++
NF1	++++	<+	+	<+	+	++	+	++
RET	++++	<+	<+	<+	+++	+++	<+	+
SDHA	++	++++	++	+	+	<+	+	+
SDHB	+++	+++	+++	+	++	<+	+++	++
SDHC	<+	+++++	<+	+	+	<+	<+	++
SDHD	++	+++++	+	+	+++	<+	+	+++
VHL	++++	<+	+	+	+++	+++	+	++++
TMEM127	++++	+	+	+	++	++	<+	+

Note: Frequencies in percentage (<+: 0–4%; +: 5–19%; ++: 20–39%; +++: 40–59%; ++++: 60–79%; +++++: 80–100%) of clinical characteristics of pheochromocytomas/paragangliomas of patients with germline mutations of the genes MAX, NF1, RET, SDHA, SDHB, SDHC, SDHD, VHL, and TMEM127; for other genes, the data are too limited to include in this summary.



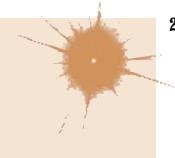
**FIGURE 387-7 Mutation distribution in the *VHL*, *RET*, *SDHB*, *SDHC*, *SDHD*, and *NF1* genes in 4156 patients with pheochromocytomas and paragangliomas from the European-American Pheochromocytoma-Paraganglioma Registry based in Freiburg, Germany, Padova, Italy, and Rochester, Minnesota, and updated as of December 20, 2020.**

**A.** Correlation with age. The bars depict the frequency of sporadic (spor) or various inherited forms of pheochromocytoma in different age groups. The inherited disorders are much more common among younger individuals presenting with pheochromocytoma. **B.** Percentages of mutated genes in hereditary pheochromocytomas and paragangliomas. **C-G.** Germline mutations according to multiple (**C**), metastatic (**D**), hereditary (**E**), extra-adrenal retroperitoneal (**F**), head and neck paragangliomas (HNP) (**G**), and thoracic (**H**). (Data from the Freiburg International Pheochromocytoma and Paraganglioma Registry, 2017. Figures courtesy of Dr. Charis Eng, Cleveland; Dr. Irina Bancos, Rochester; Dr. Birke Bausch, Freiburg; Dr. Giuseppe Oopher and Dr. Francesca Schiavi, Padova.)

to obtain an in-depth, multigenerational family history. Each of these syndromes exhibits autosomal dominant transmission with variable penetrance, but a proband with a mother affected by paraganglial tumors is not predisposed to *PLG1* and *PGL2* (*SDHD* and *SDHAF2* mutation carrier). Cutaneous neurofibromas, café au lait spots, and axillary freckling suggest neurofibromatosis. Germline mutations

in *NF1* have nearly never been reported in patients with sporadic pheochromocytomas. Thus, *NF1* testing is not needed in the absence of other clinical features of neurofibromatosis. A personal or family history of MTC or an elevation of serum calcitonin strongly suggests MEN 2 and should prompt testing for *RET* mutations. A history of visual impairment or tumors of the cerebellum, brain stem, spinal

## 388

Multiple Endocrine  
Neoplasia Syndromes

R. V. Thakker

cord or the kidney suggests the possibility of VHL. A personal and/or family history of head and neck paraganglioma suggests PGL1 or PGL4.

A single adrenal pheochromocytoma in a patient with an otherwise unremarkable history may still be associated with mutations of *VHL*, *RET*, *SDHB*, or *SDHD* (in decreasing order of frequency). Two-thirds of extra-adrenal tumors are associated with one of these syndromes, and multifocal tumors occur with decreasing frequency in carriers of *RET*, *SDHD*, *VHL*, *SDHB*, and *MAX* mutations. About 30% of head and neck paragangliomas are associated with germline mutations of one of the SDH subunit genes (most often *SDHD*) and are rare in carriers of *VHL*, *RET*, *MAX*, and *TMEM127* mutations. Immunohistochemistry is helpful in the preselection of hereditary pheochromocytoma. Negative immunostaining with antibodies to *SDHB* (Fig. 387-5*F*), *TMEM127*, and *MAX* may predict mutations of the *SDHx* (*PGL1-5*), *TMEM127*, and *MAX* genes, respectively.

Whole genome sequence analysis is increasingly replacing targeted Sanger sequencing. It is now possible to search for germline mutations in a set of genes, such that all susceptibility genes for pheochromocytoma-associated syndromes could be analyzed in one procedure. Of note, sequencing protocols may not detect large deletions of one or more exons.

Once the underlying syndrome is diagnosed, the benefit of genetic testing can be extended to relatives. For this purpose, it is necessary to identify the germline mutation in the proband and, after genetic counseling, to perform DNA sequence analyses of the responsible gene in relatives to determine whether they are affected. Other family members may benefit when individuals who carry a germline mutation are biochemically screened for paragangli tumors.

Asymptomatic paragangli tumors, now often detected in patients with hereditary tumors and their relatives, are challenging to manage. Watchful waiting strategies have been introduced. Head and neck paragangliomas—mainly carotid body, jugular, and vagal tumors—are increasingly treated by radiation, since surgery is frequently associated with permanent palsy of cranial nerves II, VII, IX, X, XI, and XII. Nevertheless, tympanic paragangliomas are symptomatic early, and most of these tumors can easily be resected, with subsequent improvement of hearing and alleviation of tinnitus.

## FURTHER READING

- Bancos I et al: Maternal and fetal outcomes in pheochromocytoma and pregnancy: A multi-center retrospective cohort study and systematic review of literature. *Lancet Diabetes Endocrinol* 9:13, 2021.
- Berends AMA et al: Incidence of pheochromocytoma and sympathetic paraganglioma in the Netherlands: A nationwide study and systematic review. *Eur Intern Med* 51:68, 2018.
- Canu L et al: CT characteristics of pheochromocytoma: Relevance for the evaluation of adrenal incidentaloma. *J Clin Endocrinol Metab* 104:312, 2019.
- Groeben H et al: International multicentre review of perioperative management and outcome for catecholamine-producing tumours. *Br J Surg* 107:e170, 2020.
- Hamidi O et al: Malignant pheochromocytoma and paraganglioma: 272 patients over 55 years. *J Clin Endocrinol Metab* 10:3296, 2017.
- Lenders JW et al: Genetics, diagnosis, management and future directions of research of phaeochromocytoma and paraganglioma: A position statement and consensus of the Working Group on Endocrine Hypertension of the European Society of Hypertension. *J Hypertens* 38:1443, 2020.
- Neumann HPH et al: Comparison of pheochromocytoma-specific morbidity and mortality among adults with bilateral pheochromocytomas undergoing total adrenalectomy vs cortical-sparing adrenalectomy. *JAMA Netw Open* 2:e198898, 2019.
- Taieb D et al: European Association of Nuclear Medicine Practice Guideline/Society of Nuclear Medicine and Molecular Imaging Procedure Standard 2019 for radionuclide imaging of phaeochromocytoma and paraganglioma. *Eur J Nucl Med Mol Imaging* 46:2112, 2019.

Multiple endocrine neoplasia (MEN) is characterized by a predilection for tumors involving two or more endocrine glands. Four major forms of MEN are recognized and referred to as MEN types 1–4 (MEN 1–4) (Table 388-1). Each type of MEN is inherited as an autosomal dominant syndrome or may occur sporadically, that is, without a family history. However, this distinction between familial and sporadic forms is often difficult because family members with the disease may have died before symptoms developed. In addition to MEN 1–4, at least six other syndromes are associated with multiple endocrine and other organ neoplasias (MEONs) (Table 388-2). These MEONs include the hyperparathyroidism-jaw tumor (HPT-JT) syndrome, Carney complex, von Hippel-Lindau disease (Chap. 387), neurofibromatosis type 1 (Chap. 90), Cowden's syndrome (CWS), and McCune-Albright syndrome (MAS) (Chap. 412); all of these are inherited as autosomal dominant disorders, except for MAS, which is caused by mosaic expression of a postzygotic somatic cell mutation (Table 388-2).

A diagnosis of a MEN or MEON syndrome may be established in an individual by one of three criteria: (1) clinical features (two or more of the associated tumors [or lesions] in an individual); (2) familial pattern (one of the associated tumors [or lesions] in a first-degree relative of a patient with a clinical diagnosis of the syndrome); and (3) genetic analysis (a germline mutation in the associated gene in an individual, who may be clinically affected or asymptomatic). Mutational analysis in MEN and MEON syndromes is helpful in clinical practice to (1) confirm the clinical diagnosis; (2) identify family members who harbor the mutation and require screening for relevant tumor detection and early/appropriate treatment; and (3) identify the ~50% of family members who do not harbor the germline mutation and can, therefore, be alleviated of the anxiety of developing associated tumors. This latter aspect also helps to reduce health care costs by reducing the need for unnecessary biochemical and radiologic investigations.

## MULTIPLE ENDOCRINE NEOPLASIA TYPE 1

**Clinical Manifestations** MEN type 1 (MEN 1), which is also referred to as Wermer's syndrome, is characterized by the triad of tumors involving the parathyroids, pancreatic islets, and anterior pituitary. In addition, adrenal cortical tumors, carcinoid tumors usually of the foregut, meningiomas, facial angiofibromas, collagenomas, and lipomas may also occur in some patients with MEN 1. Combinations of the affected glands and their pathologic features (e.g., hyperplastic adenomas of the parathyroid glands) may differ in members of the same family and even between identical twins. In addition, a nonfamilial (e.g., sporadic) form occurs in 8–14% of patients with MEN 1, and molecular genetic studies have confirmed the occurrence of de novo mutations of the *MEN1* gene in ~10% of patients with MEN 1. The prevalence of MEN 1 is ~0.25% based on randomly chosen postmortem studies but is 1–18% among patients with primary hyperparathyroidism, 16–38% among patients with pancreatic islet tumors, and <3% among patients with pituitary tumors. The disorder affects all age groups, with a reported age range of 5–81 years, with clinical and biochemical manifestations developing in the vast majority by the fifth decade. The clinical manifestations of MEN 1 are related to the sites of tumors and their hormonal products. In the absence of treatment, endocrine tumors are associated with an earlier mortality in patients with MEN 1, with a 50% probability of death by the age of 50 years. The cause of death is usually a malignant tumor, often from a pancreatic neuroendocrine tumor (NET) or foregut carcinoid. In addition, the treatment outcomes of patients with MEN 1-associated tumors are not as successful as those in patients with non-MEN 1 tumors. This is because MEN 1-associated tumors, with the exception of pituitary

**TABLE 388-1** Multiple Endocrine Neoplasia (MEN) Syndromes

TYPE (CHROMOSOMAL LOCATION)	TUMORS (ESTIMATED PENETRANCE)	GENE AND MOST FREQUENTLY MUTATED CODONS
MEN 1 (11q13)	Parathyroid adenoma (90%) Enteropancreatic tumor (30–70%) <ul style="list-style-type: none"> <li>• Gastrinoma (&gt;50%)</li> <li>• Insulinoma (10–30%)</li> <li>• Nonfunctioning and PPoma (20–55%)</li> <li>• Glucagonoma (&lt;3%)</li> <li>• VIPoma (&lt;1%)</li> </ul> Pituitary adenoma (15–50%) <ul style="list-style-type: none"> <li>• Prolactinoma (60%)</li> <li>• Somatotrophinoma (25%)</li> <li>• Corticotrophinoma (&lt;5%)</li> <li>• Nonfunctioning (&lt;5%)</li> </ul> Associated tumors <ul style="list-style-type: none"> <li>• Adrenal cortical tumor (20–70%)</li> <li>• Pheochromocytoma (&lt;1%)</li> <li>• Bronchopulmonary NET (2%)</li> <li>• Thymic NET (2%)</li> <li>• Gastric NET (10%)</li> <li>• Lipomas (&gt;33%)</li> <li>• Angiofibromas (85%)</li> <li>• Collagenomas (70%)</li> <li>• Meningiomas (8%)</li> </ul>	<i>MEN1</i> 83/84, 4-bp del ( 4%) 119, 3-bp del ( 3%) 209–211, 4-bp del ( 8%) 418, 3-bp del ( 4%) 514–516, del or ins ( 7%) Intron 4 ss ( 10%)
MEN 2 (10 cen-10q11.2)		
MEN 2A	MTC (90%) Pheochromocytoma (>50%) Parathyroid adenoma (10–25%)	<i>RET</i> 634, e.g., Cys → Arg (>85%)
MTC only	MTC (100%)	<i>RET</i> 618, missense (>50%)
MEN 2B (also known as MEN 3)	MTC (>90%) Pheochromocytoma (>50%) Associated abnormalities (40–50%) <ul style="list-style-type: none"> <li>• Mucosal neuromas</li> <li>• Marfanoid habitus</li> <li>• Medullated corneal nerve fibers</li> <li>• Megacolon</li> </ul>	<i>RET</i> 918, Met → Thr (>95%)
MEN 4 (12p13)	Parathyroid adenoma <sup>a</sup> Pituitary adenoma <sup>a</sup> Reproductive organ tumors <sup>a</sup> (e.g., testicular cancer, neuroendocrine cervical carcinoma) ?Adrenal + renal tumors <sup>a</sup>	<i>CDKN1B</i> ; no common mutations identified to date

<sup>a</sup>Insufficient numbers reported to provide prevalence information.

Note: Autosomal dominant inheritance of the MEN syndromes has been established.

Abbreviations: del, deletion; ins, insertion; MTC, medullary thyroid cancer; NET, neuroendocrine tumor; PPoma, pancreatic polypeptide–secreting tumor; VIPoma, vasoactive intestinal polypeptide–secreting tumor.

Source: Adapted with permission from RV Thakker: Multiple endocrine neoplasia—syndromes of the twentieth century. *J Clin Endocrinol Metab* 83:2617, 1998.

NETs, are usually multiple, making it difficult to achieve a successful surgical cure. Occult metastatic disease is also more prevalent in MEN 1, and the tumors may be larger, more aggressive, and resistant to treatment.

**Parathyroid Tumors (See also Chap. 410)** Primary hyperparathyroidism occurs in ~90% of patients and is the most common

**TABLE 388-2** Multiple Endocrine and Other Organ Neoplasia (MEON) Syndromes

DISEASE <sup>a</sup>	GENE PRODUCT	CHROMOSOMAL LOCATION
Hyperparathyroidism-jaw tumor (HPT-JT)	Parafibromin	1q31.2
Carney complex		
CNC1	PRAKAR1A	17q24.2
CNC2	? <sup>b</sup>	2p16
von Hippel-Lindau disease (VHL)	pVHL (elongin)	3p25
Neurofibromatosis type 1 (NF1)	Neurofibromin	17q11.2
Cowden's syndrome (CWS)		
CWS1	PTEN	10q23.31
CWS2	SDHB	1p36.13
CWS3	SDHD	11q23.1
CWS4	KLLN	10q23.31
CWS5	PIK3CA	3q26.32
CWS6	AKT1	14q32.33
CWS7	SEC23B	20p11.23
McCune-Albright syndrome (MAS)	G <sub>α</sub>	20q13.32

<sup>a</sup>The inheritance for these disorders is autosomal dominant, except MAS, which is due to mosaicism that results from the postzygotic somatic cell mutation of the *GNAS1* gene, encoding G<sub>α</sub>. <sup>b</sup>?, unknown.

feature of MEN 1. Patients may have asymptomatic hypercalcemia or vague symptoms associated with hypercalcemia (e.g., polyuria, polydipsia, constipation, malaise, or dyspepsia). Nephrolithiasis and osteitis fibrosa cystica (less commonly) may also occur. Biochemical investigations reveal hypercalcemia, usually in association with elevated circulating parathyroid hormone (PTH) (Table 388-3). The hypercalcemia is usually mild, and severe hypercalcemia or parathyroid cancer is a rare occurrence. Additional differences in the primary hyperparathyroidism of patients with MEN 1, as opposed to those without MEN 1, include an earlier age at onset (20–25 vs 55 years) and an equal male-to-female ratio (1:1 vs 1:3). Preoperative imaging (e.g., neck ultrasound with <sup>99m</sup>Tc-sestamibi parathyroid scintigraphy) is of limited benefit because all parathyroid glands may be affected, and neck exploration may be required irrespective of preoperative localization studies.

## TREATMENT

### Parathyroid Tumors

Surgical removal of the abnormally overactive parathyroids in patients with MEN 1 is the definitive treatment. However, it is controversial whether to perform subtotal (e.g., removal of 3.5 glands) or total parathyroidectomy with or without autotransplantation of parathyroid tissue in the forearm, and whether surgery should be performed at an early or late stage. Minimally invasive parathyroidectomy is not recommended because all four parathyroid glands are usually affected with multiple adenomas or hyperplasia. Surgical experience should be taken into account given the variability in pathology in MEN 1. Calcimimetics (e.g., cinacalcet), which act via the calcium-sensing receptor, have been used to treat primary hyperparathyroidism in some patients when surgery is unsuccessful or contraindicated.

**Pancreatic Tumors (See also Chap. 84)** The incidence of pancreatic islet cell tumors, which are NETs, in patients with MEN 1 ranges from 30 to 80% in different series. Most of these tumors (Table 388-1) produce excessive amounts of hormone (e.g., gastrin, insulin, glucagon, vasoactive intestinal polypeptide [VIP]) and are associated with distinct clinical syndromes, although some are nonfunctioning or

**TABLE 388-3 Biochemical and Radiologic Screening in Multiple Endocrine Neoplasia Type 1**

TUMOR	AGE TO BEGIN (YEARS)	BIOCHEMICAL TEST (PLASMA OR SERUM) ANNUALLY	IMAGING TEST (TIME INTERVAL)
Parathyroid	8	Calcium, PTH	None
Pancreatic NETs			
Gastrinoma	20	Gastrin ( $\pm$ gastric pH)	None
Insulinoma	5	Fasting glucose, insulin	None
Other pancreatic NET	<10	Chromogranin A; pancreatic polypeptide, glucagon, vasoactive intestinal peptide	MRI, CT, or EUS (annually)
Anterior pituitary	5	Prolactin, IGF-I	MRI (every 3 years)
Adrenal	<10	None unless symptoms or signs of functioning tumor and/or tumor >1 cm identified on imaging	MRI or CT (annually with pancreatic imaging)
Thymic and bronchial carcinoid	15	None	CT or MRI (every 1–2 years)

Abbreviations: CT, computed tomography; EUS, endoscopic ultrasound; IGF-I, insulin-like growth factor I; MRI, magnetic resonance imaging; PTH, parathyroid hormone.

Source: Data from PJ Newey, RV Thakker: Role of multiple endocrine neoplasia type 1 mutational analysis in clinical practice. *Endocr Pract* 17, 2011 and RV Thakker: Multiple endocrine neoplasia type 1 (MEN1). *Translational Endocrinology and Metabolism*, Vol 2. Chevy Chase, MD: The Endocrine Society; 2011.

nonsecretory. These pancreatic islet cell tumors have an earlier age at onset in patients with MEN 1 than in patients without MEN 1.

**Gastrinoma** Gastrin-secreting tumors (gastrinomas) are associated with marked gastric acid production and recurrent peptic ulcerations, a combination referred to as Zollinger-Ellison syndrome. Gastrinomas occur more often in patients with MEN 1 who are aged >30 years. Recurrent severe multiple peptic ulcers, which may perforate, and cachexia are major contributors to the high mortality. Patients with Zollinger-Ellison syndrome may also suffer from diarrhea and steatorrhea. The diagnosis is established by demonstration of an elevated fasting serum gastrin concentration in association with increased basal gastric acid secretion (Table 388-3). However, the diagnosis of Zollinger-Ellison syndrome may be difficult in hypercalcemic MEN 1 patients because hypercalcemia can also cause hypergastrinemia. Ultrasonography, endoscopic ultrasonography, computed tomography (CT), nuclear magnetic resonance imaging (MRI), selective abdominal angiography, venous sampling, and somatostatin receptor scintigraphy (SRS) are helpful in localizing the tumor prior to surgery. Gastrinomas represent >50% of all pancreatic NETs in patients with MEN 1, and ~20% of patients with gastrinomas will be found to have MEN 1. Gastrinomas, which may also occur in the duodenal mucosa, are the major cause of morbidity and mortality in patients with MEN 1.

## TREATMENT

### Gastrinoma

Medical treatment of patients with MEN 1 and Zollinger-Ellison syndrome is directed toward reducing basal acid output to <10 mmol/L. Parietal cell H<sup>+</sup>-K<sup>+</sup>-adenosine triphosphatase (ATPase) inhibitors (e.g., omeprazole or lansoprazole) reduce acid output and are the drugs of choice for gastrinomas. Some patients may also require additional treatment with the histamine H<sub>2</sub> receptor antagonists cimetidine or ranitidine. The role of surgery in the treatment of gastrinomas in patients with MEN 1 is controversial. The goal of surgery is to reduce the risk of distant metastatic disease and improve survival. For a nonmetastatic gastrinoma situated in the pancreas, surgical excision is often effective. However, the risk of hepatic metastases increases with tumor size, such that 25–40% of patients with pancreatic NETs >4 cm develop hepatic metastases, and 50–70% of patients with tumors 2–3 cm in size have lymph node metastases. Survival in MEN 1 patients with gastrinomas <2.5 cm in size is 100% at 15 years, but 52% at 15 years, if metastatic disease is present. The presence of lymph node metastases does not appear to adversely affect survival. Surgery for gastrinomas that are >2–2.5 cm has been recommended, because the disease-related survival in these patients is improved following surgery. In addition, duodenal gastrinomas, which occur more frequently in patients with MEN 1, have been treated successfully with surgery. However, in most patients with MEN 1, gastrinomas are multiple

or extrapancreatic, and with the exception of duodenal gastrinomas, surgery is rarely successful. For example, the results of one study revealed that only ~15% of patients with MEN 1 were free of disease immediately after surgery, and at 5 years, this number had decreased to ~5%; the respective outcomes in patients without MEN 1 were better, at 45 and 40%. Given these findings, most specialists recommend a nonsurgical management for gastrinomas in MEN 1, except as noted earlier for smaller, isolated lesions. Treatment of disseminated gastrinomas is difficult. Chemotherapy with streptozotocin and 5-fluorouracil; hormonal therapy with octreotide or lanreotide, which are human somatostatin analogues (SSAs); selected internal radiation therapy (SIRT); radiofrequency ablation; peptide radio receptor therapy (PRRT); hepatic artery embolization; administration of human leukocyte interferon; and removal of all resectable tumor have been successful in some patients.

**Insulinoma** These islet cell insulin-secreting tumors represent 10–30% of all pancreatic tumors in patients with MEN 1. Patients with an insulinoma present with hypoglycemic symptoms (e.g., weakness, headaches, sweating, faintness, seizures, altered behavior, weight gain) that typically develop after fasting or exertion and improve after glucose intake. The most reliable test is a supervised 72-h fast. Biochemical investigations reveal increased plasma insulin concentrations in association with hypoglycemia (Table 388-3). Circulating concentrations of C peptide and proinsulin, which are also increased, are useful in establishing the diagnosis. It also is important to demonstrate the absence of sulfonylureas in plasma and urine samples obtained during the investigation of hypoglycemia (Table 388-3). Surgical success is greatly enhanced by preoperative localization by endoscopic ultrasonography, CT scanning, or celiac axis angiography. Additional localization methods may include preoperative and perioperative percutaneous transhepatic portal venous sampling, selective intraarterial stimulation with hepatic venous sampling, and intraoperative direct pancreatic ultrasonography. Insulinomas occur in association with gastrinomas in 10% of patients with MEN 1, and the two tumors may arise at different times. Insulinomas occur more often in patients with MEN 1 who are aged <40 years, and some arise in individuals aged <20 years. In contrast, in patients without MEN 1, insulinomas generally occur in those aged >40 years. Insulinomas may be the first manifestation of MEN 1 in 10% of patients, and ~4% of patients with insulinomas will have MEN 1.

## TREATMENT

### Insulinoma

Medical treatment, which consists of frequent carbohydrate meals and diazoxide or octreotide, is not always successful, and surgery is the optimal treatment. Surgical treatment, which ranges from enucleation of a single tumor to a distal pancreatectomy or partial pancreatectomy, has been curative in many patients. Chemotherapy

(streptozotocin, 5-fluorouracil, and doxorubicin), PRRT (e.g., with  $^{177}\text{Lu}$ -DOTATATE), or hepatic artery embolization has been used for metastatic disease.

**Glucagonoma** These glucagon-secreting pancreatic NETs occur in <3% of patients with MEN 1. The characteristic clinical manifestations of a skin rash (necrolytic migratory erythema), weight loss, anemia, and stomatitis may be absent. The tumor may have been detected in an asymptomatic patient with MEN 1 undergoing pancreatic imaging or by the finding of glucose intolerance and hyperglucagonemia.

## TREATMENT

### Glucagonoma

Surgical removal of the glucagonoma is the treatment of choice. However, treatment may be difficult because ~50–80% of patients have metastases at the time of diagnosis. Medical treatment with SSAs (e.g., octreotide or lanreotide) or chemotherapy with streptozotocin and 5-fluorouracil has been successful in some patients, and hepatic artery embolization has been used to treat metastatic disease.

### Vasoactive Intestinal Peptide (VIP) Tumors (VIPomas)

VIPomas have been reported in only a few patients with MEN 1. This clinical syndrome is characterized by watery diarrhea, hypokalemia, and achlorhydria (WDHA syndrome), which is also referred to as the Verner-Morrison syndrome, or the VIPoma syndrome. The diagnosis is established by excluding laxative and diuretic abuse, confirming a stool volume in excess of 0.5–1.0 L/d during a fast, and documenting a markedly increased plasma VIP concentration.

## TREATMENT

### VIPomas

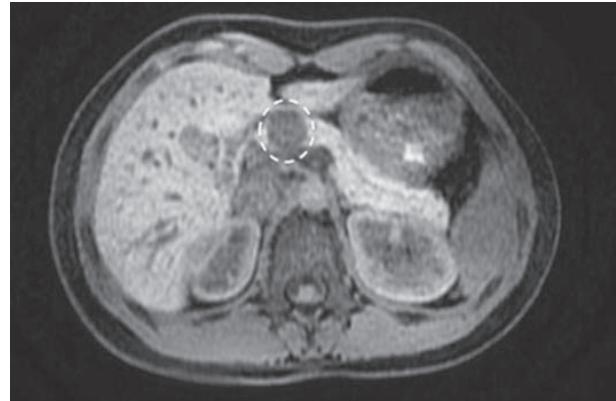
Surgical management of VIPomas, which are mostly located in the tail of the pancreas, can be curative. However, in patients with unresectable tumor, SSAs, such as octreotide and lanreotide, may be effective. Streptozotocin with 5-fluorouracil may be beneficial, along with hepatic artery embolization for the treatment of metastases.

**Pancreatic Polypeptide-Secreting Tumors (PPomas) and Nonfunctioning Pancreatic NETs** PPomas are found in a large number of patients with MEN 1. No pathologic sequelae of excessive polypeptide (PP) secretion are apparent, and the clinical significance of PP is unknown. Many PPomas may have been unrecognized or classified as nonfunctioning pancreatic NETs, which likely represent the most common enteropancreatic NET associated with MEN 1 (Fig. 388-1). The absence of both a clinical syndrome and specific biochemical abnormalities may result in a delayed diagnosis of nonfunctioning pancreatic NETs, which are associated with a worse prognosis than other functioning tumors, including insulinoma and gastrinoma. The optimum screening method and its timing interval for nonfunctioning pancreatic NETs remain to be established. At present, endoscopic ultrasound likely represents the most sensitive method of detecting small pancreatic tumors, but SRS is the most reliable method for detecting metastatic disease (Table 388-3).

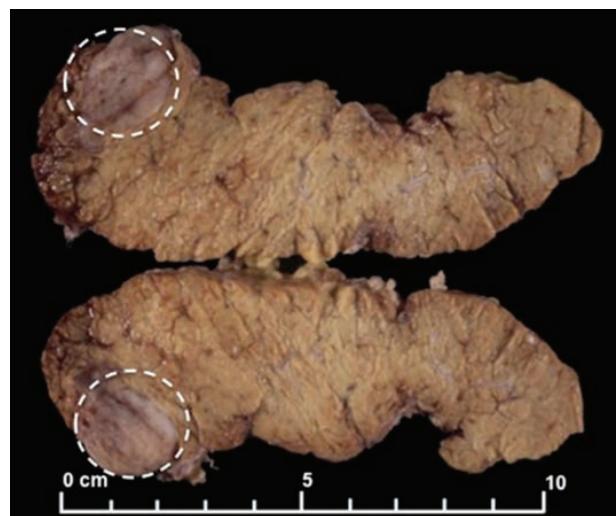
## TREATMENT

### PPomas and Nonfunctioning Pancreatic NETs

The management of nonfunctioning pancreatic NETs in the asymptomatic patient is controversial. One recommendation is to undertake surgery irrespective of tumor size after biochemical assessment is complete. Alternatively, other experts recommend surgery based on tumor size, using either >1 cm or >2 cm at different centers. Pancreatoduodenal surgery is successful in removing the tumors in 80% of patients, but >40% of patients develop complications, including



A



B

**FIGURE 388-1** Pancreatic nonfunctioning neuroendocrine tumor (NET) in a 14-year-old patient with multiple endocrine neoplasia type 1 (MEN 1). **A.** An abdominal magnetic resonance imaging scan revealed a low-intensity >2.0 cm (anteroposterior maximal diameter) tumor within the neck of pancreas. There was no evidence of invasion of adjacent structures or metastases. The tumor is indicated by white dashed circle. **B.** The pancreatic NET was removed by surgery, and macroscopic examination confirmed the location of the tumor (white dashed circles) in the neck of the pancreas. Immunohistochemistry showed the tumor to immunostain for chromogranin A, but not gastrointestinal peptides or menin, thereby confirming that it was a nonsecreting NET due to loss of menin expression. (Part A reproduced with permission from PJ Newey et al: Asymptomatic children with multiple endocrine neoplasia type 1 mutations may harbor nonfunctioning pancreatic neuroendocrine tumors. *J Clin Endocrinol Metab* 94:3640, 2009.)

diabetes mellitus, frequent steatorrhea, early and late dumping syndromes, and other gastrointestinal symptoms. However, ~50–60% of patients treated surgically survive >5 years. When considering these recommendations, it is important to consider that occult metastatic disease (e.g., tumors not detected by imaging investigations) is likely to be present in a substantial proportion of these patients at the time of presentation. Inhibitors of tyrosine kinase receptors (TKRs) and of the mammalian target of rapamycin (mTOR) signaling pathway have been reported to be effective in treating pancreatic NET metastases and in doubling the progression-free survival time. Additional treatments for metastatic disease include PRRT using  $^{177}\text{Lu}$ -DOTATATE, chemotherapy, radiofrequency ablation, transarterial chemoembolization, and SIRT.

**Other Pancreatic NETs** NETs secreting growth hormone-releasing hormone (GHRH), GHRHomas, have been reported rarely in patients with MEN 1. It is estimated that ~33% of patients with GHRHomas have other MEN 1-related tumors. GHRHomas may be diagnosed by demonstrating elevated serum concentrations of growth hormone and GHRH. More than 50% of GHRHomas occur in the lung, 30% occur in the pancreas, and 10% are found in the small intestine. Somatostatinomas secrete somatostatin, a peptide that inhibits the secretion of a variety of hormones, resulting in hyperglycemia, cholelithiasis, low acid output, steatorrhea, diarrhea, abdominal pain, anemia, and weight loss. Although 7% of pancreatic NETs secrete somatostatin, the clinical features of somatostatinoma syndrome are unusual in patients with MEN 1.

**Pituitary Tumors (See also Chap. 380)** Pituitary tumors occur in 15–50% of patients with MEN 1 (Table 388-1), and ~75% of these are microadenomas (<1 cm diameter). The tumors occur as early as 5 years of age or as late as the ninth decade. MEN 1 pituitary adenomas are more frequent in women than men, in whom they are often macroadenomas (>1 cm diameter). There are no specific histologic parameters that differentiate between MEN 1 and non-MEN 1 pituitary tumors. Approximately 60% of MEN 1-associated pituitary tumors secrete prolactin, <25% secrete growth hormone, 5% secrete adrenocorticotrophic hormone (ACTH), and the remainder appear to be nonfunctioning, with some secreting glycoprotein subunits (Table 388-1). However, pituitary tumors derived from MEN 1 patients may exhibit immunoreactivity to several hormones. In particular, there is a greater frequency of somatotropin tumors. Prolactinomas are the first manifestation of MEN 1 in ~15% of patients, whereas somatotropin tumors occur more often in patients aged >40 years. Fewer than 3% of patients with anterior pituitary tumors will have MEN 1. Clinical manifestations are similar to those in patients with sporadic pituitary tumors without MEN 1 and depend on the hormone secreted and the size of the pituitary tumor. Thus, patients may have symptoms of hyperprolactinemia (e.g., amenorrhea, infertility, and galactorrhea in women, or impotence and infertility in men) or have features of acromegaly or Cushing's disease. In addition, enlarging pituitary tumors may compress adjacent structures such as the optic chiasm or normal pituitary tissue, causing visual disturbances and/or hypopituitarism. In asymptomatic patients with MEN 1, periodic biochemical monitoring of serum prolactin and insulin-like growth factor 1 (IGF-1) levels, as well as MRI of the pituitary, can lead to early identification of pituitary tumors (Table 388-3). In patients with abnormal results, hypothalamic-pituitary testing should characterize the nature of the pituitary lesion and its effects on the secretion of other pituitary hormones.

## TREATMENT

### Pituitary Tumors

Treatment of pituitary tumors in patients with MEN 1 consists of therapies similar to those used in patients without MEN 1 and includes appropriate medical therapy (e.g., bromocriptine or cabergoline for prolactinoma; or octreotide or lanreotide for somatotropin tumors) or selective transphenoidal adenomectomy, if feasible, with radiotherapy reserved for residual unresectable tumor tissue.

**Associated Tumors** Patients with MEN 1 may also develop carcinoid tumors, adrenal cortical tumors, facial angiofibromas, collagenomas, thyroid tumors, and lipomatous tumors.

**Carcinoid Tumors (See also Chap. 84)** Carcinoid tumors occur in >3% of patients with MEN 1 (Table 388-1). The carcinoid tumor may be located in the bronchi, gastrointestinal tract, pancreas, or thymus. At the time of diagnosis, most patients are asymptomatic and do not have clinical features of the carcinoid syndrome. Importantly, no hormonal or biochemical abnormality (e.g., plasma chromogranin A) is consistently observed in individuals with thymic or bronchial carcinoid tumors. Thus, screening for these tumors is dependent on radiologic imaging. The optimum method for screening has

not been established. CT and MRI are sensitive for detecting thymic and bronchial tumors (Table 388-3), although repeated CT scanning raises concern about exposure to repeated doses of ionizing radiation. Octreotide scintigraphy may also reveal some thymic and bronchial carcinoids, although there is insufficient evidence to recommend its routine use. Gastric carcinoids, of which the type II gastric enterochromaffin-like (ECL) cell carcinoids (ECLomas) are associated with MEN 1 and Zollinger-Ellison syndrome, may be detected incidentally at the time of gastric endoscopy for dyspeptic symptoms in MEN 1 patients. These tumors, which may be found in >10% of MEN 1 patients, are usually multiple and sized <1.5 cm. Bronchial carcinoids in patients with MEN 1 occur predominantly in women (male-to-female ratio, 1:4). In contrast, thymic carcinoids in European patients with MEN 1 occur predominantly in men (male-to-female ratio, 20:1), with cigarette smokers having a higher risk for these tumors; thymic carcinoids in Japanese patients with MEN 1 have a less marked sex difference (male-to-female ratio 2:1). The course of thymic carcinoids in MEN 1 appears to be particularly aggressive. The presence of thymic tumors in patients with MEN 1 is associated with a median survival after diagnosis of ~9.5 years, with 70% of patients dying as a direct result of the tumor.

## TREATMENT

### Carcinoid Tumors

If resectable, surgical removal of carcinoid tumors is the treatment of choice. For patients with unresectable tumors and those with metastatic disease, treatment with SSAs, radiotherapy, chemotherapeutic agents (e.g., fluorouracil, temozolamide, cisplatin, etoposide), mTOR inhibitors (e.g., everolimus), or PRRT therapy has resulted in symptom improvement and regression of some tumors. Little is known about the malignant potential of gastric type II ECLomas, but treatment with SSAs has resulted in regression of these ECLomas.

**Adrenocortical Tumors (See also Chap. 386)** Asymptomatic adrenocortical tumors occur in 20–70% of patients with MEN 1 depending on the radiologic screening methods used (Table 388-1). Most of these tumors, which include cortical adenomas, hyperplasia, multiple adenomas, nodular hyperplasia, cysts, and carcinomas, are nonfunctioning. Indeed, <10% of patients with enlarged adrenal glands have hormonal hypersecretion, with primary hyperaldosteronism and ACTH-independent Cushing's syndrome being encountered most commonly. Occasionally, hyperandrogenemia may occur in association with adrenocortical carcinoma. Pheochromocytoma in association with MEN 1 is rare. Biochemical investigation (e.g., plasma renin and aldosterone concentrations, low-dose dexamethasone suppression test, urinary catecholamines, and/or metanephrines) should be undertaken in those with symptoms or signs suggestive of functioning adrenal tumors or in those with tumors >1 cm. Adrenocortical carcinoma occurs in ~1% of MEN 1 patients but increases to >10% for adrenal tumors >1 cm.

## TREATMENT

### Adrenocortical Tumors

Consensus has not been reached about the management of MEN 1-associated nonfunctioning adrenal tumors, because the majority are benign. However, the risk of malignancy increases with size, particularly for tumors with a diameter >4 cm. Indications for surgery for adrenal tumors include size >4 cm in diameter, atypical or suspicious radiologic features (e.g., increased Hounsfield unit on unenhanced CT scan) and size of 1–4 cm in diameter, or significant measurable growth over a 6-month period. The treatment of functioning (e.g., hormone-secreting) adrenal tumors is similar to that for tumors occurring in non-MEN 1 patients.

**2988 Meningioma** Central nervous system (CNS) tumors, including ependymomas, schwannomas, and meningiomas, have been reported in MEN 1 patients (Table 388-1). Meningiomas are found in <10% of patients with other clinical manifestations of MEN 1 (e.g., primary hyperparathyroidism) for >15 years. The majority of meningiomas are not associated with symptoms, and 60% do not enlarge. The treatment of MEN 1-associated meningiomas is similar to that in non-MEN 1 patients.

**Lipomas** Subcutaneous lipomas occur in >33% of patients with MEN 1 (Table 388-1) and are frequently multiple. In addition, visceral, pleural, or retroperitoneal lipomas may occur in patients with MEN 1. Management is conservative. However, when surgically removed for cosmetic reasons, they typically do not recur.

**Facial Angiofibromas and Collagenomas** The occurrence of multiple facial angiofibromas in patients with MEN 1 may range from >20 to >90%, and occurrence of collagenomas may range from 0 to >70% (Table 388-1). These cutaneous findings may allow presymptomatic diagnosis of MEN 1 in the relatives of a patient with MEN 1. Treatment for these cutaneous lesions is usually not required.

**Thyroid Tumors** Thyroid tumors, including adenomas, colloid goiters, and carcinomas, have been reported to occur in >25% of patients with MEN 1. However, the prevalence of thyroid disorders in the general population is high, and it has been suggested that the association of thyroid abnormalities in patients with MEN 1 may be incidental. The treatment of thyroid tumors in MEN 1 patients is similar to that for non-MEN 1 patients.

 **Genetics and Screening** The *MEN1* gene is located on chromosome 11q13 and consists of 10 exons, which encode a 610-amino acid protein, menin, that regulates transcription, genome stability, cell division, and proliferation. The pathophysiology of MEN 1 follows the Knudson two-hit hypothesis with a tumor-suppressor role for menin. Inheritance of a germline *MEN1* mutation predisposes an individual to developing a tumor that arises following a somatic mutation, which may be a point mutation or more commonly a deletion, leading to loss of heterozygosity (LOH) in the tumor DNA. The germline mutations of the *MEN1* gene are scattered throughout the entire 1830-bp coding region and splice sites, and there is no apparent correlation between the location of *MEN1* mutations and clinical manifestations of the disorder, in contrast with the situation in patients with MEN 2 (Table 388-1). More than 10% of *MEN1* germline mutations arise de novo and may be transmitted to subsequent generations. Some families with MEN 1 mutations develop parathyroid tumors as the sole endocrinopathy, and this condition is referred to as familial isolated hyperparathyroidism (FIHP). However, between 5 and 25% of patients with MEN 1 do not harbor germline mutations or deletions of the *MEN1* gene. Such patients with MEN 1-associated tumors but without *MEN1* mutations may represent phenocopies or have mutations involving other genes. Other genes associated with MEN 1-like features include *CDC73*, which encodes parafibromin, whose mutations result in the HPT-JT syndrome; the calcium-sensing receptor gene (*CaSR*), whose mutations result in familial benign hypocalciuric hypercalcemia (FBHH); and the aryl hydrocarbon receptor interacting protein gene (*AIP*), a tumor suppressor located on chromosome 11q13 whose mutations are associated with familial isolated pituitary adenomas (FIPA). Genetic testing to determine the *MEN1* mutation status in symptomatic family members within a MEN 1 kindred, as well as in all index cases (e.g., patients) with two or more endocrine tumors, is advisable. If a *MEN1* mutation is not identified in the index case with two or more endocrine tumors, clinical and genetic tests for other disorders such as HPT-JT syndrome, FBHH, FIPA, MEN 2, or MEN 4 should be considered because these patients may represent phenocopies for MEN 1.

The current guidelines recommend that *MEN1* mutational analysis should be undertaken in (1) an index case with two or more MEN 1-associated endocrine tumors (e.g., parathyroid, pancreatic, or pituitary tumors); (2) asymptomatic first-degree relatives of a known *MEN1* mutation carrier; and (3) first-degree relatives of a *MEN1* mutation carrier with symptoms, signs, or biochemical or radiologic evidence for one or more MEN 1-associated tumors. In addition, *MEN1* mutational

analysis should be considered in patients with suspicious or atypical MEN 1. This would include individuals with parathyroid adenomas before the age of 30 years or multigland parathyroid disease; individuals with gastrinoma or multiple pancreatic NETs at any age; or individuals who have two or more MEN 1-associated tumors that are not part of the classical triad of parathyroid, pancreatic islet, and anterior pituitary tumors (e.g., parathyroid tumor plus adrenal tumor). Family members, including asymptomatic individuals who have been identified to harbor a *MEN1* mutation, will require biochemical and radiologic screening (Table 388-3). In contrast, relatives who do not harbor the *MEN1* mutation have a risk of developing MEN 1-associated endocrine tumors that is similar to that of the general population; thus, relatives without the *MEN1* mutation do not require repeated screening.

Mutational analysis in asymptomatic individuals should be undertaken at the earliest opportunity and, if possible, in the first decade of life because tumors have developed in some children by the age of 5 years. Appropriate biochemical and radiologic investigations (Table 388-3) aimed at detecting the development of tumors should then be undertaken in affected individuals. Mutant gene carriers should undergo biochemical screening at least once per annum and also have baseline pituitary and abdominal imaging (e.g., MRI or CT), which should then be repeated at 1- to 3-year intervals (Table 388-3). Screening should commence after 5 years of age and should continue for life because the disease may develop as late as the eighth decade. The screening history and physical examination elicit the symptoms and signs of hypercalcemia; nephrolithiasis; peptic ulcer disease; neuroglycopenia; hypopituitarism; galactorrhea and amenorrhea in women; acromegaly; Cushing's disease; and visual field loss and the presence of subcutaneous lipomas, angiofibromas, and collagenomas. Biochemical screening should include measurements of serum calcium, PTH, gastrointestinal hormones (e.g., gastrin, insulin with a fasting glucose, glucagon, VIP, PP), chromogranin A, prolactin, and IGF-1 in all individuals. More specific endocrine function tests should be undertaken in individuals who have symptoms or signs suggestive of a specific clinical syndrome. Biochemical screening for the development of MEN 1 tumors in asymptomatic members of families with MEN 1 is of great importance to reduce morbidity and mortality from the associated tumors.

## MULTIPLE ENDOCRINE NEOPLASIA TYPE 2 AND TYPE 3

**Clinical Manifestations** MEN type 2 (MEN 2), which is also called Sipple's syndrome, is characterized by the association of medullary thyroid carcinoma (MTC), pheochromocytomas, and parathyroid tumors (Table 388-1). Three clinical variants of MEN 2 are recognized: MEN 2A, MEN 2B, and MTC only. MEN 2A, which is often referred to as MEN 2, is the most common variant. In MEN 2A, MTC is associated with pheochromocytomas in 50% of patients (may be bilateral) and with parathyroid tumors in 20% of patients. MEN 2A may rarely occur in association with Hirschsprung's disease, caused by the absence of autonomic ganglion cells in the terminal hindgut, resulting in colonic dilatation, severe constipation, and obstruction. MEN 2A may also be associated with cutaneous lichen amyloidosis, which is a pruritic lichenoid lesion that is usually located on the upper back. MEN 2B, which is also referred to as MEN 3, represents 5% of all cases of MEN 2 and is characterized by the occurrence of MTC and pheochromocytoma in association with a Marfanoid habitus; mucosal neuromas of the lips, tongue, and eyelids; medullated corneal fibers; and intestinal autonomic ganglion dysfunction leading to multiple diverticulae and megacolon. Parathyroid tumors do not usually occur in MEN 2B. MTC only (FMTC) is a variant in which MTC is the sole manifestation of the syndrome. However, the distinction between FMTC and MEN 2A is difficult and should only be considered if there are at least four family members aged >50 years who are affected by MTC but not pheochromocytomas or primary hyperparathyroidism. All of the MEN 2 variants are due to mutations of the rearranged during transfection (*RET*) protooncogene, which encodes a TKR. Moreover, there is a correlation between the locations of *RET* mutations and MEN 2 variants. Thus, ~95% of MEN 2A patients have mutations involving the cysteine-rich

**TABLE 388-4** Recommendations for Tests and Surgery in MEN 2 and MEN 3<sup>a</sup>

RET MUTATION, EXON (EX) LOCATION, AND CODON INVOLVED	RISK <sup>b</sup>	RECOMMENDED AGE (YEARS) FOR TEST/INTERVENTION				
		RET MUTATIONAL ANALYSIS	FIRST SERUM CALCITONIN AND NECK ULTRASOUND	PROPHYLACTIC THYROIDECTOMY	SCREENING FOR PHEOCHROMOCYTOMA	SCREENING FOR PHPT
Ex8 (533) <sup>c</sup> ; Ex10 (609, 611, 618, 620) <sup>c</sup> ; Ex11 (630, 631, 666) <sup>c</sup> ; Ex13 (768, 790) <sup>c</sup> ; Ex14 (804) <sup>c</sup> ; Ex15 (891) <sup>c</sup> ; Ex16 (912) <sup>c</sup>	+	<3–5	5	<5 <sup>d</sup>	16 <sup>e</sup>	16
Ex11 (634) <sup>c</sup> ; Ex15 (883) <sup>c</sup>	++	<3	<3	<5 <sup>f</sup>	11 <sup>e</sup>	11
Ex15 (883) <sup>b</sup> ; Ex16 (918) <sup>g</sup>	+++	ASAP and by <1	ASAP and by <0.5–1	ASAP and by <1	11 <sup>e</sup>	— <sup>h</sup>

<sup>a</sup>Data from American Thyroid Association Guidelines Task Force, RT Kloos et al: Medullary thyroid cancer: management guidelines of the American Thyroid Association. Thyroid 19:565, 2009 and revised from SA Wells Jr et al: Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. Thyroid 25:567, 2015. <sup>b</sup>Risk for early development of metastasis and aggressive growth of medullary thyroid cancer: +++, highest; ++, high; and + moderate. <sup>c</sup>Mutations associated with MEN 2A (or medullary thyroid carcinoma only). <sup>d</sup>Timing of surgery to be based on elevation of serum calcitonin and/or joint discussion with pediatrician, surgeon, and parent/family. Later surgery may be appropriate if serum calcitonin and neck ultrasound are normal. <sup>e</sup>Presence of pheochromocytoma must be excluded prior to any surgical intervention, and also in women with *RET* mutation who are planning pregnancy or are pregnant. <sup>f</sup>Surgery earlier than 5 years based on elevation of serum calcitonin. Optimal timing of surgery should be decided by the surgeon and pediatrician, in consultation with the child's parent. <sup>g</sup>Mutations associated with MEN 2B (MEN 3). <sup>h</sup>Not required because PHPT is not a feature of MEN 2B (MEN 3).

Abbreviations: ASAP, as soon as possible; MEN, multiple endocrine neoplasia; PHPT, primary hyperparathyroidism.

extracellular domain, with mutations of codon 634 accounting for ~85% of MEN 2A mutations; FMTC patients also have mutations of the cysteine-rich extracellular domain, with most mutations occurring in codon 618. In contrast, ~95% of MEN 2B/MEN 3 patients have mutations of codon 918 of the intracellular tyrosine kinase domain (Table 388-1 and Table 388-4).

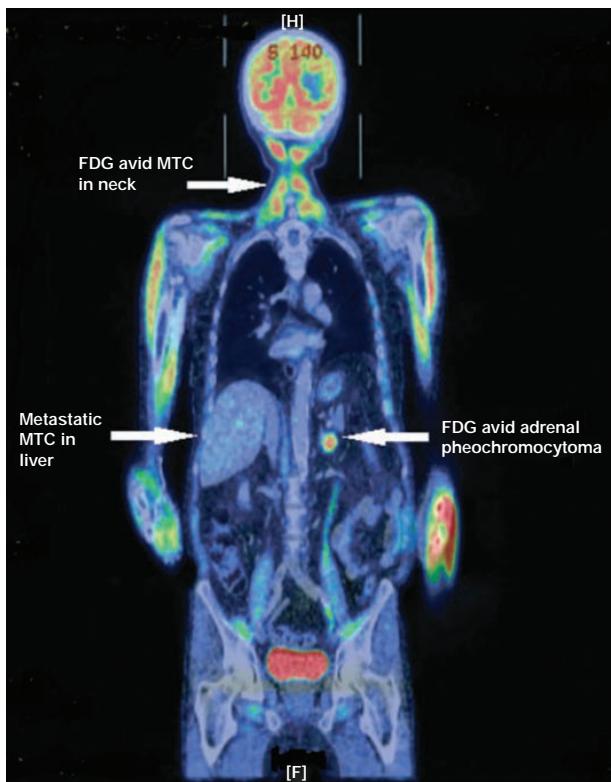
**Medullary Thyroid Carcinoma** MTC is the most common feature of MEN 2A and MEN 2B and occurs in almost all affected individuals. MTC represents 5–10% of all thyroid gland carcinomas, and 20% of MTC patients have a family history of the disorder. The use of *RET* mutational analysis to identify family members at risk for hereditary forms of MTC has altered the presentation of MTC from that of symptomatic tumors to a preclinical disease for which prophylactic thyroidectomy (Table 388-4) is undertaken to improve the prognosis and ideally result in cure. However, in patients who do not have a known family history of MEN 2A, FMTC, or MEN 2B, and therefore have not had *RET* mutational analysis, MTC may present as a palpable mass in the neck, which may be asymptomatic or associated with symptoms of pressure or dysphagia in >15% of patients. Diarrhea occurs in 30% of patients and is associated either with elevated circulating concentrations of calcitonin or tumor-related secretion of serotonin and prostaglandins. Some patients may also experience flushing. In addition, ectopic ACTH production by MTC may cause Cushing's syndrome. The diagnosis of MTC relies on the demonstration of hypercalcitoninemia (>90 pg/mL in the basal state); stimulation tests using IV pentagastrin (0.5 mg/kg) and/or calcium infusion (2 mg/kg) are rarely used now, reflecting improvements in the assay for calcitonin. Neck ultrasonography with fine-needle aspiration of the nodules can confirm the diagnosis. Radionuclide thyroid scans may reveal MTC tumors as "cold" nodules. Radiography may reveal dense irregular calcification within the involved portions of the thyroid gland and in lymph nodes involved with metastases. Positron emission tomography (PET) may help to identify the MTC and metastases (Fig. 388-2). Metastases of MTC usually occur to the cervical lymph nodes in the early stages and to the mediastinal nodes, lung, liver, trachea, adrenal, esophagus, and bone in later stages. Elevations in serum calcitonin concentrations are often the first sign of recurrence or persistent disease, and the serum calcitonin doubling time is useful for determining prognosis. MTC can have an aggressive clinical course, with early metastases and death in ~10% of patients. A family history of aggressive MTC or MEN 2B may be elicited.

## TREATMENT

### Medullary Thyroid Carcinoma

Individuals with *RET* mutations who do not have clinical manifestations of MTC should be offered prophylactic surgery between the ages of <1 and 5 years. The timing of surgery will depend on

the type of *RET* mutation and its associated risk for early development, metastasis, and aggressive growth of MTC (Table 388-4). Such patients should have a total thyroidectomy with a systematic central neck dissection to remove occult nodal metastasis, although the value of undertaking a central neck dissection has been subject to debate. Prophylactic thyroidectomy, with lifelong thyroxine replacement, has dramatically improved outcomes in patients with MEN 2 and MEN 3, such that ~90% of young patients with *RET* mutations who had a prophylactic thyroidectomy have no evidence



**FIGURE 388-2** Fluorodeoxyglucose (FDG) positron emission tomography scan in a patient with multiple endocrine neoplasia type 2A, showing medullary thyroid cancer (MTC) with hepatic and skeletal (left arm) metastasis and a left adrenal pheochromocytoma. Note the presence of excreted FDG compound in the bladder. (Reproduced with permission from A Naziat et al: Confusing genes: A patient with MEN2A and Cushing's disease. Clin Endocrinol (Oxf) 78:966, 2013.)

of persistent or recurrent MTC at 7 years after surgery. In patients with clinically evident MTC, a total thyroidectomy with bilateral central resection is recommended, and an ipsilateral lateral neck dissection should be undertaken if the primary tumor is >1 cm in size or there is evidence of nodal metastasis in the central neck. Surgery is the only curative therapy for MTC. The 10-year survival in patients with metastatic MTC is ~20%. For inoperable MTC or metastatic disease, TKR inhibitors (e.g., vandetanib, cabozantinib, selpercatinib) have improved the progression-free survival times. PRRT with <sup>177</sup>Lu-DOTATATE has been reported to be beneficial for metastatic MTCs that were found by SR<sub>S</sub> to express somatostatin receptors. Other types of chemotherapy are of limited efficacy, but radiotherapy may help to palliate local disease.

**Pheochromocytoma (See also Chap. 387)** These noradrenaline- and adrenaline-secreting tumors occur in >50% of patients with MEN 2A and MEN 2B and are a major cause of morbidity and mortality. Patients may have symptoms and signs of catecholamine secretion (e.g., headaches, palpitations, sweating, poorly controlled hypertension), or they may be asymptomatic with detection through biochemical screening based on a history of familial MEN 2A, MEN 2B, or MTC. Pheochromocytomas in patients with MEN 2A and MEN 2B differ significantly in distribution when compared with patients without MEN 2A and MEN 2B. Extra-adrenal pheochromocytomas, which occur in 10% of patients without MEN 2A and MEN 2B, are observed rarely in patients with MEN 2A and MEN 2B. Malignant pheochromocytomas are much less common in patients with MEN 2A and MEN 2B. The biochemical and radiologic investigation of pheochromocytoma in patients with MEN 2A and MEN 2B is similar to that in non-MEN 2 patients and includes the measurement of plasma (obtained from supine patients) and urinary free fractionated metanephrenes (e.g., normetanephrine and metanephrenes measured separately), CT or MRI scanning, radionuclide scanning with meta-iodo-(<sup>123</sup>I or <sup>131</sup>I)-benzyl guanidine (MIBG), and PET using (<sup>18</sup>F)-fluorodopamine or (<sup>18</sup>F)-fluoro-2-deoxy-d-glucose (Fig. 388-2).

## TREATMENT

### Pheochromocytoma

Surgical removal of pheochromocytoma, using adrenergic receptor blockade before and during the operation, is the recommended treatment. Other antihypertensive agents, including calcium channel blockers, are sometimes required for adequate blood pressure control. Endoscopic adrenal-sparing surgery, which decreases postoperative morbidity, hospital stay, and expense, as opposed to open surgery, has become the method of choice.

**Parathyroid Tumors (See also Chap. 410)** Parathyroid tumors occur in 10–25% of patients with MEN 2A. However, >50% of these patients do not have hypercalcemia. The presence of abnormally enlarged parathyroids, which are unusually hyperplastic, is often seen in the normocalcemic patient undergoing thyroidectomy for MTC. The biochemical investigation and treatment of hypercalcemic patients with MEN 2A is similar to that of patients with MEN 1.

 **Genetics and Screening** To date, ~50 different *RET* mutations have been reported, and these are located in exons 5, 8, 10, 11, 13, 14, 15, and 16. *RET* germline mutations are detected in >95% of MEN 2A, FMTC, and MEN 2B families, with Cys634Arg being most common in MEN 2A, Cys618Arg being most common in FMTC, and Met918Thr being most common in MEN 2B (Tables 388-1 and 388-4). Between 5 and 10% of patients with MTC or MEN 2A-associated tumors have de novo *RET* germline mutations, and ~50% of patients with MEN 2B have de novo *RET* germline mutations. These de novo *RET* germline mutations always occur on the paternal allele. Approximately 5% of patients with sporadic pheochromocytoma have a germline *RET* mutation, but such germline *RET* mutations do not appear to be associated with sporadic primary hyperparathyroidism. Thus, *RET* mutational analysis should be performed in (1) all patients

with MTC who have a family history of tumors associated with MEN 2, FMTC, or MEN 3, such that the diagnosis can be confirmed and genetic testing offered to asymptomatic relatives; (2) all patients with MTC and pheochromocytoma without a known family history of MEN 2 or MEN 3; (3) all patients with MTC, but without a family history of MEN 2, FMTC, or MEN 3, because these patients may have a de novo germline *RET* mutations; (4) all patients with bilateral pheochromocytoma; and (5) patients with unilateral pheochromocytoma, particularly if this occurs with increased calcitonin levels.

Screening for MEN 2/MEN 3-associated tumors in patients with *RET* germline mutations should be undertaken annually and include serum calcitonin measurements, a neck ultrasound for MTC, plasma (or 24-h urinary) fractionated metanephrenes for pheochromocytoma, and albumin-corrected serum calcium or ionized calcium with PTH for primary hyperparathyroidism. In patients with MEN 2-associated *RET* mutations, screening for MTC should begin by 1–5 years, for pheochromocytoma by 11–16 years, and for primary hyperparathyroidism by 11–16 years of age (Table 388-4).

## MULTIPLE ENDOCRINE NEOPLASIA TYPE 4

**Clinical Manifestations** Patients with MEN 1-associated tumors, such as parathyroid adenomas, pituitary adenomas, and pancreatic NETs, occurring in association with gonadal, adrenal, renal, and thyroid tumors have been reported to have mutations of the gene encoding the 196-amino acid cyclin-dependent kinase inhibitor (CK1) p27kip1 (*CDKN1B*). Such families with MEN 1-associated tumors and *CDKN1B* mutations are designated to have MEN 4 (Table 388-1). The investigations and treatments for the MEN 4-associated tumors are similar to those for MEN 1 and non-MEN 1 tumors.

 **Genetics and Screening** To date, 50 MEN patients (from <20 kindreds) with mutations of *CDKN1B*, which is located on chromosome 12p13, have been reported, and all of these are predicted to result in a loss of function. These MEN 4 patients may represent ~3% of the 5–10% of patients with MEN 1 who do not have mutations of the *MEN1* gene. Germline *CDKN1B* mutations may rarely be found in patients with sporadic (i.e., nonfamilial) forms of primary hyperparathyroidism.

## HYPERPARATHYROIDISM-JAW TUMOR SYNDROME (SEE ALSO CHAP. 410)

**Clinical Manifestations** Hyperparathyroidism-jaw tumor (HPT-JT) syndrome is an autosomal dominant disorder characterized by the development of parathyroid tumors (15% are carcinomas) and fibro-osseous jaw tumors. In addition, some patients may also develop Wilms' tumors, renal cysts, renal hamartomas, renal cortical adenomas, renal cell carcinoma (RCC), pancreatic adenocarcinomas, uterine tumors, testicular mixed germ cell tumors with a major seminoma component, and Hürthle cell thyroid adenomas. The parathyroid tumors may occur in isolation and without any evidence of jaw tumors, and this may cause confusion with other hereditary hypercalcemic disorders, such as MEN 1. However, genetic testing to identify the causative mutation will help to establish the correct diagnosis. The investigation and treatment for HPT-JT-associated tumors are similar to those in non-HPT-JT patients, except that early parathyroidectomy is advisable because of the increased frequency of parathyroid carcinoma.

 **Genetics and Screening** The gene that causes HPT-JT is located on chromosome 1q31.2 and encodes a 531-amino acid protein, parafibromin (Table 388-2). Parafibromin is also referred to as cell division cycle protein 73 (CDC73) and has a role in transcription. Genetic testing in families helps to identify mutation carriers who should be periodically screened for the development of tumors (Table 388-5).

## VON HIPPEL-LINDAU DISEASE (SEE ALSO CHAP. 387)

**Clinical Manifestations** von Hippel-Lindau (VHL) disease is an autosomal dominant disorder characterized by hemangioblastomas

**TABLE 388-5 HPT-JT Screening Guidelines**

TUMOR <sup>a</sup>	TEST	FREQUENCY <sup>b</sup>
Parathyroid	Serum Ca, PTH	6–12 months
Ossifying jaw fibroma	Panoramic jaw x-ray with neck shielding <sup>c</sup>	5 years
Renal	Abdominal MRI <sup>c,d</sup>	5 years
Uterine	Ultrasound (transvaginal or transabdominal) and additional imaging ± D&C if indicated <sup>e</sup>	Annual

<sup>a</sup>Screening for most common HPT-JT-associated tumors is considered. Assessment for other reported tumor types may be indicated (e.g., pancreatic, thyroid, testicular tumors). <sup>b</sup>Frequency of repeating test after baseline tests performed. <sup>c</sup>X-rays and imaging involving ionizing radiation should ideally be avoided to minimize risk of generating subsequent mutations. <sup>d</sup>Ultrasound scan recommended if MRI unavailable. <sup>e</sup>Such selective pelvic imaging should be considered after obtaining a detailed menstrual history.

**Abbreviations:** Ca, calcium; D&C, dilation and curettage; HPT-JT, hyperparathyroidism-jaw tumor syndrome; MRI, magnetic resonance imaging; PTH, parathyroid hormone.

**Source:** Reproduced with permission from PJ Newey et al: Cell division cycle protein 73 homolog (CDC73) mutations in the hyperparathyroidism-jaw tumor syndrome (HPT-JT) and parathyroid tumors. *Hum Mutat* 31:295, 2010.

of the retina and CNS; cysts involving the kidneys, pancreas, and epididymis; RCC; pheochromocytomas; and pancreatic islet cell tumors. The retinal and CNS hemangioblastomas are benign vascular tumors that may be multiple; those in the CNS may cause symptoms by compressing adjacent structures and/or increasing intracranial pressure. In the CNS, the cerebellum and spinal cord are the most frequently involved sites. The renal abnormalities consist of cysts and carcinomas, and the lifetime risk of RCC in VHL is 70%. The endocrine tumors in VHL consist of pheochromocytomas and pancreatic islet cell tumors. The clinical presentation of pheochromocytoma in VHL disease is similar to that in sporadic cases, except that there is a higher frequency of bilateral or multiple tumors, which may involve extra-adrenal sites in VHL disease. The most frequent pancreatic lesions in VHL are multiple cystadenomas, which rarely cause clinical disease. However, nonsecreting pancreatic islet cell tumors occur in <10% of VHL patients, who are usually asymptomatic. The pancreatic tumors in these patients are often detected by regular screening using abdominal imaging. Pheochromocytomas should be investigated and treated as described earlier for MEN 2. The pancreatic islet cell tumors frequently become malignant, and early surgery is recommended.

 **Genetics and Screening** The *VHL* gene, which is located on chromosome 3p26-p25, is widely expressed in human tissues and encodes a 213-amino acid protein (pVHL) (Table 388-2). A wide variety of germline *VHL* mutations have been identified. *VHL* acts as a tumor-suppressor gene. A correlation between the type of mutation and the clinical phenotype has been reported; large deletions and protein-truncating mutations are associated with a low incidence of pheochromocytomas, whereas some missense mutations in VHL patients are associated with pheochromocytoma (referred to as VHL type 2C). Other missense mutations may be associated with hemangioblastomas and RCC but not pheochromocytoma (referred to as VHL type 1), whereas distinct missense mutations are associated with hemangioblastomas, RCC, and pheochromocytoma (VHL type 2B). VHL type 2A, which refers to the occurrence of hemangioblastomas and pheochromocytoma without RCC, is associated with rare missense mutations. The basis for these complex genotype-phenotype relationships remains to be elucidated. One major function of pVHL, which is also referred to as elongin, is to downregulate the expression of vascular endothelial growth factor (VEGF) and other hypoxia-inducible mRNAs. Thus, pVHL, in complex with other proteins, regulates the expression of hypoxia-inducible factors (HIF-1 and HIF-2) such that loss of functional pVHL leads to a stabilization of the HIF protein complexes, resulting in VEGF overexpression and tumor angiogenesis. Screening for the development of pheochromocytomas and pancreatic islet cell tumors is as described earlier for MEN 2 and MEN 1, respectively (Tables 388-3 and 388-4).

## NEUROFIBROMATOSIS

**Clinical Manifestations** Neurofibromatosis type 1 (NF1), which is also referred to as von Recklinghausen's disease, is an autosomal dominant disorder characterized by the following manifestations: neurologic (e.g., peripheral and spinal neurofibromas); ophthalmologic (e.g., optic gliomas and iris hamartomas such as Lisch nodules); dermatologic (e.g., café au lait macules); skeletal (e.g., scoliosis, macrocephaly, short stature, pseudoarthrosis); vascular (e.g., stenoses of renal and intracranial arteries); and endocrine (e.g., pheochromocytoma, carcinoid tumors, precocious puberty). Neurofibromatosis type 2 (NF2) is also an autosomal dominant disorder but is characterized by the development of bilateral vestibular schwannomas (acoustic neuromas) that lead to deafness, tinnitus, or vertigo. Some patients with NF2 also develop meningiomas, spinal schwannomas, peripheral nerve neurofibromas, and café au lait macules. Endocrine abnormalities are not found in NF2 and are associated solely with NF1. Pheochromocytomas, carcinoid tumors, and precocious puberty occur in ~1% of patients with NF1, and growth hormone deficiency has also been reported. The features of pheochromocytomas in NF1 are similar to those in non-NF1 patients, with 90% of tumors being located within the adrenal medulla and the remaining 10% at an extra-adrenal location, which often involves the para-aortic region. Primary carcinoid tumors are often periampullary and may also occur in the ileum but rarely in the pancreas, thyroid, or lungs. Hepatic metastases are associated with symptoms of the carcinoid syndrome, which include flushing, diarrhea, bronchoconstriction, and tricuspid valve disease. Precocious puberty is usually associated with the extension of an optic glioma into the hypothalamus with resultant early activation of gonadotropin-releasing hormone secretion. Growth hormone deficiency has also been observed in some NF1 patients, who may or may not have optic chiasmal gliomas, but it is important to note that short stature is frequent in the absence of growth hormone deficiency in patients with NF1. The investigation and treatment for tumors are similar to those undertaken for each respective tumor type in non-NF1 patients.

 **Genetics and Screening** The *NF1* gene, which is located on chromosome 17q11.2 and acts as a tumor suppressor, consists of 60 exons that span >350 kb of genomic DNA (Table 388-2). Mutations in *NF1* are of diverse types and are scattered throughout the exons. The *NF1* gene product is the protein neurofibromin, which has homologies to the p120GAP (GTPase activating protein) and acts on p21ras by converting the active GTP bound form to its inactive GDP form. Mutations of *NF1* impair this downregulation of the p21ras signaling pathways, which in turn results in abnormal cell proliferation. Screening for the development of pheochromocytomas and carcinoid tumors is as described earlier for MEN 2 and MEN 1, respectively (Tables 388-3 and 388-4).

## CARNEY COMPLEX

**Clinical Manifestations** Carney complex (CNC) is an autosomal dominant disorder characterized by spotty skin pigmentation (usually of the face, labia, and conjunctiva), myxomas (usually of the eyelids and heart, but also the tongue, palate, breast, and skin), psammomatous melanotic schwannomas (usually of the sympathetic nerve chain and upper gastrointestinal tract), and endocrine tumors that involve the adrenals, Sertoli cells, somatotropes, thyroid, and ovary. Cushing's syndrome, the result of primary pigmented nodular adrenal disease (PPNAD), is the most common endocrine manifestation of CNC and may occur in one-third of patients. Patients with CNC and Cushing's syndrome often have an atypical appearance by being thin (as opposed to having truncal obesity). In addition, they may have short stature, muscle and skin wasting, and osteoporosis. These patients often have levels of urinary free cortisol that are normal or increased only marginally. Cortisol production may fluctuate periodically with days or weeks of hypercortisolism; this pattern is referred to as "periodic Cushing's syndrome." Patients with Cushing's syndrome usually have loss of the circadian rhythm of cortisol production. Acromegaly, the

result of a somatotrope tumor, affects ~10% of patients with CNC. Testicular tumors may also occur in one-third of patients with CNC. These may either be large-cell calcifying Sertoli cell tumors, adrenocortical rests, or Leydig cell tumors. The Sertoli cell tumors occasionally may be estrogen-secreting and lead to precocious puberty or gynecomastia. Some patients with CNC have been reported to develop thyroid follicular tumors, ovarian cysts, or breast duct adenomas.



**Genetics and Screening** CNC type 1 (CNC1) is due to mutations of the protein kinase A (PKA) regulatory subunit 1 (R1<sup>α</sup>) (*PRAKAR1A*), a tumor suppressor, whose gene is located on chromosome 17q24.2 (Table 388-2). The gene causing CNC type 2 (CNC2) is located on chromosome 2p16 and has not yet been identified. It is interesting to note, however, that some tumors do not show LOH of 2p16 but instead show genomic instability, suggesting that this CNC gene may not be a tumor suppressor. Screening and treatment of these endocrine tumors are similar to those described earlier for patients with MEN 1 and MEN 2 (Tables 388-3 and 388-4).

### COWDEN'S SYNDROME

**Clinical Manifestations** Multiple hamartomatous lesions, especially of the skin, mucous membranes (e.g., buccal, intestinal, colonic), breast, and thyroid, are characteristic of Cowden's syndrome (CWS), which is an autosomal dominant disorder. Thyroid abnormalities occur in two-thirds of patients with CWS, and these usually consist of multinodular goiters or benign adenomas, although <10% of patients may have a follicular thyroid carcinoma. Breast abnormalities occur in >75% of patients and consist of either fibrocystic disease or adenocarcinomas. The investigation and treatment for CWS tumors are similar to those undertaken for non-CWS patients.



**Genetics and Screening** CWS is genetically heterogeneous, and seven types (CWS1–7) are recognized (Table 388-2). CWS1 is due to mutations of the phosphate and tensin homologue deleted on chromosome 10 (*PTEN*) gene, located on chromosome 10q23.31. CWS2 is caused by mutations of the succinate dehydrogenase subunit B (*SDHB*) gene, located on chromosome 1p36.13; and CWS3 is caused by mutations of the *SDHD* gene, located on chromosome 11q13.1. *SDHB* and *SDHD* mutations are also associated with pheochromocytoma. CWS4 is caused by hypermethylation of the Killin (*KLLN*) gene, the promoter of which shares the same transcription site as *PTEN* on chromosome 10q23.31. CWS5 is caused by mutations of the phosphatidylinositol 3-kinase catalytic alpha (*PIK3CA*) gene on chromosome 3q26.32. CWS6 is caused by mutations of the V-Akt murine thymoma viral oncogene homolog 1 (*AKT1*) gene on chromosome 14q32.33, and CWS7 is caused by mutations of the *SEC23B* gene on chromosome 20p11.23. Screening for thyroid abnormalities entails neck ultrasonography and fine-needle aspiration with analysis of cell cytology.

### MCCUNE-ALBRIGHT SYNDROME

(SEE ALSO CHAP. 412)

**Clinical Manifestations** McCune-Albright syndrome (MAS) is characterized by the triad of polyostotic fibrous dysplasia, which may be associated with hypophosphatemic rickets; café au lait skin pigmentation; and peripheral precocious puberty. Other endocrine abnormalities include thyrotoxicosis, which may be associated with a multinodular goiter, somatotrope tumors, and Cushing's syndrome (due to adrenal tumors). Investigation and treatment for each endocrinopathy are similar to those used in patients without MAS.



**Genetics and Screening** MAS is a disorder of mosaicism that results from postzygotic somatic cell mutations of the G protein  $\alpha$ -stimulating subunit ( $G_s$ ), encoded by the *GNAS1* gene, located on chromosome 20q13.32 (Table 388-2). The  $G_s$  mutations, which include Arg201Cys, Arg201His, Glu227Arg, or Glu227His, are activating and are found only in cells of the abnormal tissues. Screening for hyperfunction of relevant endocrine glands and development of hypophosphatemia, which may be associated with elevated serum

fibroblast growth factor 23 (FGF23) concentrations, is undertaken in MAS patients.

### Acknowledgment

The author is grateful to the National Institute of Health Research (NIHR) Oxford Biomedical Research Centre Programme for support and to Mrs. Tracey Walker for typing the manuscript.

### FURTHER READING

- Brandi ML et al: Multiple endocrine neoplasia type 1: Latest insights. *Endocr Rev* 42:133, 2021.
- Cardoso L et al: Molecular genetics of syndromic and non-syndromic forms of parathyroid carcinoma. *Hum Mutat* 38:1621, 2017.
- Frederiksen A et al: Clinical features of multiple endocrine neoplasia type 4: Novel pathogenic variant and review of published cases. *J Clin Endocrinol Metab* 1:3637, 2019.
- Frost M et al: Current and emerging therapies for PNETs in patients with or without MEN1. *Nat Rev Endocrinol* 14:216, 2018.
- Hannan FM et al: The calcium-sensing receptor in physiology and in calcitropic and noncalcitropic diseases. *Nat Rev Endocrinol* 15:33, 2018.
- Parghane RV et al: Clinical utility of 177 Lu-DOTATATE PRRT in somatostatin receptor-positive metastatic medullary carcinoma of thyroid patients with assessment of efficacy, survival analysis, prognostic variables, and toxicity. *Head Neck* 4:401, 2020.
- Salpea P, Stratakis CA: Carney complex and McCune Albright syndrome: An overview of clinical manifestations and human molecular genetics. *Mol Cell Endocrinol* 386:85, 2014.
- Thakker RV et al: Clinical practice guidelines for multiple endocrine neoplasia type 1 (MEN1). *J Clin Endocrinol Metab* 97:2990, 2012.
- Wells SA Jr et al: Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid* 25:567, 2015.
- Wirth LJ et al: Efficacy of selpercatinib in RET-altered thyroid cancers. *N Engl J Med* 383:825, 2020.

389

## Autoimmune Polyendocrine Syndromes

Peter A. Gottlieb, Aaron W. Michels



Polyglandular deficiency syndromes have been given many different names, reflecting the wide spectrum of disorders that have been associated with these syndromes and the heterogeneity of their clinical presentations. The name used in this chapter for this group of disorders is *autoimmune polyendocrine syndrome* (APS). In general, these disorders are divided into two major categories, APS type 1 (APS-1) and APS type 2 (APS-2). Some groups have further subdivided APS-2 into APS type 3 (APS-3) and APS type 4 (APS-4) depending on the type of autoimmunity involved. For the most part, this additional classification does not clarify our understanding of disease pathogenesis or prevention of complications in individual patients. Importantly, there are many nonendocrine disease associations included in these syndromes, suggesting that although the underlying autoimmune disorder predominantly involves endocrine targets, it does not exclude other tissues. The disease associations found in APS-1 and APS-2 are summarized in Table 389-1. Understanding these syndromes and their disease manifestations can lead to early diagnosis and treatment of additional disorders in patients and their family members.

### APS-1

APS-1 (Online Mendelian Inheritance in Man [OMIM] 240300) has also been called autoimmune polyendocrinopathy–candidiasis–ectodermal

**TABLE 389-1 Disease Associations with Autoimmune Polyendocrine Syndromes**

AUTOIMMUNE POLYENDOCRINE SYNDROME TYPE 1	AUTOIMMUNE POLYENDOCRINE SYNDROME TYPE 2	OTHER AUTOIMMUNE POLYENDOCRINE DISORDERS
<b>Endocrine</b>	<b>Endocrine</b>	IPEX (immune dysfunction polyendocrinopathy X-linked)
Addison's disease	Addison's disease	Thymic tumors
Hypoparathyroidism	Type 1 diabetes	Anti-insulin receptor antibodies
Hypogonadism	<i>Graves' disease or autoimmune thyroiditis</i>	POEMS syndrome
<i>Graves' disease or autoimmune thyroiditis</i>	<i>Hypogonadism</i>	Insulin autoimmune syndrome (Hirata's syndrome)
Type 1 diabetes		Adult combined pituitary hormone deficiency (CPHD) with anti-Pit1 autoantibodies
		Kearns-Sayre syndrome
		DIDMOAD syndrome
<b>Nonendocrine</b>	<b>Nonendocrine</b>	Congenital rubella associated with thyroiditis and/or diabetes
Mucocutaneous candidiasis	Celiac disease, dermatitis herpetiformis	
Chronic active hepatitis	Pernicious anemia	
Pernicious anemia	Vitiligo	
Vitiligo	<i>Alopecia</i>	
Asplenism	<i>Myasthenia gravis</i>	
Ectodermal dysplasia	<i>IgA deficiency</i>	
<i>Alopecia</i>	<i>Parkinson's disease</i>	
<i>Malabsorption syndromes</i>	<i>Idiopathic thrombocytopenia</i>	
<i>IgA deficiency</i>		

*Abbreviations:* DIDMOAD, diabetes insipidus, diabetes mellitus, progressive bilateral optic atrophy, and sensorineural deafness; POEMS, polyneuropathy, organomegaly, endocrinopathy, M-protein, and skin changes.

*Note:* Italics denote less common disorders.

dystrophy (APECED). Mucocutaneous candidiasis, hypoparathyroidism, and Addison's disease form the three major components of this disorder. However, as summarized in Table 389-1, many other organ systems can be involved over time. APS-1 is rare, with <500 cases reported in the literature.

The classical form of APS-1 is an autosomal recessive disorder caused by mutations in the *AIRE* gene (autoimmune regulator gene) found on chromosome 21. This gene is most highly expressed in thymic medullary epithelial cells (mTECs) where it controls the expression of tissue-specific self-antigens (e.g., insulin). Deletion of this regulator leads to decreased expression of tissue-specific self-antigens and is hypothesized to allow autoreactive T cells to avoid central deletion, which normally occurs during T-cell maturation in the thymus. The *AIRE* gene is also expressed in epithelial cells found in peripheral lymphoid organs, but its role in these extrathymic cells remains controversial. To date, >100 mutations have been described in this gene, and there is a higher frequency within certain ethnic groups including Iranian Jews, Sardinians, Finns, Norwegians, and Irish. Recently, several autosomal dominant mutations have been identified and are localized primarily in the PHD1 domain of the *AIRE* gene, rather than the CARD region, where the autosomal recessive mutations have been found. Individuals with this nonclassical form of APS-1 may have a later onset of symptoms and less aggressive disease, without the full spectrum of autoimmune components being expressed.

**TABLE 389-2 Comparison of APS-1 and APS-2**

APS-1	APS-2
Early onset: infancy	Later onset
Siblings often affected and at risk	Multigenerational
Equivalent sex distribution	Females > males affected
Monogenic: <i>AIRE</i> gene, chromosome 21, autosomal recessive	Polygenic: <i>HLA</i> , <i>MICA</i> , <i>PTNP22</i> , <i>CTLA4</i>
Not HLA associated for entire syndrome, some specific component risk	DR3/DR4 associated; other HLA class III gene associations noted
Autoantibodies to type 1 interferons and IL-17 and IL-22	No autoantibodies to cytokines
Autoantibodies to specific target organs	Autoantibodies to specific target organs
Asplenism	No defined immunodeficiency
Mucocutaneous candidiasis	Association with other nonendocrine immunologic disorders like myasthenia gravis and idiopathic thrombocytopenic purpura

*Abbreviations:* APS, autoimmune polyendocrine syndrome; HLA, human leukocyte antigen; IL, interleukin.

**Clinical Manifestations** Classical APS-1 develops very early in life, often in infancy (Table 389-2). Chronic mucocutaneous candidiasis without signs of systemic disease is often the first manifestation. It affects the mouth and nails more frequently than the skin and esophagus. Chronic oral candidiasis can result in atrophic disease with areas suggestive of leukoplakia, which can pose a risk for future carcinoma. The etiology is associated with anticytokine autoantibodies (anti-interleukin [IL] 17A, IL-17F, and IL-22) related to T helper ( $T_H$ ) 17 T cells and depressed production of these cytokines by peripheral blood mononuclear cells. Hypoparathyroidism usually develops next, followed by adrenal insufficiency. The time from development of one component of the disorder to the next can be many years, and the order of disease appearance is variable.

Chronic candidiasis is nearly always present and is not very responsive to treatment. Hypoparathyroidism is found in >85% of cases, and Addison's disease is found in nearly 80%. Gonadal failure appears to affect women more than men (70 vs 25%, respectively), and hypoplasia of the dental enamel also occurs frequently (77% of patients). Other endocrine disorders that occur less frequently include type 1 diabetes (23%) and autoimmune thyroid disease (18%). Nonendocrine manifestations that present less frequently include alopecia (40%), vitiligo (26%), intestinal malabsorption (18%), pernicious anemia (31%), chronic active hepatitis (17%), and nail dystrophy. An unusual and debilitating manifestation of the disorder is the development of refractory diarrhea/obstipation that may be related to autoantibody-mediated destruction of enterochromaffin or enterochromaffin-like cells. The incidence rates for many of these disorders peak in the first or second decade of life, but the individual disease components continue to emerge over time. Therefore, prevalence rates may be higher than originally reported.

**Diagnosis** The diagnosis of APS-1 is usually made clinically when two of the three major component disorders are found in an individual patient. Siblings of individuals with APS-1 should be considered affected even if only one component disorder has been detected due to the known inheritance of the syndrome. Genetic analysis of the *AIRE* gene should be undertaken to identify mutations. Detection of anti-interferon and anti-interferon antibodies can identify nearly 100% of cases with APS-1. The autoantibody arises independent of the type of *AIRE* gene mutation and is not found in other autoimmune disorders.

Diagnosis of each underlying disorder should be done based on their typical clinical presentations (Table 389-3). Mucocutaneous candidiasis may present throughout the gastrointestinal tract, and it may be detected in the oral mucosa or from stool samples. Evaluation by a gastroenterologist to examine the esophagus for candidiasis or secondary stricture may be merited based on symptoms. Other gastrointestinal

**TABLE 389-3 Clinical Features and Recommended Follow-Up for APS-1 and APS-2**

COMPONENT DISEASE	RECOMMENDED EVALUATION
<b>APS-1</b>	
Addison's disease	Sodium, potassium, ACTH, cortisol, 21- and 17-hydroxylase autoantibodies
Diarrhea	History
Ectodermal dysplasia	Physical examination
Hypoparathyroidism	Serum calcium, phosphate, PTH
Hepatitis	Liver function tests
Hypothyroidism/Graves' disease	TSH; thyroid peroxidase and/or thyroglobulin autoantibodies and anti-TSH receptor Ab
Male hypogonadism	FSH/LH, testosterone
Malabsorption	Physical examination, anti-IL-17 and anti-IL-22 autoantibodies
Mucocutaneous candidiasis	Physical examination, mucosal swab, stool samples
Obstipation	History
Ovarian failure	FSH/LH, estradiol
Pernicious anemia	CBC, vitamin B <sub>12</sub> levels
Splenic atrophy	Blood smear for Howell-Jolly bodies; platelet count; ultrasound if positive
Type 1 diabetes	Glucose, hemoglobin A <sub>1c</sub> , diabetes-associated autoantibodies (insulin, GAD65, IA-2, ZnT8)
<b>APS-2</b>	
Addison's disease	21-Hydroxylase autoantibodies, ACTH stimulation testing if positive
Alopecia	Physical examination
Autoimmune hyper- or hypothyroidism	TSH; thyroid peroxidase and/or thyroglobulin autoantibodies, anti-TSH receptor Ab
Celiac disease	Transglutaminase autoantibodies; small intestine biopsy if positive
Cerebellar ataxia	Dictated by signs and symptoms of disease
Chronic inflammatory demyelinating polyneuropathy	Dictated by signs and symptoms of disease
Hypophysitis	Dictated by signs and symptoms of disease, anti-Pit1 autoantibody
Idiopathic heart block	Dictated by signs and symptoms of disease
IgA deficiency	IgA level
Myasthenia gravis	Dictated by signs and symptoms of disease, antiacetylcholinesterase Ab
Myocarditis	Dictated by signs and symptoms of disease
Pernicious anemia	Anti-parietal cell autoantibodies
	CBC, vitamin B <sub>12</sub> levels if positive
Serositis	Dictated by signs and symptoms of disease
Stiff man syndrome	Dictated by signs and symptoms of disease
Vitiligo	Physical examination, NALP-1 polymorphism

**Abbreviations:** Ab, antibody; ACTH, adrenocorticotrophic hormone; APS, autoimmune polyendocrine syndrome; CBC, complete blood count; FSH, follicle-stimulating hormone; IL, interleukin; LH, luteinizing hormone; PTH, parathyroid hormone; TSH, thyroid-stimulating hormone.

manifestations of APS-1, including malabsorption and obstipation, may also bring these young patients to the attention of gastroenterologists for first evaluation. Specific physical examination findings of hyperpigmentation, vitiligo, alopecia, tetany, and signs of hyper- or hypothyroidism should be considered as signs of development of component disorders.

The development of disease-specific autoantibody assays can help confirm disease and also detect risk for future disease. For example, where possible, detection of anticytokine antibodies to IL-17 and IL-22 would confirm the diagnosis of mucocutaneous candidiasis due to APS-1. The presence of anti-21-hydroxylase antibody or anti-17-hydroxylase antibody (which may be found more commonly in adrenal insufficiency associated with APS-1) would confirm the presence or risk for Addison's disease. Other autoantibodies found in type 1 diabetes (e.g.,

anti-GAD65), pernicious anemia, and other component conditions should be screened for on a regular basis (6- to 12-month intervals depending on the age of the subject).

Laboratory tests, including a complete metabolic panel, phosphorous and magnesium, thyroid-stimulating hormone (TSH), adrenocorticotrophic hormone (ACTH; morning), hemoglobin A<sub>1c</sub>, plasma vitamin B<sub>12</sub> level, and complete blood count with peripheral smear looking for Howell-Jolly bodies (asplenism), should also be performed at these time points. Detection of abnormal physical findings or test results should prompt subsequent examinations of the relevant organ system (e.g., presence of Howell-Jolly bodies indicates need for ultrasound of spleen).

## TREATMENT

### APS-1

Therapy of individual disease components is carried out as outlined in other relevant chapters. Replacement of deficient hormones (e.g., adrenal, pancreas, ovaries/testes) will treat most of the endocrinopathies noted. Several unique issues merit special emphasis. Adrenal insufficiency can be masked by primary hypothyroidism by prolonging the half-life of cortisol. The caveat therefore is that replacement therapy with thyroid hormone can precipitate an adrenal crisis in an undiagnosed individual. Hence, all patients with hypothyroidism and the possibility of APS should be screened for adrenal insufficiency to allow treatment with glucocorticoids prior to the initiation of thyroid hormone replacement. Treatment of mucocutaneous candidiasis with ketoconazole in an individual with subclinical adrenal insufficiency may also precipitate adrenal crisis. Furthermore, mucocutaneous candidiasis may be difficult to eradicate entirely. Severe cases of disease involvement may require systemic immunomodulatory therapy, but this is not commonly needed.

### APS-2

APS-2 (OMIM 269200) is more common than APS-1, with a prevalence of 1–2 in 100,000. It has a gender bias and occurs more often in female patients, with a ratio of at least 3:1 compared to male patients. In contrast to APS-1, APS-2 often has its onset in adulthood, with a peak incidence between 20 and 60 years of age. It shows a familial, multigenerational heritage (Table 389-2). The presence of two or more of the following endocrine deficiencies in the same patient defines the presence of APS-2: primary adrenal insufficiency (Addison's disease; 50–70%), Graves' disease or autoimmune thyroiditis (15–69%), type 1 diabetes mellitus (T1D; 40–50%), and primary hypogonadism. Frequently associated autoimmune conditions include celiac disease (3–15%), myasthenia gravis, vitiligo, alopecia, serositis, and pernicious anemia. These conditions occur with increased frequency in affected patients but are also found in their family members (Table 389-3).



**Genetic Considerations** The overwhelming risk factor for APS-2 has been localized to the genes in the human lymphocyte antigen (HLA) complex on chromosome 6. Primary adrenal insufficiency in APS-2, but not APS-1, is strongly associated with both HLA-DR3 and HLA-DR4. Other class I and class II genes and alleles, such as HLA-B8, HLA-DQ2 and HLA-DQ8, and HLA-DR subtypes such as DRB1\*04:04, appear to contribute to organ-specific disease susceptibility (Table 389-4). HLA-B8- and HLA-DR3-associated illnesses include selective IgA deficiency, juvenile dermatomyositis, dermatitis herpetiformis, alopecia, scleroderma, autoimmune thrombocytopenia purpura, hypophysitis, metaphyseal osteopenia, and serositis.

Several other immune genes have been proposed to be associated with Addison's disease and therefore with APS-2 (Table 389-3). The “5.1” allele of a major histocompatibility complex (MHC) gene is an atypical class I HLA molecule MIC-A. The MIC-A5.1 allele has a very strong association with Addison's disease that is not accounted for by linkage disequilibrium with DR3 or DR4. Its role is complicated because certain HLA class I genes can offset this effect. *PTPN22* codes for a polymorphism in a protein tyrosine phosphatase, which acts on

**TABLE 389-4** APS-2 and Other Polyendocrine Disorder Associations

DISEASE	HLA ASSOCIATION	INITIATING FACTOR	MECHANISM	AUTOANTIGEN
Graves' disease	DR3	Iodine Anti-CD52	Antibody	TSH receptor
Myasthenia gravis	DR3, DR7	Thymoma Penicillamine	Antibody	Acetylcholine receptor
Anti-insulin receptor	?	SLE or other autoimmune disease	Antibody	Insulin receptor
Hypoparathyroidism	?	?	Antibody	Cell surface inhibitor
Insulin autoimmune syndrome	DR4, DRB1*0406	Methimazole Sulphydryl-containing drugs	Antibody	Insulin
Celiac disease	DQ2/DQ8	Gluten diet	T cell	Transglutaminase
Type 1 diabetes	DR3/DR4 DQ2/DQ8	?	T cell	Insulin, GAD65, IA-2, ZnT8, IGRP
Addison's disease	DR3/DR4 DRB1*0404	Unknown	T cell	21-Hydroxylase P450-5cc
Thyroiditis	DR3/DQB1*0201 DQA1*0301	Iodine Interferon $\alpha$	T cell	Thyroglobulin Thyroid peroxidase
Pernicious anemia	?	?	T cell	Intrinsic factor H <sub>+</sub> /K <sub>+</sub> ATPase
Vitiligo	?	Melanoma Antigen Immunization	?	Melanocyte
Chromosome dysgenesis—trisomy 21 and Turner's syndrome	DQA1*0301	?	?	Thyroid, islet, transglutaminase
Hypophysitis	?	Pit-1, TDRD6	?	Pituitary, Pit-1

Abbreviations: APS, autoimmune polyendocrine syndrome; SLE, systemic lupus erythematosus; TSH, thyroid-stimulating hormone.

intracellular signaling pathways in both T and B lymphocytes. It has been implicated in T1D, Addison's disease, and other autoimmune conditions. *CTLA4* is a receptor on the T-cell surface that modulates the activation state of the cell as part of the signal 2 pathway (i.e., binding to CD80/86 on antigen presenting cells). Polymorphisms of this gene appear to cause downregulation of the cell surface expression of the receptor, leading to decreased T-cell activation and proliferation. This appears to contribute to Addison's disease and potentially other components of APS-2. Allelic variants of the IL-2R $\gamma$  are linked to development of T1D and autoimmune thyroid disease and could contribute to the phenotype of APS-2 in certain individuals.

**Diagnosis** When one of the component disorders is present, a second associated disorder occurs more commonly than in the general population (Table 389-3). There is controversy as to which tests to use and how often to screen individuals for disease. A strong family history of autoimmunity should raise suspicion in an individual with an initial component diagnosis. The development of a rarer form of autoimmunity, such as Addison's disease, should prompt more extensive screening for other linked disorders, as ~50% of Addison's disease patients develop another autoimmune disease during their lifetime.

Circulating autoantibodies, as previously discussed, can precede the development of clinical disease by many years but would allow the clinician to follow the patient and identify the disease onset at its earliest time point (Tables 389-3 and 389-4). For each of the endocrine components of the disorder, appropriate autoantibody assays are listed and, if positive, should prompt physiologic testing to diagnose clinical or subclinical disease. For Addison's disease, antibodies to 21-hydroxylase antibodies are highly diagnostic for risk of adrenal insufficiency. However, individuals may take many years to develop overt symptoms of hypoadrenalinism. Screening of 21-hydroxylase antibody-positive patients can be performed measuring morning ACTH and cortisol on a yearly basis. Rising ACTH values over time or low morning cortisol in association with signs or symptoms of adrenal insufficiency should prompt testing via the cosyntropin stimulation test (Chap. 386). T1D can be screened for by measuring autoantibodies

directed against insulin, GAD65, IA-2, and ZnT8. Risk for progression to disease is based on the number of antibodies (2 islet autoantibodies with normal glucose tolerance is now defined as stage 1 of T1D as the lifetime risk for developing clinical symptoms is nearly 100%) and metabolic factors (impaired oral glucose tolerance test). Many efforts are ongoing and underway to screen relatives of T1D patients and those in the general population for islet autoantibodies to identify prediabetic individuals who may qualify for intervention trials to change the course of the disease prior to clinical onset.

Screening tests for thyroid disease can include anti-thyroid peroxidase (TPO) or anti-thyroglobulin autoantibodies or anti-TSH receptor antibodies for Graves' disease. Yearly measurements of TSH can then be used to follow these individuals. Celiac disease can be screened for using the anti-tissue transglutaminase (tTG) antibody test. For those <20 years of age, testing every 1–2 years should be performed, whereas less frequent testing is indicated after the age of 20 because the majority of individuals who develop celiac disease have the antibody earlier in life. Positive tTG antibody test results should be confirmed on repeat testing, followed by small-bowel biopsy to docu-

ment pathologic changes of celiac disease. Many patients have asymptomatic celiac disease that is nevertheless associated with osteopenia and impaired growth. If left untreated, symptomatic celiac disease has been reported to be associated with an increased risk of gastrointestinal malignancy, especially lymphoma, and osteoporosis later in life.

The knowledge of the particular disease associations should guide other autoantibody or laboratory testing. A complete history and physical examination should be performed every 1–3 years including CBC, metabolic panel, TSH, and vitamin B<sub>12</sub> levels to screen for most of the possible abnormalities. More specific tests should be based on specific findings from the history and physical examination.

## TREATMENT

### APS-2

With the exception of Graves' disease, the management of each endocrine component of APS-2 involves hormone replacement and is covered in detail in the chapters on adrenal (Chap. 386), thyroid (Chap. 382), gonadal (Chaps. 391 and 392), and parathyroid diseases (Chap. 410). As noted for APS-1, adrenal insufficiency can be masked by primary hypothyroidism and should be considered and treated as discussed above. In patients with T1D, decreasing insulin requirements or hypoglycemia, without obvious secondary causes, may indicate the emergence of adrenal insufficiency. Hypocalcemia in APS-2 patients is more likely due to malabsorption, potentially from undiagnosed celiac disease, than hypoparathyroidism.

Immunotherapy for autoimmune endocrine disease has been reserved for T1D, for the most part, reflecting the lifetime burden of the disease for the individual patient and society. Although several immunotherapies (e.g., modified anti-CD3, rituximab, abatacept, alefacept, low-dose antithymocyte globulin) can prolong the honeymoon phase of T1D, none has achieved long-term success. The anti-CD3 monoclonal antibody (teplizumab) does delay the onset of clinical diabetes by an average of 2 years when administered to individuals with stage 2 T1D (e.g., those with autoantibodies and

impaired glucose tolerance). Active basic and clinical research using new approaches and combination therapy may change the treatment of this disease or other autoimmune conditions that share similar pathways.

## IMMUNE CHECKPOINT INHIBITOR-INDUCED ENDOCRINE AUTOIMMUNITY

Therapies that block immune checkpoints, such as programmed cell death protein 1 (PD-1), its ligand (PD-L1), or CTLA-4, are beneficial immunotherapies for many advanced-stage cancers. These immune checkpoint inhibitors (ICIs) block negative immune regulation, thereby allowing for an immune response directed against tumor cells. However, immune-related adverse events also occur, especially autoimmunity directed toward self-tissues. ICI-induced T1D, thyroid disease, hypophysitis, and adrenal insufficiency have all been reported with these therapies. Hypothyroidism occurs in ~8% and T1D in 1% of those receiving monoclonal antibodies directed against PD-1 or PD-L1, and hypophysitis and adrenal insufficiency occur in <1% of treated patients. These autoimmune side effects can develop during or after therapy, mostly within a few weeks to months following the start of therapy. ICI-induced T1D has a very rapid onset, presents with diabetic ketoacidosis, is permanent, and requires lifelong exogenous insulin therapy for treatment. There is a genetic association with HLA-DR4 and islet autoantibodies in ~40–50% of patients at diagnosis. The pathogenesis is immune mediated as T lymphocyte infiltration has been documented in the pancreatic islets of an ICI-T1D patient. Determining the mechanisms of autoimmune disease development following ICI therapies and developing biomarkers to stratify risk for autoimmune side effects prior to therapy are active areas of research.

### IPEX

Immune dysregulation, polyendocrinopathy, enteropathy, and X-linked disease (IPEX; OMIM 304790) is a rare X-linked recessive disorder. The disease onset is in infancy and is characterized by severe enteropathy, T1D, and skin disease, as well as variable association with several other autoimmune disorders. Many infants die within the first days of life, but the course is variable, with some children surviving for 12–15 years. Early onset of T1D, often at birth, is highly suggestive of the diagnosis because nearly 80% of IPEX patients develop T1D. Although treatment of the individual disorders can temporarily improve the situation, treatment of the underlying immune deficiency is required and includes immunosuppressive therapy generally followed by hematopoietic stem cell transplantation. Transplantation is the only life-saving form of therapy and can be fully curative by normalizing the imbalanced immune system found in this disorder.

IPEX is caused by mutations in the *FOXP3* gene, which is also mutated in the Scurvy mouse, an animal model that shares much of the same phenotype of IPEX patients. The *FOXP3* transcription factor is expressed in regulatory T cells designated CD4+CD25+FOXP3+ (Treg). Lack of this factor causes a profound deficiency of this Treg population and results in rampant autoimmunity due to the lack of peripheral tolerance normally provided by these cells. Certain mutations may lead to varying forms of expression of the full syndrome, and there are rare cases where the *FOXP3* gene is intact but other genes involved in this pathway (e.g., CD25, IL-2R $\gamma$ ) may be causative. Future therapy with autologous CD4+ T cells transfected with a functioning *FOXP3* gene may offer a better long-term outcome than has been seen in those treated with stem cell transplantation.

### THYMIC TUMORS

Thymomas and thymic hyperplasia are associated with several autoimmune diseases, with the most common being myasthenia gravis (44%) and red cell aplasia (20%). Graves' disease, T1D, and Addison's disease may also be associated with thymic tumors. Patients with myasthenia gravis and thymoma may have unique anti-acetylcholine receptor autoantibodies. Most thymomas lack AIRE expression within the thymoma, and this could be a potential factor in the development

of autoimmunity. In support of this concept, thymoma is the one other disease with "frequent" development of anticytokine antibodies and mucocutaneous candidiasis in adults. The majority of tumors are malignant, and temporary remissions of the autoimmune condition can occur with resection of the tumor.

## ANTI-INSULIN RECEPTOR ANTIBODIES

This is a very rare disorder where severe insulin resistance (type B) is caused by the presence of anti-insulin receptor antibodies. It is associated with acanthosis nigricans, which can also be associated with other forms of less severe insulin resistance. About one-third of patients have an associated autoimmune illness such as systemic lupus erythematosus or Sjögren's syndrome. Therefore, the presence of antinuclear antibodies, elevated erythrocyte sedimentation rate, hyperglobulinemia, leukopenia, and hypocomplementemia may accompany the presentation. The presence of anti-insulin receptor autoantibodies leads to marked insulin resistance, requiring >100,000 units of insulin to be given daily with only partial control of hyperglycemia. Patients can also have severe hypoglycemia due to partial activation of the insulin receptor by the antibody. The course of the disease is variable, and several patients have had spontaneous remissions. A therapeutic approach that targets B lymphocytes, including rituximab, cyclophosphamide, and pulse steroids, has been validated in follow-on case reports to induce remission of the disease.

## INSULIN AUTOIMMUNE SYNDROME (HIRATA'S SYNDROME)

The insulin autoimmune syndrome, associated with Graves' disease and methimazole therapy (or other sulphydryl-containing medications), is of particular interest due to a remarkably strong association with a specific HLA haplotype. Such patients with elevated titers of anti-insulin antibodies frequently present with hypoglycemia. In Japan, the disease is restricted to HLA-DR4-positive individuals with DRB1\*04:06, while Caucasian patients predominantly have DRB1\*04:03 (which is related to DRB1\*04:06). In Hirata's syndrome, the anti-insulin antibodies are often polyclonal. Discontinuation of the medication generally leads to resolution of the syndrome over time. There are very rare cases of insulin autoimmune syndrome not associated with sulphydryl-containing medications that result in profound, life-threatening hypoglycemia. Treatment involves treating the underlying condition that causes anti-insulin antibodies, such as a B lymphocyte lymphoma (tend to have monoclonal insulin antibodies) or systemic lupus erythematosus. As hypoglycemia is profound when elevated titers of high affinity insulin antibodies bind secreted insulin and then release it into circulation, treatment that begins with high-dose glucocorticoids and rituximab to target B lymphocytes has been shown to be effective.

## POEMS SYNDROME

POEMS (polyneuropathy, organomegaly, endocrinopathy, *M*-protein, and skin changes; also known as Crow-Fukase syndrome; OMIM 192240) patients usually present with a progressive sensorimotor polyneuropathy, diabetes mellitus (50%), primary gonadal failure (70%), and a plasma cell dyscrasia with sclerotic bony lesions. Associated findings can be hepatosplenomegaly, lymphadenopathy, and hyperpigmentation. Patients often present in the fifth to sixth decade of life and have a median survival after diagnosis of <3 years. The syndrome is assumed to be secondary to circulating immunoglobulins, but patients have excess vascular endothelial growth factor as well as elevated levels of other inflammatory cytokines such as IL1-, IL-6, and tumor necrosis factor  $\alpha$ . Patients have been treated with thalidomide, and more recently lenalidomide, leading to a decrease in vascular endothelial growth factor. Hyperglycemia responds to small, subcutaneous doses of insulin. The hypogonadism is due to primary gonadal disease with elevated plasma levels of follicle-stimulating hormone and luteinizing hormone. Temporary resolution of the features of POEMS, including normalization of blood glucose, may occur after radiotherapy for localized plasma cell lesions of bone or after chemotherapy, lenalidomide and dexamethasone, or autologous stem cell transplantation.

390 Sex Development

**Courtney Finlayson, J. Larry Jameson  
John C. Achermann**



Sex development begins in utero but continues into young adulthood with the achievement of sexual maturity and reproductive capability. The major determinants of sex development can be divided into three components: chromosomal sex, gonadal sex (sex determination), and phenotypic sex (sex differentiation) (**Fig. 390-1**). Variations at each of these stages can result in differences (or disorders) of sex development (DSDs) (**Table 390-1**). In the newborn period, ~1 in 5000 babies undergo investigation because of atypical or ambiguous genitalia. Urgent assessment is indicated, because some causes such as congenital adrenal hyperplasia (CAH) can be associated with life-threatening adrenal crises. An experienced multidisciplinary team is important for counseling, planning appropriate investigations, discussing long-term well-being, supporting parents, and providing clear communication about the diagnosis and management options. DSDs can also present at other ages and to a range of health professionals (**Table 390-2**). Subtler forms of gonadal dysfunction (e.g., Klinefelter syndrome [KS], Turner syndrome [TS]) often are diagnosed later in life by internists. Because DSDs are associated with a variety of psychological, reproductive, and potential medical consequences, an open dialogue must be established between the patient and health care providers to ensure continuity and attention to these issues across the life span. Gender variance and gender dysphoria are more common among some individuals with DSD than in the general population. Thus, attention to and comfort discussing gender identity is important. Support groups also have a valuable role to play for many patients and families.

Care of individuals with DSDs has evolved from primarily focusing on medical and surgical intervention to “genitalia” to a more holistic approach involving medical, surgical, and psychosocial care, acknowledging that the best way to care for individuals with DSD is not always clear and should be individualized. This includes many controversies, particularly concerning whether genitoplasty or prophylactic gonadectomy in selected conditions should be performed for infants and young children prior to the age of consent. Accepted nomenclature is also controversial. Previous terms such as *intersex* and *hermaphrodite* were

## **GLOBAL CONSIDERATIONS**

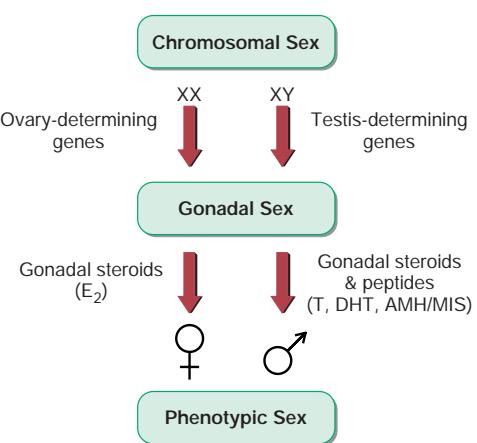
Identification of these syndromes requires access to central laboratories with the ability to detect unique autoantibodies and to sequence the specific genes that may underlie these disorders. Early recognition of the clinical features of these disorders and timely referral and/or consultation with tertiary care centers to confirm the diagnosis and initiate therapy are important to improving outcomes. The *AIRE* recessive gene mutations found in APS-1 were originally described in high frequency in several populations including Finns, Iranian Jews, Sardinians, Norwegians, and Irish. Although individuals from many other countries have now been found to have these mutations and the newly identified dominant *AIRE* gene mutations, understanding the frequency in the background population may raise the clinicians level of suspicion for these rare disorders. Hirata's syndrome was originally reported in Japanese populations but also may be found in other populations, as noted.

## FURTHER READING

- Anderson MS, Su MA: AIRE expands: New roles in immune tolerance and beyond. *Nat Rev Immunol* 16:247, 2016.

Husebye ES et al: Autoimmune polyendocrine syndromes. *N Engl J Med* 378:1132, 2018.

Postow MA et al: Immune-related adverse events associated with immune checkpoint blockade. *N Engl J Med* 378:158, 2018.



**FIGURE 390-1** Sex development can be divided into three major components: chromosomal sex, gonadal sex, and phenotypic sex. AMH, anti-müllerian hormone also known as Müllerian-inhibiting substance, MIS; DHT, dihydrotestosterone; T, testosterone.

**TABLE 390-1 Classification of Disorders of Sex Development (DSD)<sup>a</sup>**

SEX CHROMOSOME DSD	46,XY DSD (SEE TABLE 390-3)	46,XX DSD (SEE TABLE 390-4)
47,XXX (Klinefelter syndrome and variants)	<b>Disorders of gonadal (testis) development</b> Complete or partial gonadal dysgenesis (e.g., SRY, SOX9, SF1, WT1, DMRT1, DHH, GATA4, ZFPIM2, MAP3K1, ESR2, ZNRF3, SOX8, DHX37) Impaired fetal Leydig cell function (e.g., SF1/NR5A1, CXorf6/ MAMLD1, HHAT, SAMD9)	<b>Disorders of gonadal (ovary) development</b> Gonadal dysgenesis Ovotesticular DSD Testicular DSD (e.g., SRY+, dup SOX9, RSP01, SF1/NR5A1, NR2F2, WT1)
45,X (Turner syndrome and variants)	Ovotesticular DSD Testis regression	<b>Androgen excess</b> Fetal 3β-Hydroxysteroid dehydrogenase II (HSD3B2) 21-Hydroxylase (CYP21A2) P450 oxidoreductase (POR) 11β-Hydroxylase (CYP11B1)
45,X/46,XY mosaicism (mixed gonadal dysgenesis)	<b>Disorders in androgen synthesis or action</b> Disorders of androgen biosynthesis LH receptor ( <i>LHCGR</i> ) Smith-Lemli-Optiz syndrome ( <i>DHCR7</i> ) Steroidogenic acute regulatory ( <i>StAR</i> ) protein Cholesterol side-chain cleavage ( <i>CYP11A1</i> ) 3β-Hydroxysteroid dehydrogenase II (HSD3B2) 17α-Hydroxylase/17,20-lyase ( <i>CYP17A1</i> ) P450 oxidoreductase (POR) Cytochrome b5 ( <i>CYB5A</i> ) 17β-Hydroxysteroid dehydrogenase III (HSD17B3) 5α-Reductase II (SRD5A2) Aldo-keto reductase 1C2 (AKR1C2)	Fetoplacental Aromatase deficiency (CYP19) Oxidoreductase deficiency (POR) Maternal Maternal virilizing tumors (e.g., luteomas) Androgenic drugs
46,XX/46,XY (chimerism/mosaicism)	Disorders of androgen action Androgen insensitivity syndrome Drugs and environmental modulators <b>Other</b> Syndromic associations of male genital development Associated with fetal growth restriction Persistent müllerian duct syndrome Vanishing testis syndrome Isolated hypospadias Congenital hypogonadotropic hypogonadism Cryptorchidism Environmental influences	Other Syndromic associations (e.g., cloacal anomalies) Müllerian agenesis/hypoplasia (e.g., MRKH) Uterine abnormalities (e.g., MODY5) Vaginal atresia (e.g., McKusick-Kaufman) Labial adhesions

<sup>a</sup>Some experts and patient advocacy groups prefer to define DSD as *differences of sex development* rather than *disorders of sex development*.

Abbreviations: LH, luteinizing hormone; MODY, maturity-onset diabetes of the young; MRKH, Mayer-Rokitansky-Küster-Hauser syndrome.

Source: Reproduced with permission from IA Hughes et al: Consensus statement on management of intersex disorders. J Pediatr Urol 2:148, 2006.

changed by the 2006 Consensus Statement to *disorder of sex development* and *ovotesticular DSD*, but these terms are not universally accepted.

## SEX DEVELOPMENT

*Chromosomal sex*, defined by a karyotype, describes the X and/or Y chromosome complement (46,XY; 46,XX) established at the time of fertilization. The presence of a normal Y chromosome determines

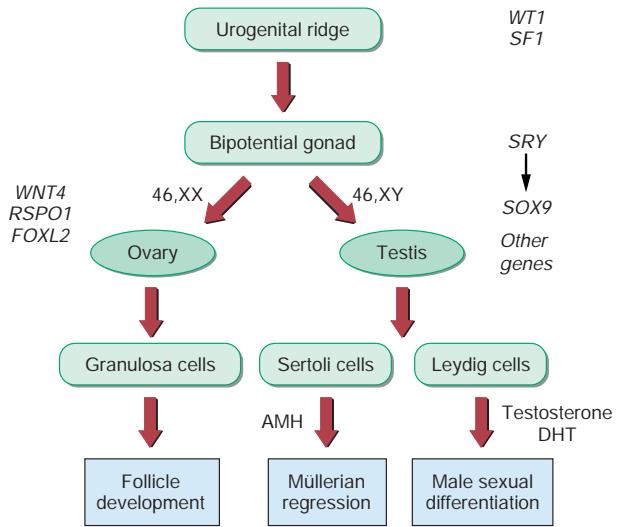
that testis development will occur even in the presence of multiple X chromosomes (e.g., 47,XXX). Loss of an X chromosome impairs gonad development (45,X or 45,X/46,XY mosaicism). Fetuses with no X chromosome (45,Y) are not viable.

*Gonadal sex* refers to the histologic and functional characteristics of gonadal tissue as testis or ovary. The embryonic gonad is initially “bisexual” and can develop (from ~42 days after conception) into

**TABLE 390-2 Presentation of Disorders of Sex Development (DSD) at Different Stages of Life**

PRESENTATION	FEATURES	PROFESSIONAL	EXAMPLES
Prenatal	Karyotype-phenotype discordance	Obstetrician; fetal medicine	Many
Neonatal	Atypical genitalia Salt-losing crisis	Obstetrician; neonatal medicine Pediatrician	Many CAH (CYP21)
Childhood	Hernia Androgenization Poor growth Associated features	Surgeon Endocrinologist Pediatrician Oncologist/nephrologist	CAIS CAH (CYP21, CYP11B1) Turner, 45,X/46,XY Wilms' tumor
Puberty	Androgenization Estrogenization Absent puberty	Endocrinologist Endocrinologist Endocrinologist	17β-HSD, 5α-reductase, SF1 Ovotestis Gonadal dysgenesis, CAH (CYP17A1), Turner
Post-puberty	Amenorrhea	Gynecologist	CAIS
Adult	Infertility	Andrologist	Klinefelter, 45,X/46,XY, SF1

Abbreviations: CAH, congenital adrenal hyperplasia; CAIS, complete androgen insensitivity syndrome; 17β-HSD, 17β-hydroxysteroid dehydrogenase deficiency; SF1, steroidogenic factor 1 (NR5A1).



**FIGURE 390-2** The genetic regulation of gonadal development. See text for additional genes involved. AMH, anti-müllerian hormone (müllerian-inhibiting substance); DHT, dihydrotestosterone; *FOXL2*, forkhead transcription factor L2; *RSP01*, R-spondin 1; *SF1*, steroidogenic factor 1 (also known as *NR5A1*); *SOX9*, *SRY*-related HMG-box gene 9; *SRY*, sex-determining region on the Y chromosome; *WNT4*, wingless-type MMTV integration site 4; *WT1*, Wilms' tumor-related gene 1.

either a testis or an ovary (Fig. 390-2). Testis development is initiated by expression of the gene *SRY* (sex-determining region on the Y chromosome). Disruption of *SRY* prevents testis development in 46,XY individuals, whereas translocation of *SRY* in 46,XX individuals induces testis development and a male phenotype. The main target of *SRY* is *SOX9* (*SRY*-related HMG-box gene 9). *SOX9* is upregulated in the developing testis but is suppressed in the ovary. Many other genes are involved in testis development, including in Sertoli cell maturation and Leydig cell differentiation/steroidogenesis. In addition to transcription factors, these genes encode an array of signaling molecules and paracrine growth factors, some of which influence other organ systems. For example, *WT1* (Wilms' tumor-related gene 1) acts early in the genetic pathway and also regulates kidney development, whereas steroidogenic factor 1 (*SF1*, *NR5A1*) influences both gonad and adrenal development. Pathogenic variants causing loss of function of *SF1* are found in ~10% of XY patients with gonadal dysgenesis and impaired androgenization. Of note, duplication of a related gene *DAX1/NR0B1* impairs testis development, revealing the exquisite sensitivity of the testis-determining pathway to gene dosage effects.

Although ovarian development once was considered a “default” genetic pathway, it is now clear that specific genes are expressed during the earliest stages of ovary development. Some of these factors may repress testis development (e.g., *WNT4*, R-spondin-1) (Fig. 390-2). Once the ovary has formed, additional factors are required for normal follicular development (e.g., follicle-stimulating hormone [FSH] receptor). Steroidogenesis in the ovary requires the development of follicles that contain granulosa cells and theca cells surrounding the oocytes (Chap. 392). Thus, there is relatively limited ovarian steroidogenesis until puberty.

Germ cells also develop in a sex dimorphic manner. In the developing ovary, primordial germ cells (PGCs) proliferate and enter meiosis, whereas they proliferate and then undergo mitotic arrest in the developing testis. PGC entry into meiosis is potentially initiated by retinoic acid. The developing testis produces high levels of CYP26B1, an enzyme that degrades retinoic acid, preventing PGC entry into meiosis. Approximately 7 million germ cells are present in the fetal ovary in the second trimester, and 1 million remain at birth. Only 400 are ovulated during a woman's reproductive life span (Chap. 392).

**Phenotypic sex** refers to the structures of the external and internal genitalia and secondary sex characteristics. In addition to bipotential

gonads in the fetuses, they also initially possess internal and external genitalia, which can develop along a male- or female-typical pathway, with sex-specific development occurring as a result of hormone action (Fig. 390-3). The developing testis releases anti-müllerian hormone (AMH; also known as müllerian-inhibiting substance [MIS]) from Sertoli cells and testosterone from Leydig cells. AMH acts through specific receptors to cause regression of the müllerian structures from 60–80 days after conception. At ~60–140 days after conception, testosterone supports the maintenance of wolffian structures, including the epididymides, vasa deferentia, and seminal vesicles. Testosterone is the precursor for dihydrotestosterone (DHT), a potent androgen that promotes development of the external genitalia, including the penis and scrotum (60–100 days, and thereafter) (Fig. 390-3). The urogenital sinus develops into the prostate and prostatic urethra in the male and into the urethra and lower portion of the vagina in the female. The genital tubercle becomes the glans penis in the male and the clitoris in the female. The urogenital swellings form the scrotum or the labia majora, and the urethral folds fuse to form the shaft of the penis and the male urethra or the labia minora. In the female, wolffian ducts regress and the müllerian ducts form the fallopian tubes, uterus, and upper segment of the vagina. A female phenotype will develop in the absence of the gonad, but estrogen is needed for maturation of the uterus and breast at puberty.

The prenatal hormone environment is likely one of many factors influencing aspects of gender identity and behavior. This is an area of ongoing research and is beyond the scope of this chapter.

## DISORDERS OF CHROMOSOMAL SEX

Variations in sex chromosome number and structure can present as DSDs (e.g., 45,X/46,XY), KS (47,XXY) and TS (45,X) do not usually present with genital ambiguity but are associated with gonadal dysfunction (Table 390-3).

### KLINEFELTER SYNDROME (47,XXY)

**Pathophysiology** The classic form of KS (47,XXY) occurs after meiotic nondisjunction of the sex chromosomes during gametogenesis (40% during spermatogenesis, 60% during oogenesis). Other forms of KS (including mosaic 46,XY/47,XXY [10–20%], 48,XXYY, and 48,XXXYY) are less common. KS has an incidence of at least 1 in 1000 men, but ~75% of cases are not diagnosed. Of those diagnosed, historically only 10% were identified prepubertally. However, the advent of noninvasive prenatal testing is leading to increased detection at an earlier age.

**Clinical Features** KS is most commonly characterized by small testes, infertility, gynecomastia, tall stature/increased leg length, and hypogonadism in phenotypic males. At birth, most infants with KS do not have clinical features, although there are higher rates of cryptorchidism and hypospadias. Most patients present in puberty with arrested pubertal development caused by testicular insufficiency. Others are diagnosed after puberty, based on low androgens, gynecomastia, or infertility. Testes are small and firm (median length 2.5 cm [4 mL volume]; almost always <3.5 cm [12 mL]) and typically seem appropriately small for the degree of androgenization. Biopsies are not usually necessary but typically reveal seminiferous tubule hyalinization and azoospermia. Other clinical features of KS are listed in Table 390-3. Plasma concentrations of FSH and luteinizing hormone (LH) are increased in most adults with 47,XXY, and plasma testosterone is decreased (50–75%), reflecting primary gonadal insufficiency. Estradiol is often increased, resulting in gynecomastia (Chap. 391). Patients with mosaic forms of KS have less severe clinical features, have larger testes, and sometimes achieve spontaneous fertility.

## TREATMENT

### Klinefelter Syndrome

Growth, endocrine function, and bone mineralization should be monitored, especially from adolescence. Educational and psychological support is important for many individuals with

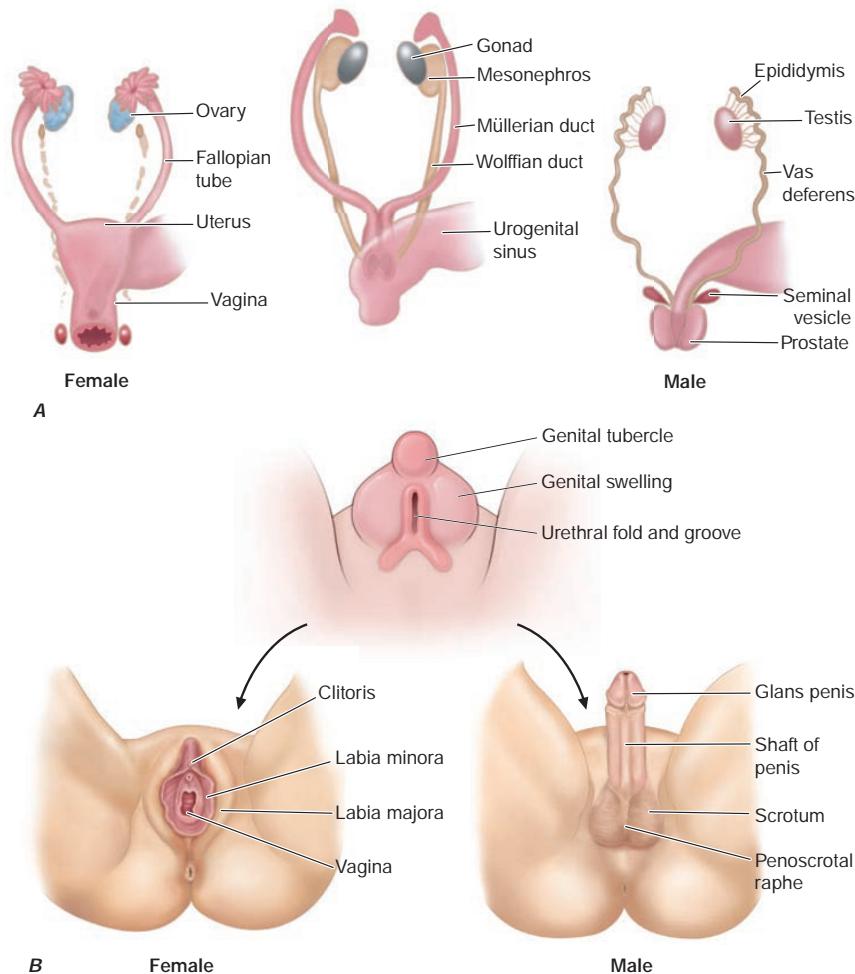


FIGURE 390-3 Sex development. **A.** Internal urogenital tract. **B.** External genitalia.

KS. Androgen supplementation improves virilization, libido, energy, hypofibrinolysis, and bone mineralization in men with low testosterone levels but may occasionally worsen gynecomastia ([Chap. 391](#)). Gynecomastia can be treated by surgical reduction if it causes concern ([Chap. 391](#)). Fertility has been achieved by using in vitro fertilization in men with oligospermia or with intracytoplasmic sperm injection (ICSI) after retrieval of spermatozoa by testicular sperm extraction techniques. In specialized centers, successful spermatozoa retrieval using this technique is possible in >50% of men with nonmosaic KS. Results may be better in younger men. After ICSI and embryo transfer, successful pregnancies can be achieved in ~50% of these cases. The risk of transmitting chromosomal anomalies needs to be considered and counseling provided, although this outcome is much less common than originally predicted. Long-term monitoring of men with KS is important given the increased risk of breast cancer, cardiovascular disease, metabolic syndrome, osteoporosis, and autoimmune disorders. Because most men with KS are never diagnosed, it is important that all internists consider this diagnosis in men with these features who might be seeking medical advice for other conditions.

## TURNER SYNDROME (GONADAL DYSGENESIS; 45,X)

**Pathophysiology** TS is caused by complete or partial loss of one X chromosome and affects ~1 in 2500 women. Approximately one-half of women with TS have a 45,X karyotype, ~20% have 45,X/46,XX

mosaicism, and the remainder have structural abnormalities of the X chromosome such as X fragments, isochromosomes, rings, or Y chromosome material. The clinical features of TS result from haploinsufficiency of multiple X chromosomal genes (e.g., short stature homeobox, *SHOX*), either directly or through effects on autosomal gene expression. However, imprinted genes are also proposed to be affected when the inherited X has different parental origins.

**Clinical Features** TS is characterized by female external genitalia, short stature, hypergonadotropic hypogonadism, infertility, and other phenotypic features ([Table 390-3](#)). Infants may present with lymphedema, nuchal folds, low hairline, or left-sided cardiac defects or later in childhood with unexplained growth failure or delayed puberty. Although limited spontaneous pubertal development occurs in up to 30% of girls with TS (10%, 45,X; 60%, 45,X/46,XX) and up to 20% have menarche, the vast majority of women with TS develop complete ovarian insufficiency. Therefore, this diagnosis should be considered in all women who present with primary or secondary amenorrhea and elevated gonadotropin levels.

## TREATMENT

### Turner Syndrome

The management of girls and women with TS requires a multidisciplinary approach to address many potentially affected organ systems according to TS practice guidelines. Individuals require long-term monitoring by an experienced cardiologist to follow

**TABLE 390-3 Possible Associated Clinical Features of Chromosomal Disorders of Sex Development (DSDs)**

DISORDER	COMMON CHROMOSOMAL COMPLEMENT	GONAD	GENITALIA		BREAST DEVELOPMENT
			EXTERNAL	INTERNAL	
Klinefelter syndrome	47,XXY or 46,XY/47,XXY	Hyalinized testes	Male	Male	Gynecomastia
<b>Clinical Features</b>					
Small testes, azoospermia, decreased facial and axillary hair, decreased libido, tall stature and increased leg length, decreased penile length, increased risk of breast tumors, thromboembolic disease, learning difficulties, anxiety, speech delay and decreased verbal IQ, obesity, diabetes mellitus, metabolic syndrome, varicose veins, hypothyroidism, systemic lupus erythematosus, epilepsy					
Turner syndrome	45,X or 45,X/46,XX	Streak gonad or immature ovary	Female	Hypoplastic female	Immature female
<b>Clinical Features</b>					
	Infancy: lymphedema, web neck, shield chest, low-set hairline, cardiac defects and coarctation of the aorta, urinary tract malformations, and horseshoe kidney				
	Childhood: short stature, cubitus valgus, short neck, short fourth metacarpals, hypoplastic nails, micrognathia, scoliosis, otitis media and sensorineural hearing loss, ptosis and amblyopia, multiple nevi and keloid formation, autoimmune thyroid disease, visuospatial learning difficulties				
	Adulthood: absent puberty and primary amenorrhea, hypertension, obesity, dyslipidemia, impaired glucose tolerance and insulin resistance, autoimmune thyroid disease, cardiovascular disease, aortic root dilation, osteoporosis, inflammatory bowel disease, chronic hepatic dysfunction, increased risk of colon cancer, hearing loss				
45,X/46,XY mosaicism	45,X/46,XY	Testis or streak gonad	Variable	Variable	Usually male
<b>Clinical Features</b>					
Short stature, increased risk of gonadal tumors, some Turner syndrome features					
Ovotesticular DSD	46,XX/46,XY	Testis and ovary or ovotestis	Variable	Variable	Gynecomastia
<b>Clinical Features</b>					
Possible increased risk of gonadal tumors					

congenital heart defects (CHDs) (30%) (bicuspid aortic valve, 30–50%; coarctation of the aorta, 30%; aortic root dilation, 5%), antibiotic prophylaxis for dental or surgical procedures, and serial magnetic resonance imaging (MRI) of aortic root dimensions, as progressive aortic root dilation is associated with increased risk of aortic dissection. Individuals found to have congenital renal and urinary tract malformations (30%) are at risk for urinary tract infections, hypertension, and nephrocalcinosis. Hypertension can occur independently of cardiac and renal malformations and should be monitored and treated as in other patients with essential hypertension. Regular assessment of thyroid function, weight, dentition, hearing, speech, vision, and educational issues should be performed during childhood. Counseling about long-term growth and fertility issues should be provided. Patient support groups are active throughout the world and can play an invaluable role.

Short stature is common, and untreated final height rarely exceeds 150 cm in nonmosaic 45,X TS. Recombinant growth hormone has been used in an attempt to increase growth, sometimes with oxandrolone in older children. Girls with evidence of ovarian insufficiency require estrogen replacement to induce breast and uterine development, support growth, and maintain bone mineralization. Most physicians now initiate low-dose estrogen therapy to induce puberty at an age-appropriate time (~11 years). Doses of estrogen are increased gradually to allow development over a 2- to 4-year period. Progestins are added later to regulate withdrawal bleeds. A very small percentage of women with TS have had spontaneous pregnancy, whereas others have achieved successful pregnancy after ovum donation and in vitro fertilization, but the risks of cardiac complications are high, and expert counseling and management are needed. Long-term follow-up of women with TS includes careful surveillance of sex hormone replacement and reproductive function, bone mineralization, cardiac function and aortic root dimensions, blood pressure, weight and glucose tolerance, hepatic and lipid profiles, thyroid function, skin examination, and hearing. This service is provided by a dedicated TS clinic in some centers.

## 45,X/46,XY MOSAICISM

The phenotype of individuals with 45,X/46,XY mosaicism (sometimes called *mixed gonadal dysgenesis*) can vary considerably. Some have a predominantly female phenotype (see TS above). Most 45,X/46,XY individuals have a male phenotype and testes, and the diagnosis is made incidentally after amniocentesis or during investigation of infertility. In practice, most newborns referred for assessment have atypical genitalia and variable somatic features. There is often marked asymmetry, with a streak gonad and hemiuterus on one side and a partially descended dysgenetic testis and hemiscrotum on the other side. Many children are raised as boys, but in some children, sex designation (whether to raise the baby as male or female) must be decided by parents and the multidisciplinary team. In these children, gender identity may be harder to predict. There is an increased risk of germ cell cancer (GCC), up to 35% in intraabdominal gonads, so prophylactic removal of intraabdominal gonads is usually considered. Individuals raised as males often have reconstructive surgery for hypospadias and removal of dysgenetic or streak gonads if the gonads cannot be brought down into the scrotum. Scrotal testes can be preserved but require regular examination for tumor development and sonography at the time of puberty. Biopsy for carcinoma in situ is recommended in adolescence, and testosterone supplementation may be required to support androgenization in puberty or if low testosterone is detected in adulthood. As 45,X/46,XY mosaicism can be associated with other features (e.g., cardiac, renal), individuals should be monitored according to TS guidelines. Infertility is typical, but non-azoospermia or focal spermatogenesis has been reported, highlighting the importance of individualized approaches to management.

## OVOTESTICULAR DSD

Ovotesticular DSD (OTDSD) is a condition in which an individual has both ovarian and testicular tissue, either by having both an ovary and a testis or by having an ovotestis. Most individuals with this diagnosis have a 46,XX karyotype (especially in individuals of African ancestry), although 46,XX/46,XY chimerism and rarely a 46,XY karyotype is also possible. OTDSD usually presents with atypical genitalia at birth

and sometimes breast development, cyclical hematuria, and/or phallic development at puberty. Progressive regression of the ovarian and/or testicular component can occur over time. Gender identity varies in OTDSD but often aligns with assigned sex. Risk of GCC is also elevated in OTDSD (~3%) and may occur in the ovarian or testicular component. Infertility is typical (especially in 46,XX testes with no Y chromosome), but births have occurred via ovum or sperm from individuals with other forms of OTDSD.

## DISORDERS OF GONADAL AND PHENOTYPIC SEX

Disorders of gonadal and phenotypic sex can result in reduced androgen production or action in individuals with a 46,XY karyotype (46,XY DSD) or excess androgen production in individuals with a 46,XX karyotype (46,XX DSD) (Table 390-1). These conditions cover a spectrum of phenotypes ranging from phenotypic females with a Y chromosome to phenotypic males with a 46,XX karyotype to individuals with atypical genitalia. Karyotype is a useful starting investigation for diagnosis but does not define an individual's gender.

### 46,XY DSD

Underandrogenization of the 46,XY fetus reflects defects in androgen production or action. It can result from disorders of testis development, defects of androgen synthesis, or resistance to testosterone and DHT (Table 390-1).

**Disorders of Testis Development • TESTICULAR DYSGENESIS** *Complete gonadal dysgenesis* (CGD, Swyer's syndrome) is associated with streak gonads, müllerian structures (due to insufficient AMH/MIS secretion), and a complete absence of androgenization. Phenotypic females with this condition often present because of absent pubertal development and are found to have a 46,XY karyotype. Serum sex steroids, AMH/MIS, and inhibin B are low, and LH and FSH are elevated. Individuals with CGD typically identify as female. The risk of GCC is high, and intraabdominal gonads should be removed. In contrast, patients with *partial gonadal dysgenesis* (PGD, dysgenetic testes) may produce enough MIS to regress the uterus and sufficient testosterone for partial androgenization and, therefore, usually present in the newborn period with atypical genitalia, highlighting the spectrum of features that are typically seen with many DSDs.

Testicular dysgenesis can result from disruption of testis-promoting genes (e.g., *WT1*, *SF1*, *SRY*, *SOX9*, *MAP3K1*, *DHH*, *DHX37*, and others) or, rarely, duplication of chromosomal loci containing "antitestis" genes (e.g., *DAX1*) (Table 390-4). Among these, deletions or mutations of *SRY* and heterozygous mutations of *SF1* (*NR5A1*) or *DHX37* appear to be most common but still account collectively for <30% of cases. Associated clinical features may be present, reflecting additional functional roles for these genes. For example, renal dysfunction occurs in patients with specific *WT1* mutations (Denys-Drash and Frasier's syndromes), primary adrenal insufficiency occurs in a minority of patients with disruption of *SF1*, and severe cartilage abnormalities (campomelic dysplasia) are the predominant clinical feature of pathogenic variants in *SOX9*. A family history of DSD, hypospadias, infertility, or early menopause is important because variations in some genes (e.g., *SF1*/*NR5A1*, *SOX8*) can be associated with a range of reproductive phenotypes. *SF1* variants are sometimes inherited from a mother in a sex-limited dominant manner (which can mimic X-linked inheritance), and a woman may later develop primary ovarian insufficiency because of the effect of *SF1* on the ovary. Gender identity can be variable in PGD. Dysgenetic testes have an increased risk of GCC. For descended testes, monitoring via physical examination is appropriate. If testes are intraabdominal and not able to be brought down, they may be removed to prevent GCC (risk up to 35% if intraabdominal). Dysgenetic testes may or may not produce sufficient testosterone for puberty. In those who identify as male, testosterone replacement may be needed. In those who identify as female, estrogen replacement will be needed for female-typical pubertal development and ongoing sex steroid requirements. *Absent (vanishing) testis syndrome (bilateral anorchia)* reflects regression of the testis during development. The absence of müllerian

structures indicates adequate secretion of AMH early in utero. Usually, androgenization of the external genitalia is normal. The etiology is often unknown but sometimes associated with pathogenic variants in *DHX37*. These individuals can be offered testicular prostheses and androgen replacement in adolescence and typically identify as male.

**Disorders of Androgen Synthesis** Defects in the pathway that regulates androgen synthesis (Fig. 390-4) cause underandrogenization of the 46,XY fetus (Table 390-1). Müllerian regression is unaffected because Sertoli cell function is preserved, and no uterus is found. These conditions can present with a spectrum of genital appearances, ranging from female-typical external genitalia or clitoromegaly in some individuals to penoscrotal hypospadias or a small phallus in others.

**LH RECEPTOR** Mutations in the LH receptor (LHCGR) cause Leydig cell hypoplasia and androgen deficiency, due to impaired actions of human chorionic gonadotropin in utero and LH late in gestation and during the neonatal period. As a result, testosterone and DHT synthesis are reduced.

**STEROIDOGENIC ENZYME PATHWAYS** Mutations in *steroidogenic acute regulatory protein* (*StAR*) and *CYP11A1* affect both adrenal and gonadal steroidogenesis (Fig. 390-4) (Chap. 386). Affected individuals (46,XY) usually have severe early-onset salt-losing adrenal failure and a female phenotype, although later-onset milder variants are increasingly reported. Defects in *3-hydroxysteroid dehydrogenase type 2* (*HSD3B2*) also cause adrenal insufficiency in severe cases, but the accumulation of dehydroepiandrosterone (DHEA) has a mild androgenizing effect, resulting in atypical genitalia or hypospadias. Salt loss occurs in many but not all children. Patients with CAH due to *17-hydroxylase* (*CYP17A1*) deficiency have variable underandrogenization and develop hypertension and hypokalemia due to the potent salt-retaining effects of corticosterone and 11-deoxycorticosterone. Patients with complete loss of *17-hydroxylase* function often present as phenotypic females who do not enter puberty and are found to have inguinal testes and hypertension in adolescence. Some mutations in *CYP17* selectively impair 17,20-lyase activity without altering 17-hydroxylase activity, leading to underandrogenization without mineralocorticoid excess and hypertension. Disruption of the coenzyme, *cytochrome b5* (*CYB5A*), can present similarly, and methemoglobinemia is usually present. Mutations in *P450 oxidoreductase* (*POR*) affect multiple steroidogenic enzymes, leading to reduced androgen production and a biochemical pattern of apparent combined 21-hydroxylase and 17-hydroxylase deficiency, sometimes with skeletal abnormalities (Antley-Bixler craniostenosis). Defects in *17-hydroxysteroid dehydrogenase type 3* (*HSD17B3*) and *5-reductase type 2* (*SRD5A2*) interfere with the synthesis of testosterone and DHT, respectively. These conditions are characterized by minimal or absent androgenization in utero, but some phallic development can occur during adolescence due to the action of other enzyme isoforms. Individuals with 5-reductase type 2 deficiency have normal wolffian structures and usually do not develop breast tissue. At puberty, the increase in testosterone induces muscle mass and other virilizing features despite DHT deficiency. DHT gel can improve prepubertal phallic growth in patients raised as male. Prevention of testosterone exposure (by gonadectomy or pubertal suppression) in adolescence and estrogen replacement at puberty can be considered in individuals who identify as female. Disruption of alternative pathways to fetal DHT production might also present with 46,XY DSD (*AKR1C2/AKR1C4*).

**Disorders of Androgen Action • androgen insensitivity syndrome** Pathogenic variants in the androgen receptor cause resistance to androgen (testosterone, DHT) action or the *androgen insensitivity syndrome* (AIS). AIS is a spectrum of disorders that affects at least 1 in 100,000 46,XY individuals. Because the androgen receptor is X-linked, only 46,XY offspring are affected. The condition is usually inherited from a mother who carries the sequence variant but can also arise de novo. XY individuals with *complete AIS* (formerly called *testicular feminization syndrome*) have a female phenotype, normal breast development (due to aromatization of testosterone), a short vagina but no uterus (because AMH/MIS production is normal), scanty pubic

**TABLE 390-4 Selected Genetic Causes of 46,XY Disorders of Sex Development (DSDs)**

GENE	INHERITANCE	GONAD	UTERUS	EXTERNAL GENITALIA	ASSOCIATED FEATURES
<b>Disorders of Testis Development</b>					
WT1	AD	Dysgenetic testis	+/-	Female or ambiguous	Wilms' tumor, renal abnormalities, gonadal tumors (WAGR, Denys-Drash and Frasier's syndromes)
SF1/NR5A1	AR/AD (SL)	Dysgenetic testis/Leydig dysfunction	+/-	Female, ambiguous or male	Primary adrenal failure (rare); hyposplenia (rare); primary ovarian insufficiency in female (46,XX) relatives
SRY	Y	Dysgenetic testis or ovotestis	+/-	Female or ambiguous	
SOX9	AD	Dysgenetic testis or ovotestis	+/-	Female or ambiguous	Campomelic dysplasia
MAP3K1	AD (SL)	Dysgenetic testis	+/-	Female or ambiguous	
DHX37	AD	Dysgenetic testis	+/-	Female, ambiguous or male	Testicular regression syndrome
DHH	AR	Dysgenetic testis/Leydig dysfunction	+	Female	Minifascicular neuropathy

Other causes of testicular dysgenesis include: *DMRT1*, *CBX2*, *SOX8*, *ZNRF3*, *GATA4*, and *ZFPM2* (congenital heart disease); *ARX* (X-linked lissencephaly); *TSPYL1* (sudden infant death); *MYRF* (diaphragmatic hernia); *ESR2/NR3A2*, *SAMD9* (MIRAGE syndrome); *MAMLD1*, dupXp21, dup1p35, del9p24, del10q23

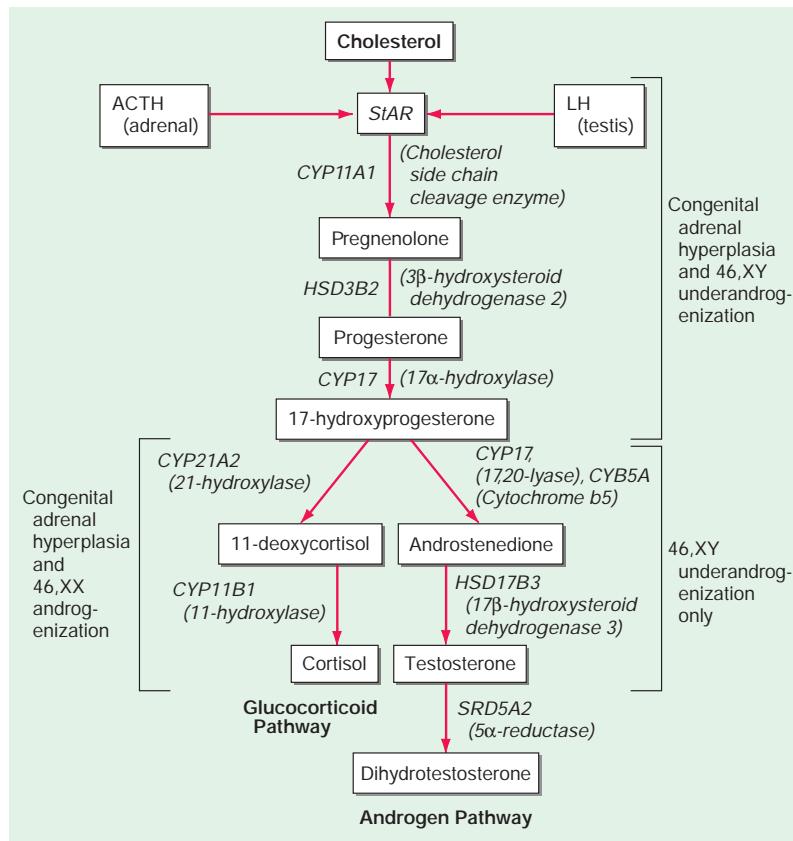
<b>Disorders of Androgen Synthesis</b>					
<i>LHCGR</i>	AR	Testis	-	Female, ambiguous or micropenis	Leydig cell hypoplasia
<i>DHCR7</i>	AR	Testis	-	Variable	Smith-Lemli-Opitz syndrome: coarse facies, second-third toe syndactyly, failure to thrive, developmental delay, cardiac and visceral abnormalities
<i>STAR</i>	AR	Testis	-	Female or ambiguous	Congenital lipid adrenal hyperplasia (primary adrenal insufficiency)
<i>CYP11A1</i>	AR	Testis	-	Female or ambiguous	Primary adrenal insufficiency
<i>HSD3B2</i>	AR	Testis	-	Ambiguous	CAH, primary adrenal insufficiency ± salt loss, partial androgenization due to ↑ DHEA
<i>CYP17A1</i>	AR	Testis	-	Female or ambiguous	CAH, hypertension due to ↑ corticosterone and 11-deoxycorticosterone, except in isolated 17,20-lyase deficiency
<i>CYB5A</i>	AR	Testis	-	Ambiguous	Apparent isolated 17,20-lyase deficiency: methemoglobinemia
<i>POR</i>	AR	Testis	-	Ambiguous or male	Mixed features of 21-hydroxylase deficiency and 17α-hydroxylase/17,20-lyase deficiency, sometimes associated with Antley-Bixler craniosynostosis
<i>HSD17B3</i>	AR	Testis	-	Female or ambiguous	Partial androgenization at puberty, ↑ androstenedione-to-testosterone ratio
<i>SRD5A2</i>	AR	Testis	-	Ambiguous or micropenis	Partial androgenization at puberty, ↑ testosterone-to-dihydrotestosterone ratio
<i>AKR1C2</i> ( <i>AKR1C4</i> )	AR	Testis	-	Female or ambiguous	Decreased fetal DHT production
<b>Disorders of Androgen Action</b>					
Androgen receptor	X	Testis	-	Female, ambiguous, micropenis or normal male	Phenotypic spectrum from complete androgen insensitivity syndrome (female external genitalia) and partial androgen insensitivity (ambiguous) to normal male genitalia and infertility

**Abbreviations:** AD, autosomal dominant; *AKR1C2*, aldo-keto reductase family 1 member 2; AR, autosomal recessive; *ARX*, aristless related homeobox, X-linked; CAH, congenital adrenal hyperplasia; *CBX2*, chromobox homolog 2; *CYB5A*, cytochrome b5; *CYP11A1*, P450 cholesterol side-chain cleavage; *CYP17A1*, cytochrome P450 family 17 subfamily A member 1; *DAX1*, dosage sensitive sex-reversal, adrenal hypoplasia congenita on the X chromosome, gene 1: DHEA, dehydroepiandrosterone; *DHCR7*, sterol 7 $\alpha$  reductase; *DHH*, desert hedgehog; *DMRT1*, doublesex and mab3-related transcription factor 1; *GATA4*, GATA binding protein 4; *HSD17B3*, 17 $\beta$ -hydroxysteroid dehydrogenase type 3; *HSD3B2*, 3 $\beta$ -hydroxysteroid dehydrogenase type 2; *LHR*, LH receptor; *MAP3K1*, mitogen-activated protein kinase kinase kinase 1; MIRAGE, myelodysplasia, infection, restriction of growth, adrenal hypoplasia, genital phenotypes, and enteropathy; *MYRF*, myelin regulatory factor; *POR*, P450 oxidoreductase; *SF1*, steroidogenic factor 1; SL, sex-limited; *SOX8*, SRY-related HMG-box gene 8; *SOX9*, SRY-related HMG-box gene 9; *SRD5A2*, 5 $\alpha$ -reductase type 2; *SRY*, sex-related gene on the Y chromosome; *STAR*, steroidogenic acute regulatory protein; *TSPYL1*, testis-specific Y-encoded-like protein 1; WAGR, Wilms' tumor, aniridia, genitourinary anomalies, and mental retardation; *WNT4*, wingless-type mouse mammary tumor virus integration site, 4; *WT1*, Wilms' tumor-related gene 1; *ZFPM2*, zinc finger protein, multitype 2; *ZNRF3*, zinc and ring finger 3.

and axillary hair, and typically a female gender identity and sex role behavior. Gonadotropins and testosterone levels can be low, normal, or elevated, depending on the degree of androgen resistance and the contribution of estradiol to feedback inhibition of the hypothalamic-pituitary-gonadal axis. AMH/MIS levels in childhood are normal or high. CAIS sometimes presents as inguinal hernias (containing testes) in childhood or more often with primary amenorrhea in late adolescence. In the past, gonadectomy was recommended in childhood, but due to the low risk of malignancy (~2%), increasingly this is delayed, and gonads are left in situ until breast development is complete. Subsequently, the adolescent or young adult should be counseled about the risk of malignancy and the option for gonadectomy (with

estrogen replacement), especially because early detection of premalignant changes by imaging or biomarkers is currently not possible. The use of graded dilators in adolescence is often sufficient to dilate the vagina for sexual activity.

**Partial AIS (Reifenstein's syndrome)** results from androgen receptor mutations that maintain residual function. Patients often present in infancy with penoscrotal hypospadias and undescended testes and with gynecomastia at the time of puberty. Gender identity can be variable. At puberty, testes produce testosterone with variable phenotypic development. For those who identify as male, high-dose testosterone has been given to support development if puberty does not progress, but long-term data are limited. For those raised as female, testosterone



**FIGURE 390-4** Simplified overview of glucocorticoid and androgen synthesis pathways. Defects in *CYP21A2* and *CYP11B1* shunt steroid precursors into the androgen pathway and cause androgenization of the 46,XX fetus. Testosterone is synthesized in the testicular Leydig cells and converted to dihydrotestosterone peripherally. Defects in enzymes involved in androgen synthesis result in underandrogenization of the 46,XY fetus. *StAR*, steroidogenic acute regulatory protein.

effects at puberty can be prevented (by gonadectomy or pubertal suppression) and female-typical puberty induced with estrogen. They also have an increased risk of GCC, again raising the question of if and when to perform gonadectomy. Azoospermia and male-factor infertility also have been described in association with mild loss-of-function mutations in the androgen receptor.

### OTHER DISORDERS AFFECTING 46,XY MALES

*Persistent müllerian duct syndrome* is the presence of a uterus in an otherwise phenotypic male. This condition can result from pathogenic variants in AMH or its receptor (AMHR2). The uterus may be removed, but only if damage to the vasa deferentia and blood supply to the testes can be avoided. *Isolated hypospadias* occurs in ~1 in 250 males. Most cases are idiopathic, although evidence of penoscrotal hypospadias, poor phallic development, and/or bilateral cryptorchidism requires investigation for an underlying DSD (e.g., partial gonadal dysgenesis, mild defect in testosterone action, or even severe forms of 46,XX CAH). Unilateral undescended testes (cryptorchidism) affect >3% of boys at birth. Orchidopexy should be considered if the testis has not descended by 6–9 months of age. Bilateral cryptorchidism occurs less frequently and should raise suspicion of gonadotropin deficiency or DSD. *Syndromic associations* and *intrauterine growth retardation* also occur relatively frequently in association with impaired testicular function or target tissue responsiveness, but the underlying etiology of many of these conditions is unknown.

### 46,XX DSD

Androgenization of the 46,XX fetus occurs when the gonad (ovary) contains androgen-secreting testicular tissue or after increased androgen exposure, which is usually adrenal in origin (Table 390-1).

**46,XX Testicular DSD** Testicular tissue can develop in 46,XX testicular DSD (46,XX males) most often following translocation of *SYR*. This may be diagnosed with karyotype/phenotype discordance or later in life during evaluation for hypogonadism or infertility. Individuals with this condition develop testes with normal testosterone production, leading to external male phenotype in utero, and produce AMH/MIS to regress müllerian structures. They have azoospermia due to lack of the AZF region of the Y chromosome. Progressive testicular regression and hypogonadism are common. Gender identity is typically male.

**46,XX OTDSD** Ovotestes (or testes) can also develop in individuals with a 46,XX karyotype following upregulation of *SOX9* or *SOX3* or defects in *RSP01*, *NR2F2*, *WT1*, or *SF1/NR5A1* (Table 390-5). OTDSD is discussed above under “Disorders of Chromosomal Sex.”

**Increased Androgen Exposure • 21-hydroxylase deficiency (congenital adrenal hyperplasia)** The classic form of 21-hydroxylase deficiency (21-OHD) is the most common cause of CAH (Chap. 386), and it is the most common cause of androgenization in chromosomal 46,XX females (incidence between 1 in 10,000 and 1 in 15,000) (Table 390-5). Affected individuals are homozygous or compound heterozygous for severely disruptive sequence variants in the gene (*CYP21A2*) encoding the enzyme 21-hydroxylase. Impaired 21-hydroxylase activity prevents adrenal glucocorticoid and mineralocorticoid synthesis, thus shunting steroid precursors into the androgen synthesis pathway (Fig. 390-4). Increased androgen synthesis in utero causes androgenization of the 46,XX fetus in the first trimester. Atypical

genitalia are seen at birth, with varying degrees of clitoral enlargement and labial fusion.

A salt-wasting crisis usually manifests between 5 and 21 days of life and is a potentially life-threatening event that requires urgent fluid resuscitation and steroid treatment. Thus, a diagnosis of 21-OHD should be considered in any baby with atypical genitalia with bilateral nonpalpable gonads. Males (46,XY) with 21-OHD have no genital abnormalities at birth but are equally susceptible to adrenal insufficiency and salt-losing crises. Excess androgen production can cause gonadotropin-independent precocious puberty in males with 21-OHD.

Patients with *nonclassic 21-OHD* produce normal amounts of cortisol and aldosterone but at the expense of producing excess androgens. Symptoms may include hirsutism, menstrual dysfunction, subfertility, and recurrent miscarriages. This is one of the most common recessive disorders in humans, with an incidence as high as 1 in 100–500 in many populations and 1 in 27 in Ashkenazi Jews of Eastern European origin.

## TREATMENT

### Congenital Adrenal Hyperplasia

Acute salt-wasting crises require fluid resuscitation, IV hydrocortisone, and correction of hypoglycemia. Once the patient is stabilized, glucocorticoids must be given to correct the cortisol insufficiency and suppress ACTH stimulation, thereby preventing further virilization, rapid skeletal maturation, adrenal rest tumor formation, and the development of polycystic ovaries. Mineralocorticoid replacement may be needed, along with salt supplements in early life. In childhood, treatment is also titrated carefully to prevent

**TABLE 390-5 Selected Genetic Causes of 46,XX Disorders of Sex Development (DSDs)**

GENE	INHERITANCE	GONAD	UTERUS	EXTERNAL GENITALIA	ASSOCIATED FEATURES
<b>Testicular/Ovotesticular DSD</b>					
SRY	Translocation	Testis or ovotestis	-	Male or ambiguous	
SOX9	dup17q24	Testis or ovotestis	-	Male or ambiguous	
SF1/NR5A1 (codon 92)	AD	Testis or ovotestis	+/-	Male or ambiguous	
WT1 (zinc finger 4)	AD	Testis or ovotestis	+/-	Male or ambiguous	
Other causes of testicular/ovotesticular DSD include: COUP-TF2/NR2F2 (congenital heart disease), RSP01 (palmar plantar hyperkeratosis, squamous cell skin carcinoma), WNT4 (SERKAL syndrome), dysregulation/duplication of SOX3 (Xq27)					
<b>Increased Androgen Synthesis</b>					
HSD3B2	AR	Ovary	+	Clitoromegaly	CAH, primary adrenal insufficiency, mild androgenization due to ↑ DHEA
CYP21A2	AR	Ovary	+	Ambiguous	CAH, phenotypic spectrum from severe salt-losing forms associated with adrenal insufficiency to simple virilizing forms with compensated adrenal function, ↑ 17-hydroxyprogesterone
POR	AR	Ovary	+	Ambiguous or female	Mixed features of 21-hydroxylase deficiency and 17α-hydroxylase/17,20-lyase deficiency, sometimes associated with Antley-Bixler craniosynostosis
CYP11B1	AR	Ovary	+	Ambiguous	CAH, hypertension due to ↑ 11-deoxycorticosterone
CYP19	AR	Ovary	+	Ambiguous	Maternal virilization during pregnancy, absent breast development at puberty

**Abbreviations:** ACTH, adrenocorticotropin; AD, autosomal dominant; AR, autosomal recessive; CAH, congenital adrenal hyperplasia; COUP-TF2, chicken ovalbumin upstream promoter transcription factor 2; CYP11B1, 11β-hydroxylase; CYP19, aromatase; CYP21A2, 21-hydroxylase; DHEA, dehydroepiandrosterone; HSD3B2, 3β-hydroxysteroid dehydrogenase type 2; POR, P450 oxidoreductase; RSP01, R-spondin 1; SERKAL, sex reversion, kidneys, adrenal and lung dysgenesis; SF1, steroidogenic factor 1; SOX3, SRY-related HMG-box gene 3; SOX9, SRY-related HMG-box gene 9; SRY, sex-related gene on the Y chromosome; WT1, Wilms' tumor-related gene 1.

impairment of linear growth. In the future, different forms of glucocorticoid replacement and multimodal therapies may improve treatment options. See Chap. 386 for detailed discussion of hormone replacement.

Individuals with 46,XX CAH due to classic 21-OHD historically underwent genitoplasty in infancy, but if and when these procedures should be performed is debated. Concerns have arisen about the importance of assent/consent by the individual for genital surgery, potential long-term side effects related to sexual function and ability to achieve orgasm, and the increased incidence of nonfemale gender identity. Surgical options include vaginoplasty and clitoroplasty. When vaginoplasty is performed in infancy, surgical revision or vaginal dilation may still be needed in adolescence or adulthood and, if deferred, may be necessary for menstrual flow or intercourse. Current clinical practice guidelines recommend that parents be informed of all surgical options including the option to defer surgery. Women with 21-OHD frequently develop polycystic ovaries and have subfertility. The latter occurs due to multiple factors including anatomic barriers, hormone imbalances, and psychological effects of the condition. Preconception genetic counseling is recommended. Due to concerns about fetal neurologic development, prenatal treatment with dexamethasone to prevent androgenization of a fetus is currently not recommended unless in a study protocol that allows long-term follow-up of all children treated.

The treatment of other forms of CAH (including in 46,XY individuals) includes mineralocorticoid and glucocorticoid replacement for salt-losing conditions (e.g., StAR, CYP11A1, HSD3B2), suppression of ACTH drive with glucocorticoids in disorders associated with hypertension (e.g., CYP11B1), and appropriate sex hormone replacement in adolescence and adulthood, when necessary.

**OTHER CAUSES** Increased androgen synthesis can also occur in CAH due to defects in POR, 11-hydroxylase (CYP11B1), and 3-hydroxysteroid dehydrogenase type 2 (HSD3B2) and with mutations in the genes encoding aromatase (CYP19). Increased androgen exposure

in utero can occur with maternal virilizing tumors, luteomas, and ingestion of androgenic compounds.

### OTHER DISORDERS AFFECTING 46,XX FEMALES

*Congenital absence of the vagina* occurs in association with müllerian agenesis or hypoplasia as part of the Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome. This diagnosis should be considered in otherwise phenotypic females with primary amenorrhea. Associated features include renal (agenesis) and cervical spinal abnormalities.

### GLOBAL CONSIDERATIONS

The approach to a child or adolescent with atypical genitalia or another DSD requires cultural sensitivity, as the concepts of sex and gender vary widely around the world. Rare genetic DSDs can occur more frequently in specific populations (e.g., 5-reductase type 2 in the Dominican Republic). Different forms of CAH also show ethnic and geographic variability. In many countries, appropriate biochemical tests may not be readily available, and access to appropriate forms of treatment and support may be limited.

### FURTHER READING

Ahmed SF et al: Turner HE. Society for Endocrinology UK Guidance on the initial evaluation of a suspected difference or disorder of sex development (Revised 2021). Clin Endocrinol (Oxf) 818, 2021.

Cools M et al: Caring for individuals with a difference of sex development (DSD): A Consensus Statement. Nat Rev Endocrinol 14:415, 2018.

Gravholt CH et al: Clinical practice guidelines for the care of girls and women with Turner syndrome: Proceedings from the 2016 Cincinnati International Turner Syndrome Meeting. Eur J Endocrinol 177:G1, 2017.

Merke DP, Auchus RG: Congenital adrenal hyperplasia due to 21-hydroxylase deficiency. N Engl J Med 383:1248, 2020.

Mongan NP et al: Androgen insensitivity syndrome. Best Pract Res Clin Endocrinol Metab 29:569, 2015.

Zitzmann M et al: European Academy of Andrology guidelines on Klinefelter syndrome: Endorsing organization: European Society of Endocrinology. Andrology 9:145, 2021.

The male reproductive system regulates sex differentiation, androgenization, and the hormonal changes that accompany puberty, ultimately leading to spermatogenesis and fertility. Under the control of the pituitary hormones—luteinizing hormone (LH) and follicle-stimulating hormone (FSH)—the Leydig cells of the testes produce testosterone and germ cells are nurtured by Sertoli cells to divide, differentiate, and mature into sperm. During embryonic development, testosterone and dihydrotestosterone (DHT) induce the Wolffian duct and virilization of the external genitalia. During puberty, testosterone promotes somatic growth and the development of secondary sex characteristics. In the adult, testosterone is necessary for spermatogenesis, libido and normal sexual function, and maintenance of muscle and bone mass. This chapter focuses on the physiology of the testes and disorders associated with decreased androgen production, which may be caused by gonadotropin deficiency or by primary testis dysfunction. Infertility occurs in ~5% of men and is increasingly amenable to treatment by hormone replacement or by using sperm transfer techniques. **For further discussion of sexual dysfunction, disorders of the prostate, and testicular cancer, see Chaps. 397, 87, and 88, respectively.**

## DEVELOPMENT AND STRUCTURE OF THE TESTIS

The fetal testis develops from a single bipotential progenitor cell population in the undifferentiated gonad after expression of a genetic cascade that is initiated by the gene encoding SRY (sex-related gene on the Y chromosome) (Chap. 390). SRY, whose expression is regulated by histone modification and DNA methylation, induces differentiation of Sertoli cells, which surround germ cells and, together with peritubular myoid cells, form testis cords that will later develop into seminiferous tubules. Fetal Leydig cells and endothelial cells migrate into the gonad from the adjacent mesonephros but may also arise from interstitial cells that reside between testis cords. Fetal Leydig cells atrophy after birth and do not contribute to the origin of adult Leydig cells, which originate from undifferentiated progenitor cells that appear in the testis after birth and acquire full steroidogenic function during puberty. Testosterone produced by the fetal Leydig cells supports the growth and differentiation of Wolffian duct structures that develop into the epididymis, vas deferens, and seminal vesicles. Testosterone is also converted to DHT (see below), which induces formation of the prostate and the external male genitalia, including the penis, urethra, and scrotum. Testicular descent through the inguinal canal is controlled in part by Leydig cell production of insulin-like factor 3 (INSL3), which acts via a receptor termed *great* (*G* protein-coupled receptor affecting testis descent). Sertoli cells produce Müllerian inhibiting substance (MIS), which causes regression of the Müllerian structures, including the fallopian tube, uterus, and upper segment of the vagina.

## NORMAL MALE PUBERTAL DEVELOPMENT

*Puberty* commonly refers to the maturation of the reproductive axis and the development of secondary sex characteristics. In addition to reproductive hormones, it requires a coordinated response of multiple hormonal systems including metabolic signals (e.g., leptin), as well as the adrenal and growth hormone (GH) axes (Fig. 391-1). The development of secondary sex characteristics is initiated by *adrenarche*, which usually occurs between 6 and 8 years of age when the adrenal gland begins to produce greater amounts of androgens from the zona reticularis, the principal site of dehydroepiandrosterone (DHEA) production. The sex maturation process is greatly accelerated by the activation of the hypothalamic-pituitary axis and the production of gonadotropin-releasing hormone (GnRH). The GnRH pulse generator in the hypothalamus

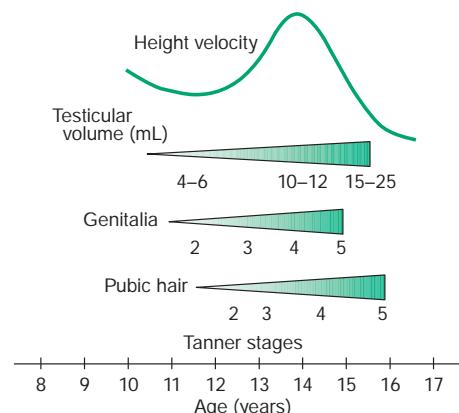


FIGURE 391-1 Pubertal events in males. Sexual maturity ratings for genitalia and pubic hair divided into five stages.

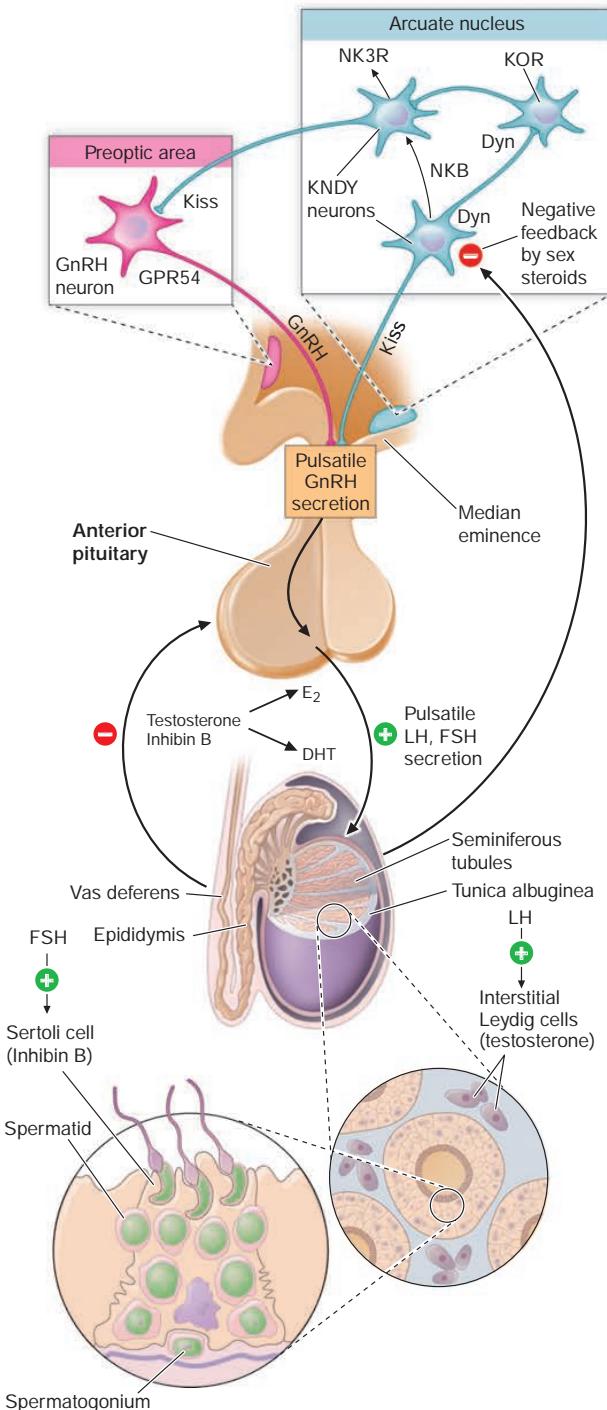
is active during fetal life and early infancy but is restrained until the early stages of puberty by a neuroendocrine brake imposed by the inhibitory actions of glutamate and  $\gamma$ -aminobutyric acid (GABA) in the mediobasal hypothalamus and neuropeptide Y. Although the pathways that initiate reactivation of the GnRH pulse generator at the onset of puberty remain incompletely understood, mounting evidence supports involvement of GPR54, a G protein-coupled receptor that binds an endogenous ligand, kisspeptin. Individuals with mutations of GPR54 fail to enter puberty, and experiments in primates demonstrate that infusion of the ligand is sufficient to induce premature puberty. Kisspeptin signaling plays an important role in mediating the feedback action of sex steroids on gonadotropin secretion and in regulating the tempo of sexual maturation at puberty. Leptin, a hormone produced by adipose cells, plays a permissive role in the resurgence of GnRH secretion at the onset of puberty, as leptin-deficient individuals also fail to enter puberty (Chap. 401). The adipocyte hormone leptin, the gut hormone ghrelin, neuropeptide Y, and kisspeptin integrate the signals originating in energy stores and metabolic tissues with mechanisms that control onset of puberty through regulation of GnRH secretion. Energy deficit and excess and metabolic stress are associated with disturbed reproductive maturation and timing of puberty.

The early stages of puberty are characterized by nocturnal surges of LH and FSH. Growth of the testes is usually the first clinical sign of puberty, reflecting an increase in seminiferous tubule volume. Increasing levels of testosterone deepen the voice and stimulate muscle growth. Conversion of testosterone to DHT leads to growth of the external genitalia and pubic hair. DHT also stimulates prostate and facial hair growth and initiates recession of the temporal hairline. The growth spurt occurs at a testicular volume of ~10–12 mL. GH increases early in puberty and is stimulated in part by the rise in gonadal steroids. GH increases the level of insulin-like growth factor 1 (IGF-1), which enhances linear bone growth. The prolonged pubertal exposure to gonadal steroids (mainly estradiol) ultimately induces epiphyseal closure and limits further bone growth.

## REGULATION OF TESTICULAR FUNCTION

### REGULATION OF THE HYPOTHALAMIC-PITUITARY-TESTIS AXIS IN ADULT MAN

Pulsatile secretion of GnRH in the hypothalamus is regulated by the KNDy neurons through the release of kisspeptin, neuropeptide B (NKB), and dynorphin (Fig. 391-2). Kisspeptin binds to the kisspeptin (GPR54) receptors in the cell bodies of the GnRH neurons as well as in the GnRH nerve terminals in the median eminence to induce pulsatile GnRH secretion into the portal blood. As a component of this autocrine/paracrine loop, NKB released by the KNDy neurons activates NK3R to stimulate kisspeptin release. KNDy neurons also produce dynorphin A, which inhibits basal as well as NKB-stimulated kisspeptin release through the mediation of K-type opioid receptor. The



**FIGURE 391-2 Hypothalamic-pituitary-gonadotropin axis, structure of testis, seminiferous tubule. DHT, dihydrotestosterone; Dyn, dynorphin A; E<sub>2</sub>, 17 $\beta$ -estradiol; FSH, follicle-stimulation hormone; GnRH, gonadotropin-releasing hormone; GPR54, G protein-coupled receptor 54 for kisspeptin; Kiss, kisspeptin; KNDY, kisspeptin, neurokinin B, dynorphin neurons; LH, luteinizing hormone; NKB, neurokinin B; NK3R, neurokinin 3 receptor.**

negative feedback effects of testosterone, estradiol, and progesterone are mediated through KNDY neurons in the preoptic area by inhibition of kisspeptin release.

Hypothalamic GnRH regulates the production of the pituitary gonadotropins LH and FSH (Fig. 391-2). GnRH is released in discrete pulses approximately every 2 h, resulting in corresponding pulses of

LH and FSH. These dynamic hormone pulses account in part for the wide variations in LH and testosterone, even within the same individual. LH acts primarily on the Leydig cell to stimulate testosterone synthesis. The regulatory control of androgen synthesis is modulated by dynamic integration of the feedforward elements exerted on the testis by LH and FSH and the feedback exerted by testosterone and estrogen on both the hypothalamus and the pituitary. FSH acts on the Sertoli cell to regulate spermatogenesis and the production of Sertoli products such as inhibin B, which acts to selectively suppress pituitary FSH. Despite these somewhat distinct Leydig and Sertoli cell-regulated pathways, testis function is integrated at several levels: GnRH regulates both gonadotropins; spermatogenesis requires high levels of testosterone; and numerous paracrine interactions between Leydig and Sertoli cells are necessary for normal testis function.

### THE LEYDIG CELL: ANDROGEN SYNTHESIS

LH binds to its seven-transmembrane G protein-coupled receptor to activate the cyclic AMP pathway. Stimulation of the LH receptor induces steroid acute regulatory (StAR) protein, along with several steroidogenic enzymes involved in androgen synthesis. LH receptor mutations cause Leydig cell hypoplasia or agenesis, underscoring the importance of this pathway for Leydig cell development and function. The rate-limiting process in testosterone synthesis is the transport of intracellular cholesterol by the StAR protein to the inner mitochondrial membrane. Mutations of the StAR protein are associated with congenital lipoid adrenal hyperplasia, a rare form of congenital adrenal hyperplasia (CAH) characterized by very low adrenal and gonadal steroids. Peripheral benzodiazepine receptor, a mitochondrial cholesterol-binding protein, is also an acute regulator of Leydig cell steroidogenesis. The major enzymatic steps involved in testosterone synthesis are summarized in Fig. 391-3. After cholesterol transport into the mitochondrion, the formation of pregnenolone by CYP11A1 (side chain cleavage enzyme) is a limiting enzymatic step. The 17 $\alpha$ -hydroxylase and the 17,20-lipase reactions are catalyzed by a single enzyme, CYP17A1; post-translational modification (phosphorylation) of this enzyme and the presence of specific enzyme cofactors, such as cytochrome P450, confer 17,20-lipase activity selectively in the testis and zona reticularis of the adrenal gland. Although CYP17A1 is able to catalyze the conversion of progesterone to 17 $\alpha$ -hydroxyprogesterone, most of 4-androstenedione in humans is not derived from 17 $\alpha$ -hydroxyprogesterone but rather from the conversion of 17 $\alpha$ -hydroxypregnenolone to DHEA in the 5 $\beta$  pathway and further conversion of DHEA to 4-androstenedione. Abiraterone is a dual inhibitor of 17 $\alpha$ -hydroxylase and 17,20-lipase activities, which play an important role in androgen synthesis in castration-resistant prostate cancers. Testosterone can be converted to the more potent DHT by a family of steroid 5 $\alpha$ -reductase enzymes, or it can be aromatized to estradiol by CYP19 (aromatase). At least two isoforms of steroid 5 $\alpha$ -reductase, SRD5A1 and SRD5A2, have been described; all known patients with 5 $\alpha$ -reductase deficiency have had mutations in SRD5A2, the predominant form in the prostate and the skin. Finasteride predominantly inhibits SRD5A2, whereas dutasteride is a dual inhibitor of both SRD5A1 and SRD5A2. DHT can also be derived through the backdoor pathway in which 17 $\alpha$ -hydroxyprogesterone is converted to androsterone and eventually to DHT. Recent reports of mutations in the AKR1C2/4 genes in undervirilized 46,XY individuals suggest that the backdoor pathway for DHT formation, which was originally described in the tammar wallaby, is active in the human fetal testis. The placental progesterone serves as a substrate for the synthesis of androsterone via the backdoor pathway, which is then converted to DHT in the genital tubercle.

**Testosterone Transport and Metabolism** In males, 95% of circulating testosterone is derived from testicular production (3–10 mg/d). Direct secretion of testosterone by the adrenal and the peripheral conversion of androstenedione to testosterone collectively account for another 0.5 mg/d of testosterone. Only a small amount of DHT (70  $\mu$ g/d) is secreted directly by the testis; most circulating DHT is derived from peripheral conversion of testosterone. Most of the daily production of estradiol (~45  $\mu$ g/d) in men is derived from aromatase-mediated peripheral conversion of testosterone and androstenedione.

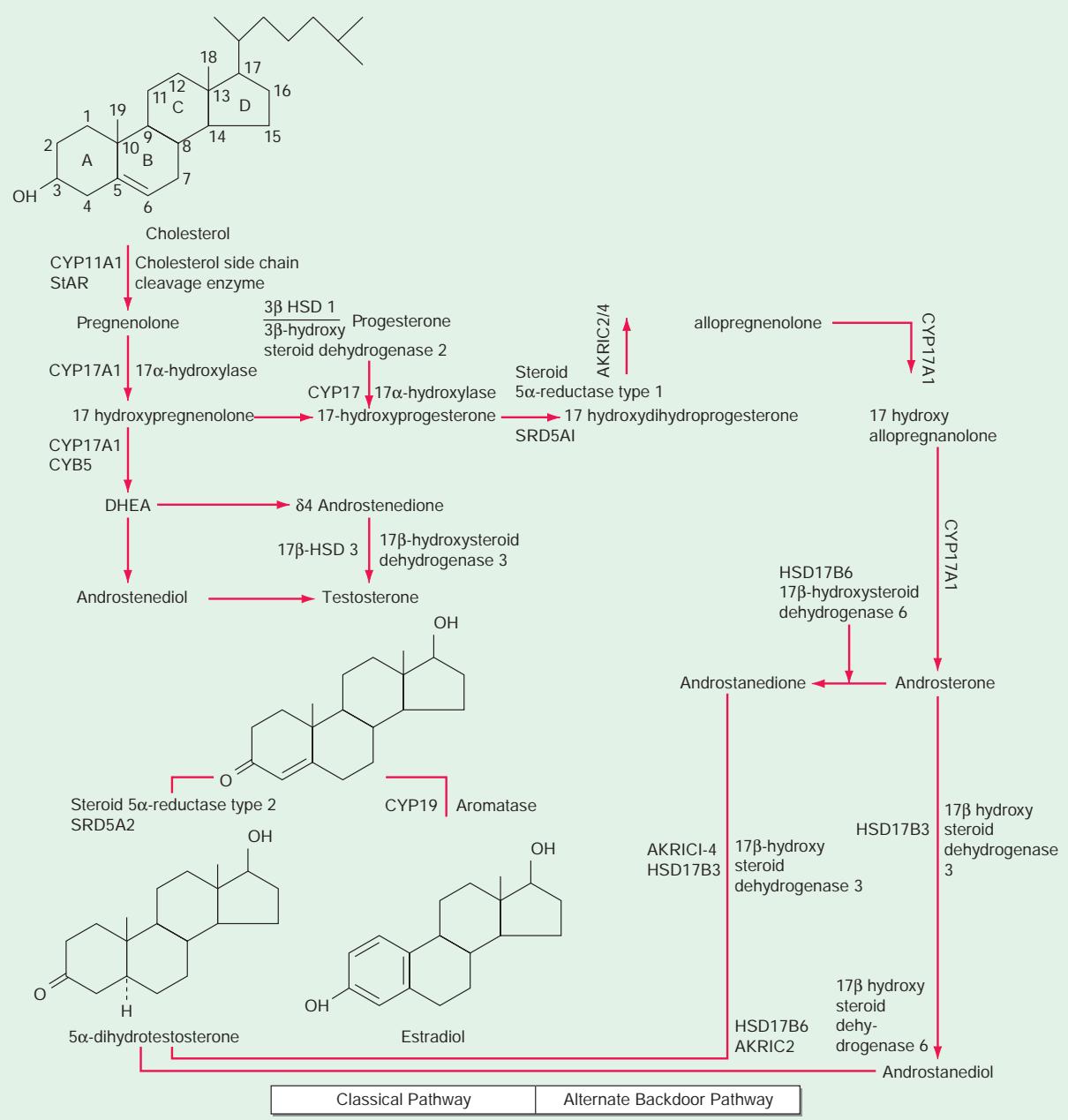
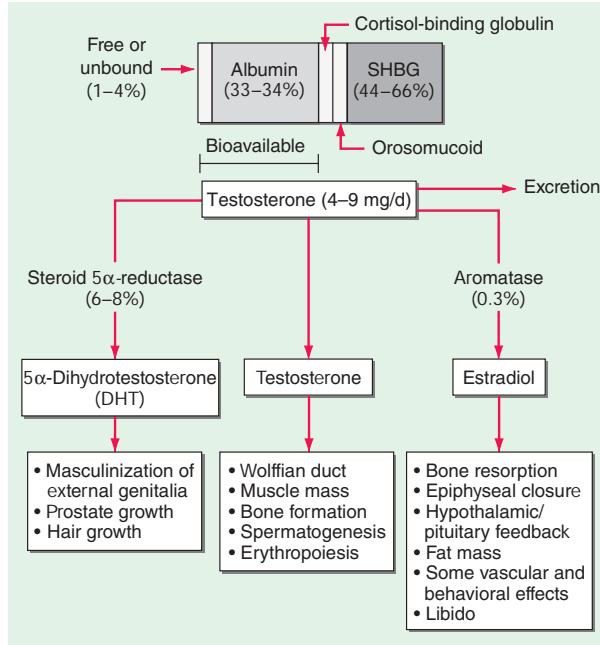


FIGURE 391-3 The biochemical pathway in the conversion of 27-carbon sterol cholesterol to androgens and estrogens.

Circulating testosterone is bound predominantly to sex hormone-binding globulin (SHBG) and albumin (Fig. 391-4) and, to a lesser extent, to cortisol-binding globulin (CBG) and orosomucoid. SHBG binds testosterone with much greater affinity than albumin, CBG, and orosomucoid. The binding proteins regulate the transport and bioavailability of testosterone. SHBG circulates as a dimer, and testosterone's binding to SHBG involves intermonomeric allostery such that neither the conformation nor the binding affinity of the two monomers is equivalent. Similarly, estradiol binding to SHBG involves bidirectional, intermonomeric allostery that changes the distribution of both monomers among various energy and conformational states. Intermonomeric allostery offers a mechanism to extend the binding range of SHBG and regulate hormone bioavailability as sex hormone concentrations vary widely during life. Human serum albumin (HSA) contains multiple, allosterically coupled binding sites for testosterone.

Testosterone shares these binding sites on HSA with free fatty acids. Commonly used drugs such as ibuprofen and antibiotics can displace testosterone from HSA under various physiologic states or disease conditions, affecting its bioavailability. SHBG concentrations are decreased by androgens, obesity, diabetes mellitus, hypothyroidism, nephrotic syndrome, and genetic factors. Conversely, estrogen administration, hyperthyroidism, many chronic inflammatory illnesses, infections such as HIV or hepatitis B and C, aging, and the use of some anticonvulsants are associated with high SHBG concentrations.

Testosterone is metabolized predominantly in the liver, although some degradation occurs in peripheral tissues, particularly the prostate and the skin. In the liver, testosterone is converted by a series of enzymatic steps that involve 5 $\alpha$ - and 5 $\beta$ -reductases, 3 $\beta$ - and 3 $\alpha$ -hydroxysteroid dehydrogenases, and 17 $\beta$ -hydroxysteroid dehydrogenase into androsterone, etiocholanolone, DHT, and 3 $\alpha$ -androstanediol.



**FIGURE 391-4** Androgen metabolism and actions. SHBG, sex hormone–binding globulin.

These compounds undergo glucuronidation or sulfation before being excreted by the kidneys.

**Mechanism of Androgen Action** Testosterone exerts some of its biologic effects by binding to androgen receptor (AR), either directly or after its conversion to DHT by the steroid 5 $\alpha$ -reductase. The actions of testosterone on the Wolffian structures, skeletal muscle, erythropoiesis, and bone in men do not require its obligatory conversion to DHT. However, the conversion of testosterone to DHT is necessary for the masculinization of the urogenital sinus and genital tubercle. Aromatization of testosterone to estradiol mediates additional effects of testosterone on bone resorption, epiphyseal closure, sexual desire, vascular endothelium, and fat. DHT can also be converted in some tissues by the combined actions of the 3 $\beta$ -hydroxysteroid dehydrogenase and 3 $\beta$ -hydroxysteroid dehydrogenase to 5 $\alpha$ -androstane-3 $,17\beta$ -diol, which is a high-affinity ligand and agonist of estrogen receptor  $\alpha$ . 5 $\alpha$ -DHT is further converted in some cell types to 5 $\alpha$ -androstane-3 $,17\beta$ -diol, a modulator of GABA $A$  receptors.

The AR is structurally related to the nuclear receptors for estrogen, glucocorticoids, and progesterone (Chap. 377). The AR, a 919-amino acid protein with a molecular mass of ~110 kDa, is encoded by a gene on the long arm of the X chromosome. A polymorphic region in the amino terminus of the receptor, which contains a variable number of glutamine and glycine repeats, modifies the transcriptional activity of the receptor. The AR protein is distributed in both the cytoplasm and the nucleus. The ligand binding to the AR induces conformational changes that allow the recruitment and assembly of tissue-specific cofactors and causes it to translocate into the nucleus, where it binds to specific androgen response elements in the DNA or other transcription factors already bound to DNA. Thus, the AR is a ligand-regulated transcription factor that regulates the expression of androgen-dependent genes in a tissue-specific manner. Testosterone binds to AR with half the affinity of DHT. The DHT-AR complex also has greater thermostability and a slower dissociation rate than the testosterone-AR complex. However, the molecular basis for selective testosterone versus DHT actions remains incompletely explained. Some androgen effects, such as those on the smooth muscle, may be mediated by nongenomic AR signal transduction pathways. The nongenomic actions of testosterone involve direct activation of kinase signaling cascades such as

mitogen-activated protein kinase and the cyclic AMP response element binding protein transcription factor. Some effects of testosterone on cell proliferation and autophagy require the mediation of GPRC6A.

## THE SEMINIFEROUS TUBULES: SPERMATOGENESIS

The seminiferous tubules are convoluted, closed loops with both ends emptying into the rete testis, a network of progressively larger efferent ducts that ultimately form the epididymis (Fig. 391-2). The seminiferous tubules total ~600 m in length and compose about two-thirds of testis volume. The walls of the tubules are formed by polarized Sertoli cells that are apposed to peritubular myoid cells. Tight junctions between Sertoli cells create the blood-testis barrier. Germ cells compose the majority of the seminiferous epithelium (~60%) and are intimately embedded within the cytoplasmic extensions of the Sertoli cells, which function as “nurse cells.” Germ cells progress through characteristic stages of mitotic and meiotic divisions. A pool of type A spermatogonia serve as stem cells capable of self-renewal. Primary spermatocytes are derived from type B spermatogonia and undergo meiosis before progressing to spermatids that undergo spermiogenesis (a differentiation process involving chromatin condensation, acquisition of an acrosome, elongation of cytoplasm, and formation of a tail) and are released from Sertoli cells as mature spermatozoa. The complete differentiation process into mature sperm requires 74 days. Peristaltic-type action by peritubular myoid cells transports sperm into the efferent ducts. The spermatozoa spend an additional 21 days in the epididymis, where they undergo further maturation and capacitation. The normal adult testes produce >100 million sperm per day.

Naturally occurring mutations in FSH or in the FSH receptor confirm an important, but not essential, role for this pathway in spermatogenesis. Females with mutations in FSH or the FSH receptor are hypogonadal and infertile because ovarian follicles do not mature; males with these mutations exhibit variable degrees of reduced spermatogenesis, presumably because of impaired Sertoli cell function. Because Sertoli cells produce inhibin B, an inhibitor of FSH, seminiferous tubule damage (e.g., by radiation) causes a selective increase of FSH. Testosterone reaches very high concentrations locally in the testis and is essential for spermatogenesis. The cooperative actions of FSH and testosterone are important in the progression of meiosis and spermatiation. In the prepubertal testis, testosterone alone is insufficient for completion of spermatogenesis; however, in men with postpubertal onset of gonadotropin deficiency, human chorionic gonadotropin (hCG) or recombinant LH can reinitiate spermatogenesis without FSH. FSH and testosterone regulate germ cell survival via the intrinsic and extrinsic apoptotic mechanisms. FSH may also play an important role in supporting spermatogonia. Gonadotropin-regulated testicular RNA helicase (GRTH/DDX25), a testis-specific gonadotropin/androgen-regulated RNA helicase, is present in germ cells and Leydig cells and may be an important factor in the paracrine regulation of germ cell development. Several cytokines and growth factors are also involved in the regulation of spermatogenesis by paracrine and autocrine mechanisms. A number of knockout mouse models exhibit impaired germ cell development or spermatogenesis, presaging possible mutations associated with male infertility.

The human Y chromosome contains two pseudoautosomal regions that are located at the two tips of Y chromosome and can recombine with homologous regions of the X chromosome (Fig. 391-5). The genes in the pseudoautosomal regions are involved in cell signaling, transcriptional regulation, and mitochondrial function. Mutations of genes in pseudoautosomal region 1 are associated with mental disorders and short stature. The euchromatic part of the Y chromosome that does not recombine with the X chromosome is referred to as the male-specific region of the Y chromosome (MSY). The MSY contains nine families of Y-specific multicity genes; many of these Y-specific genes are also testis-specific and necessary for spermatogenesis. Microdeletions in several nonoverlapping subregions of the Y chromosome—AZFa, AZFb, AZFc, and AZFd, which contain many spermatogenic genes (e.g., RNA-binding motif, RBM; deleted in azoospermia, DAZ)—are associated with oligospermia or azoospermia.

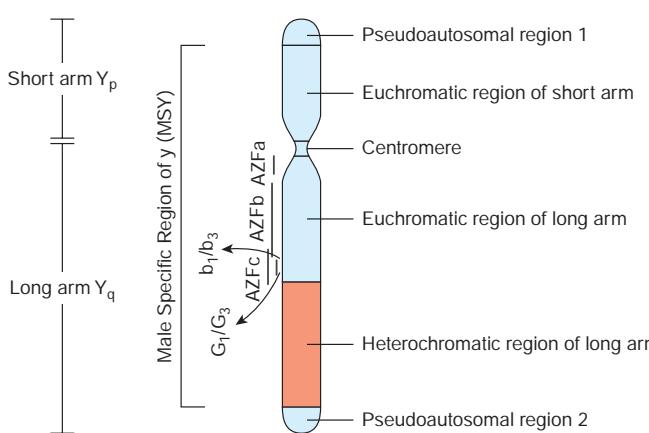


FIGURE 391-5 Structure of the Y chromosome relevant for spermatogenesis.

Approximately 15% of infertile men with azoospermia and ~6% of men with severe oligozoospermia harbor a Y microdeletion. Complete deletions of the AZFa and AZFb subregions are typically associated with Sertoli cells only and azoospermia and a poor prognosis for sperm retrieval. In contrast, AZFc subregion microdeletions are typically associated with oligozoospermia and higher success rates for sperm retrieval. Microdeletion involving the DAZ genes in the AZFc region is one of the most common Y chromosome microdeletions in infertile men. Several partial deletions of the AZFc region have been described including the gr/gr deletion, which is associated with infertility among Caucasian men in Europe and the Western Pacific region, whereas the b2/b3 deletion is associated with male infertility in African and Dravidian men.

## TREATMENT

### Male Factor Infertility

Treatment options for male factor infertility have expanded greatly in recent years. Secondary hypogonadism is highly amenable to treatment with pulsatile GnRH or gonadotropins (see below). Assisted reproductive technologies, such as in vitro fertilization (IVF) and intracytoplasmic sperm injection (ICSI), have provided new opportunities for patients with primary testicular failure and disorders of sperm transport. Choice of initial treatment options depends on sperm concentration and motility. Expectant management should be attempted initially in men with mild male factor infertility (sperm count of  $15\text{--}20 \times 10^6/\text{mL}$  and normal motility). Treatment of moderate male factor infertility ( $10\text{--}15 \times 10^6/\text{mL}$  and 20–40% motility) should begin with intrauterine insemination alone or in combination with treatment of the female partner with clomiphene or gonadotropins, but it may require IVF with or without ICSI. For men with a severe defect (sperm count of  $<10 \times 10^6/\text{mL}$ , 10% motility), IVF with ICSI or donor sperm has become the treatment of choice. Yq microdeletions will be transmitted through ICSI from the affected father to his male offspring if sperm carrying the Yq microdeletion is used.

## CLINICAL AND LABORATORY EVALUATION OF MALE REPRODUCTIVE FUNCTION

### HISTORY AND PHYSICAL EXAMINATION

The history should focus on developmental stages such as puberty and growth spurts, as well as androgen-dependent events such as early morning erections, frequency and intensity of sexual thoughts, and frequency of masturbation or intercourse. Although libido and the overall frequency of sexual acts are decreased in androgen-deficient men, young hypogonadal men can achieve erections in response to visual erotic stimuli. Men with acquired androgen deficiency often report decreased energy and low mood.

The physical examination should focus on secondary sex characteristics such as hair growth, gynecomastia, testicular volume, prostate, and height and body proportions. *Eunuchoid proportions* are defined as an arm span >2 cm greater than height and suggest that androgen deficiency occurred before epiphyseal fusion. Hair growth in the face, axilla, chest, and pubic regions is androgen-dependent; however, changes may not be noticeable unless androgen deficiency is severe and prolonged. Ethnicity also influences the intensity of hair growth (Chap. 394). Testicular volume is best assessed by using a Prader orchidometer. Testes range from 3.5 to 5.5 cm in length, which corresponds to a volume of 12–25 mL. Advanced age does not influence testicular size, although the consistency becomes less firm. Asian men generally have smaller testes than western Europeans, independent of differences in body size. Because of its possible role in infertility, the presence of varicocele should be sought by palpation while the patient is standing; it is more common on the left side. Patients with Klinefelter syndrome have markedly reduced testicular volumes (1–2 mL). In congenital hypogonadotropic hypogonadism, testicular volumes provide a good index for the degree of gonadotropin deficiency and the likelihood of response to therapy.

### GONADOTROPIN AND INHIBIN MEASUREMENTS

LH and FSH are measured using two-site immunoradiometric, immunoassay, or chemiluminescent assays, which have very low cross-reactivity with other pituitary glycoprotein hormones and hCG and have sufficient sensitivity to measure the low levels present in patients with hypogonadotropic hypogonadism. In men with a low testosterone level, an LH level can distinguish primary (high LH) versus secondary (low or inappropriately normal LH) hypogonadism. An elevated LH level indicates a primary defect at the testicular level, whereas a low or inappropriately normal LH level suggests a defect at the hypothalamic-pituitary level. LH pulses occur about every 1–3 h in normal men. Thus, gonadotropin levels fluctuate, and samples should be pooled or repeated when results are equivocal. FSH is less pulsatile than LH because it has a longer half-life. Selective increase in FSH suggests damage to the seminiferous tubules. Inhibin B, a Sertoli cell product that suppresses FSH, is reduced with seminiferous tubule damage. Inhibin B is a dimer with  $\alpha$ -B subunits and is measured by two-site immunoassays.

**GnRH Stimulation Testing** The GnRH test is performed by measuring LH and FSH concentrations at baseline and at 30 and 60 min after intravenous administration of 100  $\mu\text{g}$  of GnRH. A minimally acceptable response is a twofold LH increase and a 50% FSH increase. In the prepubertal period or with severe GnRH deficiency, the gonadotrope may not respond to a single bolus of GnRH because it has not been primed by endogenous hypothalamic GnRH; in these patients, GnRH responsiveness may be restored by chronic, pulsatile GnRH administration. With the availability of sensitive and specific LH assays, GnRH stimulation testing is used rarely.

### TESTOSTERONE ASSAYS

**Total Testosterone** Total testosterone includes both unbound and protein-bound testosterone and is measured by radioimmunoassays, immunometric assays, or liquid chromatography tandem mass spectrometry (LC-MS/MS). LC-MS/MS involves extraction of serum by organic solvents, separation of testosterone from other steroids by high-performance liquid chromatography and mass spectrometry, and quantitation of unique testosterone fragments by mass spectrometry. LC-MS/MS provides accurate and sensitive measurements of testosterone levels even in the low range and has emerged as the method of choice for testosterone measurement. The use of LC-MS/MS for the measurement of testosterone in laboratories that have been certified by the Centers for Disease Control and Prevention's (CDC) Hormone Standardization Program for Testosterone (HoST) can ensure that testosterone measurements are accurate and calibrated to an international standard. A single fasting morning sample provides a good approximation of the average testosterone concentration with the realization

that testosterone levels fluctuate because of its pulsatile, diurnal, and circannual secretory rhythms. Testosterone is generally lower in the late afternoon and is reduced by acute illness. The harmonized normal range for total testosterone, measured using LC-MS/MS in nonobese populations of European and American men aged 19–39 years, is 264–916 ng/dL. This harmonized reference range can be applied to values from laboratories that are certified by the CDC's HoST program.

Alterations in SHBG levels due to aging, obesity, diabetes mellitus, hyperthyroidism, some types of medications, or chronic illness, or on a congenital basis, can affect total testosterone levels. Heritable factors contribute substantially to the population-level variation in testosterone levels, and genome-wide association studies have revealed polymorphisms in the SHBG gene as important contributors to variation in testosterone levels.

**Measurement of Unbound Testosterone Levels** Most circulating testosterone is bound to SHBG and to albumin; only 2.0–4% of circulating testosterone is unbound, or “free.” Free testosterone should ideally be measured by equilibrium dialysis under standardized conditions using an accurate and reliable assay for total testosterone. The unbound testosterone concentration also can be calculated from total testosterone, SHBG, and albumin concentrations. Recent research has shown that testosterone binding to SHBG is a multistep process that involves complex allosteric interactions between the two binding sites within the SHBG dimer; a novel ensemble allosteric model of testosterone's binding to SHBG dimers provides good estimates of free testosterone concentrations. The previous law-of-mass-action equations based on linear models of testosterone binding to SHBG used assumptions that have been shown to be erroneous. Tracer analogue methods are relatively inexpensive and convenient, but they are inaccurate. The term *bioavailable testosterone* refers to unbound testosterone plus testosterone bound loosely to albumin and reflects the concept that albumin-bound testosterone can dissociate at the capillary level, especially in tissues with long transit time, such as the liver and the brain. *Bioavailable testosterone* can be determined by the ammonium sulfate precipitation method. However, the measurements of bioavailable testosterone using the ammonium sulfate precipitation are technically challenging, susceptible to imprecision, and not recommended.

**hCG Stimulation Test** The hCG stimulation test is performed by administering a single injection of 1500–4000 IU of hCG intramuscularly and measuring testosterone levels at baseline and 24, 48, 72, and 120 h after hCG injection. An alternative regimen involves three injections of 1500 units of hCG on successive days and measuring testosterone levels 24 h after the last dose. An acceptable response to hCG is a doubling of the testosterone concentration in adult men. In prepubertal boys, an increase in testosterone to >150 ng/dL indicates the presence of testicular tissue. No response may indicate an absence of testicular tissue or marked impairment of Leydig cell function. Measurement of MIS, a Sertoli cell product, is also used to detect the presence of testes in prepubertal boys with cryptorchidism.

### SEmen ANALYSIS

Semen analysis is the most important step in the evaluation of male infertility. Samples are collected by masturbation following a period of abstinence for 2–3 days. Semen volumes and sperm concentrations vary considerably among fertile men, and several samples may be needed before concluding that the results are abnormal. Analysis should be performed within an hour of collection. Using semen samples from >4500 men in 14 countries, whose partners had a time to pregnancy of <12 months, the World Health Organization (WHO) has generated the following one-sided reference limits for semen parameters: semen volume, 1.5 mL; total sperm number, 39 million per ejaculate; sperm concentration, 15 million per mL; vitality, 58% live; progressive motility, 32%; total (progressive + nonprogressive) motility, 40%; and morphologically normal forms, 4.0%. Some men with low sperm counts are nevertheless fertile. Some studies suggest that sperm counts have declined in recent decades. A variety of tests for sperm function can be performed in specialized laboratories, but these add relatively little to the treatment options.

### TESTICULAR BIOPSY

Testicular biopsy is useful in some patients with oligospermia or azoospermia as an aid in diagnosis and indication for the feasibility of treatment. Using fine-needle aspiration biopsy is performed under local anesthesia to aspirate tissue for histology. Alternatively, open biopsies can be performed under local or general anesthesia when more tissue is required. A normal biopsy in an azoospermic man with a normal FSH level suggests obstruction of the vas deferens, which may be correctable surgically. Biopsies are also used to harvest sperm for ICSI and to classify disorders such as hypospermatogenesis (all stages present but in reduced numbers), germ cell arrest (usually at primary spermatocyte stage), and Sertoli cell-only syndrome (absent germ cells) or hyalinization (sclerosis with absent cellular elements).

**Testing for Y Chromosome Microdeletions** Y chromosome microdeletions are detected by extracting DNA from peripheral blood leukocytes and using polymerase chain reaction (PCR) amplification using primers for ~300 sequence-tagged sites on the Y chromosome, followed by gel electrophoresis to determine whether the DNA sequences corresponding to the selected Y chromosome markers are present. However, because these ~300 Y chromosome markers account for only a small fraction of the 23 million base pairs on the Y chromosome, a negative test does not exclude microdeletions in other subregions of the Y chromosome.

### DISORDERS OF SEXUAL DIFFERENTIATION

See Chap. 390.

### DISORDERS OF PUBERTY

The onset and tempo of puberty vary greatly in the general population and are affected by genetic, nutritional, and environmental factors. Although a substantial fraction of the variance in the timing of puberty is explained by heritable factors, the genes involved remain unknown.

#### PRECOCIOUS PUBERTY

Puberty in boys aged <9 years is considered precocious. Earlier onset of puberty is associated with increased risk for several cancers, cardiovascular disease, hypertension, type 2 diabetes, hair pigmentation, and shorter life span. Genome-wide association studies for age of menarche in girls and age of voice deepening in boys have identified 389 independent loci in girls and 76 independent loci for the timing of puberty in boys.

*Isosexual precocity* refers to premature sexual development consistent with phenotypic sex and includes features such as the development of facial hair and phallic growth. *Isosexual precocity* is divided into gonadotropin-dependent and gonadotropin-independent causes of androgen excess (Table 391-1). *Heterosexual precocity* refers to the premature development of estrogenic features in boys, such as breast development.

**Gonadotropin-Dependent Precocious Puberty** This disorder, called *central precocious puberty* (CPP), is less common in boys than in girls. It is caused by premature activation of the GnRH pulse generator, sometimes because of central nervous system (CNS) lesions such as hypothalamic hamartomas, but it is often idiopathic. CPP is characterized by gonadotropin levels that are inappropriately elevated for age. Because pituitary priming has occurred, GnRH elicits LH and FSH responses typical of those seen in puberty or in adults. MRI should be performed to exclude a mass, structural defect, infection, or inflammatory process. Mutations in kisspeptin, kisspeptin receptor, and *MKRN3*, an imprinted gene encoding makorin ring finger protein 3, which is expressed only from the paternally inherited allele, have been associated with CPP. Loss-of-function mutations in *MKRN3* remove the brake that restrains pulsatile GnRH, resulting in precocious puberty.

**Gonadotropin-Independent Precocious Puberty** Androgens from the testis or the adrenal are increased but gonadotropins are low. This group of disorders includes hCG-secreting tumors; CAH; sex steroid-producing tumors of the testis, adrenal, and ovary; accidental or

**TABLE 391-1 Causes of Precocious or Delayed Puberty in Boys**

- I. Precocious puberty
  - A. Gonadotropin-dependent
    - 1. Idiopathic
    - 2. Hypothalamic hamartoma or other lesions
    - 3. CNS tumor or inflammatory state
    - 4. Mutations in genes that regulate GnRH secretion, such as kisspeptin (*KISS1*), kisspeptin receptor (*KISS1R*), and makorin ring finger protein 3 (*MKRN3*)
  - B. Gonadotropin-independent
    - 1. Congenital adrenal hyperplasia
    - 2. hCG-secreting tumor
    - 3. McCune-Albright syndrome
    - 4. Activating LH receptor mutation
    - 5. Exogenous androgens
    - 6. Androgen producing tumors of the adrenal or the testis
- II. Delayed puberty
  - A. Constitutional delay of growth and puberty
  - B. Systemic disorders
    - 1. Chronic disease
    - 2. Malnutrition
    - 3. Anorexia nervosa
  - C. CNS tumors and their treatment (radiotherapy and surgery)
  - D. Hypothalamic-pituitary causes of pubertal failure (low gonadotropins)
    - 1. Congenital disorders associated with GnRH or gonadotropin deficiency (Table 391-2)
    - 2. Acquired disorders
      - a. Pituitary tumors
      - b. Hyperprolactinemia
      - c. Infiltrative disorders, such as hemochromatosis
  - E. Gonadal causes of pubertal failure (elevated gonadotropins)
    - 1. Klinefelter syndrome
    - 2. Bilateral undescended testes
    - 3. Orchitis
    - 4. Chemotherapy or radiotherapy
    - 5. Anorchia
  - F. Androgen insensitivity

*Abbreviations:* CNS, central nervous system; GnRH, gonadotropin-releasing hormone; hCG, human chronic gonadotropin; LH, luteinizing hormone.

deliberate exogenous sex steroid administration; hypothyroidism; and activating mutations of the LH receptor or G<sub>s</sub> subunit.

**Familial Male-Limited Precocious Puberty** Also called *testotoxicosis*, familial male-limited precocious puberty is an autosomal dominant disorder caused by activating mutations in the LH receptor, leading to constitutive stimulation of the cyclic AMP pathway and testosterone production. Clinical features include premature androgenization in boys, growth acceleration in early childhood, and advanced bone age followed by premature epiphyseal fusion. Testosterone is elevated, and LH is suppressed. Treatment options include inhibitors of testosterone synthesis (e.g., ketoconazole, medroxyprogesterone acetate), AR antagonists (e.g., flutamide and bicalutamide), and aromatase inhibitors (e.g., anastrozole).

**MCCUNE-ALBRIGHT SYNDROME** This is a sporadic disorder caused by somatic (postzygotic) activating mutations in the G<sub>s</sub> subunit that links G protein-coupled receptors to intracellular signaling pathways (Chap. 412). The mutations impair the guanosine triphosphatase activity of the G<sub>s</sub> protein, leading to ligand-independent signaling of the G<sub>s</sub>-coupled receptor and constitutive activation of adenyl cyclase. Like activating LH receptor mutations, this stimulates testosterone production and causes gonadotropin-independent precocious puberty. In addition to sexual precocity, affected individuals may have autonomy in the adrenals, pituitary, and thyroid glands. *Café au lait* spots are characteristic skin lesions that reflect the onset of the somatic mutations in melanocytes during embryonic development. Constitutive

G<sub>s</sub> activation in the postnatal multipotent skeletal stem cells leads to the formation of immature woven bone and replacement of the bone marrow with fibrotic stroma (polyostotic fibrous dysplasia). Treatment is similar to that in patients with activating LH receptor mutations. Bisphosphonates have been used to treat bone lesions.

**CONGENITAL ADRENAL HYPERPLASIA** Boys with CAH who are not well controlled with glucocorticoid suppression of adrenocorticotrophic hormone (ACTH) can develop premature virilization because of excessive androgen production by the adrenal gland (Chaps. 386 and 390). LH is low, and the testes are small. Adrenal rests may develop within the testis of poorly controlled patients with CAH because of chronic ACTH stimulation; adrenal rests do not require surgical removal and regress with effective glucocorticoid therapy. Some children with CAH may develop gonadotropin-dependent precocious puberty with early maturation of the hypothalamic-pituitary-gonadal axis, elevated gonadotropins, and testicular growth.

**Heterosexual Sexual Precocity** Breast enlargement in prepuberal boys can result from familial aromatase excess, estrogen-producing tumors in the adrenal gland, Sertoli cell tumors in the testis, marijuana smoking, or exogenous estrogens or androgens. Occasionally, germ cell tumors that secrete hCG can be associated with breast enlargement due to excessive stimulation of estrogen production (see “Gynecomastia,” below).

## APPROACH TO THE PATIENT

### Precocious Puberty

After verification of precocious development, serum testosterone, LH, and FSH levels should be measured to determine whether gonadotropins are increased in relation to chronologic age (gonadotropin-dependent) or whether sex steroid secretion is occurring independent of LH and FSH (gonadotropin-independent). In children with gonadotropin-dependent precocious puberty, CNS lesions should be excluded by history, neurologic examination, and MRI scan of the head. If organic causes are not found, one is left with the diagnosis of idiopathic central precocity. Patients with high testosterone but suppressed LH concentrations have gonadotropin-independent sexual precocity; in these patients, DHEA sulfate (DHEAS) and 17-hydroxyprogesterone should be measured. High levels of testosterone and 17-hydroxyprogesterone suggest the possibility of CAH due to 21-hydroxylase or 11-hydroxylase deficiency. If testosterone and DHEAS are elevated, adrenal tumors should be excluded by obtaining a CT scan of the adrenal glands. Patients with elevated testosterone but without increased 17-hydroxyprogesterone or DHEAS should undergo careful evaluation of the testis by palpation and ultrasound to exclude a Leydig cell neoplasm. Activating mutations of the LH receptor should be considered in children with gonadotropin-independent precocious puberty in whom CAH, androgen abuse, and adrenal and testicular neoplasms have been excluded.

## TREATMENT

### Precocious Puberty

In patients with a known cause (e.g., a CNS lesion or a testicular tumor), therapy should be directed toward the underlying disorder. In patients with idiopathic CPP, treatment with long-acting GnRH analogues is indicated in boys showing rapid pubertal progression, who are more advanced in pubertal development (e.g., Tanner stage 3 or greater genital development) and experiencing rapid linear growth apparent at their first visit. GnRH analogues suppress gonadotropins and testosterone, halt early pubertal development, delay accelerated bone maturation, prevent early epiphyseal closure, promote final height gain, and mitigate the psychosocial consequences of early pubertal development without causing osteoporosis. The treatment is most effective for increasing final adult height

if it is initiated before age 6. Puberty resumes after discontinuation of the GnRH analogue. Counseling is an important aspect of the overall treatment strategy.

In children with gonadotropin-independent precocious puberty, inhibitors of steroidogenesis, such as ketoconazole, AR antagonists, and aromatase inhibitors have been used empirically. Long-term treatment with spironolactone (a weak androgen antagonist) and ketoconazole has been reported to normalize growth rate and bone maturation and to improve predicted height in small, non-randomized trials in boys with familial male-limited precocious puberty. Aromatase inhibitors, such as testolactone and letrozole, have been used as adjuncts to antiandrogen therapy for children with familial male-limited precocious puberty, CAH, and McCune-Albright syndrome. More potent novel inhibitors of testosterone synthesis, such as abiraterone, have not been evaluated in boys with gonadotropin-independent precocious puberty.

### DELAYED PUBERTY

Puberty is considered delayed in boys if it has not ensued by age 14, an age that is 2–2.5 standard deviations above the mean for healthy children. Pubertal delay is not necessarily pathologic and may be a variant of normal pubertal development in some children. Delayed puberty has been associated with lower peak bone mass, higher risk for metabolic and cardiovascular disorders, and lower risk for breast and endometrial cancer in women.

Delayed puberty is more common in boys than in girls. There are four main categories of delayed puberty: (1) constitutional delay of growth and puberty (~60% of cases); (2) functional hypogonadotropic hypogonadism caused by systemic illness or malnutrition (~20% of cases); (3) hypogonadotropic hypogonadism caused by genetic or acquired defects in the hypothalamic-pituitary region (~10% of cases); and (4) hypergonadotropic hypogonadism secondary to primary gonadal failure (~15% of cases) (Table 391-1). The constitutional delay of growth and puberty is the most common cause that accounts for nearly two-thirds of boys and one-third of girls with delayed puberty. The constitutional delay of growth and puberty clusters in families and displays a complex inheritance pattern, having an autosomal dominant pattern of inheritance in some families, but autosomal recessive, X-linked, or bilineal pattern in other families. Only rarely have mutations in genes known to disrupt the hypothalamic-pituitary-gonadal axis been identified in cases of pubertal delay; most of these mutations have been reported in relatives of patients with idiopathic hypogonadotropic hypogonadism. Functional hypogonadotropic hypogonadism is more common in girls than in boys. Permanent causes of hypogonadotropic or hypergonadotropic hypogonadism are identified in <25% of boys with delayed puberty.

### APPROACH TO THE PATIENT

#### Delayed Puberty

History of systemic illness, eating disorders, excessive exercise, social and psychological problems, and abnormal patterns of linear growth during childhood should be verified. Boys with pubertal delay may have accompanying emotional and physical immaturity relative to their peers, which can be a source of anxiety. Physical examination should focus on height; arm span; weight; visual fields; and secondary sex characteristics, including hair growth, testicular volume, phallic size, and scrotal reddening and thinning. Testicular size >2.5 cm generally indicates that the child has entered puberty.

The main diagnostic challenge is to distinguish those with constitutional delay, who will progress through puberty at a later age, from those with an underlying pathologic process. Constitutional delay should be suspected when there is a family history and when there are delayed bone age and short stature. Pituitary priming by pulsatile GnRH is required before LH and FSH are synthesized and secreted normally. Thus, blunted responses to exogenous GnRH can be seen in patients with constitutional delay, GnRH deficiency, or pituitary disorders. On the other hand, low-normal basal

gonadotropin levels or a normal response to exogenous GnRH is consistent with an early stage of puberty, which is often heralded by nocturnal GnRH secretion. Thus, constitutional delay is a diagnosis of exclusion that requires ongoing evaluation until the onset of puberty and the growth spurt.

### TREATMENT

#### Delayed Puberty

If therapy is considered appropriate, it can begin with 25–50 mg of testosterone enanthate or testosterone cypionate every 2 weeks or by using a 2.5-mg testosterone patch or 25-mg testosterone gel. Because aromatization of testosterone to estrogen is obligatory for mediating androgen effects on epiphyseal fusion, concomitant treatment with aromatase inhibitors may allow attainment of greater final adult height. Testosterone treatment should be interrupted after 6 months to determine if endogenous LH and FSH secretion have ensued. Other causes of delayed puberty should be considered when there are associated clinical features or when boys do not enter puberty spontaneously after a year of observation or treatment.

Reassurance without hormonal treatment is appropriate for many individuals with presumed constitutional delay of puberty. However, the impact of delayed growth and pubertal progression on a child's social relationships and school performance should be weighed. Boys with constitutional delay of puberty are less likely to achieve their full genetic height potential and have reduced total body bone mass as adults, mainly due to narrow limb bones and vertebrae as a result of impaired periosteal expansion during puberty. Furthermore, the time of onset of puberty is negatively associated with bone mineral content and density in boys at skeletal maturity. Judicious use of androgen therapy in carefully selected boys with constitutional delay can induce pubertal induction and progression and promote short-term growth without compromising final height, and when administered with an aromatase inhibitor, it may improve final height.

### DISORDERS OF THE MALE REPRODUCTIVE AXIS DURING ADULTHOOD

#### HYPOGONADOTROPIC HYPOGONADISM

Because LH and FSH are trophic hormones for the testes, impaired secretion of these pituitary gonadotropins results in secondary hypogonadism, which is characterized by low testosterone in the setting of low or inappropriately normal LH and FSH. Those with the most severe gonadotropin deficiency have complete absence of pubertal development, sexual infantilism, and, in some cases, hypospadias and undescended testes. Patients with partial gonadotropin deficiency have delayed or arrested sex development. The 24-h LH secretory profiles are heterogeneous in patients with hypogonadotropic hypogonadism, reflecting variable abnormalities of LH pulse frequency or amplitude. In severe cases, basal LH is low, and there are no LH pulses. A smaller subset of patients has low-amplitude LH pulses or markedly reduced pulse frequency. Occasionally, only sleep-trained LH pulses occur, reminiscent of the pattern seen in the early stages of puberty. Hypogonadotropic hypogonadism can be classified into congenital and acquired disorders. Congenital disorders most commonly involve GnRH deficiency, which leads to gonadotropin deficiency. Acquired disorders are much more common than congenital disorders and may result from a variety of sellar mass lesions or infiltrative diseases of the hypothalamus or pituitary or be due to the effects of drugs, nutritional or psychiatric disorders, or systemic diseases.

**Congenital Disorders Associated with Gonadotropin Deficiency (See Chap. 379)** Congenital hypogonadotropic hypogonadism is a heterogeneous group of disorders characterized by decreased gonadotropin secretion and testicular dysfunction either due to impaired function of the GnRH pulse generator or the gonadotrope.

**3014** The disorders characterized by GnRH deficiency represent a family of oligogenic disorders whose phenotype spans a wide spectrum. Some individuals with GnRH deficiency may suffer from complete absence of pubertal development, while others may manifest varying degrees of gonadotropin deficiency and pubertal delay, and a subset that carries the same mutations as their affected family members may even have normal reproductive function. In ~10% of men with idiopathic hypogonadotropic hypogonadism (IHH), reversal of gonadotropin deficiency may occur in adult life after sex steroid therapy. Also, a small fraction of men with IHH may present with androgen deficiency and infertility in adult life after having gone through apparently normal pubertal development. Nutritional, emotional, or metabolic stress may unmask gonadotropin deficiency and reproductive dysfunction (e.g., hypothalamic amenorrhea) in some patients who harbor mutations in the candidate genes but who previously had normal reproductive function. The clinical phenotype may include isolated anosmia or hyposmia. Oligogenicity and gene-gene and gene-environment interactions may contribute to variations in clinical phenotype.

Mutations in a number of genes involved in the development and migration of GnRH neurons or in the regulation of GnRH secretion have been linked to GnRH deficiency, although the genetic defect remains elusive in nearly two-thirds of cases. Familial hypogonadotropic hypogonadism can be transmitted as an X-linked (20%), autosomal recessive (30%), or autosomal dominant (50%) trait. Some individuals with IHH have sporadic mutations in the same genes that cause inherited forms of the disorder. The genetic defects associated with GnRH deficiency can be conveniently classified as anosmic (Kallmann syndrome) or normosmic (**Table 391-2**), although the occurrence of both anosmic and normosmic forms of GnRH deficiency in the same families suggests commonality of pathophysiologic mechanisms. *Kallmann syndrome*, the anosmic form of GnRH deficiency, can result from mutations in one or more neurodevelopmental genes associated with olfactory bulb morphogenesis or the migration of GnRH neurons from their origin in the region of the olfactory placode, along the scaffold established by the olfactory nerves, through the cribriform plate into their final location into the preoptic region of the hypothalamus. Thus, mutations in *KAL1*, NMDA receptor synaptosomal signaling and neuronal migration factor (*NSMF*), genes involved in fibroblast growth factor (FGF) signaling (*FGF8*, *FGFR1*, *FGF17*, *IL17RD*, *DUSP6*, *SPRY4*, and *FLRT3*), *NELF*, genes involved in PROK signaling (*PROK2* and *PROK2R*), *WDR11*, *SOX10*, *TUBB3* *SEMA3*, *HS6ST1*, *CHD7*, and *FEZF1* have been described in patients with Kallmann syndrome. An X-linked form of IHH is caused by mutations in the *KAL1* gene, which encodes anosmin, a protein that mediates the migration of neural progenitors of the olfactory bulb and GnRH-producing neurons. These individuals have GnRH deficiency and variable combinations of anosmia or hyposmia, renal defects, and neurologic abnormalities including mirror movements. Proteins such as those involved in FGF and prokineticin signaling and *KAL1*, which account for the great majority of Kallmann syndrome cases, interact with heparin sulfate glycosaminoglycan compounds within the extracellular matrix in supporting GnRH neuronal migration. Mutations in the *FGFR1* gene cause an autosomal dominant form of hypogonadotropic hypogonadism that clinically resembles Kallmann syndrome; mutations in its putative ligand, the *FGF8* gene product, have also been associated with IHH. Craniofacial tissues and olfactory ensheathing cells also play important roles in neurogenesis and migration of the GnRH neurons, and additional proteins that regulate these cell types may also be involved in the pathogenesis of Kallmann syndrome. The co-occurrence of tooth anomalies, cleft palate, craniofacial anomalies, pigmentation, and neurologic defects in patients with Kallmann syndrome suggests that the syndrome may be a part of the spectrum of neurocristopathies. Other dysmorphic features associated with some forms of IHH include renal agenesis, hearing loss, synkinesia, short metacarpals, eye movement abnormalities, cerebellar ataxia, and dental agenesis. The presence of these dysmorphic features can offer clues to the underlying genetic abnormality and guide genetic testing.

Normosmic GnRH deficiency results from defects in pulsatile GnRH secretion, its regulation, or its action on the gonadotrope and

has been associated with mutations in *GnRHR*, *GnRH1*, *KISS1R*, *TAC3*, *TACR3*, *NROB1 (DAX1)*, leptin, or leptin receptor. Some mutations, such as those in *PROK2*, *PROKR2*, *NSMF*, *FGFR1*, *FGF8*, *SEMA3A*, *WDR11*, and *CHD7*, have been associated with both anosmic and normosmic forms of IHH; it is possible that these genes are involved in GnRH neuronal migration as well in regulation of GnRH secretion. *GnRHR* mutations, the most frequent identifiable cause of normosmic IHH, account for ~40% of autosomal recessive and 10% of sporadic cases of hypogonadotropic hypogonadism. These patients have decreased LH response to exogenous GnRH. Some receptor mutations alter GnRH binding affinity, allowing apparently normal responses to pharmacologic doses of exogenous GnRH, whereas other mutations may alter signal transduction downstream of hormone binding. Mutations of the *GnRH1* gene have also been reported in patients with hypogonadotropic hypogonadism, although they are rare. The G protein-coupled receptor *KISS1R (GPR54)* and its cognate ligand, kisspeptin (*KISS1*), are important regulators of sexual maturation in primates. Recessive mutations in *GPR54* cause gonadotropin deficiency without anosmia. Patients retain responsiveness to exogenous GnRH, suggesting an abnormality in the neural pathways controlling GnRH release. The genes encoding NKB (*TAC3*), which is involved in preferential activation of GnRH release in early development, and its receptor (*TAC3R*) have been implicated in some families with normosmic IHH. The neurokinin pathway plays an important role in GnRH activation during “mini-puberty” as well as in puberty. Prokineticin 2 (*PROK2*) and its receptor (*PROK2R*) are highly expressed in the olfactory ventricle and subventricular zone of the lateral ventricle and are associated with neurogenesis of the olfactory bulbs and the migration of the olfactory neuronal cells. Mutations in the *CHD7* gene that encodes for the chromodomain helicase DNA binding protein 7 causes CHARGE syndrome characterized by eye coloboma, heart anomalies, choanal atresia, growth and developmental retardation, genitourinary anomalies, hypogonadism, and ear abnormalities. X-linked hypogonadotropic hypogonadism also occurs in *adrenal hypoplasia congenita*, a disorder caused by mutations in the *DAX1* gene, which encodes a nuclear receptor in the adrenal gland and reproductive axis. Adrenal hypoplasia congenita is characterized by absent development of the adult zone of the adrenal cortex, leading to neonatal adrenal insufficiency. Puberty usually does not occur or is arrested, reflecting variable degrees of gonadotropin deficiency. Although sexual differentiation is normal, some patients have testicular dysgenesis and impaired spermatogenesis despite gonadotropin replacement. Less commonly, adrenal hypoplasia congenita, sex reversal, and hypogonadotropic hypogonadism can be caused by mutations of steroidogenic factor 1 (SF1). Rarely, recessive mutations in the *LH* or *FSH* genes have been described in patients with selective deficiencies of these gonadotropins.

A number of homeodomain transcription factors are involved in the development and differentiation of the specialized hormone-producing cells within the pituitary gland (**Table 391-2**). Patients with mutations of *PROP1* have combined pituitary hormone deficiency that includes GH, prolactin (PRL), thyroid-stimulating hormone (TSH), LH, and FSH, but not ACTH. *LHX3* mutations cause combined pituitary hormone deficiency in association with cervical spine rigidity. *HEX1* mutations cause septo-optic dysplasia and combined pituitary hormone deficiency. Mutations of *ARNT1*, inherited as an autosomal recessive disorder, are associated with diabetes insipidus; ACTH, GH, LH, and FSH deficiency; anterior pituitary hypoplasia; hypoplastic frontal and temporal lobes; thin corpus callosum; prominent forehead; and retrognathia. Patients with *SOX2* mutations can have gonadotropin deficiency, variable deficiencies of TSH and ACTH, pituitary hypoplasia, microphthalmia, and intellectual disability.

*Prader-Willi syndrome (PWS)* is characterized by obesity, hypotonic musculature, intellectual disability, hypogonadism, short stature, and small hands and feet. Prader-Willi syndrome is a genomic imprinting disorder caused by deletions of the proximal portion of the paternally derived chromosome 15q11-15q13 region, which contains a bipartite imprinting center; uniparental disomy of the maternal alleles; or mutations of the genes/loci involved in imprinting (**Chap. 466**). Recent studies suggest that at least some of the major

**TABLE 391-2 Causes of Congenital Hypogonadotropic Hypogonadism**

GENE	LOCUS	INHERITANCE	ASSOCIATED FEATURES
<b>A. Hypogonadotropic Hypogonadism due to GnRH Deficiency</b>			
<b>A1. GnRH Deficiency Associated with Hyposmia or Anosmia</b>			
<i>KAL1</i>	Xp22	X-linked	Anosmia, renal agenesis, synkinesia, cleft lip/palate, oculomotor/visuospatial defects, gut malformations
<i>NELF</i>	9q34.3	AR	Anosmia, hypogonadotropic hypogonadism
<i>FGF8</i>	10q24	AR	Anosmia (some patients may be normosmic), skeletal abnormalities
<i>FGF17</i>	8p21.3	AR	Anosmia (some patients may be normosmic)
<i>FGFR1</i>	8p11-p12	AD	Anosmia, cleft lip/palate, synkinesia, syndactyly
<i>PROK2</i>	3p21	AR	Anosmia/sleep dysregulation
<i>PROK2R</i>	20p12.3	AR	Variable
<i>CHD7</i>	8q12.1		Anosmia, other features of CHARGE syndrome
<i>FEZ1</i>	8p22	AR	Anosmia, olfactory bulb aplasia
<i>WDR11</i>	10q26	AD	Anosmia
<i>SOX10</i>	22q13		Deafness
<i>SEMA3A</i>	7q21		Anosmia; some persons with mutations are normal
<i>HS6ST1</i>	2q14	Complex	Anosmia
<i>TUBB3</i>	Tubulin β 3	AR	Anosmia
<i>NSMF</i>	9q34.3	AR	Anosmia (some patients may be normosmic)
<i>DUSP6</i>	12q21.33	AR	Anosmia
<i>GLCE</i>	15q23	AR	Anosmia (some patients may be normosmic)
<i>FLRT3</i>	20p12.1	AR	Anosmia (some patients may be normosmic)
<i>SPRY4</i>	5q31.3	AR	Anosmia (some patients may be normosmic)
<i>IL17RD</i>	3p14.3	AR	Anosmia
<b>A2. GnRH Deficiency with Normal Sense of Smell</b>			
<i>GNRHR</i>	4q21	AR	None
<i>GnRH1</i>	8p21	AR	None
<i>KISS1R</i>	19p13	AR	None
<i>TAC3</i>	12q13	AR	Microphallus, cryptorchidism, reversal of GnRH deficiency
<i>TAC3R</i>	4q25	AR	Microphallus, cryptorchidism, reversal of GnRH deficiency
<i>LEPR</i>	1p31	AR	Obesity
<i>LEP</i>	7q31	AR	Obesity
<i>DMXL2</i>	15q21.2	AR	Polyendocrine polyneuropathy syndrome
<i>OTUD4</i>	4q31.21	AR	Ataxia
<i>RNF216</i>	7p22.1	AR	Ataxia
<i>STUB1</i>	16p13.3	AR	Ataxia
<i>POLR3B</i>	12q23.3	AR	Ataxia
<i>PNPLA6</i>	19p13.2	AR	Ataxia
<i>NR0B1</i>	Xp21.2	X-linked	Primary adrenal failure
<b>B. Hypogonadotropic Hypogonadism Not Due to GnRH Deficiency</b>			
<i>PC1</i>	5q15-21	AR	Obesity, diabetes mellitus, ACTH deficiency
<i>HESX1</i>	3p21	AR	Septo-optic dysplasia, CPHD
		AD	Isolated GH insufficiency
<i>LHX3</i>	9q34	AR	CPHD (ACTH spared), cervical spine rigidity
<i>PROP1</i>	5q35	AR	CPHD (ACTH usually spared)
<i>FSHβ</i>	11p13	AR	↑ LH
<i>LHβ</i>	19q13	AR	↑ FSH
<i>SF1 (NR5A1)</i>	9p33	AD/AR	Primary adrenal failure, XY sex reversal

**Abbreviations:** ACTH, adrenocorticotrophic hormone; AD, autosomal dominant; AR, autosomal recessive; CHARGE syndrome, eye coloboma, heart defects, choanal atresia, growth and developmental retardation, genitourinary anomalies, ear anomalies; CPHD, combined pituitary hormone deficiency; *DAX1*, dosage-sensitive sex-reversal, adrenal hypoplasia congenita, X chromosome; *DMXL2*, DMX like 2; *DUSP6*, dual specificity phosphatase 6; *FGFR1*, fibroblast growth factor receptor 1; *FGF17*, fibroblast growth factor 17; *FSHβ*, follicle-stimulating hormone β-subunit; *FLRT3*, fibronectin like domain containing leucine rich transmembrane protein 3; GH, growth hormone; *GLCE*, glucuronic acid epimerase; *GNRHR*, gonadotropin-releasing hormone receptor; *GPR54*, G protein-coupled receptor 54; *HESX1*, homeobox gene expressed in embryonic stem cells 1; *KAL1*, Kallmann syndrome interval gene 1, also known as anosmin 1; *LEP*, leptin; *LEPR*, leptin receptor; *LHX3*, LIM homeobox gene 3; *LHβ*, luteinizing hormone β-subunit; *NELF*, nasal embryonic luteinizing hormone-releasing hormone factor; *NSMF*, NMDA receptor synaptosomal signaling and neuronal migration factor; *NR0B1*, nuclear receptor subfamily 0, group B, member 1; *OTUD4*, OUT domain containing protein 4; *PNPLA6*, patatin-like phospholipase domain-containing protein 6; *PC1*, prohormone convertase 1; *PROK2*, prokineticin 2; *PROP1*, prophet of pit 1; *RNF216*, ring finger protein 216; *POLR3B*, polymerase III RNA subunit B; *SF1*, steroidogenic factor 1; *SPRY4*, sprouty RTK signalling antagonist 4; *STUB1*, srip 1 homologous and U box containing protein 1; *TUBB3*, tubulin beta 3; *IL17RD*, interleukin 17 receptor D.

**3016** manifestations of PWS may be due to reduced expression of prohormone convertase 1.

*Laurence-Moon syndrome* is an autosomal recessive disorder characterized by obesity, hypogonadism, mental retardation, polydactyly, and retinitis pigmentosa. Recessive mutations of leptin, or its receptor, cause severe obesity and pubertal arrest, apparently because of hypothalamic GnRH deficiency ([Chap. 401](#)).

**Acquired Hypogonadotropic Disorders • SEVERE ILLNESS, STRESS, MALNUTRITION, AND EXERCISE** These may cause reversible gonadotropin deficiency. Although gonadotropin deficiency and reproductive dysfunction are well documented in these conditions in women, men exhibit similar but less pronounced responses. Unlike women, most male runners and other endurance athletes have normal gonadotropin and sex steroid levels, despite low body fat and frequent intensive exercise. Testosterone levels fall at the onset of illness and recover during recuperation. The magnitude of gonadotropin suppression generally correlates with the severity of illness. Although hypogonadotropic hypogonadism is the most common cause of androgen deficiency in patients with acute illness, some have elevated levels of LH and FSH, which suggest primary gonadal dysfunction. The pathophysiology of reproductive dysfunction during acute illness is unknown but likely involves a combination of cytokine and/or glucocorticoid effects. There is a high frequency of low testosterone levels in patients with chronic illnesses such as HIV infection, end-stage renal disease, chronic obstructive lung disease, and many types of cancer and in patients receiving glucocorticoids. About 20% of HIV-infected men with low testosterone levels have elevated LH and FSH levels; these patients presumably have primary testicular dysfunction. The remaining 80% have either normal or low LH and FSH levels; these men have a central hypothalamic-pituitary defect or a dual defect involving both the testis and the hypothalamic-pituitary centers. Muscle wasting is common in chronic diseases associated with hypogonadism, which also leads to debility, poor quality of life, and adverse outcome of disease. There is great interest in exploring strategies that can reverse androgen deficiency or attenuate the sarcopenia associated with chronic illness.

Men using opioids for relief of cancer or noncancerous pain or because of addiction often have suppressed testosterone and LH levels and high prevalence of sexual dysfunction and osteoporosis; the degree of suppression is dose-related and particularly severe with long-acting opioids such as methadone. Opioids suppress GnRH secretion and alter the sensitivity to feedback inhibition by gonadal steroids. Men who are heavy users of marijuana have decreased testosterone secretion and sperm production. The mechanism of marijuana-induced hypogonadism is decreased GnRH secretion. Gynecomastia observed in marijuana users can also be caused by plant estrogens in crude preparations. Androgen deprivation therapy in men with prostate cancer has been associated with increased risk of bone fractures, diabetes mellitus, cardiovascular events, fatigue, sexual dysfunction, tender gynecomastia, and poor quality of life.

**OBESITY** In men with mild to moderate obesity, SHBG levels decrease in proportion to the degree of obesity, resulting in lower total testosterone levels. However, free testosterone levels usually remain within the normal range. SHBG production in the liver is inhibited by hepatic lipids and by tumor necrosis factor  $\alpha$  and interleukin 1, but it is not affected by insulin. Thus, the low SHBG levels seen in obesity and diabetes are likely the result of low-grade inflammation and the increased amount of hepatic lipids rather than high insulin levels. Estradiol levels are higher in obese men compared to healthy, nonobese controls, because of aromatization of testosterone to estradiol in adipose tissue. Weight loss is associated with reversal of these abnormalities including an increase in total and free testosterone levels and a decrease in estradiol levels. A subset of obese men with moderate to severe obesity may have a defect in the hypothalamic-pituitary axis as suggested by low free testosterone in the absence of elevated gonadotropins. Weight gain in adult men can accelerate the rate of age-related decline in testosterone levels.

**HYPERPROLACTINEMIA** ([See also Chap. 380](#)) Elevated PRL levels are associated with hypogonadotropic hypogonadism. PRL inhibits hypothalamic GnRH secretion either directly or through modulation of tuberoinfundibular dopaminergic pathways. A PRL-secreting tumor may also destroy the surrounding gonadotropes by invasion or compression of the pituitary stalk. Treatment with dopamine agonists reverses gonadotropin deficiency, although there may be a delay relative to PRL suppression.

**SELLAR MASS LESIONS** Neoplastic and nonneoplastic lesions in the hypothalamus or pituitary can directly or indirectly affect gonadotrope function. In adults, pituitary adenomas constitute the largest category of space-occupying lesions affecting gonadotropin and other pituitary hormone production. Pituitary adenomas that extend into the suprasellar region can impair GnRH secretion and mildly increase PRL secretion (usually <50  $\mu$ g/L) because of impaired tonic inhibition by dopaminergic pathways. These tumors that cause hyperprolactinemia by stalk compression should be distinguished from prolactinomas, which typically are associated with higher PRL levels. The presence of diabetes insipidus suggests the possibility of a craniopharyngioma, infiltrative disorder, or other hypothalamic lesions ([Chap. 381](#)).

**HEMOCHROMATOSIS** ([See also Chap. 414](#)) Both the pituitary and testis can be affected by excessive iron deposition. However, the pituitary defect is the predominant lesion in most patients with hemochromatosis and hypogonadism. The diagnosis of hemochromatosis is suggested by the association of characteristic skin discoloration, hepatic enlargement or dysfunction, diabetes mellitus, arthritis, cardiac conduction defects, and hypogonadism.

### PRIMARY TESTICULAR CAUSES OF HYPOGONADISM

Common causes of primary testicular dysfunction include Klinefelter syndrome, uncorrected cryptorchidism, cancer chemotherapy, radiation to the testes, trauma, torsion, infectious orchitis, HIV infection, anorchia syndrome, and myotonic dystrophy. Primary testicular disorders may be associated with impaired spermatogenesis, decreased androgen production, or both. [See Chap. 390 for disorders of testis development, androgen synthesis, and androgen action.](#)

**Klinefelter Syndrome** ([See also Chap. 390](#)) Klinefelter syndrome is the most common chromosomal disorder associated with testicular dysfunction and male infertility. It occurs in about 1 in 600 live-born males. Azoospermia is the rule in men with Klinefelter syndrome who have the 47,XXY karyotype due to the progressive loss of 47,XXY spermatogonial stem cells; however, men with mosaicism may have germ cells, especially at a younger age. The clinical phenotype of Klinefelter syndrome can be variable, possibly because of mosaicism, polymorphisms in AR gene, the parental origin of the X chromosome, X-linked copy number variations, gene-dosage effects in conjunction with X chromosome inactivation, variable testosterone levels, or other genetic factors. Testicular histology shows hyalinization of seminiferous tubules and germ cell aplasia. However, spermatogenesis can be observed in a small number of tubules from which sperm can be harvested during testicular sperm extraction for IVF. Although their function is impaired, the number of Leydig cells appears to increase. Testosterone is decreased and estradiol is increased, leading to clinical features of undervirilization and gynecomastia. Men with Klinefelter syndrome are at increased risk of systemic lupus erythematosus, Sjögren's syndrome, breast cancer, diabetes mellitus, osteoporosis, non-Hodgkin's lymphoma, and some types of lung cancer and at reduced risk of prostate cancer. Periodic mammography for breast cancer surveillance is recommended for men with Klinefelter syndrome. Fertility can be achieved by intracytoplasmic injection of sperm retrieved surgically from the testes of men with Klinefelter syndrome, including some men with nonmosaic form of Klinefelter syndrome. Although sperm retrieval for fertility preservation offers no benefit over harvesting in adulthood, fertility counseling, including the potential for sperm retrieval, should be offered prior to starting testosterone replacement therapy. The karyotypes 48,XXX and 49,XXXX are associated with

a more severe phenotype, increased risk of congenital malformations, and lower intelligence than 47,XXY individuals.

**Cryptorchidism** Cryptorchidism occurs when there is incomplete descent of the testis from the abdominal cavity into the scrotum. About 1–4% of full-term and 30% of premature male infants have at least one undescended testis at birth, but descent is usually complete by the first few weeks of life. Fifty percent of undescended testes at birth will descend spontaneously within the first 6–18 months of life; consequently, the incidence of cryptorchidism is <1% by 9 months of age. Cryptorchidism should be distinguished from retractile testes that can be pulled down into the scrotum during physical examination and require no treatment.

Androgens regulate predominantly the inguinoscrotal descent of the testes through degeneration of the crano-suspensory ligament and a shortening of the gubernaculum, respectively. Mutations in *INSL3* and leucine-rich repeat family of G-protein-coupled receptor 8 (*LGR8*), which regulate the transabdominal portion of testicular descent, have been found in some patients with cryptorchidism.

Cryptorchidism is associated with increased risk of malignancy, infertility, inguinal hernia, and torsion. Unilateral cryptorchidism, even when corrected before puberty, is associated with decreased sperm count, possibly reflecting unrecognized damage to the fully descended testis or other genetic factors. Therefore, surgical correction is usually performed between 6 and 18 months of age depending on the location of the testes, the child's body size, and parental preference. Epidemiologic, clinical, and molecular evidence supports the idea that cryptorchidism, hypospadias, impaired spermatogenesis, and testicular cancer may be causally related to common genetic and environmental perturbations and are components of the testicular dysgenesis syndrome.

**Acquired Testicular Defects** *Viral orchitis* may be caused by the mumps virus, echovirus, lymphocytic choriomeningitis virus, and group B arboviruses. Orchitis occurs in as many as one-fourth of adult men with mumps; the orchitis is unilateral in about two-thirds and bilateral in the remainder. Orchitis usually develops a few days after the onset of parotitis but may precede it. The testis may return to normal size and function or undergo atrophy. Semen analysis returns to normal for three-fourths of men with unilateral involvement but for only one-third of men with bilateral orchitis. *Trauma*, including testicular torsion, can also cause secondary atrophy of the testes. The exposed position of the testes in the scrotum renders them susceptible to both thermal and physical trauma, particularly in men with hazardous occupations.

The late-term adverse effects of cancer treatment on reproductive health have emerged as an important concern among cancer survivors. Many professional cancer societies have published guidelines on the fertility preservation in patients with cancer and endorsed formal consideration of fertility preservation measures prior to initiation of cancer treatment. The testes are sensitive to *radiation damage* due to the direct effects of ionized radioactive particles as well as indirect effects of free radicals generated from water. Although radiation doses as low as 0.75 Gy can transiently raise LH, only doses >20 Gy are associated with Leydig cell dysfunction and increased FSH and LH levels. Radiation doses <1.0 Gy are associated with only a transient decline in sperm density; doses between 1.0 and 2.0 Gy are associated with temporary azoospermia, and doses >2.0 Gy are generally associated with permanent azoospermia. Permanent androgen deficiency in adult men is uncommon after therapeutic radiation; however, most boys given direct testicular radiation therapy for acute lymphoblastic leukemia have permanently low testosterone levels. Direct testicular radiation and whole-body radiation before bone marrow transplantation pose the greatest risk of permanent testicular damage.

Combination chemotherapy for acute leukemia, Hodgkin's disease, and testicular and other cancers may impair Leydig cell function and cause infertility. The degree of gonadal dysfunction depends on the type of chemotherapeutic agent and the dose and duration of therapy. Because of the high response rates and the young age of these

men, infertility and androgen deficiency have emerged as important long-term complications of cancer chemotherapy. Cyclophosphamide and combination regimens containing alkylating agents such as procarbazine are particularly toxic to germ cells. Thus, 90% of men with Hodgkin's lymphoma receiving MOPP (mechlorethamine, oncovin, procarbazine, prednisone) therapy develop azoospermia or extreme oligozoospermia; newer regimens that do not include procarbazine, such as ABVD (doxorubicin, bleomycin, vinblastine, dacarbazine), are less toxic to germ cells. Azoospermia is uncommon with cyclophosphamide equivalent doses of <4000 mg/m<sup>2</sup>, and higher doses of alkylating agents are generally associated with varying degree of damage to germ cells. The outcomes of assisted reproductive technologies in cancer survivors using cryopreserved sperm collected prior to cancer treatment are similar to those in infertile men who do not have cancer. A smaller subset of cancer survivors treated with large doses of alkylating agents may also suffer from testosterone deficiency and sexual dysfunction.

Drugs interfere with testicular function by several mechanisms, including inhibition of testosterone synthesis (e.g., ketoconazole), blockade of androgen action (e.g., spironolactone), increased estrogen (e.g., marijuana), and toxic effects on spermatogenesis (e.g., chemotherapy).

Alcohol, when consumed in excess for prolonged periods, decreases testosterone, independent of liver disease or malnutrition. Elevated estradiol and decreased testosterone levels may occur in men taking digitalis.

The occupational and recreational history should be carefully evaluated in all men with infertility because of the toxic effects of many *chemical agents* on spermatogenesis. Known environmental hazards include pesticides (e.g., vinclozolin, dicofol, atrazine), sewage contaminants (e.g., ethinyl estradiol in birth control pills, surfactants such as octylphenol, nonylphenol), plasticizers (e.g., phthalates), flame retardants (e.g., polychlorinated biphenyls, polybrominated diphenyl ethers), industrial pollutants (e.g., heavy metals such as cadmium and lead, dioxins, polycyclic aromatic hydrocarbons), microwaves, and ultrasound. In some populations, sperm density is said to have declined by as much as 40% in the past 50 years. Environmental estrogens or antiandrogens may be partly responsible.

Testicular failure also occurs as a part of *polyglandular autoimmune insufficiency* (Chap. 388). Sperm antibodies can cause isolated male infertility. In some instances, these antibodies are secondary phenomena resulting from duct obstruction or vasectomy. Granulomatous diseases can affect the testes, and testicular atrophy occurs in 10–20% of men with lepromatous leprosy because of direct tissue invasion by the mycobacteria. The tubules are involved initially, followed by endarteritis and destruction of Leydig cells.

*Systemic disease* can cause primary testis dysfunction in addition to suppressing gonadotropin production. In cirrhosis, a combined testicular and pituitary abnormality leads to decreased testosterone production independent of the direct toxic effects of ethanol. Impaired hepatic extraction of adrenal androstenedione leads to extraglandular conversion to estrone and estradiol, which partially suppresses LH. Testicular atrophy and gynecomastia are present in approximately one-half of men with cirrhosis. In chronic renal failure, androgen synthesis and sperm production decrease despite elevated gonadotropins. The elevated LH level is due to reduced clearance, but it does not restore normal testosterone production. About one-fourth of men with renal failure have hyperprolactinemia. Improvement in testosterone production with hemodialysis is incomplete, but successful renal transplantation may return testicular function to normal. Testicular atrophy is present in one-third of men with sickle cell anemia. The defect may be at either the testicular or the hypothalamic-pituitary level. Sperm density can decrease temporarily after acute febrile illness in the absence of a change in testosterone production. Infertility in men with celiac disease is associated with a hormonal pattern typical of androgen resistance, namely elevated testosterone and LH levels.

Neurologic diseases associated with altered testicular function include myotonic dystrophy, spinobulbar muscular atrophy, and paraplegia. In myotonic dystrophy, small testes may be associated with

**3018** impairment of both spermatogenesis and Leydig cell function. Spinobulbar muscular atrophy is caused by an expansion of the glutamine repeat sequences in the amino-terminal region of the AR; this expansion impairs function of the AR, but it is unclear how the alteration is related to the neurologic manifestations. Men with spinobulbar muscular atrophy often have undervirilization and infertility as a late manifestation. Spinal cord injury that causes paraplegia is often associated with low testosterone levels and may cause persistent defects in spermatogenesis; some patients retain the capacity for penile erection and ejaculation.

## ANDROGEN INSENSITIVITY SYNDROMES

Mutations in the AR cause resistance to the action of testosterone and DHT. These X-linked mutations are associated with variable degrees of defective male phenotypic development and undervirilization (*Chap. 390*). Although not technically hormone-insensitivity syndromes, two genetic disorders impair testosterone conversion to active sex steroids. Mutations in the *SRD5A2* gene, which encodes 5'-reductase type 2, prevent the conversion of testosterone to DHT, which is necessary for the normal development of the male external genitalia. Mutations in the *CYP19* gene, which encodes aromatase, prevent testosterone conversion to estradiol. Males with *CYP19* mutations have delayed epiphyseal fusion, tall stature, eunuchoid proportions, visceral adiposity, and osteoporosis, consistent with evidence from an estrogen receptor-deficient individual that these testosterone actions are mediated via estrogen.

## GYNECOMASTIA

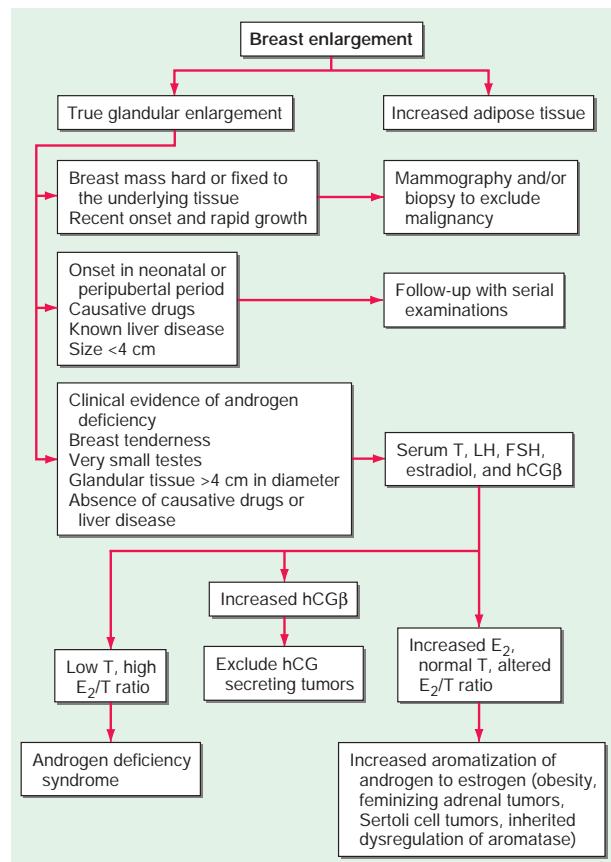
Gynecomastia refers to enlargement of the male breast. It is caused by excess estrogen action and is usually the result of an increased estrogen/androgen ratio. True gynecomastia is associated with glandular breast tissue that is >4 cm in diameter and often tender. Glandular tissue enlargement should be distinguished from excess adipose tissue: glandular tissue is firmer and contains fibrous-like cords. Gynecomastia occurs as a normal physiologic phenomenon in the newborn (due to transplacental transfer of maternal and placental estrogens), during puberty (high estrogen-to-androgen ratio in early stages of puberty), and with aging (increased fat tissue and increased aromatase activity along with the age-related decline in testosterone levels), but it can also result from pathologic conditions associated with androgen deficiency or estrogen excess. The prevalence of gynecomastia increases with age and body mass index (BMI), likely because of increased aromatase activity in adipose tissue. Medications that alter androgen metabolism or action may also cause gynecomastia. The relative risk of breast cancer is increased in men with gynecomastia, although the absolute risk is relatively small.

## PATHOLOGIC GYNECOMASTIA

Any cause of androgen deficiency can lead to gynecomastia, reflecting an increased estrogen/androgen ratio, as estrogen synthesis still occurs by aromatization of residual adrenal and gonadal androgens. Gynecomastia is a characteristic feature of Klinefelter syndrome (*Chap. 390*). Androgen insensitivity disorders also cause gynecomastia. Excess estrogen production may be caused by tumors, including Sertoli cell tumors in isolation or in association with Peutz-Jeghers syndrome or Carney complex. Tumors that produce hCG, including some testicular tumors, stimulate Leydig cell estrogen synthesis. Increased conversion of androgens to estrogens can be a result of increased availability of substrate (androstenedione) for extraglandular estrogen formation (CAH, hyperthyroidism, and most feminizing adrenal tumors) or of diminished catabolism of androstenedione (liver disease) so that estrogen precursors are shunted to aromatase in peripheral sites. Obesity is associated with increased aromatization of androgen precursors to estrogens. Extraglandular aromatase activity can also be increased in tumors of the liver or adrenal gland or rarely as an inherited disorder. Several families with increased peripheral aromatase activity inherited as an autosomal dominant or as an X-linked disorder have been described. In some families with this disorder, an inversion in chromosome 15q21.2-3 causes the *CYP19* gene to be

activated by the regulatory elements of contiguous genes, resulting in excessive estrogen production in the fat and other extragonadal tissues. The familial aromatase excess syndrome due to *CYP19* mutation or chromosomal rearrangement is characterized by pre- or peripubertal onset of gynecomastia, advanced bone age, short adult height due to premature epiphyseal closure, and hypogonadotropic hypogonadism. Peutz-Jeghers syndrome is characterized by intestinal hamartomas, mucocutaneous pigmentation, calcifying Sertoli cell tumors, and prepubertal gynecomastia due to increased aromatization and premature epiphyseal closure. Drugs can cause gynecomastia by acting directly as estrogenic substances (e.g., oral contraceptives, phytoestrogens, digitalis) or inhibiting androgen synthesis (e.g., ketoconazole) or action (e.g., spironolactone, AR blockers such as enzalutamide); for many drugs, such as cimetidine, imatinib, or some antiretroviral drugs for HIV, the precise mechanism is unknown. Unintentional exposure to estrogenic agents in skin care products has been reported as a cause of gynecomastia in prepubertal children.

Because up to two-thirds of pubertal boys and about half of hospitalized men have palpable glandular tissue that is benign, detailed investigation or intervention is not indicated in all men presenting with gynecomastia (*Fig. 391-6*). In addition to the extent of gynecomastia, recent onset, rapid growth, tender tissue, and occurrence in a lean subject should prompt more extensive evaluation. This should include a careful drug history, measurement and examination of the testes, assessment of virilization, evaluation of liver function, and hormonal measurements including testosterone, estradiol, androstenedione, LH, and hCG. Markedly elevated estradiol concentrations along with suppressed LH should prompt a search for a testicular or adrenal estrogen-secreting tumor. A karyotype should be obtained in men with very small testes to exclude Klinefelter syndrome. Despite



**FIGURE 391-6 Evaluation of gynecomastia.**  $E_2$ , 17 $\beta$ -estradiol; FSH, follicle-stimulating hormones; hCG $\beta$ , human chorionic gonadotropin  $\beta$ ; LH, luteinizing hormone; T, testosterone.

extensive evaluation, the etiology is established in fewer than one-half of patients.

## TREATMENT

### Gynecomastia

When the primary cause can be identified and corrected shortly after the onset of gynecomastia, breast enlargement usually subsides over several months. However, if gynecomastia is of long duration, surgery is the most effective therapy. Indications for surgery include severe psychological and/or cosmetic problems, continued growth or tenderness, or suspected malignancy. In patients who have painful gynecomastia and in whom surgery cannot be performed, treatment with antiestrogens such as tamoxifen (20 mg/d) can reduce pain and breast tissue size in over half the patients. The estrogen receptor antagonists tamoxifen and raloxifene have been reported in small trials to reduce breast size in men with pubertal gynecomastia, although complete regression of breast enlargement is unusual with the use of estrogen receptor antagonists. Aromatase inhibitors can be effective in the early proliferative phase of the disorder. However, in a randomized trial in men with established gynecomastia, anastrozole proved no more effective than placebo in reducing breast size. Tamoxifen is effective in prevention and treatment of breast enlargement and breast pain in men with prostate cancer who are receiving antiandrogen therapy.

## AGING-RELATED CHANGES IN MALE REPRODUCTIVE FUNCTION

A number of cross-sectional and longitudinal studies (e.g., the Baltimore Longitudinal Study of Aging, the Framingham Heart Study, the Massachusetts Male Aging Study, and the European Male Aging Study [EMAS]) have established that testosterone concentrations decrease with advancing age. This age-related decline starts in the third decade of life and progresses slowly; the rate of decline in testosterone concentrations is greater in obese men, in men with chronic illness, and in those taking medications. Because SHBG concentrations are higher in older men than in younger men, free or bioavailable testosterone concentrations decline with aging to a greater extent than total testosterone concentrations. The age-related decline in testosterone is due to defects at all levels of the hypothalamic-pituitary-testicular axis: pulsatile GnRH secretion is attenuated, LH response to GnRH is reduced, and testicular response to LH is impaired. However, the gradual rise of LH with aging suggests that testis dysfunction is the main cause of declining androgen levels. The term *andropause* has been used to denote age-related decline in testosterone concentrations; this term is a misnomer because there is no discrete time when testosterone concentrations decline abruptly.

Several epidemiologic studies, such as the Framingham Heart Study, the EMAS, and the Study of Osteoporotic Fractures in Men (MrOS), that used mass spectrometry for measuring testosterone levels have reported ~10% prevalence of low testosterone levels in middle-aged and older men; the prevalence of unequivocally low testosterone and sexual symptoms in men aged 40–70 years in the EMAS was 2.1% and increased with age from 0.1% for men aged 40–49 years of age to 5.1% for those aged 70–79 years. The age-related decline in testosterone should be distinguished from classical hypogonadism due to diseases of the testes, the pituitary, and the hypothalamus. Low total and bioavailable testosterone concentrations have been associated with decreased appendicular skeletal muscle mass and strength, decreased self-reported physical function, higher visceral fat mass, insulin resistance, and increased risk of coronary artery disease and mortality. An analysis of signs and symptoms in older men in the EMAS revealed a syndromic association of sexual symptoms with total testosterone levels <320 ng/dL and free testosterone levels <64 pg/mL in community-dwelling older men.

A series of placebo-controlled testosterone trials have provided important information about the efficacy of testosterone in improving outcomes in older men. Testosterone replacement in older men, aged

65, with sexual symptoms improved sexual activity, sexual desire, and erectile function more than placebo. Testosterone replacement did not improve fatigue or cognitive function and had only a small effect on mood and mobility. Among older men with low testosterone and age-associated memory impairment, testosterone replacement did not improve memory or other measures of cognition relative to placebo. Testosterone replacement was associated with significantly greater increase in vertebral as well as femoral volumetric bone mineral density and estimated bone strength relative to placebo. Testosterone replacement was associated with a greater increase in hemoglobin levels and corrected anemia in a greater proportion of men who had unexplained anemia of aging. Testosterone administration was associated with a significantly greater increase in coronary artery noncalcified plaque volume, as measured by coronary artery computed tomography angiography. Neither the testosterone trials nor a randomized trial of the effects of testosterone on atherosclerosis progression in aging men (TEAAM trial) with low or low normal testosterone levels found significant differences between testosterone and placebo arms in the rates of change in either the coronary artery calcium scores or the common carotid artery intima-media thickness. Neither of the trials was long enough or large enough to determine the effects of testosterone replacement therapy on prostate or major adverse cardiovascular events. In systematic reviews of randomized controlled trials, testosterone therapy of healthy older men with low or low-normal testosterone levels was associated with greater increments in lean body mass, grip strength, and self-reported physical function than that associated with placebo. Testosterone therapy has not been shown to improve clinical depression, fracture risk, progression to dementia, progression from prediabetes to diabetes, or response to phosphodiesterase inhibitors in older men.

The long-term risks of testosterone therapy remain largely unknown. While there is no evidence that testosterone causes prostate cancer, there is concern that testosterone therapy might cause subclinical prostate cancers to grow. Testosterone therapy is associated with increased risk of detection of prostate events.

The data relating cardiovascular disease (CVD) and venous thromboembolic (VTE) risk with the use of testosterone supplementation in men with low testosterone levels and hypogonadal symptoms are few and inconclusive. The relationship between testosterone and cardiovascular events in cross-sectional and prospective cohort studies has been inconsistent. A small number of epidemiologic studies have reported an inverse relationship between testosterone concentrations and common carotid artery intima-media thickness. Low testosterone level has been associated with increased risk of all-cause mortality, especially cardiovascular mortality. It is possible that testosterone is a marker of health; older men with multiple comorbid conditions who are at increased risk of death may have low testosterone levels as a result of comorbid conditions.

Most meta-analyses have not shown a statistically significant association between testosterone and cardiovascular events, major adverse cardiovascular events, or deaths. No adequately powered randomized trials have been conducted to determine the effects of testosterone replacement on major adverse cardiovascular events. Thus, there are insufficient data to establish a causal link between testosterone therapy and cardiovascular events.

Population screening of all older men for low testosterone levels is not recommended, and testing should be restricted to men who have symptoms or signs attributable to androgen deficiency. Testosterone therapy is not recommended for all older men with low testosterone levels. In older men with significant symptoms of androgen deficiency who have unequivocally low testosterone levels, testosterone therapy may be considered on an individualized basis and should be instituted after careful discussion of the risks and benefits (see “Testosterone Replacement,” below).

Testicular morphology, semen production, and fertility are maintained up to a very old age in men. Although concern has been expressed about age-related increases in germ cell mutations and impairment of DNA repair mechanisms, there is no clear evidence that the frequency of chromosomal aneuploidy is increased in the

sperm of older men. However, the incidence of autosomal dominant diseases, such as achondroplasia, polyposis coli, Marfan syndrome, and Apert syndrome, increases in the offspring of men who are advanced in age, consistent with transmission of sporadic missense mutations. Advanced paternal age may be associated with increased rates of de novo mutations, which may contribute to an increased risk of neurodevelopmental diseases such as schizophrenia and autism. The somatic mutations in male germ cells that enhance the proliferation of germ cells could lead to within-testis expansion of mutant clonal lines, thus favoring the propagation of germ cells carrying these pathogenic mutations and increasing the risk of mutations in the offspring of older fathers (the “selfish spermatogonial selection” hypothesis).

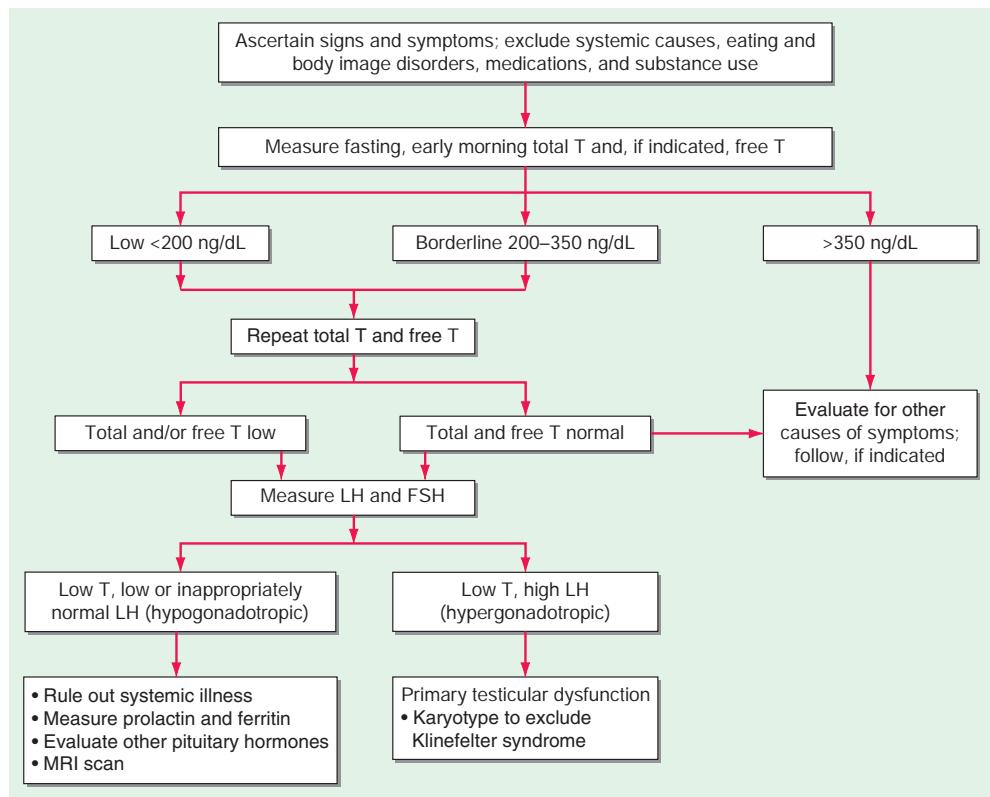
## APPROACH TO THE PATIENT

### Androgen Deficiency

Hypogonadism is often characterized by decreased sex drive, reduced frequency of sexual activity, inability to maintain erections, reduced beard growth, loss of muscle mass, decreased testicular size, and gynecomastia. Erectile dysfunction and androgen deficiency are two distinct clinical disorders that can coexist in middle-aged and older men. Some, but not all, patients with erectile dysfunction have testosterone deficiency. Thus, it is useful to evaluate men presenting with erectile dysfunction for androgen deficiency. Except when extreme, these clinical features of androgen deficiency may be difficult to distinguish from changes that occur with normal aging. Moreover, androgen deficiency may develop gradually. When symptoms or clinical features suggest possible androgen deficiency, the laboratory evaluation is initiated by the measurement of total testosterone in a fasting specimen, preferably in the morning using a reliable assay, such as LC-MS/MS that has been calibrated to an international testosterone standard (Fig. 391-7). A consistently low total testosterone level below the lower limit

of the normal male range, measured by an LC-MS/MS assay in a CDC-certified laboratory, in association with symptoms, is evidence of testosterone deficiency. An early-morning testosterone level  $>400$  ng/dL makes the diagnosis of androgen deficiency unlikely. In men with testosterone levels between 200 and 400 ng/dL, the total testosterone level should be repeated, and a free testosterone level should be measured. In older men and in patients with other clinical states that are associated with alterations in SHBG levels, a direct measurement of free testosterone level by equilibrium dialysis can be useful in unmasking testosterone deficiency.

When androgen deficiency has been confirmed by the consistently low testosterone concentrations, LH should be measured to classify the patient as having primary (high LH) or secondary (low or inappropriately normal LH) hypogonadism. An elevated LH level indicates that the defect is at the testicular level. Common causes of primary testicular failure include Klinefelter syndrome, HIV infection, uncorrected cryptorchidism, cancer chemotherapeutic agents, radiation, surgical orchectomy, or prior infectious orchitis. Unless causes of primary testicular failure are known, a karyotype should be performed in men with low testosterone and elevated LH to diagnose Klinefelter syndrome. Men who have a low testosterone but “inappropriately normal” or low LH levels have secondary hypogonadism; their defect resides at the hypothalamic-pituitary level. Common causes of acquired secondary hypogonadism include space-occupying lesions of the sella, hyperprolactinemia, chronic illness, hemochromatosis, excessive exercise, and the use of anabolic-androgenic steroids, opiates, marijuana, glucocorticoids, and alcohol. Measurement of PRL and MRI scan of the hypothalamic-pituitary region can help exclude the presence of a space-occupying lesion. Patients in whom known causes of hypogonadotropic hypogonadism have been excluded are classified as having IHH. It is not unusual for congenital causes of hypogonadotropic hypogonadism, such as Kallmann syndrome, to be diagnosed in young adults.



**FIGURE 391-7** Evaluation of hypogonadism. FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; LH, luteinizing hormone; T, testosterone.

## TREATMENT

### Androgen Deficiency

#### GONADOTROPINS

Gonadotropin therapy is used to establish or restore fertility in patients with gonadotropin deficiency of any cause. Several gonadotropin preparations are available. Human menopausal gonadotropin (hMG; purified from the urine of postmenopausal women) contains 75 IU FSH and 75 IU LH per vial. hCC (purified from the urine of pregnant women) has little FSH activity and resembles LH in its ability to stimulate testosterone production by Leydig cells. Recombinant LH is also available. Treatment is usually begun with hCG alone, and hMG is added later to promote the FSH-dependent stages of spermatid development. Recombinant human FSH (rhFSH) is available and is indistinguishable from purified urinary rhFSH in its biologic activity and pharmacokinetics *in vitro* and *in vivo*, although the mature subunit of recombinant rhFSH has seven fewer amino acids. Recombinant rhFSH is available in ampoules containing 75 IU (~7.5 µg FSH), which accounts for >99% of protein content. Once spermatogenesis is restored using combined FSH and LH therapy, hCG alone is often sufficient to maintain spermatogenesis.

Although a variety of treatment regimens are used, 1000 IU of hCG or recombinant human LH (rhLH) administered intramuscularly three times weekly is a reasonable starting dose. Testosterone levels should be measured 6–8 weeks later and 48–72 h after the hCG or rhLH injection; the hCG/rhLH dose should be adjusted to achieve testosterone levels in the mid-normal range. Sperm counts should be monitored on a monthly basis. It may take several months for spermatogenesis to be restored; therefore, it is important to forearm patients about the potential length and expense of the treatment and to provide conservative estimates of success rates. If testosterone levels are in the mid-normal range but the sperm concentrations are low after 6 months of therapy with hCG alone, FSH should be added. This can be done by using hMG, highly purified urinary rhFSH, or recombinant rhFSH. The selection of FSH dose is empirical. A common practice is to start with the addition of 75 IU FSH three times each week in conjunction with the hCG/rhLH injections. If sperm densities are still low after 3 months of combined treatment, the FSH dose should be increased to 150 IU. Occasionally, it may take 18–24 months for spermatogenesis to be restored.

The two best predictors of the success of gonadotropin therapy in hypogonadotropic men are testicular volume at presentation and time of onset of gonadotropin deficiency. In general, men with testicular volumes >8 mL have better response rates than those who have testicular volumes <4 mL. Patients who become hypogonadotropic after puberty experience higher success rates than those who have never undergone pubertal changes. Spermatogenesis can usually be reinitiated by hCC alone, with high rates of success for men with postpubertal onset of hypogonadotropism. The presence of a primary testicular abnormality, such as cryptorchidism, will attenuate testicular response to gonadotropin therapy. Prior androgen therapy does not preclude subsequent response to gonadotropin therapy, although some studies suggest that it may attenuate response to subsequent gonadotropin therapy.

#### TESTOSTERONE REPLACEMENT

Androgen therapy is indicated to restore testosterone levels to normal to correct features of androgen deficiency in men with organic hypogonadism due to known diseases of the testes, pituitary, and the hypothalamus. Testosterone replacement induces secondary sex characteristics; improves libido and overall sexual activity; increases lean muscle mass, hemoglobin, hematocrit, and bone mineral density; and decreases fat mass. The benefits of testosterone replacement therapy have only been proven in men who have documented symptomatic androgen deficiency, as demonstrated by testosterone levels that are well below the lower limit of normal.

Testosterone is available in a variety of formulations with distinct pharmacokinetics (**Table 391-3**). Testosterone serves as a

prohormone and is converted to 17 $\alpha$ -estradiol by aromatase and to 5 $\alpha$ -DHT by steroid 5 $\alpha$ -reductase. Therefore, when evaluating testosterone formulations, it is important to consider whether the formulation being used can achieve physiologic estradiol and DHT concentrations, in addition to normal testosterone concentrations. The current recommendation is to restore testosterone levels to the mid-normal range.

**Oral Derivatives of Testosterone** Testosterone is well absorbed after oral administration but is quickly degraded during the first pass through the liver. Therefore, it is difficult to achieve sustained blood levels of testosterone after oral administration of crystalline testosterone. 17 $\alpha$ -Alkylated derivatives of testosterone (e.g., 17 $\alpha$ -methyl testosterone, oxandrolone, fluoxymesterone) are relatively resistant to hepatic degradation and can be administered orally; however, because of the potential for hepatotoxicity, including cholestatic jaundice, peliosis, and hepatoma, these formulations should not be used for testosterone replacement. Hereditary angioedema due to C1 esterase deficiency is the only exception to this general recommendation; in this condition, oral 17 $\alpha$ -alkylated androgens are useful because they stimulate hepatic synthesis of the C1 esterase inhibitor.

**Injectable Forms of Testosterone** The esterification of testosterone at the 17 $\alpha$ -hydroxy position makes the molecule hydrophobic and extends its duration of action. The slow release of testosterone ester from an oily depot in the muscle accounts for its extended duration of action. The longer the side chain, the greater is the hydrophobicity of the ester and longer the duration of action. Thus, testosterone enanthate, cypionate, and undecanoate with longer side chains have longer durations of action than testosterone propionate. Within 24 h after intramuscular administration of 200 mg testosterone enanthate or cypionate, testosterone levels rise into the high-normal or supraphysiologic range and then gradually decline into the hypogonadal range over the next 2 weeks. A bimonthly regimen of testosterone enanthate or cypionate therefore results in peaks and troughs in testosterone levels that may be accompanied by changes in a patient's mood, sexual desire, and energy level; weekly administration of testosterone enanthate or cypionate can reduce these variations in testosterone levels during the dosing interval. The kinetics of testosterone enanthate and cypionate are similar. Estradiol and DHT levels are normal if testosterone replacement is physiologic.

A long-acting testosterone undecanoate in oil, administered at an initial priming dose of 750 mg intramuscularly followed by a second dose of 750 mg 4 weeks later, and then at a maintenance dose of 750 mg every 10 weeks, maintains serum testosterone, estradiol, and DHT in the normal male range and corrects symptoms of androgen deficiency in a majority of treated men. However, its relative drawbacks are the large injection volume and the risk of pulmonary oil microembolism (POME) reaction in a very small proportion of patients.

**Transdermal Testosterone Patch** The nongenital testosterone patch, when applied in an appropriate dose, can normalize testosterone, DHT, and estradiol levels 4–12 h after application. Sexual function and well-being are restored in androgen-deficient men treated with the nongenital patch. One 4-mg patch may not be sufficient to increase testosterone into the mid-normal male range in all hypogonadal men; many patients may need two 4-mg patches daily to achieve the targeted testosterone concentrations. The use of testosterone patches may be associated with skin irritation in some individuals.

**Testosterone Gel** Several transdermal testosterone gels, Androgel, Testim, Fortesta, and Axiron, and some generic versions, when applied topically to the skin in appropriate doses (**Table 391-3**), can maintain total and free testosterone concentrations in the normal range in hypogonadal men. The current recommendations are to begin with an initial U.S. Food and Drug Administration-recommended dose and adjust the dose based on testosterone levels.

**TABLE 391-3 Clinical Pharmacology of Some Testosterone Formulations**

FORMULATION	REGIMEN	PHARMACOKINETIC PROFILE	DHT AND E2	ADVANTAGES	DISADVANTAGES
T enanthate or cypionate	150–200 mg IM q2wk or 70–100 mg/wk	After a single IM injection, serum T levels rise into the supraphysiologic range, then decline gradually into the low-normal or the hypogonadal range by the end of the dosing interval	DHT and E2 levels rise in proportion to the increase in T levels; T:DHT and T:E2 ratios do not change	Corrects symptoms of androgen deficiency; relatively inexpensive if self-administered; flexibility of dosing	Requires IM injection; peaks and valleys in serum T levels that are associated with fluctuations in patient's mood, energy level, and sex drive
Topical T gels and axillary T solution	Available in sachets, tubes, and pumps	When used in appropriate doses, these topical formulations restore serum T and E2 levels to the physiologic male range	Serum DHT levels and DHT:T ratio are higher in hypogonadal men treated with the transdermal gels than in healthy eugonadal men	Corrects symptoms of androgen deficiency, ease of application, good skin tolerability	Potential of transfer to a female partner or child by direct skin-to-skin contact; skin irritation in a small proportion of treated men; moderately high DHT levels; considerable interindividual and intraindividual variation in on-treatment testosterone levels
Transdermal T patch	1 or 2 patches, designed to nominally deliver 4–8 mg T over 24 h applied daily on nonpressure areas	Restores serum T, DHT, and E2 levels to the physiologic male range	T:DHT and T:E2 levels are in the physiologic male range	Ease of application, corrects symptoms of androgen deficiency	Serum T levels in some androgen-deficient men may be in the low-normal range; these men may need application of 2 patches daily; skin irritation at the application site occurs frequently in many patients
Buccal, bioadhesive, T tablets	30-mg, controlled-release, bioadhesive tablets bid	Absorbed from the buccal mucosa	Normalizes serum T and DHT levels in hypogonadal men	Corrects symptoms of androgen deficiency	Gum-related adverse events in 16% of treated men
T pellets	Several pellets implanted SC; dose and regimen vary with formulation	Serum T peaks at 1 month and then is sustained in normal range for 3–4 months, depending on formulation	T:DHT and T:E2 ratios do not change	Corrects symptoms of androgen deficiency	Requires surgical incision for insertions; pellets may extrude spontaneously
17 $\alpha$ -Methyl T	This 17 $\alpha$ -alkylated compound should <i>not</i> be used because of potential for liver toxicity	Orally active			Clinical responses are variable; potential for liver toxicity; should <i>not</i> be used for treatment of androgen deficiency
Oral T undecanoate (TU)	237 mg PO bid with food	TU formulated in a self-emulsifying drug delivery system that includes hydrophilic and lipophilic excipients to enable the solubilization of TU and its absorption through the lymphatics after oral ingestion with a typical meal. After each administration, serum T levels rise and return to baseline by 12 h. When administered at the recommended dose, average serum T levels are maintained in the normal range in a majority of treated men	High DHT:T ratio	Convenience of oral administration	High DHT:T ratio
Injectable long-acting TU in oil	U.S. regimen 750 mg IM, followed by 750 mg at 4 weeks, and 750 mg every 10 weeks	When administered at the recommended dose, serum T levels are maintained in the normal range in a majority of treated men	DHT and E2 levels rise in proportion to the increase in T levels; T:DHT and T:E2 ratios do not change	Corrects symptoms of androgen deficiency; requires infrequent administration	Requires IM injection of a large volume; serious pulmonary oil microembolism (POME) reactions, characterized by cough, dyspnea, throat tightening, chest pain, dizziness, and syncope, and episodes of anaphylaxis have been reported to occur during or immediately after the injection in a very small number of patients; patients should be watched for POME reaction for 30 min after each injection
T-in-adhesive matrix patch <sup>a</sup>	2 × 60 cm <sup>2</sup> patches delivering ~4.8 mg of T/d	Restores serum T, DHT, and E2 to the physiologic range	T:DHT and T:E2 are in the physiologic range	Lasts 2 d	Some skin irritation
Intranasal T	2 actuations of the metered-dose pump (11 mg) applied into the nostrils 3 times daily	Restores T into the normal male range	T:DHT and T:E2 ratio in the physiologic range		Requires 3 times daily application; nasal irritation, epistaxis, nasopharyngitis

<sup>a</sup>These formulations are not approved for clinical use in the United States but are available outside the United States in many countries. Physicians in those countries where these formulations are available should follow the approved drug regimens.

Abbreviations: DHT, dihydrotestosterone; E2, estradiol; T, testosterone.

The advantages of the testosterone gel include the ease of application. A major concern is the potential for inadvertent transfer of the gel to a sexual partner or to children who may come in close contact with the patient. The ratio of DHT to testosterone concentrations is higher in men treated with the testosterone gel than in healthy men. Also, there is considerable intra- and interindividual variation in serum testosterone levels in men treated with the transdermal gel due to variations in transdermal absorption and plasma clearance of testosterone. Therefore, monitoring of serum testosterone levels and multiple dose adjustments may be required to achieve and maintain testosterone levels in the target range.

**Buccal Adhesive Testosterone** A buccal testosterone tablet, which adheres to the buccal mucosa and releases testosterone as it is slowly dissolved, has been approved. After twice-daily application of 30-mg tablets, serum testosterone levels are maintained within the normal male range in a majority of treated hypogonadal men. The adverse effects include buccal ulceration and gum problems in a few subjects. The effects of food and brushing on absorption have not been studied in detail.

Pellets of crystalline testosterone can be inserted in the subcutaneous tissue through a small skin incision. Testosterone is released by surface erosion of the implant and absorbed into the systemic circulation, and testosterone levels can be maintained in the normal range for 3–4 months. Potential drawbacks include incising the skin for insertion and removal and spontaneous extrusions and fibrosis at the site of the implant.

Testosterone undecanoate, formulated in a self-emulsifying drug delivery system that includes hydrophilic and lipophilic excipients to enable its solubilization in the gut, is absorbed through the lymphatics after oral ingestion with a typical meal and is spared the first-pass degradation in the liver. After each administration, serum testosterone levels rise and return to baseline by 12 h. When administered at the recommended dose, average serum testosterone levels are maintained in the normal range in a majority of treated men, but DHT-to-testosterone ratios are higher in hypogonadal men treated with oral testosterone undecanoate, as compared to eugonalad men.

An intranasal testosterone gel is now available as a metered-dose pump and is administered typically at a starting dose of 11 mg testosterone in the form of two pump actuations, one in each nostril three times daily. Formulation-specific adverse effects include rhinorrhea, nasal discomfort, epistaxis, nasopharyngitis, and nasal scab.

**Novel Androgen Formulations** A number of androgen formulations with better pharmacokinetics or more selective activity profiles are under development. Long-acting biodegradable microsphere formulations have also been investigated. 7 -Methyl-19-nortestosterone is an androgen that cannot be 5 -reduced; therefore, compared to testosterone, it has relatively greater agonist activity in muscle and gonadotropin suppression but lesser activity on the prostate.

Selective AR modulators (SARMs) are a class of AR ligands that bind the AR and display tissue-selective actions. A number of nonsteroidal SARMs that act as agonists on the muscle and bone and that spare the prostate to varying degrees have advanced to phase 3 human trials. Nonsteroidal SARMs do not serve as substrates for either the steroid 5 -reductase or the CYP19 (aromatase). SARM binding to AR induces specific conformational changes in the AR protein, which then modulates protein–protein interactions between AR and its coregulators, resulting in tissue-specific regulation of gene expression. SARMs that are strong agonists for the muscle, bone, and sexual function and antagonists for the prostate may be valuable in treating men with prostate cancer who are receiving androgen deprivation therapy.

**Pharmacologic Uses of Androgens** Androgens and SARMs are being evaluated as anabolic therapies for functional limitations associated with aging and chronic illness. Testosterone supplementation increases skeletal muscle mass, maximal voluntary strength, and muscle power in healthy men, hypogonadal men, older men with

low testosterone levels, HIV-infected men with weight loss, and men receiving glucocorticoids. These anabolic effects of testosterone are related to testosterone dose and circulating concentrations. Systematic reviews have confirmed that testosterone therapy of HIV-infected men with weight loss promotes improvements in body weight, lean body mass, muscle strength, and depression indices, leading to the recommendation that testosterone be considered as an adjunctive therapy in HIV-infected men who are experiencing unexplained weight loss and who have low testosterone levels. It is unknown whether testosterone therapy of older men with functional limitations is safe and effective in improving physical function, vitality, and health-related quality of life, and reducing disability. Concerns about potential adverse effects of testosterone on the prostate and cardiovascular event rates have encouraged the development of SARMs that are preferentially anabolic and spare the prostate.

Testosterone administration induces hypertrophy of both type 1 and 2 fibers and increases satellite cell (muscle progenitor cells) and myonuclear number. Androgens promote the differentiation of mesenchymal, multipotent progenitor cells into the myogenic lineage and inhibit their differentiation into the adipogenic lineage. Testosterone binding to AR promotes the association of liganded AR with -catenin and its translocation into the nucleus where it binds TCF-4 and activates Wnt-target genes, including follistatin, which blocks signaling through the transforming growth factor pathway, thereby promoting myogenic differentiation of muscle progenitor cells. Testosterone may have additional effects on satellite cell replication and polyamine pathway, which may contribute to an increase in skeletal muscle mass.

Other indications for androgen therapy are in selected patients with anemia due to bone marrow failure (an indication largely supplanted by erythropoietin) or for hereditary angioedema.

**Male Hormonal Contraception Based on Combined Administration of Testosterone and Gonadotropin Inhibitors** Supraphysiologic doses of testosterone (200 mg testosterone enanthate weekly) suppress LH and FSH secretion and induce azoospermia in 50% of Caucasian men and >95% of Chinese men. The WHO-supported multicenter efficacy trials have demonstrated that suppression of spermatogenesis to azoospermia or severe oligozoospermia (<3 million/mL) by administration of supraphysiologic doses of testosterone enanthate to men results in highly effective contraception. Because of concern about long-term adverse effects of supraphysiologic testosterone doses, regimens that combine other gonadotropin inhibitors, such as GnRH antagonists and progestins, with replacement doses of testosterone have been investigated. Regimens containing an androgen plus a progestin such as depo-medroxyprogesterone acetate, etonogestrel, or norethisterone enanthate have been highly effective in inducing azoospermia or severe oligozoospermia (sperm density <1 million/mL) in nearly 99% of treated men over a 1-year period. The combined regimens of testosterone plus a progestin have been associated with weight gain, acne, mood changes including depressed mood, libido changes, and decreased plasma high-density lipoprotein (HDL) cholesterol, and their long-term safety has not been demonstrated. One such trial of a combined regimen of testosterone undecanoate plus norethisterone enanthate was stopped early due to adverse events. SARMs, which are more potent inhibitors of gonadotropins than testosterone and spare the prostate, hold promise for their contraceptive potential.

**Recommended Regimens for Androgen Replacement** Testosterone esters are administered typically at doses of 70–100 mg intramuscularly every week or 140–200 mg every 2 weeks. Testosterone undecanoate is administered at an initial dose of 750 mg followed 4 weeks later by a second injection of 750 mg and then 750 mg every 10 weeks. Testosterone gels are typically applied over a covered area of skin at initial doses that vary with the formulation. Patients should wash their hands after gel application and keep the area of gel application covered with clothing to minimize the risk of gel transfer to another person. One or two 4-mg nongenital

testosterone patches are applied daily over the skin of the back, thigh, or upper arm away from pressure areas. Bioadhesive buccal testosterone tablets at a dose of 30 mg are applied twice daily on the buccal mucosa. Oral testosterone undecanoate is taken twice daily with meals at a starting dose of 237 mg. Intranasal testosterone is administered as a spray in each nostril three times a day (33 mg/d).

**Evaluating Efficacy of Testosterone Replacement Therapy** Because a clinically useful marker of androgen action is not available, correction of symptoms, induction and maintenance of secondary sex characteristics, and restoration of testosterone levels into the mid-normal range remain the goals of therapy. Measurements of LH and FSH are not useful in assessing the adequacy of testosterone replacement. Testosterone should be measured 3 months after initiating therapy to assess adequacy of therapy. There is substantial interindividual variability in serum testosterone levels, especially with transdermal gels, presumably due to genetic differences in testosterone clearance and substantial variation in transdermal absorption. In patients who are treated with testosterone enanthate or cypionate, testosterone levels should be 350–600 ng/dL 1 week after the injection. If testosterone levels are outside this range, adjustments should be made either in the dose or in the interval between injections. In men on transdermal patch, gel, or buccal testosterone therapy, testosterone levels should be in the mid-normal range (400–750 ng/dL) 4–12 h after application. If testosterone levels are outside this range, the dose should be adjusted. Multiple dose adjustments are often necessary to achieve testosterone levels in the desired therapeutic range.

Restoration of sexual function, induction and maintenance of secondary sex characteristics, well-being, and maintenance of muscle and bone health are important objectives of testosterone replacement therapy. The patient should be asked about sexual desire and activity, the presence of early morning erections, and the ability to achieve and maintain erections adequate for sexual intercourse. The hair growth in response to androgen replacement is variable and depends on ethnicity. Hypogonadal men with prepubertal onset of androgen deficiency who begin testosterone therapy in their late twenties or thirties may find it difficult to adjust to their newly found sexuality and may benefit from counseling. If the patient has a sexual partner, the partner should be included in counseling because of the dramatic physical and sexual changes that occur with androgen treatment.

**Contraindications for Androgen Administration** Testosterone administration is contraindicated in men with prostate or breast cancer (**Table 391-4**). Testosterone therapy should not be administered without further urologic evaluation to men with a palpable prostate nodule or induration, a prostate-specific antigen >3 ng/mL, or severe

**TABLE 391-4 Conditions in Which Testosterone Administration Is Associated with an Increased Risk of Adverse Outcomes**

Conditions in which testosterone administration is associated with very high risk of serious adverse outcomes:
Metastatic prostate cancer
Breast cancer
Conditions in which testosterone administration is associated with moderate to high risk of adverse outcomes:
Undiagnosed prostate nodule or induration
PSA >3
Erythrocytosis (hematocrit >50%)
Severe lower urinary tract symptoms associated with benign prostatic hypertrophy as indicated by American Urological Association/International prostate symptom score >19
Uncontrolled or poorly controlled congestive heart failure
Myocardial infarction, stroke, or acute coronary syndrome in the preceding 3 months

Abbreviation: PSA, prostate-specific antigen.

Source: Modified with permission from S Bhasin et al: Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 95:2536, 2010.

lower urinary tract symptoms (American Urological Association lower urinary tract symptom score >19). Testosterone replacement should not be administered to men with baseline hematocrit 50%, severe untreated obstructive sleep apnea, or uncontrolled or poorly controlled congestive heart failure, or with myocardial infarction, stroke, or acute coronary syndrome in the preceding 3 months.

**Monitoring Potential Adverse Experiences** The clinical effectiveness and safety of testosterone replacement therapy should be assessed 3–6 months after initiating testosterone therapy and annually thereafter (**Table 391-5**). Potential adverse effects include acne, oiliness of skin, erythrocytosis, breast tenderness and enlargement, leg edema, and increased risk of detection of prostate events. In addition, there may be formulation-specific adverse effects such as skin irritation with transdermal patch; risk of gel transfer to a sexual partner with testosterone gels; buccal ulceration and gum problems with buccal testosterone; pain and mood fluctuation with injectable testosterone esters; cough and injection site pain with long-acting testosterone undecanoate; and nasal irritation, epistaxis, and nasal scab with intranasal formulation.

**Hemoglobin Levels** Administration of testosterone to androgen-deficient men is typically associated with only a small (<3%) increase in hemoglobin levels, due to direct effects of testosterone on hematopoietic progenitors in the bone marrow, stimulation of erythropoietin, suppression of hepcidin, and increased iron availability for erythropoiesis. The magnitude of hemoglobin increase during testosterone therapy is greater in older men than younger men and in men who have sleep apnea, a significant smoking history, or chronic obstructive lung disease or who live at high altitude. The frequency of erythrocytosis is higher in hypogonadal men treated with injectable testosterone esters than in those treated with transdermal formulations, presumably due to the higher testosterone dose delivered by the typical regimens of testosterone esters. Erythrocytosis is the most frequent adverse event reported in testosterone trials in middle-aged and older men and is also the most frequent cause of treatment discontinuation in these trials. If hematocrit rises above 54%, testosterone therapy should be stopped until hematocrit has fallen to <50%. After evaluation of the patient for hypoxia and sleep apnea, testosterone therapy may be reinitiated at a lower dose.

**Prostate and Serum Prostate-Specific Antigen (PSA) Levels** Testosterone replacement therapy increases prostate volume to the size seen in age-matched controls but does not increase prostate volume beyond that expected for age. There is no evidence that testosterone therapy causes prostate cancer. However, androgen administration can exacerbate preexisting metastatic prostate cancer. Many older men harbor microscopic foci of cancer in their prostates. It is not known whether long-term testosterone administration will induce these microscopic foci to grow into clinically significant cancers.

PSA levels are lower in testosterone-deficient men and are restored to normal after testosterone replacement. There is considerable test-retest variability in PSA measurements. Increments in PSA levels after testosterone supplementation in androgen-deficient men are generally <0.5 ng/mL, and increments >1.0 ng/mL over a 3–6-month period are unusual. The 90% confidence interval for the change in PSA values in men with benign prostatic hyper trophy, measured 3–6 months apart, is 1.4 ng/mL. Therefore, the Endocrine Society expert panel suggested that an increase in PSA >1.4 ng/mL in any 1 year after starting testosterone therapy, if confirmed, should lead to urologic evaluation. PSA velocity criterion can be used for patients who have sequential PSA measurements for >2 years; a change of >0.40 ng/mL per year merits closer urologic follow-up. PSA level >4 ng/mL during treatment, if confirmed by repeat testing, requires further urologic evaluation.

**Cardiovascular Risk** As discussed above, there is insufficient evidence to determine whether testosterone replacement therapy increases the risk of major adverse cardiovascular events in hypogonadal men. In two randomized, placebo-controlled trials, the

**TABLE 391-5 Monitoring Men Receiving Testosterone Therapy**

- Evaluate the patient 3–6 months after treatment initiation and then annually to assess whether symptoms have responded to treatment and whether the patient is suffering from any adverse effects.
- Monitor testosterone level 3–6 months after initiation of testosterone therapy:
  - Therapy should aim to raise serum testosterone level into the mid-normal range.
  - Injectable testosterone enanthate or cypionate: Measure serum testosterone level midway between injections. If testosterone is >600 ng/dL (20.9 nmol/L) or <350 ng/dL (12.2 nmol/L), adjust dose or frequency.
  - Transdermal patches: Assess testosterone level 3–12 h after application of the patch; adjust dose to achieve testosterone level in the mid-normal range.
  - Buccal testosterone bioadhesive tablet: Assess level immediately before application of fresh system.
  - Transdermal gels and solution: Assess testosterone level 2–12 h after patient has been on treatment for at least 2 weeks; adjust dose to achieve serum testosterone level in the mid-normal range.
  - Testosterone pellets: Measure testosterone levels at the end of the dosing interval. Adjust the number of pellets and/or the dosing interval to achieve serum testosterone levels in the normal range.
  - Oral testosterone undecanoate: Measure testosterone levels 4–6 h after an oral dose.
  - Injectable testosterone undecanoate: Measure serum testosterone level just prior to each subsequent injection and adjust the dosing interval to maintain serum testosterone in mid-normal range.
- Check hematocrit at baseline, at 3–6 months, and then annually. If hematocrit is >54%, stop therapy until hematocrit decreases to a safe level; evaluate the patient for hypoxia and sleep apnea; reinitiate therapy with a reduced dose.
- Measure bone mineral density of lumbar spine and/or femoral neck after 1–2 years of testosterone therapy in hypogonadal men with osteoporosis or low trauma fracture, consistent with regional standard of care.
- In men aged >40 years with baseline PSA >0.6 ng/mL, perform digital rectal examination and check PSA level before initiating treatment, at 3–6 months, and then in accordance with guidelines for prostate cancer screening depending on the age and race of the patient.
- Obtain urologic consultation if there is:
  - An increase in serum PSA concentration >1.4 ng/mL within any 12-month period of testosterone treatment, confirmed by repeating the test.
  - A PSA level >4 ng/mL any time during treatment, confirmed by repeating the test.
  - Detection of a prostatic abnormality on digital rectal examination.
  - An AUA/IPSS prostate symptom score of >19 along with an increase in IPSS score of 5 points above baseline.
- Evaluate formulation-specific adverse effects at each visit:
  - Buccal testosterone tablets<sup>a</sup>: Inquire about alterations in taste and examine the gums and oral mucosa for irritation.
  - Injectable testosterone esters (enanthate, cypionate, and undecanoate): Ask about fluctuations in mood or libido, and rarely cough after injections.
  - Testosterone patches: Look for skin reaction at the application site.
  - Testosterone gels: Advise patients to cover the application sites with a shirt and to wash the skin with soap and water before having skin-to-skin contact because testosterone gels leave a testosterone residue on the skin that can be transferred to a woman or child who might come in close contact. Serum testosterone levels are maintained when the application site is washed 4–6 h after application of the testosterone gel.
  - Testosterone undecanoate injection: Observe patients for POME reaction for 30 min after each injection.
  - Testosterone pellets: Look for signs of infection, fibrosis, or pellet extrusion.
  - Intranasal testosterone: Look for signs of nasal irritation or scab.

<sup>a</sup>Not approved for clinical use in the United States.

Abbreviations: AUA/IPSS, American Urological Association International Prostate Symptom Score; POME, pulmonary oil microembolism; PSA, prostate-specific antigen.

Source: Modified with permission from S Bhasin et al: Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 95:2536, 2010.

rate of progression of coronary artery atherosclerosis did not differ between the testosterone- and placebo-treated men. In another randomized trial, compared to placebo, testosterone treatment was associated with greater increase in noncalcified plaque volume in

the coronary arteries, assessed using CT coronary angiography. An ongoing large prospective randomized trial (TRAVERSE trial) will determine the effects of testosterone replacement therapy on major adverse cardiovascular events in middle-aged and older hypogonadal men at increased risk of CVD.

**Androgen Abuse by Athletes and Recreational Bodybuilders** The illicit use of androgenic-anabolic steroids (AAS) to enhance athletic performance first surfaced in the 1950s among powerlifters and spread rapidly to other sports, professional as well as high school athletes, and recreational bodybuilders. In the early 1980s, the use of AAS spread beyond the athletic community into the general population, and now, as many as 3–4 million Americans—most of them men—have likely used these compounds. Most AAS users are not athletes, but rather recreational weightlifters, almost all men, who use these drugs to look lean and more muscular. A subset of AAS users suffers from muscle dysmorphia, a form of body image disorder characterized by excessive preoccupation with leanness and muscularity and poor functioning in social and occupational life. A secular transformation in idealized body image toward greater muscularity and leanness has contributed to increasing prevalence of body image disorders and the use of muscle-building anabolic drugs in young men.

The most commonly used AAS include testosterone esters, nandrolone, stanozolol, methandienone, and methenolol. AAS users generally use large doses of multiple steroids in cycles, a practice known as stacking. AAS users may also typically use other drugs that are perceived to be muscle building or performance enhancing, such as human GH; erythropoiesis-stimulating agents; insulin; stimulants such as amphetamine, clenbuterol, cocaine, ephedrine, and thyroxine; and drugs perceived to reduce adverse effects such as hCG, aromatase inhibitors, or estrogen antagonists. Recent years have witnessed increasing use of unapproved nonsteroidal SARMs and GH secretagogues purchased from internet sites.

Most of the information about the adverse effects of AAS has emerged from case reports, uncontrolled studies, or clinical trials that used replacement doses of testosterone. The adverse event data from clinical trials using physiologic replacement doses of testosterone have been extrapolated unjustifiably to AAS users, who may administer 10–100 times the replacement doses of testosterone over many years, to support the claim that AAS use is safe and manageable. The adverse events associated with AAS use may be due to AAS themselves, concomitant use of other drugs, high-risk behaviors, and host characteristics that may render these individuals more susceptible to AAS use or to other high-risk behaviors.

The high rates of premature mortality and morbidities observed in AAS users are alarming. One Finnish study reported 4.6 times the risk of death among elite powerlifters compared with age-matched men from the general population. The causes of death among powerlifters included suicides, myocardial infarction, and liver failure. A retrospective review of patient records in Sweden also reported higher standardized mortality ratios for AAS users than for nonusers and increased death rates due to suicide, homicide, and accidents.

Four categories of adverse events associated with AAS abuse are of particular concern: cardiovascular events, psychiatric, prolonged suppression of the hypothalamic-pituitary-testicular axis, and potential neurotoxicity. Numerous reports of premature cardiac death among young AAS users raise concerns about the adverse cardiovascular effects of AAS. High doses of AAS may induce proatherogenic dyslipidemia, accelerate atherogenesis, increase thrombosis risk via effects on clotting factors and platelets, and induce vasospasm through their effects on vascular nitric oxide. Long-term AAS use may be associated with myocardial hypertrophy and fibrosis. Myocardial tissue of powerlifters using AAS has been shown to be infiltrated with fibrous tissue and fat droplets. Current AAS users display significantly reduced left ventricular systolic and diastolic function compared to previous users and nonusers. Additionally, studies using CT angiography have reported higher coronary artery plaque volume in AAS users than

in nonusers. Lifetime AAS dose is strongly associated with coronary atherosclerotic burden. Power athletes using AAS often have short QT intervals but increased QT dispersion, which may predispose them to ventricular arrhythmias.

Unlike replacement doses of testosterone, which are associated with only a small decrease in HDL cholesterol and little or no effect on total cholesterol, low-density lipoprotein (LDL) cholesterol, and triglyceride levels, supraphysiologic doses of testosterone and orally administered 17'-alkylated, nonaromatizable AAS are associated with marked reductions in HDL cholesterol and increases in LDL cholesterol.

Some AAS users develop hypomanic and manic symptoms (irritability, aggressiveness, reckless behavior, and occasional psychotic symptoms, sometimes associated with violence) during AAS exposure, and depression, sometimes associated with suicidality, during AAS withdrawal. Users may also be susceptible to other forms of illicit drug use.

Long-term AAS use suppresses LH, FSH, and testosterone production and spermatogenesis. Men who have used AAS for more than a few months experience marked suppression of the hypothalamic-pituitary-testicular (HPT) axis after stopping AAS that may be associated with sexual dysfunction, fatigue, infertility, depressed mood, and even suicidality. In some long-term AAS users, recovery of the HPT axis may take a long time, may be incomplete, or may never occur. The symptoms of androgen deficiency caused by androgen withdrawal may cause some men to revert back to using AAS, leading to continued use and AAS dependence. As many as 30% of AAS users develop a syndrome of AAS dependence, characterized by long-term AAS use despite adverse medical and psychiatric effects. AAS withdrawal hypogonadism has emerged as an important cause of androgen deficiency, accounting for a substantial fraction of testosterone prescriptions in many men's health clinics; therefore, AAS use should be considered in the differential diagnosis of hypogonadism in young men.

Supraphysiologic doses of testosterone may also impair insulin sensitivity. Orally administered androgens also have been associated with insulin resistance and diabetes.

AAS users are more likely to engage in high-risk behaviors such as unsafe injection practices and have increased rates of incarceration that may render them at increased risk of HIV and hepatitis B and C. AAS users are more likely to report high-risk unprotected anal sex than nonusers.

Elevated liver enzymes, cholestatic jaundice, hepatic neoplasms, and peliosis hepatitis have been reported with oral, 17'-alkylated AAS. AAS use may cause muscle hypertrophy without compensatory adaptations in tendons, ligaments, and joints, thus increasing the risk of tendon and joint injuries. Upper extremity tendon ruptures are observed almost exclusively among weightlifters who use AAS. AAS use is associated with acne, baldness, and increased body hair.

## APPROACH TO THE PATIENT

### Androgenic-Anabolic Steroids Use

The suspicion of AAS use should be raised by increased hemoglobin and hematocrit levels, suppressed LH and FSH and testosterone levels, low HDL cholesterol, and low testicular volume and sperm density in a person who looks highly muscular. In AAS users seeking medical attention, evaluation using the Appearance and Performance Enhancing Drug Use Schedule (APEDUS), a validated semi-structured interview, is sufficient to assess the associated body image or eating disorder, psychiatric symptoms, and the use of AAS and other substances; formal testing for AAS usually is not needed. History of AAS use should be obtained in all young men being evaluated for hypogonadism. As AAS use is often associated with the use of other substances, a urine drug screen for other substances is helpful in guiding treatment. If needed, accredited laboratories use gas chromatography-mass spectrometry or liquid chromatography-mass spectrometry to detect anabolic steroid abuse. High-resolution mass spectrometry and tandem mass

spectrometry have further improved the sensitivity of detecting AAS use. Illicit testosterone use is detected generally by the measurement of urinary testosterone-to-epitestosterone ratio and further confirmed by the  $^{13}\text{C}:\text{C}^{12}$  ratio in testosterone using isotope ratio combustion mass spectrometry. Exogenous testosterone administration increases urinary testosterone glucuronide excretion and consequently the testosterone-to-epitestosterone ratio. Ratios  $>4$  suggest exogenous testosterone use but can also reflect genetic variation. Genetic variations in the uridine diphospho-glucuronyltransferase 2B17 (*UGT2B17*), the major enzyme for testosterone glucuronidation, affect the testosterone-to-epitestosterone ratio. Synthetic testosterone has a lower  $^{13}\text{C}:\text{C}^{12}$  ratio than endogenously produced testosterone, and these differences in  $^{13}\text{C}:\text{C}^{12}$  ratio can be detected by isotope ratio combustion mass spectrometry, which is used to confirm exogenous testosterone use in individuals with a high testosterone-to-epitestosterone ratio.

The treatment of AAS use disorder requires a multidisciplinary team that includes an endocrinologist or an internist to treat the AAS withdrawal hypogonadism and other medical problems; a mental health expert to treat the substance use disorder and depressive symptoms and to address suicide risk and body image disorder; and sometimes a social worker for care coordination. In patients who are willing to stop or who have already stopped AAS use, the initial step is to restore the hypothalamic-pituitary-gonadal axis by administering either clomiphene (or its enantiomer trans clomiphene), a partial estrogen agonist, at an initial dose of 50 mg daily or hCG at a dose of 750–1000 IU three times weekly. Some men may not respond to clomiphene and may require switching to hCG. AAS users also need evaluation and treatment of the underlying body image disorder. Mirror exposure therapy in which the patient stands in front of a mirror and describes his body appearance to the mental health provider has been moderately efficacious in small, randomized trials. Body dysmorphism may require cognitive-behavioral therapy or pharmacotherapy using selective serotonin uptake inhibitors or tricyclic antidepressants.

## FURTHER READING

- Bangalore KK et al: Use of gonadotropin-releasing hormone analogs in children: Update by an International Consortium. *Horm Res Paediatr* 91:357, 2019.
- Bhasin S: Testosterone replacement in aging men: An evidence-based patient-centric perspective. *J Clin Invest* 131: e146607, 2021.
- Bhasin S et al: Testosterone therapy in men with hypogonadism: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 103:1715, 2018.
- Finkelstein JS et al: Gonadal steroids and body composition, strength, and sexual function in men. *N Engl J Med* 369:1011, 2013.
- Hildebrandt T et al: Body image disturbance in 1000 male appearance and performance enhancing drug users. *J Psychiatr Res* 44:841, 2010.
- Hollis B et al: Genomic analysis of male puberty timing highlights shared genetic basis with hair color and lifespan. *Nat Commun* 11:1536, 2020.
- Hughes JF, Page DC: The biology and evolution of mammalian Y chromosomes. *Annu Rev Genet* 49:507, 2015.
- Jasuja R et al: Estradiol binding induces bidirectional allosteric coupling and repartitioning of sex hormone binding globulin monomers among various conformational states. *iScience* 24:102414, 2021.
- O'Shaughnessy PJ et al: Alternative (backdoor) androgen production and masculinization in the human fetus. *PLoS Biol* 17:e3000002, 2019.
- Pope HG Jr et al: Adverse health consequences of performance-enhancing drugs: An Endocrine Society scientific statement. *Endocr Rev* 35:341, 2014.
- Sedlmeier IL et al: Delayed puberty: Analysis of a large case series from an academic center. *J Clin Endocrinol Metab* 87:1613, 2002.
- Snyder PJ et al: Effects of testosterone treatment in older men. *N Engl J Med* 74:611, 2016.
- Stamou MI et al: Kallmann syndrome: Phenotype and genotype of hypogonadotropic hypogonadism. *Metabolism* 86:124, 2018.

# 392

## Disorders of the Female Reproductive System

Janet E. Hall, Anuja Dokras

The female reproductive system regulates the hormonal changes responsible for puberty and adult reproductive function. Normal reproductive function in women requires the dynamic integration of hormonal signals from the hypothalamus, pituitary, and ovary, resulting in repetitive cycles of follicle development, ovulation, and preparation of the endometrial lining of the uterus for implantation should conception occur.

For further discussion of related topics, see the following chapters: amenorrhea and pelvic pain (Chap. 393), infertility and contraception (Chap. 396), menopause (Chap. 395), disorders of sex development (Chap. 390), and disorders of the male reproductive system (Chap. 391).

### DEVELOPMENT OF THE OVARY AND EARLY FOLLICULAR GROWTH

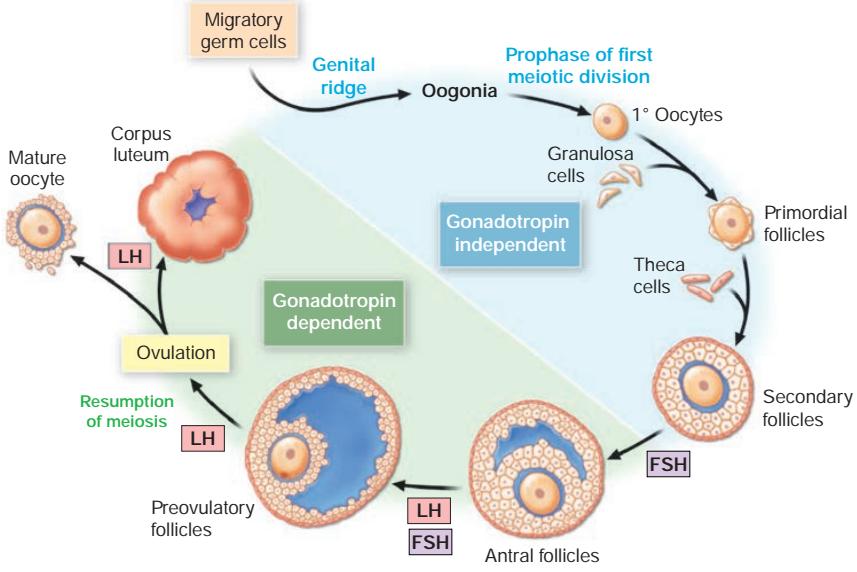
The ovary orchestrates the development and release of a mature oocyte and secretes hormones (e.g., estrogen, progesterone, inhibins A and B, relaxin) that play critical roles in a variety of target tissues, including breast, bone, and uterus, in addition to the hypothalamus and pituitary. To achieve these functions in repeated monthly cycles, the ovary undergoes some of the most dynamic changes of any organ in the body. Primordial germ cells can be identified by the third week of gestation, and their migration to the genital ridge is complete by 6 weeks of gestation. Germ cells persist within the genital ridge, are then referred to as *oogonia*, and are essential for induction of ovarian development. In patients with 45,X Turner syndrome, primordial germ cells proliferate and migrate to the genital ridge but do not persist because their survival requires pregranulosa cells that are dependent on the presence of both X chromosomes (Chap. 390).

The germ cell population expands, and starting at ~8 weeks of gestation, oogonia begin to enter prophase of the first meiotic division and become primary oocytes. This allows the oocyte to be surrounded by a single layer of flattened granulosa cells to form a primordial follicle (Fig. 392-1). Granulosa cells are derived from mesonephric cells that migrate into the ovary early in its development, pushing the germ cells to the periphery. Although there is evidence that both oocyte-like cells and

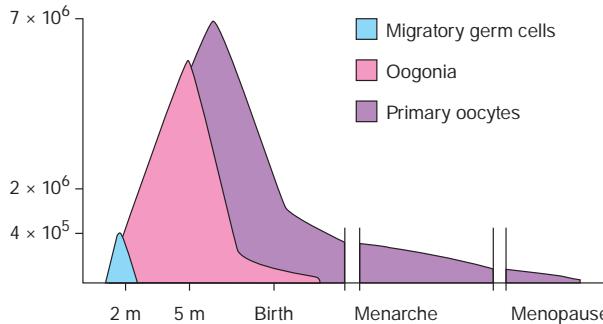
follicle-like structures can form from embryonic stem cells in culture, there is, as yet, no clear evidence that this occurs *in vivo*, and thus, the ovary appears to contain a nonrenewable pool of germ cells. Through the combined processes of mitosis, meiosis, and atresia, the population of oogonia reaches its maximum of 6–7 million by 20 weeks in the fetus, after which there is a progressive loss of both oogonia and primordial follicles through the process of atresia. It appears that entry into meiosis provides some degree of protection from programmed cell death. At birth, oogonia are no longer present in the ovary, and only 1–2 million germ cells remain in the form of primordial follicles (Fig. 392-2). The oocyte persists in prophase of the first meiotic division until just before ovulation, when meiosis resumes.

The quiescent primordial follicles are recruited to further growth and differentiation through a highly regulated process that limits the size of the developing cohort to ensure that folliculogenesis can continue throughout the reproductive life span. This initial recruitment of primordial follicles to form primary follicles (Fig. 392-1) is characterized by growth of the oocyte and the transition from squamous to cuboidal granulosa cells. The theca interna cells that surround the developing follicle begin to form as the primary follicle grows. Acquisition of a zona pellucida by the oocyte and the presence of several layers of surrounding cuboidal granulosa cells mark the development of secondary follicles. It is at this stage that granulosa cells develop follicle-stimulating hormone (FSH), estradiol, and androgen receptors and communicate with one another through the development of gap junctions.

Bidirectional signaling between the germ cells and the somatic cells in the ovary is a necessary component underlying the maturation of the oocyte and the capacity for hormone secretion. For example, oocyte-derived growth differentiation factor 9 (GDF-9) and bone morphogenic protein-15 (BMP-15), also known as GDF-9b, are required for migration of pregranulosa and pretheca cells to the outer surface of the developing follicle and, hence, initial follicle formation. GDF-9 is also required for formation of secondary follicles, as are granulosa cell-derived KIT ligand (KITL) and the forkhead transcription factor (FOXL2). A significant number of genes have been identified that are required for development of the normal complement of oogonia in the ovary, initial follicle development, and resistance to follicle loss; all are candidates for premature ovarian insufficiency (POI), and mutations in >50 genes have been identified in patients with POI, with even more that have been associated with an earlier age at natural menopause.



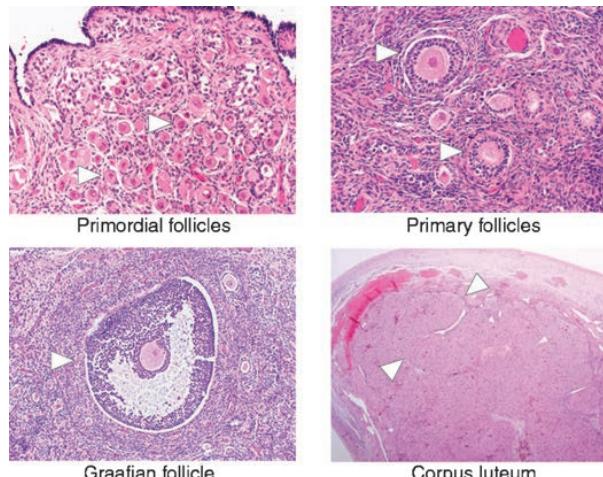
**FIGURE 392-1 Stages of ovarian development** from the arrival of the migratory germ cells at the genital ridge through gonadotropin-independent and gonadotropin-dependent phases that ultimately result in ovulation of a mature oocyte. FSH, follicle-stimulating hormone; LH, luteinizing hormone.



**FIGURE 392-2** Ovarian germ cell number is maximal at mid-gestation and decreases precipitously thereafter.

## DEVELOPMENT OF A MATURE FOLLICLE

The early stages of follicle growth are primarily driven by intraovarian factors. Further maturation to the preovulatory stage, including the resumption of meiosis in the oocyte, requires the combined stimulus of FSH and luteinizing hormone (LH) (Fig. 392-1). Recruitment of secondary follicles from the resting follicle pool requires the direct action of FSH, whereas anti-müllerian hormone (AMH) produced from small growing follicles (preantral) restrains this effect of FSH, controlling the number of follicles entering the actively growing pool. Accumulation of follicular fluid between the layers of granulosa cells creates an antrum that divides the granulosa cells into two functionally distinct groups: mural cells that line the follicle wall and cumulus cells that surround the oocyte (Fig. 392-3). In addition to its role in normal development of the müllerian system, the WNT signaling pathway is required for normal antral follicle development and may also play a role in ovarian steroidogenesis. Recruitment to the small antral stage generally occurs over several cycles with further growth to follicle sizes of >4–7 mm in waves during a single cycle. A single dominant follicle emerges from the growing follicle pool within the first 5–7 days after the onset of menses while the majority of follicles fall off their growth trajectory and become atretic. Autocrine actions of activin and BMP-6, derived from the granulosa cells, and paracrine actions of GDF-9, BMP-15, BMP-6, and Gpr149, derived from the oocyte, are involved in granulosa cell proliferation and modulation of FSH responsiveness. Differential exposure to these factors, and to vascular endothelial growth factor (VEGF), can alter vascular density and permeability, likely explaining the mechanism whereby a given follicle is selected for continued growth to the preovulatory stage. The dominant follicle can be distinguished by its size, evidence of granulosa cell proliferation,



**FIGURE 392-3** Development of ovarian follicles. The Graafian follicle is also known as a tertiary or preovulatory follicle. (Courtesy of JH Eichhorn and D Roberts, Massachusetts General Hospital; with permission.)

large number of FSH receptors, high aromatase activity, and elevated concentrations of estradiol and inhibin A in follicular fluid. In addition, secretion of estradiol and inhibin from the dominant follicle inhibits FSH and the growth of other follicles.

The dominant follicle undergoes rapid expansion during the 5–6 days prior to ovulation, reflecting granulosa cell proliferation and accumulation of follicular fluid. FSH induces LH receptors on the granulosa cells, and the preovulatory, or Graafian, follicle moves to the outer ovarian surface in preparation for ovulation. The LH surge triggers the resumption of meiosis, the suppression of granulosa cell proliferation, and the induction of cyclooxygenase 2 (COX-2), prostaglandins, the progesterone receptor (PR), and the epidermal growth factor (EGF)-like growth factors amphiregulin, epiregulin, betacellulin, and neuregulin 1, all of which are required for ovulation. Ovulation requires production of extracellular matrix, leading to expansion of the cumulus cell population that surrounds the oocyte and the controlled expulsion of the egg and follicular fluid. Both progesterone and prostaglandins (induced by the ovulatory stimulus) are essential for this process, as are members of the matrix metalloproteinase family. After ovulation, luteinization of theca and granulosa cells is induced by LH in conjunction with the acquisition of a rich vascular network in response to VEGF and basic fibroblast growth factor (FGF). Traditional regulators of central reproductive control, gonadotropin-releasing hormone (GnRH) and its receptor (GnRHR), as well as kisspeptin, are also produced in the ovary and may be involved in corpus luteum function.

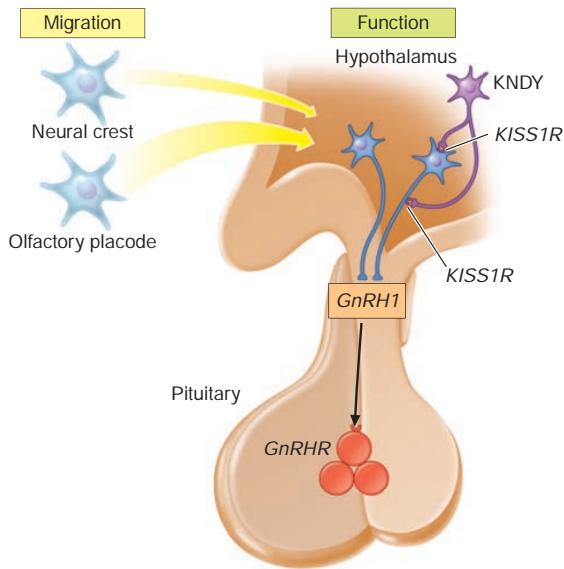
## REGULATION OF OVARIAN FUNCTION

### HYPOTHALAMIC AND PITUITARY SECRETION

GnRH neurons derive from cells in the olfactory placode and, to a lesser extent, the neural crest. They migrate along the scaffold of the olfactory neurons across the cribriform plate to the hypothalamus where they separate from the olfactory neurons. Studies in GnRH-deficient patients who fail to undergo puberty have provided insights into genes that control the ontogeny and function of GnRH neurons (Fig. 392-4). *KAL1*, *FCF8/FCFR1*, *PROK2/PROKR2*, *NSMF*, *HS6SD1*, and *CDH7*, among others (Chap. 391), have been implicated in the migration of GnRH neurons to the hypothalamus while *KISS*, *TAC3*, *Dyn* and their receptors are involved in the upstream regulation of GnRH secretion. Approximately 7000 GnRH neurons, scattered throughout the medial basal hypothalamus, establish contacts with capillaries of the pituitary portal system in the median eminence. GnRH is secreted into the pituitary portal system in discrete pulses to stimulate synthesis and secretion of LH and FSH from pituitary gonadotropes, which comprise ~10% of cells in the pituitary (Chap. 378). Functional connections of GnRH neurons with the portal system are established by the end of the first trimester, coinciding with the production of pituitary gonadotropins. Thus, like the ovary, the hypothalamic and pituitary components of the reproductive system are present before birth. However, the high levels of estradiol and progesterone produced by the placenta suppress hypothalamic-pituitary stimulation of ovarian hormonal secretion in the fetus.

After birth and the loss of placenta-derived steroids, gonadotropin levels rise. FSH levels are much higher in girls than in boys. This rise in FSH results in circulating estradiol and increased inhibin B, but without terminal follicle maturation or ovulation. Studies that have identified mutations in *TAC3*, which encodes neurokinin B, and its receptor, *TAC3R*, in patients with GnRH deficiency indicate that both are involved in control of GnRH secretion and may be particularly important at this early stage of development. By 12–20 months of age, the reproductive axis is again suppressed, and a period of relative quiescence persists until puberty (Fig. 392-5). At the onset of puberty, pulsatile GnRH secretion induces pituitary gonadotropin production. In the early stages of puberty, LH and FSH secretion are apparent only during sleep, but as puberty develops, pulsatile LH secretion occurs throughout the day and night.

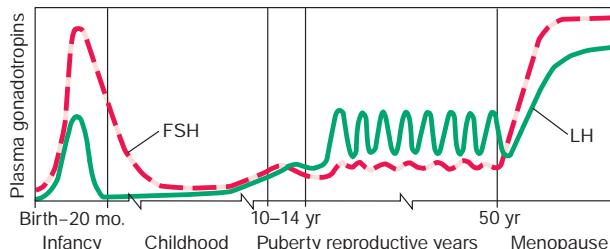
The mechanisms responsible for the childhood quiescence and pubertal reactivation of the reproductive axis remain incompletely understood. GnRH neurons in the hypothalamus respond to both excitatory and inhibitory factors. Increased sensitivity to the inhibitory



**FIGURE 392-4** Genetic studies in patients with congenital forms of hypogonadotropic hypogonadism have expanded our understanding of the development and migration of gonadotropin-releasing hormone (GnRH) neurons from the olfactory placode and neural crest to the hypothalamus as well as the upstream regulation of GnRH secretion by kisspeptin (*KISS1*), neurokinin B (*TAC3*) and dynorphin (*Dyn*) which are co-expressed in the KNDY neurons.

influence of gonadal steroids has long been implicated in the inhibition of GnRH secretion during childhood but has not been definitively established in the human. Metabolic signals, including adipocyte-derived leptin, play a permissive role in reproductive function (Chap. 401). Studies of patients with isolated GnRH deficiency reveal that mutations in the G protein-coupled receptor 54 (*GPR54*) gene (now known as *KISS1R*) preclude the onset of puberty. The ligand for this receptor is derived from the parent peptide, kisspeptin-1 (*KISS1*), and is a powerful stimulant for GnRH release. A potential role for kisspeptin in the onset of puberty has been suggested by upregulation of *KISS1* and *KISS1R* transcripts in the hypothalamus at the time of puberty. *TAC3*, which stimulates GnRH secretion through kisspeptin signaling, and dynorphin (*Dyn*), which plays an inhibitory role in GnRH control, are frequently co-expressed with *KISS1* in KNDY neurons of the median eminence that project to GnRH neurons. This system is intimately involved in both estrogen and progesterone negative feedback regulation of GnRH secretion.

RFamide-related peptides (RFRPs) are the mammalian orthologues of gonadotropin inhibitory hormone (GnIH), which was initially discovered in the quail. While RFRP-1 and RFRP-3 neurons send axonal projections to GnRH neurons in humans and RFRPs are secreted into the pituitary portal system, further studies are required to determine their potential physiologic role in the human.

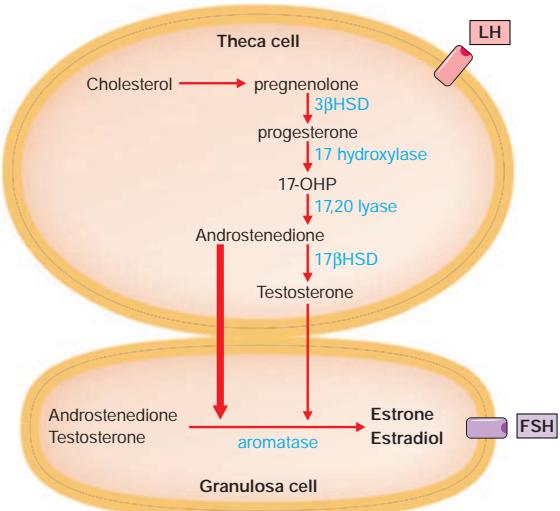


**FIGURE 392-5** Follicle-stimulating hormone (FSH) and luteinizing hormone (LH) are increased during the neonatal years but go through a period of childhood quiescence before increasing again during puberty. Gonadotropin levels are cyclic during the reproductive years and increase dramatically with the loss of negative feedback that accompanies menopause.

## OVARIAN STEROIDS

Ovarian steroid-producing cells do not store hormones but produce them in response to FSH and LH during the normal menstrual cycle. The sequence of steps and the enzymes involved in the synthesis of steroid hormones are similar in the ovary, adrenal, and testis. However, the enzymes required to catalyze specific steps are compartmentalized and may not be abundant or even present in all cell types. Within the developing ovarian follicle, estrogen synthesis from cholesterol requires close integration between theca and granulosa cells—sometimes called the *two-cell model for steroidogenesis* (Fig. 392-6). FSH receptors are confined to the granulosa cells, whereas LH receptors are restricted to the theca cells until the late stages of follicular development, when they are also found on granulosa cells. The theca cells surrounding the follicle are highly vascularized and use cholesterol, derived primarily from circulating lipoproteins, as the starting point for the synthesis of androstenedione and testosterone under the control of LH. These steroid precursors cross the basal lamina to the granulosa cells, which receive no direct blood supply. The mural granulosa cells are particularly rich in aromatase and, under the control of FSH, produce estradiol, the primary steroid secreted from the follicular phase ovary and the most potent estrogen. Theca cell-produced androstenedione and, to a lesser extent, testosterone are also secreted into peripheral blood, where they can be converted to dihydrotestosterone in skin and to estrogens in adipose tissue. The hilar interstitial cells of the ovary are functionally similar to Leydig cells and are also capable of secreting androgens. Stromal cells proliferate in response to androgens (as in polycystic ovary syndrome [PCOS]) but do not secrete androgens. However, high levels of androgens may be produced by luteinized theca cells in women with hyperthecosis.

Development of the rich capillary network following rupture of the follicle at the time of ovulation makes it possible for large molecules such as low-density lipoprotein (LDL) to reach the luteinized granulosa and theca lutein cells. As in the follicle, both cell types are required for steroidogenesis in the corpus luteum. The luteinized granulosa cells are the main source of progesterone production, whereas the smaller theca lutein cells produce 17-hydroxyprogesterone and androgenic substrates for aromatization to estradiol by the luteinized granulosa cells. Production of estrogen metabolites by the corpus luteum plays a significant role in maintenance of the vascularization required for its function. LH is critical for formation and maintenance of corpus luteum structure and function. LH and human chorionic gonadotropin (hCG) bind to a common receptor; thus, in conception cycles, hCG produced upon fertilization rescues the declining function of the



**FIGURE 392-6** Estrogen production in the ovary requires the cooperative function of the theca and granulosa cells under the control of luteinizing hormone (LH) and follicle-stimulating hormone (FSH). HSD, hydroxysteroid dehydrogenase; OHP, hydroxyprogesterone.

**3030** corpus luteum, maintaining steroid and peptide secretion for the first 10 weeks of pregnancy. hCG is commonly used for luteal phase support in the treatment of infertility.

**Steroid Hormone Actions** Both estrogen and progesterone play critical roles in the expression of secondary sexual characteristics in women ([Chap. 377](#)). Estrogen promotes development of the ductule system in the breast, whereas progesterone is responsible for glandular development. In the reproductive tract, estrogens create a receptive environment for fertilization and support pregnancy and parturition through carefully coordinated changes in the endometrium, thickening of the vaginal mucosa, thinning of the cervical mucus, and uterine growth and contractions. Progesterone induces secretory activity in the estrogen-primed endometrium, increases the viscosity of cervical mucus, and inhibits uterine contractions. Both gonadal steroids play critical roles in negative and positive feedback of gonadotropin secretion. Progesterone also increases basal body temperature, which is used clinically as a marker of ovulation.

The vast majority of circulating estrogens and androgens are carried in the blood bound to carrier proteins, which restrain their free diffusion into cells and prolong their clearance, serving as a reservoir. High-affinity binding proteins include sex hormone-binding globulin (SHBG), which binds androgens with somewhat greater affinity than estrogens, and corticosteroid-binding globulin (CBG), which also binds progesterone. Modulations in binding protein levels by insulin, androgens, and estrogens contribute to high bioavailable testosterone levels in PCOS and to high circulating total estrogen and progesterone levels during pregnancy.

Estrogens act primarily through binding to the nuclear receptors, estrogen receptor (ER) and . Transcriptional coactivators and co-repressors modulate ER action ([Chap. 377](#)). Both ER subtypes are present in the hypothalamus, pituitary, ovary, and reproductive tract. Although ER and exhibit some functional redundancy, there is also a high degree of specificity, particularly in expression within cell types. For example, ER functions in ovarian theca cells, whereas ER is critical for granulosa cell function. There is also evidence for membrane-initiated signaling by estrogen. Similar signaling mechanisms pertain for progesterone with evidence of transcriptional regulation through PR A and B protein isoforms, as well as rapid membrane signaling.

### OVARIAN PEPTIDES

Inhibin was initially isolated from gonadal fluids based on its ability to selectively inhibit FSH secretion from pituitary cells. Inhibin is a heterodimer composed of an subunit and a A or B subunit to form inhibin A or inhibin B, both of which are secreted from the ovary. Activin is a homodimer of inhibin subunits with the capacity to stimulate the synthesis and secretion of FSH. Inhibins and activins are members of the transforming growth factor (TGF-) superfamily of growth and differentiation factors. During the purification of inhibin, follistatin, an unrelated monomeric protein that inhibits FSH secretion, was discovered. Within the pituitary, follistatin inhibits FSH secretion indirectly by binding and neutralizing activin.

Inhibin B is constitutively secreted from the granulosa cells of small antral follicles, and its serum levels increase in conjunction with granulosa cell proliferation during recruitment of secondary follicles under the control of FSH. Inhibin B is an important inhibitor of FSH, independent of estradiol, during the menstrual cycle. Inhibin A is present in both granulosa and theca cells and is secreted by the dominant follicle. Inhibin A is also present in luteinized granulosa cells and is a major secretory product of the corpus luteum. Synthesis and secretion of inhibin A are directly controlled by FSH and LH. Although activin is also secreted from the ovary, the excess of follistatin in serum, combined with its nearly irreversible binding of activin, make it unlikely that ovarian activin plays an endocrine role in FSH regulation. However, there is evidence that activin plays an autocrine/paracrine role in the ovary, in addition to its intrapituitary role in modulation of FSH production.

AMH (also known as müllerian-inhibiting substance) is important in ovarian biology in addition to the function from which it derived its name (i.e., promotion of the degeneration of the müllerian system

during embryogenesis in the male). AMH is produced by granulosa cells from small preantral and early antral follicles and is a marker of ovarian reserve with advantages over inhibin B because of its relative stability across the menstrual cycle. AMH inhibits the recruitment of primordial follicles into the follicle pool and counters FSH stimulation of aromatase expression. AMH levels are highest in the early twenties and decrease markedly by menopause. AMH is increased in PCOS in conjunction with the abundance of small follicles in this disorder.

Gonadotropin surge attenuating factor (GnSAF) is an ovarian factor that attenuates GnRH-induced gonadotropin secretion. Its role is not yet fully understood, but there is an inverse relationship between GnSAF and follicle size, suggesting that its primary role involves the early stages of follicle development rather than curtailing the gonadotropin surge as its name implies.

Relaxin is produced primarily by the theca lutein cells of the corpus luteum. Both relaxin and its receptor, RXFP1, are highly expressed in the uterus during the peri-implantation period in the marmoset, and its primary role appears to be in promoting decidualization and vascularization of the endometrium prior to implantation. Relaxin was named for its ability to suppress myometrial contractility in pigs and rodents, but it does not appear to exert this activity in women.

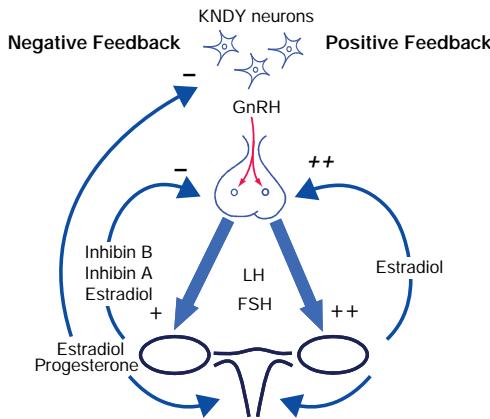
## HORMONAL INTEGRATION OF THE NORMAL MENSTRUAL CYCLE

The sequence of changes responsible for mature reproductive function is coordinated through a series of negative and positive feedback loops that alter pulsatile GnRH secretion, the pituitary response to GnRH, and the relative secretion of LH and FSH from the gonadotrope. The frequency and amplitude of pulsatile GnRH secretion differentially modulate the synthesis and secretion of LH and FSH. Slow GnRH pulse frequencies favor FSH synthesis, whereas increased GnRH pulse frequency and amplitude favor LH synthesis. Activin is produced in both pituitary gonadotropes and folliculostellate cells and stimulates the synthesis and secretion of FSH through autocrine-paracrine mechanisms that are modulated by follistatin. Inhibins function as potent antagonists of activins through sequestration of the activin receptors. Although inhibin is expressed in the pituitary, gonadal inhibin is the principal source of feedback inhibition of FSH.

For the majority of the cycle, the reproductive system functions in a classic endocrine negative feedback mode. Estradiol and progesterone inhibit GnRH secretion, acting through kisspeptin and dynorphin in the KNDy neurons, while the inhibins act at the pituitary to selectively inhibit FSH synthesis and secretion ([Fig. 392-7](#)). Estradiol also contributes to negative feedback at the pituitary with an effect that is greater for FSH than LH. This tightly regulated negative feedback control of FSH is critical for development of the single mature oocyte that characterizes normal reproductive function in women. In addition to these negative feedback controls, the menstrual cycle is uniquely dependent on estrogen-induced positive feedback to produce an LH surge that is essential for ovulation of a mature follicle. Estrogen negative feedback in women occurs primarily at the hypothalamus with a small pituitary contribution, whereas estrogen positive feedback occurs at the pituitary in women with upregulation of GnRH signaling and responsiveness. In women, hypothalamic GnRH secretion plays a permissive role in generating the preovulatory gonadotropin surge, a mechanism that differs significantly from that in rodents and other species that rely on seasonal and circadian cues, in which a surge of GnRH also occurs.

### THE FOLLICULAR PHASE

The follicular phase is characterized by recruitment of a cohort of secondary follicles and the ultimate selection of a dominant preovulatory follicle ([Fig. 392-8](#)). The follicular phase begins, by convention, on the first day of menses. However, follicle recruitment is initiated by the rise in FSH that begins in the late luteal phase of the previous cycle in conjunction with the loss of negative feedback of gonadal steroids and likely inhibin A. The fact that a 20–30% increase in FSH is adequate for follicular recruitment speaks to the marked sensitivity of the resting follicle pool to FSH. The resultant granulosa cell proliferation is responsible for increasing early follicular phase levels of inhibin B. Inhibin B, in conjunction with rising

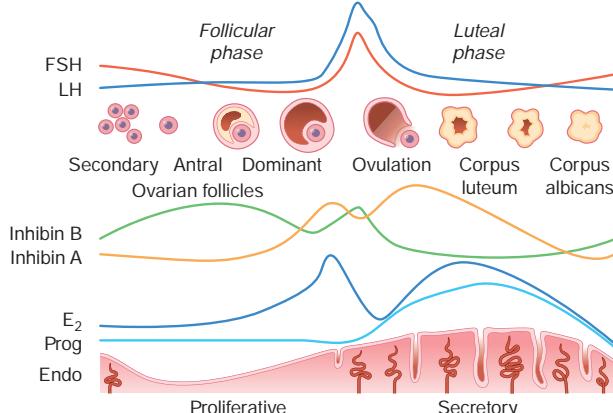


**FIGURE 392-7** The reproductive system in women is critically dependent on both negative feedback of gonadal steroids and inhibin to modulate follicle-stimulating hormone (FSH) secretion and on estrogen positive feedback to generate the preovulatory luteinizing hormone (LH) surge. GnRH, gonadotropin-releasing hormone.

levels of estradiol and inhibin A, restrains FSH secretion during this critical period such that only a single follicle matures in the vast majority of cycles. The increased risk of multiple gestation associated with the increased levels of FSH characteristic of advanced maternal age or with exogenous gonadotropin administration in the treatment of infertility attests to the importance of negative feedback regulation of FSH. With further growth of the dominant follicle, estradiol and inhibin A increase and the follicle acquires LH receptors. Increasing levels of estradiol are responsible for proliferative changes in the endometrium. The exponential rise in estradiol results in positive feedback on the pituitary, leading to the generation of an LH surge (and a smaller FSH surge), thereby triggering ovulation and luteinization of granulosa and theca cells.

### THE LUTEAL PHASE

The luteal phase begins with the formation of the corpus luteum from the ruptured follicle (Fig. 392-8). Progesterone and inhibin A are produced from the luteinized granulosa cells, which continue to aromatize theca-derived androgen precursors, producing estradiol. The combined actions of estrogen and progesterone are responsible for the secretory changes in the endometrium that are necessary for implantation. The corpus luteum is supported by LH but has a finite life span because of diminished sensitivity to LH. The demise of the corpus luteum results in a progressive decline in hormonal support of the endometrium. Inflammation or local hypoxia and ischemia result in vascular changes in the endometrium, leading to the release of cytokines, cell death, and shedding of the endometrium.



**FIGURE 392-8** Relationship between gonadotropins, follicle development, gonadal secretion, and endometrial changes during the normal menstrual cycle. E<sub>2</sub>, estradiol; Endo, endometrium; FSH, follicle-stimulating hormone; LH, luteinizing hormone; Prog, progesterone.

If conception occurs, hCG produced by the trophoblast binds to LH receptors on the corpus luteum, maintaining steroid hormone production and preventing involution of the corpus luteum until its hormonal function is taken over by the placenta 6–10 weeks after conception.

## CLINICAL ASSESSMENT OF OVARIAN FUNCTION

Menstrual bleeding should become regular within 2–4 years of menarche, although anovulatory and irregular cycles are common before that. For the remainder of adult reproductive life, the cycle length counted from the first day of menses to the day preceding subsequent menses is ~28 days, with a range of 25–35 days. However, cycle-to-cycle variability for an individual woman is ±2 days. Luteal phase length is relatively constant between 12 and 14 days in normal cycles; thus, the major variability in cycle length is due to variations in follicular phase length. The duration of menstrual bleeding in ovulatory cycles varies between 4 and 6 days. There is a gradual shortening of cycle length with age such that women aged >35 years have cycles that are shorter than during their younger reproductive years. Anovulatory cycles increase as women approach menopause, and bleeding patterns may be erratic.

Women who report regular monthly bleeding generally have ovulatory cycles, but several other clinical signs can be used to assess the likelihood of ovulation. Some women experience *mittelschmerz*, described as midcycle pelvic discomfort that is thought to be caused by the rapid expansion of the dominant follicle at the time of ovulation. A constellation of premenstrual moliminal symptoms such as bloating, breast tenderness, mood changes, and food cravings often occur several days before menses in ovulatory cycles, but their absence cannot be used as evidence of anovulation. Methods that can be used to determine whether ovulation occurred include a serum progesterone level >3 ng/mL ~7 days after ovulation, an increase in basal body temperature of 0.24°C (>0.5°F) in the second half of the cycle due to the thermoregulatory effect of progesterone, or detection of the urinary LH surge using ovulation predictor kits. Because ovulation occurs ~36 h after the LH surge, urinary LH can be helpful in timing intercourse to coincide with ovulation.

Ultrasound can be used to detect the growth of the fluid-filled antrum of the developing follicle and to assess endometrial thickness in response to increasing estradiol levels in the follicular phase. It can also be used to provide evidence of ovulation by documenting collapse of the dominant follicle and/or the presence of a corpus luteum as well as the characteristic echogenicity of the secretory endometrium of the luteal phase.

### PUBERTY

#### NORMAL PUBERTAL DEVELOPMENT IN GIRLS

The first menstrual period (*menarche*) occurs relatively late in the series of developmental milestones that characterize normal pubertal development (Table 392-1). Menarche is preceded by the appearance of pubic and then axillary hair (*adrenarche*) as a result of maturation of the zona reticularis in the adrenal gland and increased adrenal androgen secretion, particularly dehydroepiandrosterone (DHEA). The triggers for adrenarche remain unknown but may involve increases in body mass index, as well as in utero and neonatal factors. Menarche is also preceded by breast development (*thelarche*). The breast is exquisitely sensitive to the very low levels of estrogen that result from peripheral conversion of adrenal androgens and the low levels of estrogen secreted

**TABLE 392-1** Mean Age (Years) of Pubertal Milestones in Girls

	ONSET OF BREAST/ PUBIC HAIR DEVELOPMENT	AGE OF PEAK HEIGHT VELOCITY	MENARCHE	FINAL BREAST/ PUBIC HAIR DEVELOPMENT	ADULT HEIGHT
White	10.2	11.9	12.6	14.3	17.1
Black	9.6	11.5	12	13.6	16.5

Source: Adapted with permission from FM Biro et al: Pubertal correlates in black and white girls. J Pediatr 148:234, 2006.

3032 from the ovary early in pubertal maturation. This estrogen sensitivity also explains why infants occasionally develop breast tissue in response to exogenous or environmental estrogens. Breast development precedes the appearance of pubic and axillary hair in ~60% of girls. The interval between the onset of breast development and menarche is ~2 years. There has been a gradual decline in the age of menarche over the past century, attributed in large part to improvement in nutrition, and there is a relationship between adiposity and earlier sexual maturation in girls. In the United States, menarche occurs at an average age of 12.5 years (Table 392-1).

Much of the variation in the timing of puberty is due to genetic factors. Heritability estimates from twin studies range between 50 and 80%. Adrenarche and thelarche occur ~1 year earlier in black girls compared with white girls, although the difference in the timing of menarche is less pronounced. Genome-wide association studies have identified over a hundred genes associated with pubertal timing in boys and girls attesting to the high degree of coordination of this reproductive and growth milestone. These findings include genes involved in GnRH secretion (e.g., *TACR3*, and the maternally imprinted gene, *MKRN3*, that has been associated with familial precocious puberty), pituitary development and function (e.g., *POU1F1*), hormone synthesis and bioactivity (e.g., *STARD4*, *ESR1*, *RXRG*), gonadal feedback (e.g., *INHBA*, *ESRI*), and energy homeostasis and growth, including *LIN28B*, a sentinel puberty gene, which is a potent regulator of microRNA processing.

Other important hormonal changes also occur in conjunction with puberty. Growth hormone (GH) levels increase early in puberty, stimulated in part by the pubertal increase in estrogen secretion. GH increases insulin-like growth factor-1 (IGF-1), which enhances linear growth. The growth spurt is generally less pronounced in girls than in boys, with a peak growth velocity of ~7 cm/year. Linear growth is ultimately limited by closure of epiphyses in the long bones as a result of prolonged exposure to estrogen. Puberty is also associated with mild insulin resistance.

## DISORDERS OF PUBERTY

The differential diagnosis of precocious and delayed puberty is similar in boys ([Chap. 391](#)) and girls. However, there are differences in the timing of normal puberty and differences in the relative frequency of specific disorders in girls compared with boys.

**Precocious Puberty** Traditionally, precocious puberty has been defined as the development of secondary sexual characteristics before the age of 8 in girls based on data from Marshall and Tanner in British girls studied in the 1960s. More recent studies led to recommendations that girls be evaluated for precocious puberty if breast development or pubic hair is present at <7 years of age for white girls or <6 years for black girls; however, these guidelines have not been widely accepted in favor of careful follow-up in girls presenting at <8 years.

Precocious puberty in girls is most often centrally mediated ([Table 392-2](#)), resulting from early activation of the hypothalamic-pituitary-ovarian axis. It is characterized by pulsatile LH secretion (which is initially associated with deep sleep) and an enhanced LH and FSH response to exogenous GnRH or a GnRH agonist (two- to threefold stimulation) ([Table 392-3](#)). True precocity is marked by advancement in bone age of >2 standard deviations, a recent history of growth acceleration, and progression of secondary sexual characteristics. In girls, centrally mediated precocious puberty (CPP) is idiopathic in ~85% of cases; however, neurogenetic causes must be considered. Activating mutations in *KISS1*, *KISS1R*, and *KISS1R* have been found in a small number of patients with CPP, and loss-of-function mutations in *MKRN3* have been reported in familial CPP. However, the frequency of these mutations is insufficient to justify their use in routine clinical testing. GnRH agonists that induce pituitary desensitization are the mainstay of treatment to prevent premature epiphyseal closure and preserve adult height, as well as to manage psychosocial repercussions of precocious puberty.

Peripherally mediated precocious puberty does not involve activation of the hypothalamic-pituitary-ovarian axis and is characterized by suppressed gonadotropins in the presence of elevated estradiol. Management of peripheral precocious puberty involves treating the underlying disorder ([Table 392-2](#)) and limiting the effects of gonadal steroids using

**TABLE 392-2 Differential Diagnosis of Precocious Puberty**

CENTRAL (GnRH DEPENDENT)	PERIPHERAL (GnRH INDEPENDENT)
Idiopathic	Congenital adrenal hyperplasia
CNS tumors	Estrogen-producing tumors
Hamartomas	Adrenal tumors
Astrocytomas	Ovarian tumors
Adenomyomas	Gonadotropin/hCG-producing tumors
Gliomas	Exogenous exposure to estrogen or androgen or lavender or tea-tree oil
Germinomas	McCune-Albright syndrome
CNS infection	Aromatase excess syndrome
Genetic, i.e., <i>KISS1</i> , <i>KISS1R</i> , <i>MKRN3</i> , <i>DLK1</i>	
Head trauma	
Iatrogenic	
Radiation	
Chemotherapy	
Surgical	
CNS malformation	
Arachnoid or suprasellar cysts	
Septo-optic dysplasia	
Hydrocephalus	

Abbreviations: CNS, central nervous system; *DLK1*, delta-like 1 homolog gene; GnRH, gonadotropin-releasing hormone; hCG, human chorionic gonadotropin; *KISS1*, kisspeptin gene; *KISS1R*, kisspeptin receptor gene; *MKRN3*, makorin ring finger protein 3 gene.

aromatase inhibitors, inhibitors of steroidogenesis, and ER blockers. It is important to be aware that central precocious puberty can also develop in girls whose precocity was initially peripherally mediated, as in McCune-Albright syndrome and congenital adrenal hyperplasia.

**TABLE 392-3 Evaluation of Precocious and Delayed Puberty**

	PRECOCIOUS	DELAYED
<b>Initial Screening Tests</b>		
History and physical	x	x
Assessment of growth velocity	x	x
Bone age	x	x
LH, FSH	x	x
Estradiol, testosterone	x	x
DHEAS	x	x
17-Hydroxyprogesterone	x	
TSH, $T_4$	x	x
Complete blood count		x
Sedimentation rate, C-reactive protein		x
Electrolytes, renal function		x
Liver enzymes		x
IGF-1, IGFBP-3		x
Urinalysis		x
<b>Secondary Tests</b>		
Pelvic ultrasound	x	x
Cranial MRI	x	x
$\beta$ -hCG	x	
GnRH/agonist stimulation test	x	x
ACTH stimulation test	x	
Inflammatory bowel disease panel	x	x
Celiac disease panel		x
Prolactin		x
Karyotype		x

Abbreviations: ACTH, adrenocorticotrophic hormone; DHEAS, dehydroepiandrosterone sulfate; FSH, follicle-stimulating hormone; GnRH, gonadotropin-releasing hormone; hCG, human chorionic gonadotropin; IGF-1, insulin-like growth factor 1; IGFBP-3, IGF-binding protein 3; LH, luteinizing hormone; MRI, magnetic resonance imaging; TSH, thyroid-stimulating hormone;  $T_4$ , thyroxine.

Incomplete and intermittent forms of precocious puberty may also occur. For example, premature breast development may occur in girls before the age of 2 years, with no further progression and without significant advancement in bone age, estrogen production, or compromised height. Premature adrenarche can also occur in the absence of progressive pubertal development, but it must be distinguished from late-onset congenital adrenal hyperplasia and androgen-secreting tumors, in which case it may be termed *heterosexual precocity*. Premature adrenarche may be associated with obesity, hyperinsulinemia, and the subsequent predisposition to PCOS.

**Delayed Puberty** Delayed puberty (**Table 392-4**) is defined as the absence of secondary sexual characteristics by age 13 in girls. The diagnostic considerations are very similar to those for primary amenorrhea (**Chap. 393**). Between 25 and 40% of delayed puberty in girls is of ovarian origin, with Turner syndrome accounting for the majority of such

patients. Delayed puberty may occur in the setting of systemic illnesses, including celiac disease and chronic renal disease, and endocrinopathies such as diabetes and hypothyroidism. In addition, girls appear to be particularly susceptible to the adverse effects of decreased energy balance resulting from exercise, dieting, and/or eating disorders, and thus, functional hypothalamic amenorrhea (HA) can present with primary amenorrhea. Together, these reversible conditions account for ~25% of delayed puberty in girls. Congenital hypogonadotropic hypogonadism in girls or boys can be caused by mutations in several different genes or combinations of genes (Fig. 392-4, **Chap. 391**, Table 392-2). Approximately 50% of girls with congenital hypogonadotropic hypogonadism, with or without anosmia, have a history of some degree of breast development, and 10% report one to two episodes of vaginal bleeding. Family studies suggest that genes identified in association with absent puberty may also cause delayed puberty, and recent reports have further suggested that a genetic susceptibility to environmental stresses such as diet and exercise may account for at least some cases of functional HA, including in girls who present with primary amenorrhea. Although neuroanatomic causes of delayed puberty are considerably less common in girls than in boys, it is always important to rule these out in the setting of hypogonadotropic hypogonadism.

**TABLE 392-4 Differential Diagnosis of Delayed Puberty**

#### Hypergonadotropic

- Ovarian
  - Turner's syndrome
  - Gonadal dysgenesis
  - Chemotherapy/radiation therapy
  - Galactosemia
  - Autoimmune oophoritis
  - Congenital lipoïd hyperplasia
  - Steroidogenic enzyme abnormalities
  - 17 $\alpha$ -Hydroxylase deficiency
  - Aromatase deficiency
  - Gonadotropin/receptor mutations
  - FSH $\beta$ , LHR, FSHR*
  - Androgen resistance syndrome

#### Hypogonadotropic

- Genetic
  - Hypothalamic syndromes
    - Leptin/leptin receptor
    - HESX1* (septo-optic dysplasia)
    - PC1* (prohormone convertase)
  - IHH and Kallmann's syndrome
    - KAL1, FGF8, FGFR1, NSMF, PROK2, PROKR2, SEM3A, HS6ST1, WDR11, CHD7, KISS1, KISS1R, TAC3, TAC3R, GnRH1, GnRHR*, and others
  - Abnormalities of pituitary development/function
    - PROP1*
- CNS tumors/infiltrative disorders
  - Craniopharyngioma
  - Astrocytoma, germinoma, glioma
  - Prolactinomas, other pituitary tumors
  - Histiocytosis X
- Chemotherapy/radiation
- Functional
  - Chronic diseases
  - Malnutrition
  - Excessive exercise
  - Eating disorders

**Abbreviations:** *CHD7*, chromodomain-helicase-DNA-binding protein 7; CNS, central nervous system; *FGF8*, fibroblast growth factor 8; *FGFR1*, fibroblast growth factor 1 receptor; *FSH*, follicle-stimulating hormone  $\beta$  chain; *FSHR*, FSH receptor; *GnRH*, gonadotropin-releasing hormone receptor; *HESX1*, homeobox, embryonic stem cell expressed 1; *HS6ST1*, heparin sulfate 6-O sulfotransferase 1; IHH, idiopathic hypogonadotropic hypogonadism; *KAL*, Kallmann; *KISS1*, kisspeptin 1; *KISSR1*, Kiss1 receptor; *LHR*, luteinizing hormone receptor; *NSMF*, NMDA receptor synaptosomal signaling and neuronal migration factor; *PROK2*, prokineticin 2; *PROKR2*, prokineticin receptor 2; *PROP1*, prophet of Pit1, paired-like homeodomain transcription factor; *SEMA3A*, semaphorin-3A; *WDR11*, WD repeat-containing protein 11.

#### FURTHER READING

- Bradley SH et al: Precocious puberty. *BMJ* 368:l6597, 2020.
- Euster EA: Update on precocious puberty. *J Pediatr Adolesc Gynecol* 32:455, 2019.
- George JT et al: Effect of gonadotropin-inhibitory hormone on luteinizing hormone secretion in humans. *Clin Endocrinol (Oxf)* 86:731, 2017.
- Kaplowitz P, Bloch C: The section on endocrinology. Evaluation and referral of children with signs of early puberty. *Pediatrics* 137:e20153732, 2016.
- Mishra GD et al: EMAS position statement: Predictors of premature and early natural menopause. *Maturitas* 123:8, 2019.
- Neocleous V et al: GnRH deficient patients with congenital hypogonadotropic hypogonadism: Novel genetic findings in *ANOS1*, *RNF216*, *WDR11*, *FGFR1*, *CHD7*, and *POLR3A* genes in a case series and review of the literature. *Front Endocrinol (Lausanne)* 11:626, 2020.
- Richards JS: The ovarian cycle. *Vitam Horm* 107:1, 2018.
- Stamou MI et al: Discovering genes essential to the hypothalamic regulation of human reproduction using a human disease model: Adjusting to life in the “-omics” era. *Endocr Rev* 36:603, 2015.
- Tucker EJ et al: Premature ovarian insufficiency: New perspectives on genetic cause and phenotypic spectrum. *Endocr Rev* 37:609, 2016.
- Willemesen RH, Dunger DB: Normal variation in pubertal timing: Genetic determinants in relation to growth and adiposity. *Endocr Dev* 29:17, 2016.

393

## Menstrual Disorders and Pelvic Pain

Janet E. Hall, Anuja Dokras



Menstrual dysfunction can signal an underlying abnormality that may have long-term health consequences. Although frequent or prolonged bleeding usually prompts a woman to seek medical attention, infrequent or absent bleeding may seem less troubling, and the patient may not bring it to the attention of the physician. Thus, a focused menstrual history is a critical part of every encounter with a female patient. Pelvic pain is a common complaint that may relate to an abnormality of the reproductive organs but also may be of gastrointestinal, urinary

**3034** tract, or musculoskeletal origin. Depending on its cause, pelvic pain may require urgent surgical attention. Recent guidelines no longer recommend routine pelvic examination in asymptomatic, average-risk women other than periodic cervical cancer screening. However, pelvic examination is an important part of the evaluation of amenorrhea, abnormal uterine bleeding, and pelvic pain.

## MENSTRUAL DISORDERS

### DEFINITION AND PREVALENCE

Amenorrhea refers to the absence of menstrual periods. Amenorrhea is classified as *primary* if menstrual bleeding has never occurred in the absence of hormonal treatment or *secondary* if menstrual periods cease for 3–6 months. Primary amenorrhea is a rare disorder that occurs in <1% of the female population. However, between 3 and 5% of women experience at least 3 months of secondary amenorrhea in any specific year. There is no evidence that race or ethnicity influences the prevalence of amenorrhea. However, because of the importance of adequate nutrition for normal reproductive function, both the age at menarche and the prevalence of secondary amenorrhea vary significantly in different parts of the world.

*Oligomenorrhea* is defined as a cycle length >35 days or <10 menses per year. Both the frequency and the amount of vaginal bleeding are irregular in oligomenorrhea, and moliminal symptoms (premenstrual breast tenderness, food cravings, mood lability), suggestive of ovulation, are variably present. Anovulation can also present with intermenstrual intervals <24 days. Frequent or heavy irregular bleeding is termed *dysfunctional uterine bleeding* if anatomic uterine and outflow tract lesions or a bleeding diathesis have been excluded. Oligo- and anovulation are most frequently associated with polycystic ovarian syndrome (PCOS).

**Primary Amenorrhea** The absence of menarche (the first menstrual period) by age 16 has been used traditionally to define primary amenorrhea. However, other factors, such as growth, secondary sexual characteristics, and the presence of cyclic pelvic pain, also influence the age at which primary amenorrhea should be investigated. Recent studies suggest that puberty is occurring at an earlier age, particularly in obese girls. However, it is important to note that these data reflect earlier breast development alone with minimal change in the age of menarche. Thus, an evaluation for amenorrhea should be initiated by age 15 or 16 in the presence of normal growth and secondary sexual characteristics; age 13 in the absence of secondary sexual characteristics or if height is less than the third percentile; age 12 or 13 in the presence of breast development and cyclic pelvic pain; or within 2 years of breast development if menarche has not occurred.

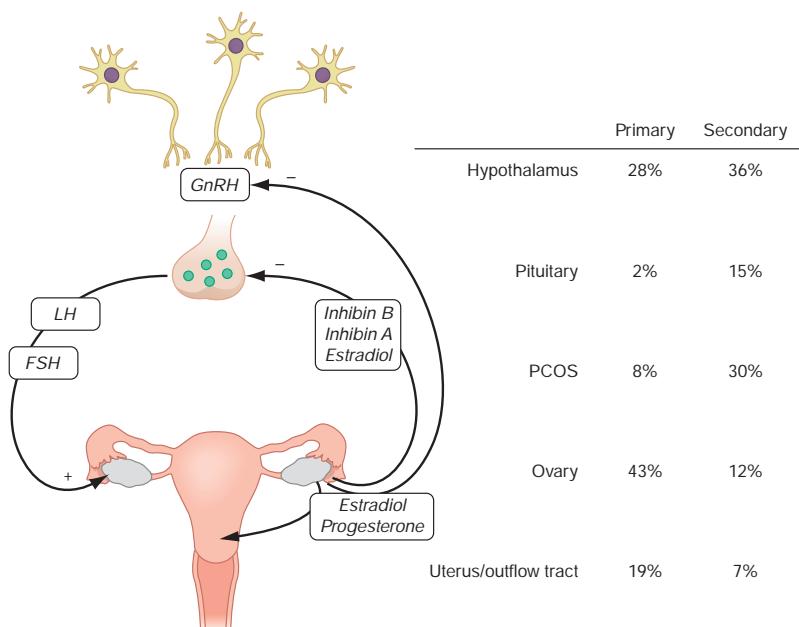
**Secondary Amenorrhea or Oligomenorrhea** Irregular cycles are relatively common for up to 3 years after menarche and for 1–2 years before the final menstrual period. In the intervening years, menstrual cycle length is ~28 days, with an intermenstrual interval normally ranging between 25 and 35 days. Cycle-to-cycle variability in an individual woman who is ovulating consistently is generally  $\pm 2$  days. Pregnancy is the most common cause of amenorrhea and should be excluded early in any evaluation of menstrual irregularity. However, many women occasionally miss a single period. Three months of secondary amenorrhea, or 6 months in women with previously irregular cycles, should prompt an evaluation, as should a history of intermenstrual intervals >35 or <21 days or bleeding that persists for >7 days.

### DIAGNOSIS

Pregnancy is the most common cause of amenorrhea and must be excluded in all cases, regardless of patient history. Evaluation of menstrual dysfunction depends on understanding the interrelationships between the four critical components of the reproductive tract: (1) the hypothalamus, (2) the pituitary, (3) the ovaries, and (4) the uterus and outflow tract (Fig. 393-1; Chap. 392). This system is maintained by complex negative and positive feedback loops involving the ovarian steroids (estradiol and progesterone) and peptides (inhibin B and inhibin A) and the hypothalamic (gonadotropin-releasing hormone [GnRH]) and pituitary (follicle-stimulating hormone [FSH] and luteinizing hormone [LH]) components of this system (Fig. 393-1).

Disorders of menstrual function can be thought of in two main categories: disorders of the uterus and outflow tract and disorders of ovulation. Many of the conditions that cause primary amenorrhea are congenital but go unrecognized until the time of normal puberty (e.g., genetic, chromosomal, and anatomic abnormalities). All causes of secondary amenorrhea also can cause primary amenorrhea.

**Disorders of the Uterus or Outflow Tract** Abnormalities of the uterus and outflow tract typically present as primary amenorrhea. In patients with normal pubertal development and a blind vagina, the differential diagnosis includes *obstruction* by a transverse vaginal septum or imperforate hymen; *Müllerian agenesis* (Mayer-Rokitansky-Kuster-Hauser syndrome), which can be caused by mutations in the *WNT4* gene, and *androgen insensitivity syndrome* (AIS), which is an X-linked recessive disorder that accounts for ~10% of all cases of primary amenorrhea (Chap. 391). Patients with AIS have a 46,XY karyotype, but because of the lack of androgen receptor responsiveness, those with complete AIS lack features of androgenization and have female external genitalia. The absence of pubic and axillary hair distinguishes them clinically from patients with Müllerian agenesis, as does a testosterone level in the male range. The rare patient with 5 $\alpha$ -reductase type 2 enzyme deficiency has a similar presentation but undergoes virilization at the time of puberty. *Asherman's syndrome*



**FIGURE 393-1** Role of the hypothalamic-pituitary-gonadal axis in the etiology of amenorrhea. Gonadotropin-releasing hormone (GnRH) secretion from the hypothalamus stimulates follicle-stimulating hormone (FSH) and luteinizing hormone (LH) secretion from the pituitary to induce ovarian folliculogenesis and steroidogenesis. Ovarian secretion of estradiol and progesterone controls the shedding of the endometrium, resulting in menses, and, in combination with the inhibins, provides feedback regulation of the hypothalamus and pituitary to control secretion of FSH and LH. The prevalence of amenorrhea resulting from abnormalities at each level of the reproductive system (hypothalamus, pituitary, ovary, uterus, and outflow tract) varies depending on whether amenorrhea is primary or secondary. PCOS, polycystic ovarian syndrome.

presents as secondary amenorrhea or hypomenorrhea and results from partial or complete obliteration of the uterine cavity by adhesions that prevent normal growth and shedding of the endometrium. Curettage performed for pregnancy complications accounts for >90% of cases; genital tuberculosis is an important cause in regions where it is endemic.

## TREATMENT

### Disorders of the Uterus or Outflow Tract

**Obstruction** of the outflow tract usually presents as dysmenorrhea or lower abdominal cyclic pain with no menses. Evaluation of the patient includes a medical history, physical examination including a perineal examination, and ultrasound imaging. In some cases, an MRI can more accurately identify the reproductive tract anomaly prior to surgery. It is important that surgery be performed as soon as the diagnosis is made as the risk of endometriosis is increased with retrograde menstrual flow. *Müllerian agenesis* may require surgical intervention to allow sexual intercourse, although vaginal dilatation is adequate in some patients. Because ovarian function is normal, assisted reproductive techniques can be used with a surrogate carrier. More recently, there have been a few cases of successful uterine transplantation in women with müllerian agenesis. AIS (Chap. 390) requires gonadectomy because there is risk of gonadoblastoma in the dysgenetic gonads, although surgery is generally delayed until after breast development and the pubertal growth spurt. Estrogen replacement is indicated after gonadectomy, and vaginal dilatation may be required to allow sexual intercourse.

**Disorders of Ovulation** Once uterus and outflow tract abnormalities have been excluded, other causes of amenorrhea involve disorders of ovulation. The differential diagnosis is based on the results of initial tests, including a pregnancy test, an FSH level (to determine whether the cause is likely to be ovarian or central), and assessment of hyperandrogenism (Fig. 393-2).

**HYPOGONADOTROPIC HYPOGONADISM** Low estrogen levels in combination with normal or low levels of LH and FSH are seen with anatomic, genetic, or functional abnormalities that interfere with hypothalamic GnRH secretion or normal pituitary responsiveness to GnRH. Although relatively uncommon, tumors and infiltrative diseases should be considered in the differential diagnosis of hypogonadotrophic hypogonadism (Chap. 380). These disorders may present with primary or secondary amenorrhea. They may occur in association with other features suggestive of hypothalamic or pituitary dysfunction, such as short stature, diabetes insipidus, galactorrhea, and headache. Hypogonadotrophic hypogonadism also may be seen after cranial irradiation. In the postpartum period, amenorrhea occurs normally in association with breast feeding but may also be caused by pituitary necrosis (Sheehan's syndrome) or lymphocytic hypophysitis. Because reproductive dysfunction is commonly associated with hyperprolactinemia from neuroanatomic lesions or medications, prolactin should be measured in all patients with hypogonadotrophic hypogonadism (Chap. 380).

Isolated hypogonadotropic hypogonadism (IHH) occurs in women, although it is three times more common in men. IHH generally presents with primary amenorrhea, although 50% have some degree of breast development, and ~10% report one to two menses. IHH is associated with anosmia in half of women (termed Kallmann's syndrome). Genetic causes of IHH have been identified in ~50% of patients (Chaps. 391 and 392).

Functional hypothalamic amenorrhea (HA) is a diagnosis of exclusion of other causes of hypogonadotropic hypogonadism including chronic diseases (type 1 diabetes, celiac disease, hyperthyroidism, Cushing's syndrome) and use of opioids, glucocorticoids, or psychotropic medications that increase prolactin levels. Functional HA is most commonly associated with conditions causing a mismatch between energy expenditure and energy intake and/or significant stress. Variants in genes associated with IHH may increase susceptibility to these environmental inputs, accounting in part for the clinical variability in this disorder. Metabolic and stress signaling is transduced to the reproductive axis, at least in part, through leptin signaling from the periphery and via hypothalamic kisspeptin control of GnRH. The diagnosis

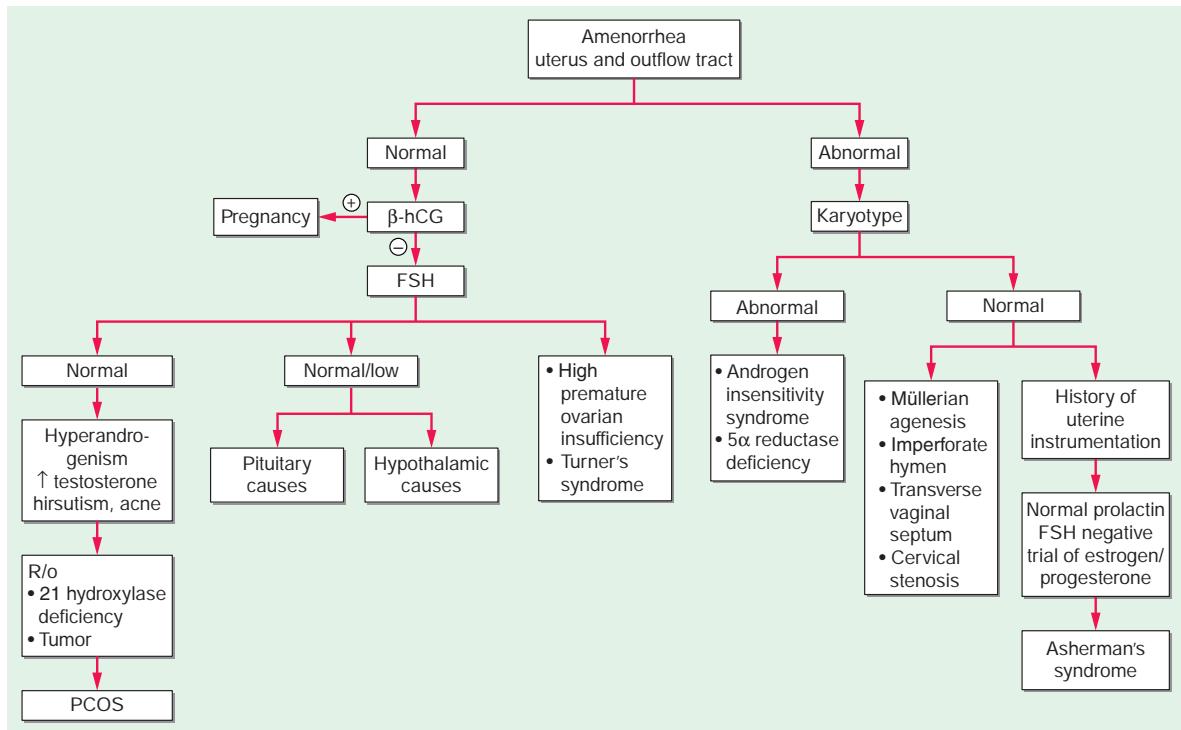


FIGURE 393-2 Algorithm for evaluation of amenorrhea. β-hCG, β-human chorionic gonadotropin; FSH, follicle-stimulating hormone; GYN, gynecologist; MRI, magnetic resonance imaging; PRL, prolactin; R/O, rule out; TSH, thyroid-stimulating hormone.

of HA generally can be made on the basis of a careful history, a physical examination, and the demonstration of low levels of gonadotropins and normal prolactin levels. Eating disorders, excessive exercise, and chronic disease must be specifically excluded. An atypical history, headache, signs of other hypothalamic dysfunction, or hyperprolactinemia, even if mild, necessitates cranial magnetic resonance imaging (MRI) to exclude a neuroanatomic cause. Up to 10% of women with HA may have some features of PCOS (irregular menses, increased ovarian volume with polycystic appearing ovaries, higher anti-müllerian hormone [AMH] levels, and slightly elevated androgen levels).

**HYPERGONADOTROPIC HYPOGONADISM** Ovarian failure is considered premature when it occurs in women <40 years old and accounts for ~10% of secondary amenorrhea. *Primary ovarian insufficiency* (POI) has replaced the terms *premature menopause* and *premature ovarian failure* in recognition of the continuum of impaired ovarian function encompassed by this disorder. Ovarian insufficiency is associated with the loss of negative feedback restraint on the hypothalamus and pituitary, resulting in increased FSH and LH levels. FSH is a better marker of ovarian failure because of loss of negative feedback effects of both estradiol and the inhibins and because its levels are less variable than those of LH. AMH levels will also be low in patients with POI but are more frequently used in management of infertility. As with natural menopause, POI may wax and wane, and serial measurements may be necessary to establish the diagnosis. The presentation may include irregular menses or complete cessation of menses, hot flashes, and vaginal dryness.

Once the diagnosis of POI has been established, further evaluation is indicated because of other health problems that may be associated with POI. Although POI is most commonly of unknown cause, it also occurs in association with a variety of chromosomal abnormalities (most often Turner's syndrome), autoimmune polyglandular failure syndromes, and other rare disorders. Radiotherapy and chemotherapy may reduce ovarian reserve, with effects on both the oocytes and the supporting granulosa cells. New approaches, including ovarian, oocyte, and embryo cryopreservation, should be offered to women of reproductive age prior to gonadotoxic chemotherapy or pelvic radiation treatment. The recognition that early ovarian insufficiency occurs in premutation carriers of the fragile X syndrome is important because of the increased risk of severe intellectual disability in male children with *FMR1* mutations. Thus, follow-up testing should include a karyotype in all POI patients, serum anti-cortical and 21-hydroxylase antibodies (specific but not sensitive for subsequent adrenal insufficiency), thyroid function and thyroid peroxidase antibodies, *FMR1* premutation screening, and assessment of bone mineral density. Ovarian biopsy is not indicated. Although the number of genetic causes of POI is increasing, routine testing for mutations other than *FMR1* is currently not recommended.

Hypergonadotropic hypogonadism occurs rarely in other disorders, such as mutations in the FSH or LH receptors. Aromatase deficiency and 17<sup>-</sup>hydroxylase deficiency are associated with decreased estrogen and elevated gonadotropins and with hyperandrogenism and hypertension, respectively. Gonadotropin-secreting tumors in women of reproductive age generally present with high, rather than low, estrogen levels and cause ovarian hyperstimulation or dysfunctional bleeding.

## TREATMENT

### Hypo- and Hypergonadotropic Causes of Amenorrhea

Amenorrhea almost always is associated with chronically low levels of estrogen, whether it is caused by hypogonadotropic hypogonadism or ovarian insufficiency. Development of secondary sexual characteristics requires gradual titration of estradiol replacement with eventual addition of progestin. Hormone replacement with either low-dose estrogen/progesterone regimens or oral contraceptive pills is recommended until the usual age of menopause for bone and cardiovascular protection. In women with functional HA or anorexia nervosa, hormone replacement alone may not be sufficient to restore or maintain bone density. Patients with

hypogonadotropic hypogonadism who are interested in fertility require treatment with both exogenous FSH and LH. Patients with POI can consider oocyte donation, which has a high rate of success in this population, although its use in women with Turner's syndrome is limited by significant cardiovascular risk in pregnancy.

**POLYCYSTIC OVARIAN SYNDROME** The diagnosis of PCOS is made in adult women using the Rotterdam criteria: irregular menses (<8 menses per year), clinical or biochemical hyperandrogenism (elevated total or free testosterone, modified Ferriman-Gallwey score >4–6 depending on ethnicity, see Chap. 394), and polycystic-appearing ovaries on ultrasound ( $\geq 20$  antral follicles or ovarian volume  $\geq 10 \text{ cm}^3$  in at least one ovary). The presence of two of the three criteria will confirm the diagnosis, resulting in different phenotypes, namely, hyperandrogenic or non-hyperandrogenic. PCOS is a diagnosis of exclusion, and other etiologies for irregular menses and hyperandrogenism should be excluded (hypothyroidism, hyperprolactinemia, adrenal sources for hyperandrogenism). Diagnosis in adolescents maybe difficult to establish, and it is recommended to wait at least 3 years after menarche before confirming the diagnosis. In adolescents, the diagnosis is based on irregular menses and hyperandrogenism criteria only, as the ultrasound criteria are not established for this age group. The prevalence of obesity is high in PCOS and significantly increases the risk of comorbidities including metabolic syndrome, type 2 diabetes, dyslipidemia, and hypertension. The failure of regular ovulation results in irregular menses and increased risk of endometrial hyperplasia and endometrial cancer (two- to sixfold increased risk). Abnormalities in GnRH pulsatility result in elevated LH and increased production of ovarian androgens. Insulin resistance associated with PCOS may also contribute to increased insulin-stimulated ovarian androgen production. An alternate source of androgens, namely 11-oxygenated androgens, has also been shown to be elevated in women with PCOS. Genome-wide association studies in diverse populations and PCOS phenotypes have identified several loci associated with PCOS. Symptoms generally begin in adolescence and are modified by obesity and age, such that by the fourth decade of life, most women with PCOS will have regular menses and normal serum androgens. Lean oligo-ovulatory patients with PCOS generally have high LH levels in the presence of normal to low levels of FSH and estradiol, although given the pulsatility of LH secretion, a random serum LH/FSH ratio is not included in the diagnostic criteria.

## TREATMENT

### Polycystic Ovarian Syndrome

The first-line treatment of women with PCOS not attempting pregnancy is combined hormonal contraceptives to regulate menstrual cycles and decrease serum androgens by increasing sex hormone-binding globulin levels. Although serum androgens decrease by 3 months after initiating hormonal therapy, it may take longer to observe the beneficial effects on hirsutism and acne. Patients should be evaluated for metabolic comorbidities and given hormonal contraceptives containing the lowest effective dose of estrogen, either in a cyclic or continuous manner. If there is an inadequate response to hormonal contraceptives for management of hyperandrogenic symptoms, antiandrogens, such as spironolactone and flutamide, can be considered (Chap. 394). Endometrial protection can also be achieved with the use of progestins (medroxyprogesterone acetate, 5–10 mg, or Prometrium [progesterone], 200 mg daily for 10–14 days at least every 3 months, or a levonorgestrel intrauterine device [IUD]). All women with PCOS should be screened for obesity, hypertension, and glycemic control at the time of diagnosis and then at regular intervals. Overweight and obese women should also have a fasting lipid profile at the time of diagnosis. Lifestyle management should be recommended in all women with PCOS, and metformin should be considered for managing cardiometabolic risk factors (Chap. 408). Women with PCOS are at an increased risk of gestational diabetes, gestational hypertension, and preeclampsia.

Lifestyle management is the first-line treatment prior to attempting pregnancy (**Chap. 396**). Letrozole, an aromatase inhibitor, and clomiphene citrate, a selective estrogen response modulator, are effective first-line treatments for ovulation induction. Exogenous gonadotropins can be used by experienced practitioners; a diagnosis of polycystic ovaries increases the risk of hyperstimulation, even in women with regular, ovulatory menstrual cycles. Metformin is frequently used in patients with PCOS and is appropriate as an adjunct with diet and exercise for obese women with PCOS or for treatment of diabetes or impaired glucose tolerance, as in non-PCOS patients. However, metformin alone is not recommended for endometrial protection or treatment of hyperandrogenic symptoms, infertility, pregnancy loss, or prevention of gestational diabetes.

## PELVIC PAIN

The mechanisms that cause pelvic pain are similar to those that cause abdominal pain (**Chap. 15**) and include inflammation of the parietal peritoneum, obstruction of hollow viscera, vascular disturbances, and pain originating in the abdominal wall. Pelvic pain may reflect pelvic disease *per se* but also may reflect extrapelvic disorders that refer pain to the pelvis. In up to 60% of cases, pelvic pain can be attributed to gastrointestinal problems, including appendicitis, cholecystitis, infections, intestinal obstruction, diverticulitis, and inflammatory bowel disease. Urinary tract and musculoskeletal disorders are also common causes of pelvic pain.

## APPROACH TO THE PATIENT

### Pelvic Pain

As with all types of abdominal pain, the first priority is to identify life-threatening conditions (shock, peritoneal signs) that may require emergent surgical management. The possibility of pregnancy should be identified as soon as possible by menstrual history and -human chorionic gonadotropin (-hCG) testing. A thorough history that includes the type, location, radiation, and recurrence can help identify the cause of acute pelvic pain. Specific associations with vaginal bleeding, sexual activity, defecation, urination, movement, or eating should be specifically sought. Determination of whether the pain is acute versus chronic and cyclic versus noncyclic will direct further investigation (**Table 393-1**). However, disorders that cause cyclic pain occasionally may cause noncyclic pain, and the converse is also true.

### ACUTE PELVIC PAIN

*Pelvic inflammatory disease* (PID) refers to infection of the upper genital tract and may present with a spectrum of symptoms. In the acute setting, the most common presentation is bilateral lower abdominal pain of recent onset that may be exacerbated with sexual activity. Risk

factors for PID include history of multiple sexual partners, prior sexually transmitted infections (STIs), history of recent uterine procedures, and age <25 years. However, any sexually active woman can be at risk for PID. PID associated with tubo-ovarian abscess or peritonitis may present with severe pain, fever, and peritoneal signs. Abnormal uterine bleeding may occur in about one-third of patients. Cervical motion tenderness, uterine and adnexal pain, and vaginal discharge are common findings on pelvic examination. The presence of right upper quadrant pain is suggestive of perihepatitis (Fitz-Hugh-Curtis syndrome).

The diagnosis of PID is established based on symptoms and clinical examination and can be aided by a wet mount preparation of vaginal discharge and nucleic acid amplification tests for *Chlamydia trachomatis* and *Neisseria gonorrhoeae*. Of note, a presumptive clinical diagnosis is sufficient to prescribe treatment even in the absence of positive test results, as PID can occur due to other vaginal and enteric pathogens. Pelvic imaging can be obtained based on symptoms, findings of the pelvic examination, or if there is lack of response to therapy. With public health efforts to control STIs, the incidence and severity of PID have declined in the United States and Europe; however, this is not the case in the developing world. Subclinical PID with its attendant risks of infertility and ectopic pregnancy remains a significant problem worldwide. Public health and professional organizations recommend annual testing for *C. trachomatis* in all sexually active women <25 years old and both *C. trachomatis* and *N. gonorrhoeae* in all women at increased risk. *Adnexal pathology* can present acutely and may be due to rupture, bleeding, or torsion of ovarian cysts or, much less commonly, the fallopian tubes. Rupture of an ovarian cyst may be diagnosed based on the acute presentation in a reproductive-age woman and pelvic ultrasound findings of a simple, collapsed or hemorrhagic cyst, with or without free fluid in the pelvis. Ovarian torsion typically presents as acute onset of unilateral, intermittent pain and is a diagnosis of exclusion unless absent blood flow to the ovary is demonstrated via Doppler ultrasound imaging. Neoplasms of the ovary or fallopian tube are much less common causes of acute pain. *Ectopic pregnancy* represents 2% of all pregnancies and most commonly occurs in the fallopian tubes. It may present with acute lower abdominal pain, hemodynamic instability, and peritoneal signs. The index of suspicion should be high in any reproductive-age woman presenting with abdominal pain or vaginal bleeding irrespective of current use of contraception. Risk factors for an ectopic pregnancy include history of tubal disease, pelvic infection, tubal surgery, previous ectopic pregnancy, infertility, smoking, and current use of IUD, although a large proportion may have no risk factors. Rupture of the fallopian tube remains a life-threatening emergency; the incidence depends on access to care but is ~18% in developed countries. Diagnosis of an ectopic pregnancy can be established by assessing the patient's menstrual history and symptoms, measuring -hCG levels, and performing pelvic ultrasound imaging. The discriminatory zone refers to -hCG values above which the landmarks of a normal intrauterine pregnancy should be seen on ultrasound. Absence of an intrauterine pregnancy and presence of an adnexal mass or free fluid increase the likelihood of an ectopic pregnancy. *Threatened abortion* may also present with amenorrhea, abdominal pain, and vaginal bleeding in the setting of an intrauterine pregnancy with cardiac activity in the first trimester of pregnancy. Although more common than ectopic pregnancy, it is rarely associated with systemic signs. *Uterine pathology* includes endometritis and, less frequently, degenerating leiomyomas (fibroids) present with acute pain. Endometritis often is associated with vaginal bleeding and systemic signs of infection. It occurs in the setting of STIs, uterine instrumentation, or postpartum infection.

## TREATMENT

### Acute Pelvic Pain

Treatment of acute pelvic pain depends on the suspected etiology but may require surgical or gynecologic intervention. Immediate treatment of PID is indicated upon diagnosis, even if the diagnosis is presumed or the symptoms are mild, due to long-term complications resulting in increased risk of ectopic pregnancy and infertility. Treatment in patients eligible for outpatient management

**TABLE 393-1 Gynecologic Causes of Pelvic Pain**

	ACUTE	CHRONIC
Cyclic pelvic pain		Mittelschmerz Dysmenorrhea
Noncyclic pelvic pain	Pelvic inflammatory disease  Ruptured or hemorrhagic ovarian cyst, endometrioma, or ovarian torsion  Ectopic pregnancy  Endometritis  Acute growth or degeneration of uterine myoma  Threatened abortion	Endometriosis Fibroids Adenomyosis Adhesions and retroversion of the uterus  Pelvic malignancy Vulvodynia Chronic pelvic inflammatory disease  Tuberculous salpingitis History of sexual abuse  Pelvic congestion syndrome

includes 250 mg IM ceftriaxone and a 14-day course of oral doxycycline 100 mg twice daily. If the presentation is acute with high fever, nausea, vomiting, severe abdominal pain, or presence of tubo-ovarian abscess, inpatient therapy is recommended (**Chap. 136**). Conservative management is an important consideration for ovarian cysts, if torsion is not suspected, to avoid unnecessary surgery and associated risks of reduced fertility due to cystectomy or adhesions. If surgery is performed, it is preferable to perform a cystectomy, removing the cyst wall and leaving the remaining ovary, in a reproductive-age woman. Combined hormonal contraceptives are recommended in women with a history of recurrent ovarian cyst formation. Surgical treatment may be required for *ectopic pregnancies* when the patient presents with acute pain, is hemodynamically unstable, or has signs of intraperitoneal bleeding. The choice of salpingectomy versus salpingostomy is based on patient's presentation, desire for future child bearing, and prior pelvic infections. Clinically stable women presenting with unruptured ectopic pregnancies may be appropriate for treatment with methotrexate, which is effective in ~90% of cases when multiple doses are used. Threatened abortion is managed conservatively even in the presence of a subchorionic hemorrhage. The treatment of endometritis is similar to PID. Pain from a degenerating fibroid, if visualized on pelvic sonography, can be managed with nonsteroidal anti-inflammatory drugs (NSAIDs).

## CHRONIC PELVIC PAIN

Chronic pelvic pain is a complex condition resulting from gynecologic, urologic, or gastrointestinal organs and contributes to significant frustration and burden of disease. Common gynecologic conditions contributing to chronic pain are endometriosis, fibroids, adenomyosis, and adnexal pathology. Estimated prevalence rates range from 5 to 20% for cyclic and noncyclic pain. In addition to a detailed history and physical exam, the evaluation of chronic pelvic pain typically includes a pelvic ultrasound. As causes other than those related to the female reproductive system are common, referral should be made to other specialists, as appropriate. Neuromuscular and psychosomatic etiologies should also be considered.

Some women experience discomfort at the time of ovulation (*mittelschmerz* or *ovulation pain*). The pain can be quite intense but is generally of short duration. The mechanism is thought to involve rapid expansion of the dominant follicle, although it also may be caused by peritoneal irritation by follicular fluid released at the time of ovulation.

*Dysmenorrhea* typically refers to the crampy lower abdominal midline discomfort that begins with the onset of menstrual bleeding and gradually decreases over 12–72 h. It may be associated with nausea, diarrhea, fatigue, and headache and occurs in 60–93% of adolescents, beginning with the establishment of regular ovulatory cycles. Its prevalence decreases after pregnancy and with the use of oral contraceptives. *Primary dysmenorrhea* results, in a majority of cases, from hormone-dependent prostaglandin (PG) pathway mechanisms that cause intense uterine contractions, decreased blood flow, and increased peripheral nerve hypersensitivity, resulting in pain. However, variability in response to cyclooxygenase inhibitors suggests that PG-independent pathways, such as platelet activating factor, may also mediate inflammation. *Secondary dysmenorrhea* is caused by underlying pelvic pathology.

*Endometriosis* results from the presence of endometrial glands and stroma outside the uterus. These deposits of ectopic endometrium respond to hormonal stimulation and are associated with dysmenorrhea, painful intercourse, painful bowel movements, and tender nodules that may be palpated along the uterosacral ligaments during pelvic exam. The stage/severity of endometriosis does not always correlate with the extent of pain, and pain associated with endometriosis can be cyclic or continuous. Transvaginal pelvic ultrasound is part of the initial workup and may detect an endometrioma within the ovary or, in severe cases, rectovaginal or bladder nodules. The CA125 level may be increased, but it has low negative predictive value. Diagnostic laparoscopy is performed when patients do not respond to empiric treatment. If endometriosis is detected, the severity can be staged and

the endometriotic lesions ablated or excised. The prevalence is lower in black and Hispanic women than in Caucasians and Asians.

Large *fibroids* can cause chronic pelvic pain or pressure, and submucosal fibroids may be associated with dysmenorrhea. Other secondary causes of dysmenorrhea include adenomyosis, a condition caused by the presence of ectopic endometrial glands and stroma within the myometrium. Chronic PID may be associated with ongoing pelvic pain and is associated with tuberculosis or actinomycosis. *Pelvic congestion syndrome* is associated with pelvic varicosities with low blood flow, resulting in pelvic venous congestion. However, this is no clear evidence to indicate that this finding is associated with chronic pelvic pain.

## TREATMENT

### Chronic Pelvic Pain

#### DYSMENORRHEA

Local application of heat is of some benefit. Exercise, sexual activity, a vegetarian diet, use of vitamins D, B<sub>1</sub>, B<sub>6</sub>, and E and fish oil, acupuncture, and yoga have all been suggested to be of benefit, but studies are not adequate to provide recommendations. However, NSAIDs are very effective and provide >80% sustained response rates. Ibuprofen, naproxen, ketoprofen, mefanamic acid, and nimesulide are all superior to placebo. For best response, treatment should be initiated prior to the onset of menses and continued for at least 2–3 days. Combined oral contraceptives taken cyclically or continuously effectively reduce symptoms of dysmenorrhea.

#### ENDOMETRIOSIS

Combined hormonal contraceptives or continuous progestin (either orally or a levonorgestrel IUD) are used for the treatment of endometriosis. Evidence of an endometrioma on ultrasound imaging can be medically managed and does not require surgical removal unless symptomatic. Patients who do not respond to medical management and laparoscopic resection of endometriotic lesions can be offered GnRH agonist suppression with add-back therapy or aromatase inhibitors.

Chronic pain and dysmenorrhea associated with *fibroids* can be managed surgically depending on the number and location of fibroids and associated symptoms. Chronic pain and dysmenorrhea associated with adenomyosis can be managed with combined hormonal treatment, levonorgestrel IUD, or hysterectomy after child bearing is complete.

#### FURTHER READING

- Bloomfield H et al: Screening pelvic examinations in asymptomatic average risk adult women. WA-ESP Project #09-009; 2013.
- Bouilly J et al: Identification of multiple gene mutations accounts for the new genetic architecture of ovarian insufficiency. J Clin Endocrinol Metab 101:4541, 2016.
- Brunham RC et al: Pelvic inflammatory disease. N Engl J Med 372:2039, 2015.
- Fourman LR, Fazeli PK: Neuroendocrine causes of amenorrhea—An update. J Clin Endocrinol Metab 100:812, 2015.
- Ju H et al: The prevalence and risk factors of dysmenorrhea. Epidemiol Rev 36:104, 2014.
- Morley LC et al: On behalf of the Royal College of Obstetricians and Gynaecologists. Metformin therapy for the management of infertility in women with polycystic ovary syndrome. Scientific Impact Paper No. 13. BJOG 124:e306, 2017.
- Oladosu FA et al: Nonsteroidal anti-inflammatory drug resistance in dysmenorrhea: Epidemiology, causes, and treatment. Am J Obstet Gynecol 218:390, 2018.
- Teede HJ et al: Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. International PCOS Network. Fertil Steril 110:364, 2018.
- Vercellini P et al: Endometriosis: Pathogenesis and treatment. Nat Rev Endocrinol 10:261, 2014.



## DEFINING HIRSUTISM

Body hair can be categorized as either *vellus* (fine, soft, and not pigmented) or *terminal* (long, coarse, and pigmented). Approximately 10% of reproductive age women have hirsutism, defined by the presence of excessive terminal hair growth. Hirsutism is most often idiopathic or the consequence of androgen excess associated with polycystic ovary syndrome (PCOS). Less frequently, it results from adrenal androgen overproduction as occurs in nonclassic congenital adrenal hyperplasia (CAH) (Table 394-1). *Androgenization* or *virilization* refers to a condition in which androgen levels are sufficiently high to cause deepening of the voice, breast atrophy, increased muscle bulk, clitoromegaly, and increased libido. Androgenization may be caused by benign hyperplasia of ovarian theca and stroma cells (e.g., *hyperthecosis*); it may also be a harbinger of a serious underlying condition, such as an ovarian or adrenal neoplasm. Cutaneous manifestations commonly associated with hirsutism include acne and hair thinning or pattern hair loss (androgenic alopecia).

## HAIR FOLLICLE GROWTH AND DIFFERENTIATION

The number of hair follicles remains unchanged over the life span, but follicle size and the type of hair can change in response to numerous

**TABLE 394-1 Causes of Hirsutism**

Gonadal hyperandrogenism
Ovarian hyperandrogenism
Polycystic ovary syndrome/functional ovarian hyperandrogenism
Ovarian steroidogenic blocks
Syndromes of extreme insulin resistance
Ovarian neoplasms
Hyperthecosis
Adrenal hyperandrogenism
Premature adrenarche
Functional adrenal hyperandrogenism
Congenital adrenal hyperplasia (nonclassic and classic)
Abnormal cortisol action/metabolism
Adrenal neoplasms
Other endocrine disorders
Cushing's syndrome
Hyperprolactinemia
Acromegaly
Peripheral androgen overproduction
Obesity
Idiopathic
Pregnancy-related hyperandrogenism
Hyperreactio luteinalis
Thecoma of pregnancy
Drugs
Androgens
Oral contraceptives containing androgenic progestins
Minoxidil
Phenytoin
Diazoxide
Cyclosporine
Valproic Acid
Ovotesticular disorders of sex development

factors, particularly androgens. Androgens are necessary for terminal hair and sebaceous gland development and mediate differentiation of pilosebaceous units (PSUs) into a terminal hair follicle and/or a sebaceous gland. In the former case, androgens transform the vellus hair into a terminal hair; in the latter case, the sebaceous component proliferates and the hair remains vellus.

There are three phases in the cycle of hair growth: (1) *anagen* (growth phase), (2) *catagen* (involution phase), and (3) *telogen* (rest phase). Depending on the body site, hormonal regulation may play an important role in the hair growth cycle. Hair growth on the face, chest, upper abdomen, and back typically requires elevated androgen concentrations. However, there is only a modest correlation between androgen levels and the quantity of hair growth. This is due to the fact that hair growth from the follicle also depends on local growth factors, and the variability in end-organ (PSU) sensitivity to androgens. Genetic factors and ethnic background also influence hair growth. Androgen excess in women may result in hair thinning or loss because androgens cause scalp hairs to spend less time in the anagen phase.

In general, dark-haired individuals tend to be more hirsute than blond or fair individuals. Asians and Native Americans have relatively sparse hair in regions sensitive to high androgen levels, whereas people of Mediterranean descent are more hirsute.

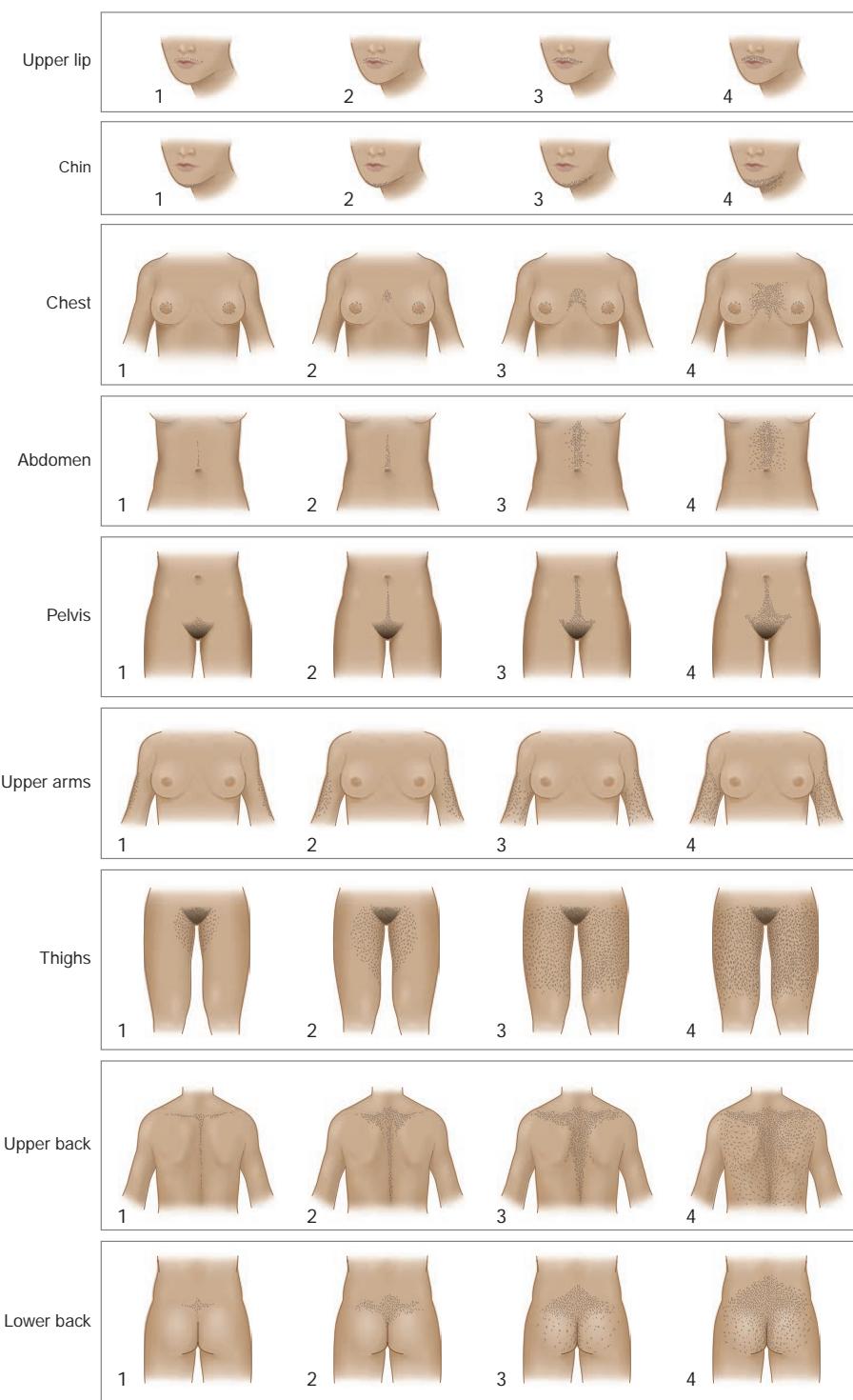
## CLINICAL ASSESSMENT

Historic elements relevant to the assessment of hirsutism include the age at onset and rate of progression of hair growth and associated symptoms or signs (e.g., menstrual irregularity and acne). Depending on the cause, excess hair growth typically is first noted during the second and third decades of life. The growth is usually slow but progressive. Sudden development and rapid progression of hirsutism suggest the possibility of an androgen-secreting neoplasm, in which case androgenization may also be present.

The age at onset of menstrual cycles (menarche) and the pattern of the menstrual cycle should be ascertained; oligomenorrhea (<8 cycles per calendar year) from the time of menarche onward is more likely to result from ovarian than adrenal androgen excess. Associated symptoms such as galactorrhea should prompt evaluation for hyperprolactinemia (Chap. 380) or possibly hypothyroidism (Chap. 382). Hypertension, striae, easy bruising, and centripetal weight gain suggest hypercortisolism (Cushing's syndrome; Chap. 386). Rarely, patients with acromegaly present with hirsutism. Medications such as phenytoin, minoxidil, and cyclosporine may be associated with androgen-independent excess hair growth (i.e., hypertrichosis). A family history of infertility and/or hirsutism may indicate inherited disorders such as nonclassic CAH (Chap. 386).

Physical examination should include measurement of height and weight and calculation of body mass index (BMI). A BMI >25 kg/m<sup>2</sup> is indicative of excess weight for height, and values >30 kg/m<sup>2</sup> are often seen in association with hirsutism, probably the result of increased conversion of androgen precursors to testosterone. Notation should be made of blood pressure, as adrenal causes may be associated with hypertension. Cutaneous signs sometimes associated with androgen excess and insulin resistance include acanthosis nigricans and skin tags.

An objective clinical assessment of hair distribution and quantity is central to the evaluation in any woman presenting with concerns about excessive hair growth. This assessment permits the distinction between hirsutism and hypertrichosis and provides a baseline reference point to gauge the response to treatment. A simple and commonly used method to grade hair growth is the modified scale of Ferriman and Gallwey (Fig. 394-1), in which each of nine androgen-sensitive sites is graded from 0 (no hair growth) to 4 (hair growth typically seen in adult men). Although it is normal for most women to have some hair growth in androgen-sensitive sites, ~95% of non-Hispanic white and African-American women have a score <8 on this scale. Scores >8 suggest excess androgen-mediated hair growth, a finding that should be assessed further by means of hormonal evaluation (see below). Asian and Native American women are less likely to manifest hirsutism, and the only cutaneous evidence of androgen excess may be pustular acne and thinning scalp hair.



**FIGURE 394-1 Hirsutism scoring scale of Ferriman and Gallwey.** The nine body areas that have androgen-sensitive areas are graded from 0 (no terminal hair) to 4 (frankly virile) to obtain a total score. A normal hirsutism score is <8. (Modified with permission from LJ DeGroot, JL Jameson: *Endocrinology*, 5th ed. Philadelphia, PA: Saunders; 2006.)

### HORMONAL EVALUATION

Androgens are secreted by the ovaries and adrenal glands in response to their respective tropic hormones: luteinizing hormone (LH) and adrenocorticotrophic hormone (ACTH). Testosterone is the principal circulating steroid involved in the etiology of hirsutism; other steroids that may contribute to the development of hirsutism include

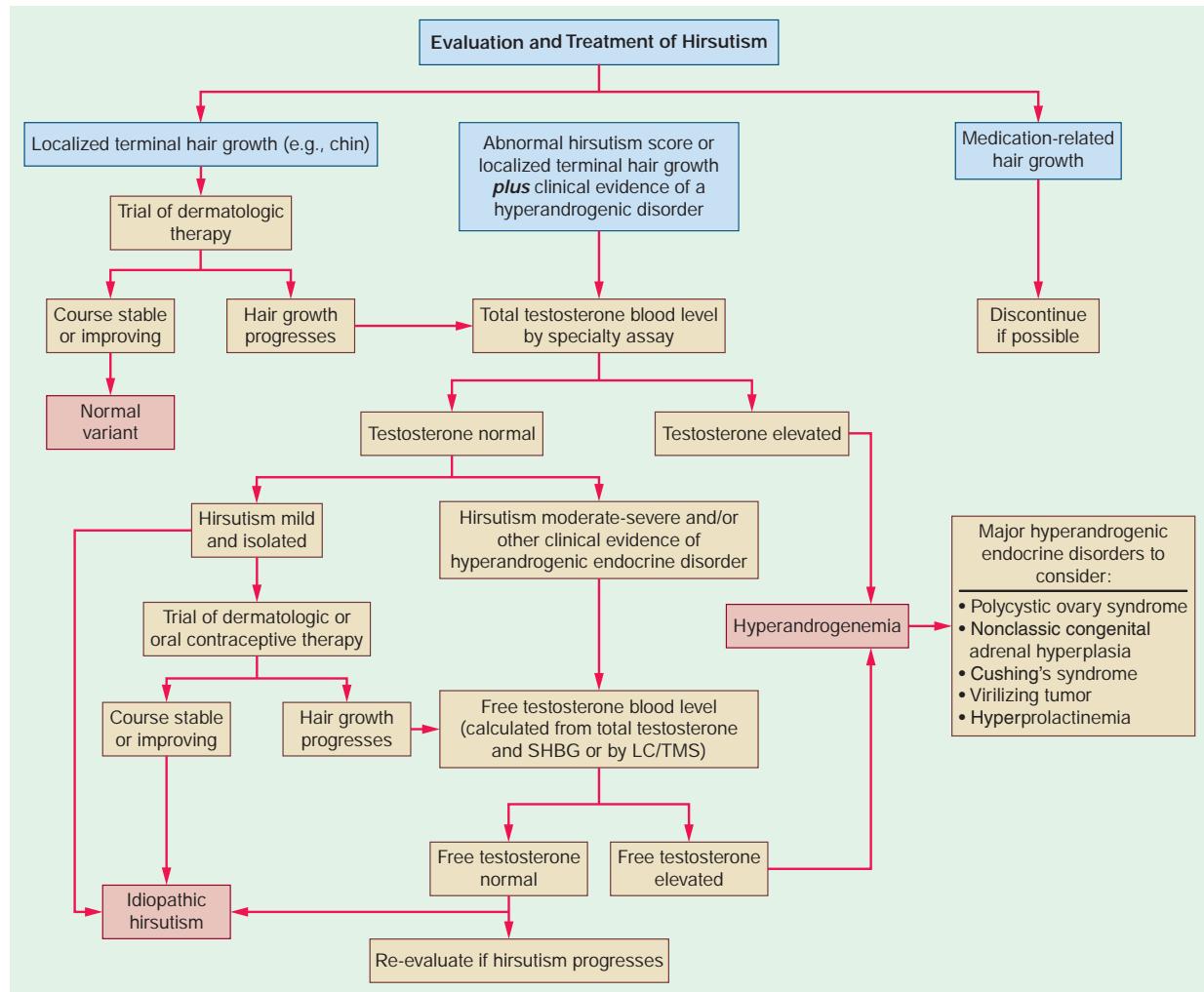
androstenedione and dehydroepiandrosterone (DHEA) and its sulfated form (DHEAS). The ovaries and adrenal glands normally contribute about equally to testosterone production. Approximately half of the total testosterone originates from direct glandular secretion, and the remainder is derived from the peripheral conversion of androstenedione and DHEA (Chap. 381).

Testosterone is the most important circulating androgen, but it is a precursor hormone in mediating hirsutism. Testosterone is converted to dihydrotestosterone (DHT) by the enzyme 5'-reductase, which is located in the PSU. DHT is more potent than testosterone as it has a higher affinity for, and slower dissociation from, the androgen receptor. The local production of DHT allows it to serve as the primary mediator of androgen action at the level of the PSU. There are two isoenzymes of 5'-reductase: type 2 is found in the prostate gland and in hair follicles, and type 1 is found primarily in sebaceous glands.

One approach to the evaluation and treatment of hirsutism is depicted in Fig. 394-2. In addition to measuring blood levels of testosterone and DHEAS, it is often important to measure the level of free (or unbound) testosterone, i.e., the fraction of testosterone that is not bound to its carrier protein, sex hormone-binding globulin (SHBG). Unbound testosterone is biologically available for conversion to DHT and binding to androgen receptors. Both hyperinsulinemia and androgen excess decrease hepatic production of SHBG, resulting in levels of total testosterone within the high-normal range, whereas the unbound hormone is elevated more substantially. Although there is a decline in ovarian testosterone production after menopause, ovarian estrogen production decreases to an even greater extent, and the concentration of SHBG is reduced. Consequently, there is an increase in the relative proportion of unbound testosterone, and it may exacerbate hirsutism after menopause.

A baseline plasma total testosterone level >12 nmol/L (>3.5 ng/mL) usually indicates an androgen-producing tumor, whereas a level >7 nmol/L (>2 ng/mL) is suggestive of tumor but may also be observed in women with hyperthecosis. A basal DHEAS level >18.5 µmol/L (>7000 µg/L) suggests an adrenal tumor. Although DHEAS has been proposed as a "marker" of predominant adrenal androgen excess, it is not unusual to find modest elevations in DHEAS among women with PCOS. Computed tomography (CT) or magnetic resonance imaging (MRI) should be used to localize an adrenal mass, and ultrasound usually suffices to identify an ovarian mass if clinical evaluation and hormonal levels suggest these possibilities.

PCOS is the most common cause of ovarian androgen excess (Chap. 392). An increased ratio of LH to follicle-stimulating hormone (FSH) is characteristic in carefully studied patients with PCOS. However, because of the pulsatile nature of gonadotropin secretion, a random measurement of LH and FSH may be misleading and is not recommended. Transvaginal ultrasound classically shows enlarged ovaries, increased stroma, and multiple "cysts" in women with PCOS. These so-called cysts are, in fact, preantral and early antral follicles that result from abnormal follicular maturation. "Cystic" ovaries also may be found in women with hypothalamic amenorrhea (Chap. 392) and even among women without clinical or laboratory features of PCOS. Thus, ultrasonography is often not needed to diagnose PCOS given its relatively low specificity and its high degree of operator dependence.



**FIGURE 394-2** Algorithm for the evaluation and treatment of hirsutism. LC/TMS, liquid chromatography/tandem mass spectrometry; SHBG, sex hormone-binding globulin. (Reproduced with permission from KA Martin et al: Evaluation and treatment of hirsutism in premenopausal women: An endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 103:1233, 2018.)

Because adrenal androgens are readily suppressed by low doses of glucocorticoids, the dexamethasone androgen-suppression test may broadly distinguish ovarian from adrenal androgen overproduction. A blood sample is obtained before and after the administration of dexamethasone (0.5 mg orally every 6 h for 4 days). An adrenal source is suggested by suppression of unbound testosterone into the normal range; incomplete suppression suggests ovarian androgen excess. An overnight 1-mg dexamethasone suppression test, with measurement of 8:00 a.m. serum cortisol, is useful when there is clinical suspicion of Cushing's syndrome ([Chap. 386](#)).

Nonclassic CAH is most commonly due to 21-hydroxylase deficiency but also can be caused by autosomal recessive defects in other steroidogenic enzymes necessary for adrenal corticosteroid synthesis ([Chap. 386](#)). Because of the enzyme defect, the adrenal gland cannot secrete glucocorticoids (especially cortisol) efficiently. This results in diminished negative feedback inhibition of ACTH, leading to compensatory adrenal hyperplasia and the accumulation of steroid precursors that subsequently are converted to androgen. Deficiency of 21-hydroxylase can be reliably excluded by determining a morning 17-hydroxyprogesterone level  $<6$  nmol/L ( $<2$  µg/L) (drawn in the follicular phase). Alternatively, 21-hydroxylase deficiency can be diagnosed by measurement of 17-hydroxyprogesterone 1 h after the administration of 250 µg of synthetic ACTH (cosyntropin) intravenously.

## TREATMENT

### Hirsutism

Treatment of hirsutism may be accomplished pharmacologically or by mechanical means of hair removal. Nonpharmacologic treatments should be considered in all patients either as the only treatment or as an adjunct to drug therapy.

Nonpharmacologic treatments include (1) bleaching, (2) depilatory (removal from the skin surface) such as shaving and chemical treatments, and (3) epilatory (removal of the hair including the root) such as plucking, waxing, electrolysis, laser, and intense pulsed light (IPL). Despite perceptions to the contrary, shaving does not increase the rate or density of hair growth. Chemical depilatory treatments may be useful for mild hirsutism that affects only limited skin areas, although they can cause skin irritation. Wax treatment removes hair temporarily but is uncomfortable. Electrolysis is effective for more permanent hair removal, particularly in the hands of a skilled electrologist. Laser and IPL are used to treat large areas of pigmented, terminal hair. Light of specific wavelength, duration, and energy is absorbed by melanin in the hair shaft and follicle leading to photothermolysis. Properly delivered, this treatment delays hair regrowth and causes permanent hair removal in many patients.

Pharmacologic therapy is directed at interrupting one or more of the steps in the pathway of androgen synthesis and action: (1) suppression of adrenal and/or ovarian androgen production, (2) enhancement of androgen-binding to plasma-binding proteins, particularly SHBG, (3) impairment of the peripheral conversion of androgen precursors to active androgen, and (4) inhibition of androgen action at the target tissue level. Attenuation of hair growth is typically not evident until 4–6 months after initiation of medical treatment and, in most cases, leads to only a modest reduction in hair growth.

Combination estrogen-progestin therapy in the form of an oral contraceptive is usually the first-line endocrine treatment for hirsutism and acne, after cosmetic and dermatologic management. The estrogenic component of most oral contraceptives currently in use is either ethynodiol diacetate or mestranol. The suppression of LH leads to reduced production of ovarian androgens. The reduced androgen levels also result in a dose-related increase in SHBG, thus lowering the fraction of unbound plasma testosterone. Estrogens also have a direct, dose-dependent suppressive effect on sebaceous cell function.

The choice of a specific oral contraceptive should be predicated on the progestational component, as progestins vary in their suppressive effect on SHBG levels and in their androgenic

potential. Ethynodiol diacetate has relatively low androgenic potential, whereas progestins such as norgestrel and levonorgestrel are particularly androgenic, as judged from their attenuation of the estrogen-induced increase in SHBG. Norgestimate exemplifies the newer generation of progestins that are virtually nonandrogenic. Drosperone, an analogue of spironolactone that has both antimineralocorticoid and antiandrogenic activities, is commonly used as a progestational agent in combination with ethynodiol diacetate.

Oral contraceptives are contraindicated in women with a history of thromboembolic disease and women with increased risk of breast or other estrogen-dependent cancers ([Chap. 395](#)). There is a relative contraindication to the use of oral contraceptives in smokers and those with hypertension or a history of migraine headaches. In most trials, estrogen-progestin therapy alone improves the extent of acne by a maximum of 50–70%. The effect on hair growth may not be evident for 6 months, and the maximum effect may require 9–12 months owing to the length of the hair growth cycle. Improvements in hirsutism are typically in the range of 20%, but there may be an arrest of further progression of hair growth.

Because oral contraceptives are efficacious and have fewer side effects, they are recommended over glucocorticoids as first-line treatment of hirsutism in CAH. If the response to oral contraceptives is inadequate, glucocorticoids may be used. The lowest effective dose of glucocorticoid should be used (e.g., dexamethasone [0.2–0.5 mg] or prednisone [5–10 mg]) taken at bedtime to achieve maximal suppression by inhibiting the nocturnal surge of ACTH.

Cyproterone acetate is the prototypic antiandrogen. It acts mainly by competitive inhibition of the binding of testosterone and DHT to the androgen receptor. In addition, it may enhance the metabolic clearance of testosterone by inducing hepatic enzymes. Although not available for use in the United States, cyproterone acetate is widely used in Canada, Mexico, and Europe. Cyproterone (50–100 mg) is given on days 1–15, and ethynodiol diacetate (50 µg) is given on days 5–26 of the menstrual cycle. Side effects include irregular uterine bleeding, nausea, headache, fatigue, weight gain, and decreased libido.

Spirostanolactone, which usually is used as a mineralocorticoid antagonist, is also a weak antiandrogen. It is almost as effective as cyproterone acetate when used at high enough doses (100–200 mg daily). Patients should be monitored intermittently for hyperkalemia or hypotension, though these side effects are uncommon. Pregnancy should be avoided because of the risk of feminization of a male fetus. Spirostanolactone can also cause menstrual irregularity. It often is used in combination with an oral contraceptive, which suppresses ovarian androgen production and helps prevent pregnancy.

Flutamide is a potent nonsteroidal antiandrogen that is effective in treating hirsutism, but concerns about the induction of hepatocellular dysfunction preclude its use. Finasteride is a competitive inhibitor of 5'-reductase type 2. Beneficial effects on hirsutism have been reported, but the predominance of 5'-reductase type 1 in the PSU appears to account for its limited efficacy. Finasteride would also be expected to impair sexual differentiation in a male fetus, and it should not be used in women who may become pregnant. Although studies of dutasteride are limited in number, it appears that this agent may have efficacy in treating scalp hair thinning and loss as well as hirsutism. Dutasteride differs from finasteride as it targets 5'-reductase types 1 and 2.

Ultimately, the choice of any specific agent(s) must be tailored to the unique needs of the patient being treated. As noted previously, pharmacologic treatments for hirsutism should be used in conjunction with nonpharmacologic approaches. It is also helpful to review the pattern of female hair distribution in the normal population to dispel unrealistic expectations.

## FURTHER READING

Azarchi S et al: Androgens in women: Hormone-modulating therapies for skin disease. *J Am Acad Derm* 80:1509, 2019.

## 395 Menopause and Postmenopausal Hormone Therapy

JoAnn E. Manson, Shari S. Bassuk

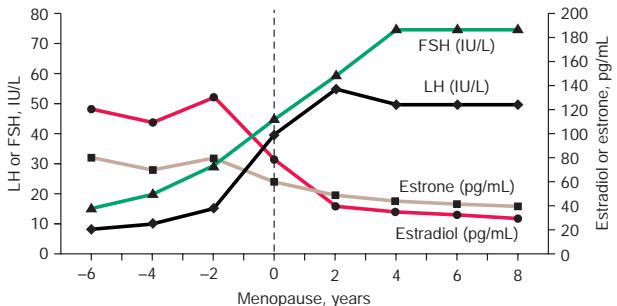
Menopause is the permanent cessation of menstruation due to loss of ovarian follicular function. It is diagnosed retrospectively after 12 months of amenorrhea. The average age at menopause is 51 years among U.S. women. *Perimenopause* refers to the time period preceding menopause, when fertility wanes and menstrual cycle irregularity increases, until the first year after cessation of menses. The onset of perimenopause precedes the final menses by 2–8 years, with a mean duration of 4 years. Smoking accelerates the menopausal transition by 2 years.

Although the peri- and postmenopausal transitions share many symptoms, the physiology and clinical management of the two differ. Low-dose oral contraceptives have become a therapeutic mainstay in perimenopause, whereas postmenopausal hormone therapy (HT) has been a common method of symptom alleviation after menstruation ceases.

### PERIMENOPAUSE

#### PHYSIOLOGY

Ovarian mass and fertility decline sharply after age 35 and even more precipitously during perimenopause; depletion of primary follicles, a process that begins before birth, occurs steadily until menopause (Chap. 392). In perimenopause, intermenstrual intervals shorten significantly (typically by 3 days) as a result of an accelerated follicular phase. Follicle-stimulating hormone (FSH) levels rise because of altered folliculogenesis and reduced inhibin secretion. In contrast to the consistently high FSH and low estradiol levels seen in menopause, perimenopause is characterized by “irregularly irregular” hormone levels. The propensity for anovulatory cycles can produce a hyperestrogenic, hypoprogesteragenic environment that may account for the increased incidence of endometrial hyperplasia or carcinoma, uterine polyps, and leiomyoma observed among women of perimenopausal age. Mean serum levels of selected ovarian and pituitary hormones during the menopausal transition are shown in Fig. 395-1. With transition into menopause, estradiol levels fall markedly, whereas estrone levels are relatively preserved, a pattern reflecting peripheral aromatization of adrenal and ovarian androgens. Levels of FSH increase more than those of luteinizing hormone, presumably because of the loss of inhibin as well as estrogen feedback.



**FIGURE 395-1** Mean serum levels of ovarian and pituitary hormones during the menopausal transition. FSH, follicle-stimulating hormone; LH, luteinizing hormone. (Data from G Rannevik et al: A longitudinal study of the perimenopausal transition: altered profiles of steroid and pituitary hormones, SHBG and bone mineral density. *Maturitas* 21:103, 1995.)

### DIAGNOSTIC TESTS

The Stages of Reproductive Aging Workshop +10 (STRAW+10) classification provides a comprehensive framework for the clinical assessment of ovarian aging. As shown in Fig. 395-2, menstrual cycle characteristics are the principal criteria for characterizing the menopausal transition, with biomarker measures as supportive criteria. Because of their extreme intraindividual variability, FSH and estradiol levels are imperfect diagnostic indicators of perimenopause in menstruating women. However, a consistently low FSH level in the early follicular phase (days 2–5) of the menstrual cycle does not support a diagnosis of perimenopause, while levels >25 IU/L in a random blood sample are characteristic of the late menopause transition. FSH measurement can also aid in assessing fertility; levels of <20 IU/L, 20 to <30 IU/L, and >30 IU/L measured on day 3 of the cycle indicate a good, fair, and poor likelihood of achieving pregnancy, respectively. Anti-müllerian hormone and inhibin B may also be useful for assessing reproductive aging.

### SYMPTOMS

Determining whether symptoms that develop in midlife are due to ovarian senescence or to other age-related changes is difficult. There is strong evidence that the menopausal transition can cause hot flashes, night sweats, irregular bleeding, and vaginal dryness, and there is moderate evidence that it can cause sleep disturbances in some women. There is inconclusive or insufficient evidence that ovarian aging is a major cause of mood swings, depression, impaired memory or concentration, somatic symptoms, urinary incontinence, or sexual dysfunction. In one U.S. study, nearly 60% of women reported hot flashes in the 2 years before their final menses. Symptom intensity, duration, frequency, and effects on quality of life are highly variable.

### TREATMENT

#### Perimenopause

#### PERIMENOPAUSAL THERAPY

For women with irregular or heavy menses or hormone-related symptoms that impair quality of life, low-dose combined oral contraceptives are a staple of therapy. Static doses of estrogen and progestin (e.g., 20 µg of ethinyl estradiol and 1 mg of norethindrone acetate daily for 21 days each month) can eliminate vasomotor symptoms and restore regular cyclicity. Oral contraceptives provide other benefits, including protection against ovarian and endometrial cancers and increased bone density, although it is not clear whether use during perimenopause decreases fracture risk later in life. Moreover, the contraceptive benefit is important, given that the unintentional pregnancy rate among women in their forties rivals that of adolescents. Contraindications to oral contraceptive use include cigarette smoking, liver disease, a history

Menarche							FMP (0)			
Stage	-5	-4	-3b	-3a	-2	-1	+1a	+1b	+1c	+2
Terminology	<b>Reproductive</b>				<b>Menopausal transition</b>		<b>Postmenopause</b>			
	Early	Peak	Late		Early	Late	Early		Late	
			<b>Perimenopause</b>							
Duration	<b>Variable</b>			Variable	1–3 years	2 years (1+1)	3–6 years	<b>Remaining lifespan</b>		
<b>Principal criteria</b>										
Menstrual cycle	Variable to regular	Regular	Regular	Subtle changes in flow/length	Variable Length Persistent ≥7-day difference in length of consecutive cycles	Interval of amenorrhea of ≥60 days				
<b>Supportive criteria</b>										
<i>Endocrine</i> FSH AMH Inhibin B			Low Low	Variable* Low Low	↑ Variable* Low Low	↑ >25 IU/L** Low Low	↑ Variable Low Low	Stabilizes Very low Very low		
<i>Antral follicle count</i>			Low	Low	Low	Low	Very low	Very low		
<b>Descriptive characteristics</b>										
Symptoms						Vasomotor symptoms <i>Likely</i>	Vasomotor symptoms <i>Most likely</i>		Increasing symptoms of urogenital atrophy	

\*Blood draw on cycle days 2–5 ↑ = elevated.

\*\*Approximate expected level based on assays using current international pituitary standard.

**FIGURE 395-2 The Stages of Reproductive Aging Workshop +10 (STRAW +10) staging system for reproductive aging in women.** AMH, anti-müllerian hormone; FSH, follicle-stimulating hormone. (Reproduced with permission from SD Harlow et al: Executive summary of the Stages of Reproductive Aging Workshop + 10: addressing the unfinished agenda of staging reproductive aging. *Menopause* 19:387, 2012.)

of thromboembolism or cardiovascular disease, breast cancer, or unexplained vaginal bleeding. Progestin-only formulations (e.g., 0.35 mg of norethindrone daily) or medroxyprogesterone (Depo-Provera) injections (e.g., 150 mg IM every 3 months) may provide an alternative for the treatment of perimenopausal menorrhagia in women who smoke or have cardiovascular risk factors. Although progestins neither regularize cycles nor reduce the number of bleeding days, they reduce the volume of menstrual flow.

Nonhormonal strategies to reduce menstrual flow include the use of nonsteroidal anti-inflammatory agents such as mefenamic acid (an initial dose of 500 mg at the start of menses, then 250 mg qid for 2–3 days) or, when medical approaches fail, endometrial ablation. It should be noted that menorrhagia requires an evaluation to rule out uterine disorders. Transvaginal ultrasound with saline enhancement is useful for detecting leiomyomata or polyps, and endometrial aspiration can identify hyperplastic changes.

#### TRANSITION TO MENOPAUSE

For sexually active women using contraceptive hormones to alleviate perimenopausal symptoms, the question of when and if to switch to HT must be individualized. Doses of estrogen and progestogen (either synthetic progestins or natural forms of progesterone) in HT are lower than those in oral contraceptives and have not been documented to prevent pregnancy. Although a 1-year absence of spontaneous menses reliably indicates ovulation cessation, it is not possible to assess the natural menstrual pattern while a woman is taking an oral contraceptive. Women willing to switch to a barrier method of contraception should do so; if menses occur spontaneously, oral contraceptive use can be resumed. The average age of final menses among relatives can serve as a guide for when to initiate this process, which can be repeated yearly until menopause has occurred.

## MENOPAUSE AND POSTMENOPAUSAL HT

One of the most complex health care decisions facing women is whether to use postmenopausal HT. Once prescribed primarily to relieve vasomotor symptoms, HT has been promoted as a strategy to forestall various disorders that accelerate after menopause, including osteoporosis and cardiovascular disease. In 2000, nearly 40% of postmenopausal women aged 50–74 in the United States had used HT. This widespread use occurred despite the paucity of conclusive data, until recently, on the health consequences of such therapy. Although many women rely on their health care providers for a definitive answer to the question of whether to use postmenopausal hormones, balancing the benefits and risks for an individual patient is challenging.

Although observational studies suggest that HT prevents cardiovascular and other chronic diseases, the apparent benefits may result at least in part from differences between women who opt to take postmenopausal hormones and women who do not. Those choosing HT tend to be healthier, have greater access to medical care, are more compliant with prescribed treatments, and maintain a more health-promoting lifestyle. Randomized trials, which eliminate these confounding factors, have not consistently confirmed the benefits found in observational studies. Indeed, the largest HT trial to date, the Women's Health Initiative (WHI), which examined >27,000 postmenopausal women aged 50–79 (mean age, 63) for an average of 5–7 years, was stopped early because of an overall unfavorable benefit-risk ratio in the estrogen-progestin arm and an excess risk of stroke that was not offset by a reduced risk of coronary heart disease (CHD) in the estrogen-only arm.

The following summary offers a decision-making guide based on a synthesis of currently available evidence. Prevention of cardiovascular disease is eliminated from the equation due to lack of evidence for such benefits in randomized clinical trials.

## BENEFITS AND RISKS OF POSTMENOPAUSAL HT

**See Table 395-1.**

**Definite Benefits • SYMPTOMS OF MENOPAUSE** Compelling evidence, including data from randomized clinical trials, indicates that estrogen therapy is highly effective for controlling vasomotor and genitourinary symptoms. Alternative approaches, including the use of antidepressants (such as paroxetine, 10–25 mg/d; paroxetine salt, 7.5 mg/d; or venlafaxine, 37.5–75 mg/d), -aminobutyric acid analogues (such as gabapentin, 300 mg nightly, up to 900 mg in divided doses; or pregabalin, 75–150 mg/d twice per day), or clonidine patch (0.1–0.3 mg weekly), may also alleviate vasomotor symptoms, although they are less effective than HT. Paroxetine is the only non-hormonal drug approved by the U.S. Food and Drug Administration for treatment of vasomotor symptoms. Bazedoxifene, an estrogen agonist/antagonist, in combination with conjugated estrogens has also received approval for this use. Cognitive behavioral therapy and clinical hypnosis have been shown in randomized trials to help with vasomotor symptom management. Weight loss, mindfulness-based stress reduction, stellate ganglion block, and the consumption of S-equol soy derivatives are also promising strategies, although more trials are needed. For genitourinary syndrome of menopause, the efficacy of low-dose vaginal estrogen is similar to that of oral or transdermal estrogen; oral ospemifene or vaginal prasterone are additional options.

### OSTEOPOROSIS (See also Chap. 411)

**Bone density** By reducing bone turnover and resorption rates, estrogen slows the aging-related bone loss experienced by most postmenopausal women. More than 50 randomized trials have demonstrated that postmenopausal estrogen therapy, with or without a progestogen, rapidly increases bone mineral density at the spine by 4–6% and at the hip by 2–3% and that those increases are maintained during treatment.

**Fractures** Data from observational studies indicate a 50–80% lower risk of vertebral fracture and a 25–30% lower risk of hip, wrist, and other peripheral fractures among current estrogen users; addition of a progestogen does not appear to modify this benefit. In the WHI, 5–7 years of either combined estrogen-progestin or estrogen-only therapy was associated with a 33% reduction in hip fractures and 25–30% fewer total fractures among a population unselected for osteoporosis. Bisphosphonates (such as alendronate, 10 mg/d or 70 mg once per week; risedronate, 5 mg/d or 35 mg once per week; ibandronate, 2.5 mg/d or 150 mg once per month or 3 mg every 3 months IV; or zoledronic acid, 5 mg once per year IV) and denosumab (60 mg twice per year SC) increase bone mass density by reducing bone resorption and have been shown in randomized trials to decrease fracture rates. Other treatment options include bazedoxifene in combination with conjugated estrogens; the selective estrogen receptor modulator (SERM) raloxifene (60 mg/d); and parathyroid hormone (teriparatide, 20 µg/d SC). Unlike estrogen, these alternative therapies do not appear to have adverse effects on the endometrium or breast. Increased weight-bearing and resistance exercise; adequate calcium intake (1000–1200 mg/d through diet or supplements in two or three divided doses); and adequate vitamin D intake (600–1000 IU/d) may also reduce the risk of osteoporosis-related fractures. According to a 2011 report by the Institute of Medicine (now the National Academy of Medicine), 25-hydroxyvitamin D blood levels of 50 nmol/L are sufficient for bone-density maintenance and fracture prevention. The Fracture Risk Assessment (FRAX®) score, an algorithm that combines an individual's bone-density score with age and other risk factors to predict her 10-year risk of hip and major osteoporotic fracture, may be of use in guiding decisions about pharmacologic treatment (see <https://www.sheffield.ac.uk/FRAX/>).

**Definite Risks • ENDOMETRIAL CANCER (WITH ESTROGEN ALONE)** A combined analysis of 30 observational studies found a tripling of endometrial cancer risk among short-term users (1–5 years) of unopposed estrogen and a nearly tenfold increased risk among long-term users (10 years). These findings are supported by results from the randomized Postmenopausal Estrogen/Progestin Interventions

(PEPI) trial, in which 24% of women assigned to unopposed estrogen for 3 years developed atypical endometrial hyperplasia—a premalignant lesion—as opposed to only 1% of women assigned to placebo. Use of a progestogen, which opposes the effects of estrogen on the endometrium, eliminates these risks and may even reduce risk (see later).

**VENOUS THROMBOEMBOLISM** A meta-analysis of observational studies found that current oral estrogen use was associated with a 2.5-fold increase in risk of venous thromboembolism in postmenopausal women. A meta-analysis of randomized trials, including the WHI, found a 2.1-fold increase in risk. Results from the WHI indicate a nearly twofold increase in risk of pulmonary embolism and deep-vein thrombosis with estrogen-progestin and a 35–50% increase in these risks with estrogen-only therapy. Transdermal estrogen, taken alone or with certain progestogens (micronized progesterone or pregnane derivatives), appears to be a safer alternative with respect to thrombotic risk.

**BREAST CANCER (WITH ESTROGEN-PROGESTIN)** An increased risk of breast cancer has been found among current or recent estrogen users in observational studies; this risk is directly related to duration of use. In a meta-analysis of 51 case-control and cohort studies, short-term use (<5 years) of postmenopausal HT did not appreciably elevate breast cancer incidence, whereas long-term use (5 years) was associated with a 35% increase in risk. In contrast to findings for endometrial cancer, combined estrogen-progestin regimens appear to increase breast cancer risk more than estrogen alone. Data from randomized trials also indicate that estrogen-progestin raises breast cancer risk. In the WHI, women assigned to receive combination hormones for an average of 5.6 years were 24% more likely to develop breast cancer than women assigned to placebo, but 7.1 years of estrogen-only therapy did not increase risk. Indeed, the WHI showed a trend toward a reduction in breast cancer risk with estrogen alone, although it is unclear whether this finding would pertain to formulations of estrogen other than conjugated equine estrogens or to treatment durations of >7 years. In the Heart and Estrogen/Progestin Replacement Study (HERS), combination therapy for 4 years was associated with a 27% increase in breast cancer risk. Although the latter finding was not statistically significant, the totality of evidence strongly implicates estrogen-progestin therapy in breast carcinogenesis.

Some observational data suggest that the length of the interval between menopause onset and HT initiation may influence the association between such therapy and breast cancer risk, with a “gap time” of <3–5 years conferring a higher HT-associated breast cancer risk. (This pattern of findings contrasts with that for CHD, as discussed later in this chapter.) However, this association remains inconclusive and may be a spurious finding attributable to higher rates of screening mammography and thus earlier cancer detection in HT users than in nonusers, especially in early menopause. Indeed, in the WHI trial, hazard ratios for HT and breast cancer risk did not differ among women 50–59, those 60–69, and those 70–79 years of age at trial entry. (There was insufficient power to examine finer age categories.) Additional research is needed to clarify the issue.

**GALLBLADDER DISEASE** Large observational studies report a two- to threefold increased risk of gallstones or cholecystectomy among postmenopausal women taking oral estrogen. In the WHI, women randomized to estrogen-progestin or estrogen alone were ~55% more likely to develop gallbladder disease than those assigned to placebo. Risks were also increased in HERS. Transdermal HT might be a safer alternative, but further research is needed.

**Probable or Uncertain Risks and Benefits • CORONARY HEART DISEASE/STROKE** Until recently, HT had been enthusiastically recommended as a possible cardioprotective agent. In the past three decades, multiple observational studies suggested, in the aggregate, that estrogen use leads to a 35–50% reduction in CHD incidence among postmenopausal women. The biologic plausibility of such an association is supported by data from randomized trials demonstrating that exogenous estrogen lowers plasma low-density lipoprotein (LDL) cholesterol levels and raises high-density lipoprotein (HDL) cholesterol

TABLE 395-1 Benefits and Risks of Postmenopausal Hormone Therapy in the Overall Study Population of Women aged 50–79 Years in the Intervention Phase of the Women's Health Initiative (WHI) Estrogen-Progestin and Estrogen-Alone Trials<sup>a</sup>

OUTCOME	EFFECT	ESTROGEN-PROGESTIN		ESTROGEN ALONE	
		RELATIVE BENEFIT OR RISK	ABSOLUTE BENEFIT OR RISK <sup>b</sup>	RELATIVE BENEFIT OR RISK	ABSOLUTE BENEFIT OR RISK <sup>b</sup>
<b>Definite Benefits</b>					
Symptoms of menopause	Definite improvement	↓65–90% decreased risk <sup>c</sup>		↓65–90% decreased risk <sup>c</sup>	
Osteoporosis	Definite increase in bone mineral density and decrease in fracture risk	↓33% decreased risk for hip fracture	6 fewer cases (11 vs 17) of hip fracture	↓33% decreased risk for hip fracture	6 fewer cases (13 vs 19) of hip fracture
<b>Definite Risks<sup>h</sup></b>					
Endometrial cancer	Definite increase in risk with estrogen alone (see below for estrogen-progestin)	See below	See below		4.6 excess cases (observational studies)
Pulmonary embolism	Definite increase in risk	↑98% increased risk	9 excess cases (18 vs 9)	↑35% increased risk (n.s.)	4 excess cases (14 vs 10)
Deep-vein thrombosis	Definite increase in risk	↑87% increased risk	11.5 excess cases (25 vs 14)	↑48% increased risk	7.5 excess cases (23 vs 15)
Breast cancer	Definite increase in risk with long-term use (5 years) of estrogen-progestin	↑24% increased risk	8.5 excess cases (43 vs 35)	↓21% decreased risk (n.s.)	7 fewer cases (28 vs 35)
Gallbladder disease	Definite increase in risk	↑57% increased risk	47 excess cases (131 vs 84)	↑55% increased risk	58 excess cases (164 vs 106)
<b>Probable or Uncertain Risks and Benefits<sup>h</sup></b>					
Coronary heart disease <sup>d</sup>	Probable increase in risk among older women and women many years past menopause; possible decrease in risk or no effect in younger or recently menopausal women <sup>e</sup>	↑18% increased risk (n.s.)	6 excess cases (41 vs 35)	No increase in risk	No difference in risk
Myocardial infarction	Significant interaction by age group for estrogen alone, with reduced risk in younger—but not older—women ( <i>p</i> for trend by age = .02)	↑24% increased risk (n.s.)	6 excess cases (35 vs 29)	No increase in risk <sup>e</sup>	No difference in risk <sup>e</sup>
Stroke	Probable increase in risk	↑37% increased risk	9 excess cases (33 vs 24)	↑35% increased risk	11 excess cases (45 vs 34)
Ovarian cancer	Probable increase in risk with long-term use (5 years)	↑41% increased risk (n.s.)	1 excess case (5 vs 4)	Not available	Not available
Endometrial cancer	Probable decrease in risk with estrogen-progestin during long-term follow-up (see above for estrogen alone)	↓33% decreased risk <sup>f</sup>	3 fewer cases (7 vs 10)	See above	See above
Urinary incontinence	Probable increase in risk	↑49% increased risk	549 excess cases (1661 vs 1112)	↑61% increased risk	852 excess cases (2255 vs 1403)
Colorectal cancer	Probable decrease in risk with estrogen-progestin; possible increase in risk in older women with estrogen alone ( <i>p</i> for trend by age = .02 for estrogen alone)	↓38% decreased risk	6.5 fewer cases (10 vs 17)	No increase or decrease in risk <sup>e</sup>	No difference in risk <sup>e</sup>
Type 2 diabetes	Probable decrease in risk	↓19% decreased risk	16 fewer cases (72 vs 88)	↓14% decreased risk	21 fewer cases (134 vs 155)
Dementia (age ≥ 65)	Increase in risk in older women (but inconsistent data from observational studies and randomized trials)	↑101% increased risk	23 excess cases (46 vs 23)	↑47% increased risk (n.s.)	15 excess cases (44 vs 29)
Total mortality	Possible increase in risk among older women and women many years past menopause; possible decrease in risk or no effect in younger or recently menopausal women ( <i>p</i> for trend by age <.05 for both trials combined)	No increase in risk	No difference in risk	No increase in risk <sup>e</sup>	No difference in risk <sup>e</sup>
Global index <sup>g</sup>	Probable increase in risk or no effect among older women and women many years past menopause; possible decrease in risk or no effect in younger or recently menopausal women ( <i>p</i> for trend by age = .02 for estrogen alone)	↑12% increased risk	20.5 excess cases (189 vs 168)	No increase in risk <sup>e</sup>	No difference in risk <sup>e</sup>

<sup>a</sup>The estrogen-progestin arm of the WHI assessed 5.6 years of conjugated equine estrogens (0.625 mg/d) plus medroxyprogesterone acetate (2.5 mg/d) versus placebo. The estrogen-alone arm of the WHI assessed 7.1 years of conjugated equine estrogens (0.625 mg/d) versus placebo. <sup>b</sup>Number of cases per 10,000 women per year. <sup>c</sup>The WHI was not designed to assess the effect of hormone therapy (HT) on menopausal symptoms. Data from other randomized trials suggest that HT reduces risk for menopausal symptoms by 65–90%. <sup>d</sup>Coronary heart disease is defined as nonfatal myocardial infarction or coronary death. <sup>e</sup>There was a significant interaction by age; that is, the association between HT and the specified outcome was different in younger women and older women. <sup>f</sup>This is the risk reduction that was observed during a cumulative 13-year follow-up period (5.6 years of treatment plus 8.2 years of postintervention observation). <sup>g</sup>The global index is a composite outcome representing the first event for each participant from among the following: coronary heart disease, stroke, pulmonary embolism, breast cancer, colorectal cancer, endometrial cancer (estrogen-progestin arm only), hip fracture, and death. Because participants can experience more than one type of event, the global index cannot be derived by a simple summing of the component events. <sup>h</sup>Includes some outcomes where results were divergent between the estrogen-progestin arm and the estrogen-alone arm.

Abbreviation: n.s., not statistically significant.

Source: Data from JE Manson et al: Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. JAMA 310:1353, 2013.

levels by 10–15%. Administration of estrogen also favorably affects lipoprotein(a) levels, LDL oxidation, endothelial vascular function, fibrinogen, and plasminogen activator inhibitor 1. However, estrogen therapy has unfavorable effects on other biomarkers of cardiovascular risk: it boosts triglyceride levels; promotes coagulation via factor VII, prothrombin fragments 1 and 2, and fibrinopeptide A elevations; and raises levels of the inflammatory marker C-reactive protein.

Randomized trials of estrogen or combined estrogen-progestin in women with preexisting cardiovascular disease have not confirmed the benefits reported in observational studies. In HERS (a secondary-prevention trial designed to test the efficacy and safety of estrogen-progestin therapy with regard to clinical cardiovascular outcomes), the 4-year incidence of coronary death and nonfatal myocardial infarction was similar in the active-treatment and placebo groups, and a 50% increase in risk of coronary events was noted during the first year among participants assigned to the active-treatment group. Although it is possible that progestin may mitigate estrogen's benefits, the Estrogen Replacement and Atherosclerosis (ERA) trial indicated that angiographically determined progression of coronary atherosclerosis was unaffected by either opposed or unopposed estrogen treatment. Moreover, no cardiovascular benefit was found in the Papworth Hormone Replacement Therapy Atherosclerosis Study, a trial of transdermal estradiol with and without norethindrone; the Women's Estrogen for Stroke Trial (WEST), a trial of oral 17 $\alpha$ -estradiol; or the Estrogen in the Prevention of Reinfarction Trial (ESPRIT), a trial of oral estradiol valerate. Thus, in clinical trials, HT has not proved effective for the secondary prevention of cardiovascular disease in postmenopausal women.

Primary-prevention trials also suggest an early increase in cardiovascular risk and an absence of cardioprotection with postmenopausal HT. In the WHI, women assigned to 5.6 years of estrogen-progestin therapy were 18% more likely to develop CHD (defined in primary analyses as nonfatal myocardial infarction or coronary death) than those assigned to placebo, although this risk elevation was not statistically significant. However, during the trial's first year, there was a significant 80% increase in risk, which diminished in subsequent years ( $p$  for trend by time = .03). In the estrogen-only arm of the WHI, no overall effect on CHD was observed during the 7.1 years of the trial or in any specific year of follow-up. This pattern of results was similar to that for the outcome of total myocardial infarction.

However, a closer look at available data suggests that timing of initiation of HT may critically influence the association between such therapy and CHD. Estrogen may slow early stages of atherosclerosis but have adverse effects on advanced atherosclerotic lesions. It has been hypothesized that the prothrombotic and proinflammatory effects of estrogen manifest themselves predominantly among women with subclinical lesions who initiate HT well after the menopausal transition, whereas women with less arterial damage who start HT early in menopause may derive cardiovascular benefit because they have not yet developed advanced lesions. Data from experiments in nonhuman primates and from some recent randomized trials in humans support this concept. Conjugated estrogens had no effect on the extent of coronary artery plaque in cynomolgus monkeys assigned to receive estrogen alone or combined with progestin starting 2 years (~6 years in human terms) after oophorectomy and well after the establishment of atherosclerosis. However, administration of exogenous hormones immediately after oophorectomy, during the early stages of atherosclerosis, reduced the extent of plaque by 70%. In the Early versus Late Intervention Trial with Estradiol (ELITE), a 6-year trial among 643 healthy postmenopausal women that was designed to test whether effects of estrogen on the development and progression of atherosclerosis depend on age at initiation of therapy, oral 17 $\alpha$ -estradiol administered with or without vaginal micronized progesterone significantly slowed carotid atherosclerotic progression in women within 6 years of menopause onset (mean age, 55.4 years) but not in women >10 years past menopause onset (mean age, 65.4 years) ( $p$  for interaction = .007). On the other hand, in the Kronos Early Estrogen Prevention Study (KEEPS), a 4-year trial among 729 healthy postmenopausal women within 3 years of menopause onset at trial entry (mean age,

53 years), neither oral conjugated estrogens nor transdermal estradiol, administered with oral micronized progesterone, affected carotid atherosclerotic progression. However, the low prevalence of this endpoint in the overall study population may have curtailed power to detect a treatment difference.

Lending further credence to the timing hypothesis are results of subgroup analyses of data from observational studies and large clinical trials. For example, among women who entered the WHI trial with a relatively favorable cholesterol profile, estrogen with or without progestin led to a 40% lower risk of incident CHD. Among women who entered with a worse cholesterol profile, therapy resulted in a 73% higher risk ( $p$  for interaction = .02). The presence or absence of the metabolic syndrome (Chap. 408) also strongly influenced the relation between HT and incident CHD. Among women with the metabolic syndrome, HT more than doubled CHD risk, whereas no association was observed among women without the syndrome. Moreover, although there was no association between estrogen-only therapy and CHD in the WHI trial cohort as a whole, such therapy was associated with a CHD risk reduction of 40% among participants aged 50–59; in contrast, a risk reduction of only 5% was observed among those aged 60–69, and a risk increase of 9% was found among those aged 70–79 ( $p$  for trend by age = .08). For the outcome of total myocardial infarction, estrogen alone was associated with a borderline-significant 45% reduction and a nonsignificant 24% increase in risk among the youngest and oldest women, respectively ( $p$  for trend by age = .02). Estrogen was also associated with lower levels of coronary artery calcified plaque in the younger age group. Although age did not have a similar effect in the estrogen-progestin arm of the WHI, CHD risks increased with years since menopause ( $p$  for trend = .08), with a significantly elevated risk among women who were 20 years past menopause. For the outcome of total myocardial infarction, estrogen-progestin was associated with a 9% risk reduction among women <10 years past menopause as opposed to a 16% increase in risk among women 10–19 years past menopause and a twofold increase in risk among women >20 years past menopause ( $p$  for trend = .01). In the large observational Nurses' Health Study, women who chose to start HT within 4 years of menopause experienced a lower risk of CHD than did nonusers, whereas those who began therapy 10 years after menopause appeared to receive little coronary benefit. Observational studies include a high proportion of women who begin HT within 3–4 years of menopause, whereas clinical trials include a high proportion of women 12 years past menopause; this difference helps to reconcile some of the apparent discrepancies between the two types of studies.

For the outcome of stroke, WHI participants assigned to estrogen-progestin or estrogen alone were ~35% more likely to suffer a stroke than those assigned to placebo. Whether or not age at initiation of HT influences stroke risk is not well understood. In the WHI and the Nurses' Health Study, HT was associated with an excess risk of stroke in all age groups. Further research is needed on age, time since menopause, and other individual characteristics (including biomarkers) that predict increases or decreases in cardiovascular risk associated with exogenous HT. Furthermore, it remains uncertain whether different doses, formulations, or routes of administration of HT will produce different cardiovascular effects.

**COLORECTAL CANCER** Observational studies have suggested that HT reduces risks of colon and rectal cancer, although the estimated magnitudes of the relative benefits have ranged from 8 to 34% in various meta-analyses. In the WHI (the sole trial to examine the issue), estrogen-progestin was associated with a significant 38% reduction in colorectal cancer over a 5.6-year period, although no benefit was seen with 7 years of estrogen-only therapy. However, a modifying effect of age was observed, with a doubling of risk with HT in women aged 70–79 but no risk elevation in younger women ( $p$  for trend by age = .02).

**COGNITIVE DECLINE AND DEMENTIA** A meta-analysis of 10 case-control and two cohort studies suggested that postmenopausal HT is associated with a 34% decreased risk of dementia. Subsequent randomized trials (including the WHI), however, have failed to demonstrate any benefit of estrogen or estrogen-progestin therapy on the progression of

mild to moderate Alzheimer's disease and/or have indicated a potential adverse effect of HT on the incidence of dementia, at least in women 65 years of age. Among women randomized to HT (as opposed to placebo) at age 50–55 in the WHI, no effect on cognition was observed during the postintervention phase. Determining whether timing of initiation of HT influences cognitive outcomes will require further study.

**OVARIAN CANCER AND OTHER DISORDERS** On the basis of limited observational and randomized data, it has been hypothesized that HT increases the risk of ovarian cancer and reduces the risk of type 2 diabetes mellitus. Results from the WHI support these hypotheses. The WHI also found that HT use was associated with an increased risk of urinary incontinence and that estrogen-progestin was associated with increased rates of lung cancer mortality.

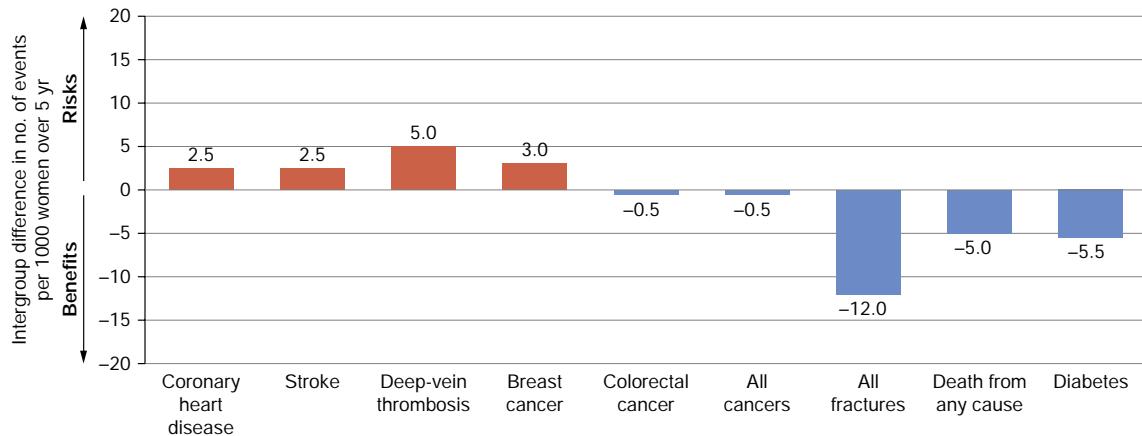
**ENDOMETRIAL CANCER (WITH ESTROGEN-PROGESTIN)** In the WHI, use of estrogen-progestin was associated with a nonsignificant 17% reduction in risk of endometrial cancer. A significant reduction in risk emerged during the postintervention period (see later).

**ALL-CAUSE MORTALITY** In the overall WHI cohort, estrogen with or without progestin was not associated with all-cause mortality. However, there was a trend toward reduced mortality in younger women, particularly with estrogen alone. For women aged 50–59, 60–69, and 70–79 years, relative risks (RRs) associated with estrogen-only therapy were 0.70, 1.01, and 1.21, respectively ( $p$  for trend = .04).

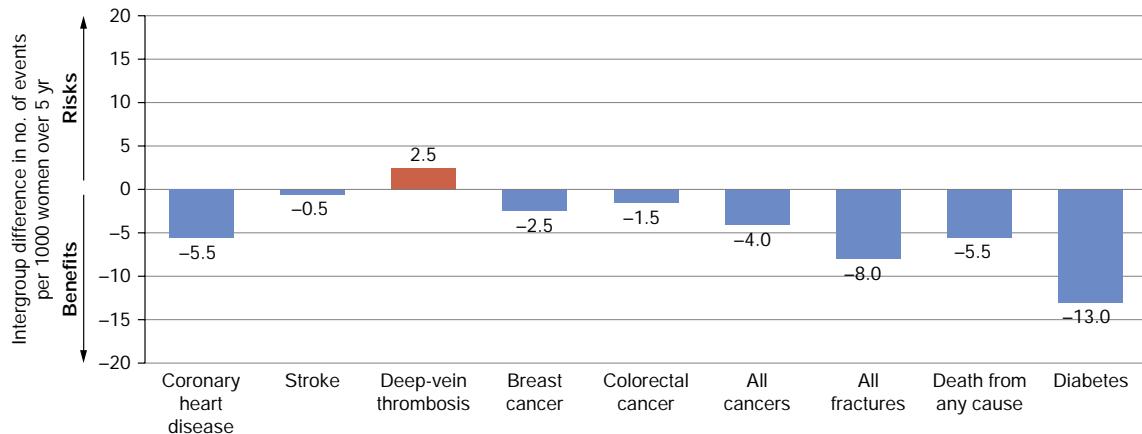
**OVERALL BENEFIT-RISK PROFILE** Estrogen-progestin was associated with an unfavorable benefit-risk profile (excluding relief from menopausal symptoms) as measured by a “global index”—a composite outcome including CHD, stroke, pulmonary embolism, breast cancer, colorectal cancer, endometrial cancer, hip fracture, and death (Table 395-1)—in the WHI cohort as a whole, and this association did not vary by 10-year age group. Estrogen-only therapy was associated with a neutral benefit-risk profile in the WHI cohort as a whole. However, there was a significant trend toward a more favorable benefit-risk profile among younger women and a less favorable profile among older women, with RRs of 0.84, 0.99, and 1.17 for women aged 50–59, 60–69, and 70–79 years, respectively ( $p$  for trend by age = .02). The balance of benefits and risks of estrogen with and without progestin among women aged 50–59 is shown in Fig. 395-3.

**CHANGES IN HEALTH STATUS AFTER DISCONTINUATION OF HT** In the WHI, many but not all risks and benefits associated with active use of HT dissipated within 5–7 years after discontinuation of therapy. For estrogen-progestin, an elevated risk of breast cancer persisted (RR = 1.28; 95% confidence interval [CI], 1.11–1.48) during a median cumulative 13-year follow-up period (5.6 years of treatment plus 8.2 years of postintervention observation), but most cardiovascular disease risks became neutral. A reduction in hip fracture risk persisted (RR = 0.81; 95% CI, 0.68–0.97), and a significant reduction in endometrial cancer risk emerged (RR = 0.67; 95% CI, 0.49–0.91). For estrogen alone,

A CEE+MPA Trial



B CEE+Alone Trial



**FIGURE 395-3** Benefits and risks of the two hormone therapy (HT) formulations evaluated in the Women's Health Initiative, in women aged 50–59 years. Results are shown for the two formulations, conjugated equine estrogens (CEE) alone or in combination with medroxyprogesterone acetate (MPA). Risks and benefits are expressed as the difference in number of events (number in the HT group minus the number in the placebo group) per 1000 women over 5 years. (Reproduced with permission from JE Manson, AM Kaunitz. Menopause Management—Getting Clinical Care Back on Track. *N Engl J Med* 374:803, 2016.)

the reduction in breast cancer risk became statistically significant ( $RR = 0.79$ ; 95% CI, 0.65–0.97) during a median cumulative 13-year follow-up period (6.8 years of treatment plus 6.6 years of postintervention observation), and significant differences by age group persisted for total myocardial infarction and the global index, with more favorable results for younger women. During a median cumulative 18-year follow-up, estrogen alone was associated with a significant reduction in all-cause mortality in women aged 50–59 years ( $RR = 0.79$ ; 95% CI, 0.64–0.96); the protective effect was seen primarily in those with bilateral oophorectomy ( $RR = 0.68$ ; 95% CI, 0.48–0.96).

## APPROACH TO THE PATIENT

### Postmenopausal HT

The rational use of postmenopausal HT requires balancing the potential benefits and risks. **Table 395-2** provides one approach to decision-making. This approach applies to women with menopausal symptoms who are age 45 years and older or to women who have had removal of both ovaries, regardless of age. Women below age 45 years or those with uncertain menopausal status may need additional clinical evaluation before determining a management plan. The clinician should first assess whether the patient has moderate to severe hot flashes and/or night sweats—the primary indication for initiation of systemic HT—that do not subside in response to behavioral/lifestyle modifications. (A patient handout with suggested lifestyle modifications can be found at <http://www.menopause.org/docs/for-women/mnflashes.pdf>.) Systemic HT may also be used to prevent osteoporosis in women at high risk of fracture who

cannot tolerate alternative osteoporosis therapies. (Vaginal estrogen or other medications may be used to treat genitourinary syndrome of menopause in the absence of vasomotor symptoms [see below].) The benefits and risks of such therapy should be reviewed with the patient, giving more emphasis to absolute than to relative measures of effect and pointing out uncertainties in clinical knowledge where relevant. Because chronic disease rates generally increase with age, absolute risks tend to be greater in older women, even when RRs remain similar. Potential side effects—especially vaginal bleeding that may result from use of the combined estrogen-progestogen formulations recommended for women with an intact uterus—should be noted. The patient's own preference regarding therapy should be elicited and factored into the decision. Contraindications should be assessed routinely and include unexplained vaginal bleeding; liver dysfunction or disease; venous thromboembolism; known blood clotting disorder or thrombophilia (transdermal estrogen may be an option); untreated hypertension; history of endometrial cancer (except stage 1 without deep invasion), breast cancer, or other estrogen-dependent cancer; and history of CHD, stroke, or transient ischemic attack. Relative contraindications to systemic HT include an elevated risk of breast cancer (e.g., women who have one or more first-degree relatives with breast cancer, susceptibility genes such as *BRCA1* or *BRCA2*, a personal history of cellular atypia detected by breast biopsy); hypertriglyceridemia (>400 mg/dL); an elevated risk of cardiovascular disease; and active gallbladder disease (transdermal estrogen may be an option in the latter three cases because it has a less adverse effect on triglyceride levels, clotting factors, and inflammation factors than oral HT). Primary prevention of heart disease should not be viewed as an expected benefit of HT, and an increase in the risk of stroke as well as a small early increase in the risk of coronary artery disease should be considered. Nevertheless, such therapy may be appropriate if the noncoronary benefits of treatment clearly outweigh the risks. Reassess benefits and risks at least once every 6–12 months, assuming the patient's continued preference for HT, or if the patient's health status changes. A woman who suffers an acute coronary event or stroke while taking HT should discontinue therapy immediately.

Many options for systemic HT are available. Estrogen alone is recommended for women with hysterectomy, whereas estrogen plus progestogen is recommended for women with a uterus. In the United States, the most commonly prescribed oral estrogens for systemic treatment of vasomotor symptoms are 17 $\alpha$ -estradiol (1.0 or 0.5 mg/d or other doses) and conjugated equine estrogens (CEE; 0.625, 0.45, or 0.3 mg/d or other doses). The most commonly prescribed transdermal estrogen products are 17 $\alpha$ -estradiol skin patches (0.035 or 0.05 mg/d or other doses). The most commonly prescribed progestogens are medroxyprogesterone acetate (MPA; 2.5, 5, or 10 mg/d) and micronized progesterone (100 or 200 mg/d). Also available are oral estrogen-progestin combinations, such as oral CEE and MPA, oral 17 $\alpha$ -estradiol or ethynodiol diacetate with norethindrone acetate, oral estradiol with progesterone, and other options. CEE/bazedoxifene may be an option for women with a uterus, especially those with concerns about breast tenderness, breast density, or uterine bleeding. Contraindications to CEE/bazedoxifene are similar to those for systemic HT.

*Short-term use* (<5 years for estrogen-progestogen and <7 years for estrogen alone) is appropriate for relief of menopausal symptoms among women without contraindications to such use. However, such therapy should be avoided by women with an elevated baseline risk of future cardiovascular events. Women who have contraindications for or are opposed to HT may derive benefit from the use of certain antidepressants (including venlafaxine, fluoxetine, or paroxetine), gabapentin or pregabalin, or clonidine.

*Long-term use* (>5 years for estrogen-progestogen and >7 years for estrogen alone) is more problematic because a heightened risk of breast cancer must be factored into the decision, especially for estrogen-progestogen. Reasonable candidates for such use include postmenopausal women who have persistent severe vasomotor symptoms

**TABLE 395-2 Approach to Initiating Menopausal Hormone Therapy for Vasomotor Symptom Management**

1. Vasomotor symptom assessment		
Confirm that hot flashes and/or night sweats are adversely affecting sleep, daytime functioning, or quality of life		
2. Risk factor assessment		
Confirm that there are no absolute contraindications to menopausal hormone therapy		
Breast, endometrial, or other estrogen-dependent cancer		
Cardiovascular disease (heart disease, stroke, transient ischemic attack)		
Active liver disease		
Undiagnosed vaginal bleeding		
3. Menopausal hormone therapy initiation		
RECOMMEND	CONSIDER WITH CAUTION	AVOID
Age <60 years and Menopause onset within 10 years and Low risk of breast cancer <sup>a</sup> and cardiovascular disease <sup>b</sup>	Age 60 years OR Menopause onset >10 years prior OR Moderate risk of breast cancer <sup>a</sup> or cardiovascular disease <sup>a</sup>	High risk of breast cancer <sup>a</sup> or cardiovascular disease <sup>b</sup> OR Age 60 years or menopause onset >10 years prior and Moderate risk of breast cancer <sup>a</sup> or cardiovascular disease <sup>b</sup>

<sup>a</sup>For online tools to assess breast cancer risk, see AH McClintock et al: Breast cancer risk assessment: A step-wise approach for primary care providers on the front lines of shared decision making. Mayo Clin Proc 95:1268, 2020. <sup>b</sup>For online tools to assess cardiovascular disease risk, see D Lloyd-Jones et al: Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: A special report from the American Heart Association and American College of Cardiology. Circulation 139:e1162, 2019.

Source: Data from AM Kaunitz, JE Manson: Management of menopausal symptoms. Obstet Gynecol 126:859, 2015 and Manson JE et al: Algorithm and mobile app for menopausal symptom management and hormonal/non-hormonal therapy decision making: A clinical decision-support tool from The North American Menopause Society. Menopause 22:247, 2015.

along with an increased risk of osteoporosis (e.g., those with osteopenia, a personal or family history of nontraumatic fracture, or a weight <125 lb), who also have no personal or family history of breast cancer in a first-degree relative or other contraindications, and who have a strong personal preference for therapy. Poor candidates are women with elevated cardiovascular risk, those at increased risk of breast cancer, and those at low risk of osteoporosis. Even for reasonable candidates, strategies to minimize dose and duration of use should be employed. For example, women using HT to relieve intense vasomotor symptoms in early postmenopause should consider discontinuing therapy within 5 years, resuming it only if such symptoms persist. Because of the role of progestogens in increasing breast cancer risk, regimens that employ cyclic rather than continuous progestogen exposure as well as formulations other than MPA should be considered if treatment is extended. For prevention of osteoporosis, alternative therapies such as bisphosphonates or SERMs should be considered. Research on alternative progestogens and androgen-containing preparations has been limited, particularly with respect to long-term safety. Additional research on the effects of these agents on cardiovascular disease, glucose tolerance, and breast cancer will be of particular interest.

For genitourinary symptoms such as vaginal dryness or pain with intercourse/sexual activity, intravaginal estrogen creams, tablets, or rings; prasterone (vaginal dehydroepiandrosterone); and ospemifene are options. Contraindications to low-dose vaginal estrogen include unexplained vaginal bleeding or breast cancer, endometrial cancer, or other estrogen-dependent cancer. Contraindications to ospemifene and prasterone are the same as those for low-dose vaginal estrogen, and contraindications for ospemifene additionally include venous or arterial thromboembolic disease, severe liver disease, and use of estrogens or estrogen agonists-antagonists.

In addition to HT, lifestyle choices such as smoking abstention, adequate physical activity, and a healthy diet can play a role in controlling symptoms and preventing chronic disease. An expanding array of pharmacologic options (e.g., bisphosphonates, SERMs, and other agents for osteoporosis; cholesterol-lowering or antihypertensive agents for cardiovascular disease) should also reduce the widespread reliance on hormone use. However, short-term HT may still benefit some women.

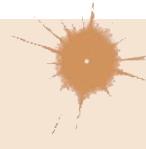
## FURTHER READING

- Bassuk SS, Manson JE: Menopausal hormone therapy and cardiovascular disease risk: Utility of biomarkers and clinical factors for risk stratification. *Clin Chem* 60:68, 2014.
- Canonico M et al: Hormone replacement therapy and risk of venous thromboembolism in postmenopausal women: Systematic review and meta-analysis. *BMJ* 336:1227, 2008.
- Kaunitz AM, Manson JE: Management of menopausal symptoms. *Obstet Gynecol* 126:859, 2015.
- Manson JE, Bassuk SS: *Hot Flashes, Hormones and Your Health*. New York, McGraw-Hill, 2007.
- Manson JE et al: Menopausal hormone therapy and health outcomes during the intervention and extended poststopping phases of the Women's Health Initiative randomized trials. *JAMA* 310:1353, 2013.
- Manson JE et al: The Women's Health Initiative trials of menopausal hormone therapy: Lessons learned. *Menopause* 27:918, 2020.
- North American Menopause Society: The 2017 hormone therapy position statement of the North American Menopause Society. *Menopause* 24:728, 2017.
- North American Menopause Society: The 2020 genitourinary syndrome of menopause position statement of the North American Menopause Society. *Menopause* 27:976, 2020.
- Pinkerton JV: Hormone therapy for postmenopausal women. *N Engl J Med* 382:446, 2020.
- Shifren JL et al: Menopausal hormone therapy. *JAMA* 321:2458, 2019.

396

## Infertility and Contraception

Anuja Dokras, Janet E. Hall



## **INFERTILITY**

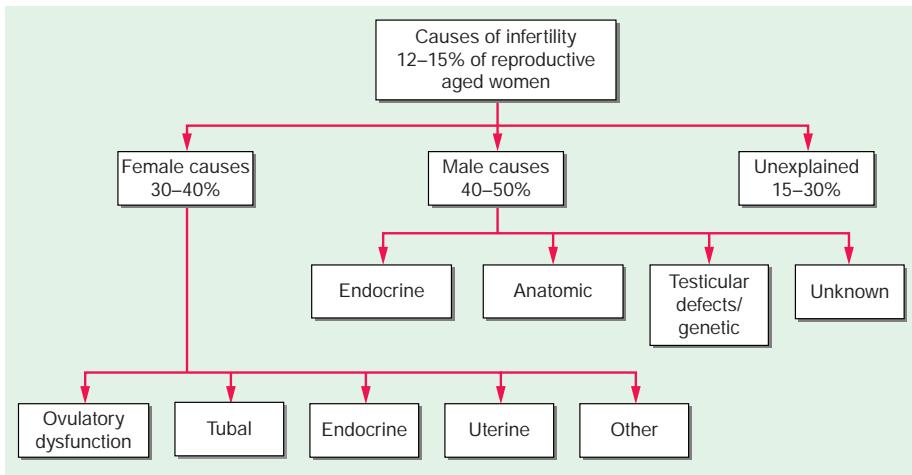
The World Health Organization (WHO) categorizes infertility as a disease of the reproductive system. Infertility is the third most common disease worldwide, affecting ~48 million couples. It is defined as the inability to achieve a pregnancy over 12 months of unprotected intercourse. The prevalence of infertility, ~15% globally, has remained relatively stable over the past few decades. Primary infertility occurs in couples who have never achieved a pregnancy, whereas secondary infertility refers to infertility after achieving at least one pregnancy. During the first year of attempting pregnancy, the fecundability rate, defined as the ability to achieve a pregnancy within one menstrual cycle, is highest in the first 3 months and declines over the next 9 months. Approximately 85% of couples will achieve pregnancy after 12 months, and 95% will achieve pregnancy after 24 months. Increasing trends toward later childbearing can have significant implications due to age-related decrease in the fecundability rate. Compared to women aged 30–31 years of age, fecundability is reduced by 14% in women aged 34–35 years, 19% in women aged 36–37 years, 53% in women aged 40–41 years, and 59% in women aged 42–44 years.

## **ETIOLOGY**

The causes for infertility are generally classified as female, male, and unexplained (Fig. 396-1). The female causes include tubal factors (pelvic inflammatory disease, salpingitis isthmica nodosum, endometriosis, prior surgery), uterine etiology (congenital malformations, fibroids, uterine scarring), ovulatory dysfunction (polycystic ovary syndrome [PCOS], diminished ovarian reserve, premature ovarian insufficiency), and endocrine dysfunction (hypothyroidism, hyperprolactinemia). Although the probability of achieving a pregnancy decreases after the age of 35 in women, primarily due to chromosomal abnormalities in the oocyte during meiosis, a similar decline has not been observed in men aged <50 years. The male causes of infertility include anatomic factors in the reproductive system (vasectomy, infection, absence of the vas), endocrine factors (hypogonadotropic hypogonadism, hypothyroidism, hyperprolactinemia, morbid obesity, medications), sexual dysfunction (erectile or ejaculatory dysfunction, decreased libido), and genetic factors contributing to primary testicular dysfunction, including defects in spermatogenesis (Klinefelter's syndrome, Y chromosome microdeletions). The distribution of these causes varies significantly in couples across the world. Overall, female factors are present in 30–40% of couples with infertility, male factors are present in 40–50%, and both male and female factors are identified in 20–30%. Unexplained infertility refers to the absence of any identified abnormality after completing the fertility workup and occurs in 10–15% of couples. As a result, a complete workup of both partners is recommended in all couples presenting with infertility.

## **FERTILITY EVALUATION**

Diagnostic evaluation for infertility is typically initiated after 1 year of unprotected intercourse because 80–85% of couples will achieve a pregnancy over this time period. Evaluation of the couple can be initiated even prior to meeting the definition of infertility, especially if they have risk factors for infertility. If the female partner's age is >35 years, it is recommended to initiate evaluation after 6 months of attempting pregnancy. If the age of the female partner is >40 years, it is recommended to start evaluating the couple immediately. The initial evaluation should include detailed medical history, laboratory testing, and preconception counseling for both partners. As multiple causes for infertility may be identified, it is best to perform the complete diagnostic evaluation prior to initiating treatment.



**FIGURE 396-1** Causes of infertility. FSH, follicle-stimulating hormone; LH, luteinizing hormone.

**History and Physical Exam** A detailed history obtained from both partners is essential to identify risk factors for infertility. In the female partner, gynecologic history (menstrual frequency, menorrhagia, dysmenorrhea, history of sexually transmitted infections, endometriosis), medical and endocrine history, exposure to pelvic radiation, abdominal or pelvic surgeries, tobacco and alcohol use, medication use including cytotoxic drugs, family history of early menopause, and prior history of pregnancy should be assessed. In addition, frequency of intercourse, timing of intercourse, use of methods to detect ovulation, and concerns regarding sexual dysfunction over the past several months should be ascertained. Physical exam in the female partner should include assessment of weight and blood pressure (BP), thyroid and breast exam, assessment for signs of hyperandrogenism, and pelvic exam to assess uterine size, adnexal masses, and factors that might impact intercourse. Similarly, a detailed history should be obtained in the male partner with specific questions regarding injuries and surgery in the male reproductive tract; mumps orchitis; exposure to pelvic radiation; use of androgens, cytotoxic drugs, and other medications; and fertility with any prior partner. The exam in the male partner should include body mass index (BMI), BP, and complete physical exam including testicular exam.

**Ultrasound** An abdominal and transvaginal pelvic ultrasound can assess uterine (myomas, adenomyosis, müllerian anomalies) and adnexal abnormalities (endometriosis, polycystic-appearing ovaries) and evaluate ovarian reserve (number of antral follicles in both ovaries).

**Ovulation Assessment** Women who have regular menstrual cycles between 25 and 35 days will typically have ovulatory cycles. Ovulation can be assessed by using ovulation detection strips at home to detect urinary luteinizing hormone (LH) or by measuring a serum progesterone level 7 days after ovulation. Basal body temperatures can also be used to confirm ovulation when a rise in temperature is noted in the luteal phase. However, basal body temperature measurements are less reliable than the above methods.

**Hysterosalpingogram** An hysterosalpingogram (HSG) is performed during the follicular phase to assess the patency of fallopian tubes by injecting radiopaque contrast through the cervix into the uterus and imaging the flow of contrast through one or both tubes. In addition to identifying tubal pathology, an HSG may identify intrauterine abnormalities such as polyps, submucosal myomas, and adhesions. Although the negative predictive value of HSG for assessing tubal patency is high, the positive predictive value is relatively low. Interestingly, pregnancy rates have been shown to be higher in women after an HSG test compared to those who did not have the test, likely related to tubal flushing. Alternate options that are increasingly used include injection of

agitated saline contrast through the cervix into the uterus. Tubal patency is assessed by demonstrating passage of agitated saline contrast through the tubes or accumulation in the cul de sac as visualized by ultrasonography. A saline infusion sonogram is more accurate in assessing intrauterine pathology such as polyps and intrauterine scarring compared to HSG and can be combined with ultrasound assessment of the pelvis.

**Ovarian Reserve Evaluation** Assessment of ovarian reserve includes measurement of serum FSH and estradiol on day 2 or 3 of the menstrual cycle and serum anti-müllerian hormone (AMH). These screening tests combined with age of the female partner and antral follicle counts measured by ultrasound can identify diminished ovarian reserve and provide information on the urgency to initiate treatment. AMH and antral follicle counts are also used to determine starting doses of gonadotropins for fertility treatments. These markers of ovarian reserve, however, do not predict the likelihood of pregnancy and live birth.

**Endocrine Tests** In women with irregular menses, serum TSH, prolactin, and androgens should be measured to identify other causes for anovulation.

**Semen Analysis (see Chap. 391)** The semen sample is collected after 2–7 days of abstinence and provides an assessment of sperm count, motility, morphology, volume, and pH. Although there is significant overlap between semen parameters of fertile and infertile men, those with abnormal sperm parameters based on the WHO criteria (*oligoasthenozoospermia* is defined as sperm counts <15 million/mL, motility <40%, and normal morphology <4%) should have a physical exam and further endocrine (serum follicle-stimulating hormone [FSH], LH, prolactin, and thyroid-stimulating hormone [TSH]) and genetic evaluation (karyotype and Y chromosome microdeletion).

**Genetic Screening** All couples can be offered preconception genetic screening based on ethnicity, family history, or common autosomal recessive conditions.

Of note, diagnostic laparoscopy, postcoital test, endometrial biopsy, thrombophilia, and immunologic testing and karyotype are not indicated as part of the initial workup of infertility.

## COUNSELING AND TREATMENT

**Preconception Counseling** All patients seeking fertility care should be provided with preconception counseling. This includes counseling about eating disorders or lifestyle modifications for weight management as obesity in women is associated with an increase in anovulatory cycles, miscarriage rates, and maternal and fetal complications in pregnancy. Obesity in men is associated with abnormal sperm parameters. Preconception counseling regarding smoking cessation is

important as there is evidence to suggest that smoking cessation can reverse the detrimental impact of smoking on fecundity. Smoking decreases fertility rates by a direct impact on oocyte DNA and also increases the risk of miscarriage and ectopic pregnancy. In addition, smoking during pregnancy is associated with an increased risk of placental abruption and intrauterine growth restriction (IUGR). Moreover, the impact of smoking on ovarian reserve has been shown to accelerate the time to menopause by 1–4 years. As high levels of caffeine consumption increase the risk of infertility and miscarriage rates, women should be counseled to restrict caffeine consumption to 2 cups while attempting pregnancy and during pregnancy. Use of testosterone products, which are widely used for the treatment of hypoandrogenism and sexual dysfunction in men, should be stopped. Inquiries should be made about possible misuse of androgens for physical appearance or performance enhancement (Chap. 399). As part of the preconception counseling, patients should be informed that the fertile window is typically 5–6 days prior to ovulation, and therefore, intercourse every 1–2 days during this time period will increase the chance of pregnancy. Various methods are used by women to detect ovulation, including basal body temperature measurements, assessment of changes in cervical mucus, and urinary LH kits. A rise in basal body temperatures indicates that ovulation has occurred and therefore cannot be used to time intercourse. LH kits can be used to detect the start of ovulation and subsequently time intercourse on the day of the LH surge and the following day.

**Treatment** Treatment recommendations depend on the results of the fertility evaluation described above (Table 396-1). The success of different treatments depends on several factors including age of the female partner, assessment of ovarian reserve, history of smoking, BMI, and race.

**Tubal Factor Infertility** Tubal factor infertility constitutes 30–35% of cases of female infertility, and a large majority are secondary to tubal obstruction resulting from STIs. In vitro fertilization (IVF) was first developed as a treatment for tubal factor infertility as it bypasses the fallopian tubes and allows fertilization of oocytes in the laboratory prior to transfer into the uterus. IVF offers the highest success rates for couples with tubal factor infertility. Tubal repair or reconstruction is not recommended in most cases associated with underlying tubal infections or hydrosalpinx due to both the low success rate in achieving tubal patency and increased risk of ectopic pregnancy. In fact, removal of hydrosalpinges by salpingectomy will improve pregnancy rates in subsequent IVF treatments. If a proximal tubal blockage is observed on HSG, radiographically guided cannulation of fallopian tubes can be attempted. In women with bilateral tubal ligation, the decision between microsurgical reanastomosis versus IVF will depend on a number of

factors including patient's age, ovarian reserve, number of children desired, partner's semen parameters, and experience of the surgeon.

**Ovulatory Dysfunction** Endocrine conditions such as hypothyroidism and hyperprolactinemia should be treated prior to use of ovulation induction medications. Lifestyle modifications should be recommended in patients with low BMI or obesity. Weight loss in obese women has been shown to increase the likelihood of spontaneous or drug-induced ovulation. First-line treatment for women with anovulatory infertility (most common etiology is PCOS) includes use of medications such as letrozole and clomiphene citrate to induce ovulation. A large majority of women with PCOS (60–80%) respond to these oral medications, and the addition of metformin, as a second-line agent, may further increase the chance of ovulation, particularly in obese women. In women with hypothalamic amenorrhea, behavioral modifications such as weight gain and decreased exercise may resume ovulation. If there is no response, judicious use of low-dose injectable gonadotropins can induce monofollicular growth. In women with diminished ovarian reserve, treatment can be escalated from ovulation induction with oral medications and intrauterine insemination (IUI) to IVF as the overall pregnancy rates are lower. In both women with diminished ovarian reserve and women with premature ovarian insufficiency, the option of using donor oocytes can be offered. In that case, the egg donor will undergo the IVF procedure, the harvested eggs are fertilized with the male partner's sperm, and the fertilized embryos will be transferred to the patient's uterus.

**Male Infertility** Given the high prevalence of male factor infertility (40–50%), timely evaluation and treatment are recommended. In men with no sperm (azoospermia) in the ejaculate, further evaluation including a physical examination, endocrine tests, and genetics studies should be performed to identify obstructive (40% prevalence among men with azoospermia) versus nonobstructive etiology. First-line treatment for mild to moderate male factor infertility includes IUI alone or IUI combined with ovulation induction, depending on the female partner's age and other causes of infertility. In men with severe male factor infertility, IVF with intracytoplasmic sperm injection is recommended. In men with obstructive azoospermia, sperm can be procured by direct aspiration from the epididymis or testis. In men with congenital bilateral absence of the vas deferens (CBAVD), testing for *CFTR* mutations and genetic counseling are indicated. In men with nonobstructive azoospermia, sperm retrieval from the testes may be less successful, and the use of donor sperm for IUI is an alternate option. Men with hypogonadotropic hypogonadism (e.g., Kallmann's syndrome) can be treated with gonadotropins to initiate spermatogenesis followed by IUI or IVF. Treatment of male sexual dysfunction and avoidance of exogenous androgens are effective strategies for addressing male factor infertility. Repair of a moderate to large varicocele is recommended when associated with abnormal semen parameters or if the patient is symptomatic from the varicocele.

**Unexplained Infertility** In 15–30% of couples, no clear causes of infertility are identified. In such cases, it is appropriate to initiate ovarian stimulation with oral medications to increase the number of developing oocytes and combine this with IUI timed to ovulation in order to increase the number of motile sperm in the reproductive tract. Depending on the age of the female partner, this approach offers modest success rates and can be used for 3–6 months before recommending IVF. Overall, IVF is associated with a low risk of complications; the risk of ovarian hyperstimulation syndrome has been significantly decreased by judiciously monitoring stimulation and using alternate protocols. Multiple pregnancy remains the highest risk associated with IVF despite improvements in cryopreservation of embryos and age-based guidelines for the number of embryos to transfer. In some couples, the IVF treatment may reveal an underlying cause of infertility such as lower fertilization or embryo cleavage rates. Of note, guidelines from different medical societies around the world vary in the rapidity of offering IVF for unexplained infertility.

**Uterine Factors** Fibroids are the most common benign tumors of the reproductive tract and occur in 50–70% of reproductive-age

**TABLE 396-1 Assisted Reproductive Technologies**

Ovulation induction	Clomiphene citrate (selective estrogen response modulator)
Oral agents	Letrozole (aromatase inhibitor)
Injectable hormones	FSH, LH (gonadotropins)
Intrauterine insemination (IUI)	Office-based procedure by which washed and concentrated ejaculated sperm is deposited in the uterine cavity via a soft catheter passed through the cervix
In vitro fertilization (IVF)	Oocytes are harvested transvaginally under local anesthesia or intravenous sedation and incubated with sperm to facilitate fertilization. The fertilized embryos are cultured for 3 days (cleavage stage) or 5 days (blastocyst stage) prior to transcervical placement of one or more embryos, depending on the age of the female patient, into the uterine cavity under ultrasound guidance.
Intracytoplasmic sperm injection (ICSI)	In cases of severe male factor infertility, a single motile sperm is injected into the oocyte in order to facilitate fertilization.

Abbreviations: FSH, follicle-stimulating hormone; LH, luteinizing hormone.

women. It is not clear whether fibroids decrease the likelihood of pregnancy; submucosal fibroids and intramural fibroids that distort the endometrial cavity may lower pregnancy rates and increase the risk of pregnancy loss. Removal of submucosal fibroids, uterine polyps, and adhesions hysteroscopically may improve subsequent pregnancy rates.

**Endometriosis** Endometriosis is a common gynecologic condition associated with pelvic pain and dysmenorrhea, and in severe cases, it is associated with tubo-ovarian infertility. Approximately 25–50% of infertile women have endometriosis, and 30–50% of women with endometriosis have infertility. Prolonged medical management to suppress endometriotic lesions and surgical treatment of stage 1 and 2 endometriosis have not been shown to improve subsequent fertility rates. Surgical removal of endometriotic lesions in women with stage 3 or 4 endometriosis may improve subsequent pregnancy rates. First-line treatment of infertility associated with endometriosis alone includes use of oral ovulation induction medications and IUI.

### PSYCHOLOGICAL ASPECTS OF INFERTILITY

It is well recognized that infertility is associated with psychological stress related not only to the diagnostic and therapeutic procedures themselves but also to repeated cycles of hope and loss associated with each new procedure or cycle of treatment that does not result in the birth of a child. These feelings are often combined with a sense of isolation from friends and family. Counseling and stress-management techniques should be offered early in the evaluation of infertility. Importantly, infertility and its treatment do not appear to be associated with long-term psychological sequelae.

### CONTRACEPTION

The desired ideal number of children per family varies around the globe and is approximately 2.6 in the United States. Couples not using any form of contraception have an 85% chance of achieving a pregnancy over 1 year. Based on these data, couples spend most of their reproductive life preventing a pregnancy and a much smaller proportion attempting to become or being pregnant. It is therefore not surprising that a majority of women who have been sexually active will have used some form of contraception to prevent a pregnancy. Unintended pregnancies primarily occur due to lack of use or inconsistent use of contraceptives rather than failure of the contraceptive method used. Of the different forms of contraception used worldwide in 2019, tubal sterilization was the most common (~219 million) followed by use of male condom (189 million), intrauterine device (IUD) (159 million), and the (birth control) pill (151 million). The rates of female sterilization increased steadily in the last century and now show a slight decrease, likely due to the increasing use of long-acting reversible contraceptive (LARC) agents, such as IUDs and implants, which are as effective as sterilization. The convenience of use of contraceptives determines their compliance and efficacy; contraceptives requiring daily and coitus-related use have higher failure rates compared to long-acting reversible and permanent methods. The U.S. Medical Eligibility Criteria (USMEC) for contraceptive use are evidence-based guidelines to help health care providers recommend appropriate contraceptives to women with chronic medical conditions (**Table 396-2**). This excellent resource is adapted from the WHO guidance and is kept up to date through continual review of published literature.

### TYPES OF CONTRACEPTION

These can be classified in a number of ways, such as permanent versus reversible, hormonal versus nonhormonal, or barrier versus nonbarrier (**Table 396-3**).

**Permanent Contraception** The permanent forms of contraception include tubal sterilization and vasectomy, with twice as many women choosing permanent sterilization compared to men. Vasectomy is a low-risk procedure typically performed in an outpatient setting with a very low failure rate of 0.1 pregnancies per 100 women per year. It is not immediately effective, and patients should be told to use other forms of contraception for a minimum of 3 months after the procedure. Tubal sterilization can be performed in the

**TABLE 396-2 U.S. Medical Eligibility Criteria (USMEC) for Contraceptive Use**

**USMEC Category 4 (a condition that represents an unacceptable health risk if the contraceptive method is used)**

- Women age >35 years who smoke 15 cigarettes per day
- Known ischemic heart disease or multiple risk factors for cardiovascular disease (older age, smoking, diabetes, and hypertension)
- Acute DVT
- Previous thromboembolic event: high risk of recurrent DVT
- Stroke or known thrombogenic mutations
- Complicated valvular heart disease
- Peripartum cardiomyopathy
- Complicated solid organ transplantation
- Hypertension (systolic 160 mmHg or diastolic 100 mmHg, vascular disease)
- Systemic lupus erythematosus (positive or unknown antiphospholipid antibodies)
- Cirrhosis, hepatic adenoma or hepatoma
- Viral hepatitis, acute flare
- Pregnancy and early postpartum (<21 days)
- Breast-feeding <21 days postpartum
- Breast cancer

**USMEC Category 3 (a condition for which the theoretical or proven risks outweigh the advantages for using the method)**

- Previous thromboembolic event: lower risk of recurrent DVT
- Past history of breast cancer and no evidence for 5 years
- Hypertension (adequately controlled or systolic 140–159 mmHg or diastolic 90–99 mmHg)
- Women receiving anticonvulsant drug therapy
- Women receiving antiretroviral therapy for prevention or treatment of HIV
- Women following bariatric surgery (Roux-en-Y gastric bypass or biliopancreatic diversion)
- Breast-feeding 21–42 days postpartum

*Abbreviation:* DVT, deep-vein thrombosis.

postpartum period or as an interval procedure and has a failure rate of 0.5 pregnancies per 100 women per year. Postpartum sterilization can be performed during a cesarean section or after a vaginal delivery via mini-laparotomy. Interval procedures can be performed laparoscopically or via mini-laparotomy and include partial or complete salpingectomy or occlusion of the fallopian tubes using electrocoagulation or mechanical devices such as clips. These permanent methods of contraception are highly effective as they avoid the need for user-dependent contraception. All patients should undergo preprocedure counseling regarding risk of failure, permanence of the procedure, regret, and alternatives.

**Hormonal Contraceptives • COMBINED ESTROGEN- AND PROGESTIN-CONTAINING CONTRACEPTIVES** The mechanism of action of the hormonal contraceptives involves negative feedback from continuous estrogen administration, thereby decreasing FSH secretion, follicular development, and formation of a dominant follicle. The continuous progestin suppresses LH secretion and inhibits ovulation, alters endometrial receptivity, thickens the cervical mucus, and impairs tubal motility. These hormones can be delivered via oral pills to be taken daily, as a transdermal patch that is changed weekly, or a vaginal ring that is replaced monthly or annually. There are numerous *pills* available containing different doses of estrogen (<50 µg) and types of progestins and varying doses within a pack (monophasic vs multiphasic); the pills can be taken in a cyclic or extended cycle schedule. The contraceptive efficacy is similar with varying doses of estrogen and progestin. Decreasing the duration of hormone-free days may decrease some side effects associated with menses, such as menstrual migraines and dysmenorrhea. The overall failure rate for combined hormonal contraceptives is 8 pregnancies per 100 women per year, although compliance with daily use of pills may be lower, affecting efficacy. The contraceptive patch and vaginal ring have higher compliance compared to daily pills. Use of the *contraceptive patch* is associated with a low risk

TABLE 396-3 Effectiveness of Different Forms of Contraception

METHOD OF CONTRACEPTION	THEORETICAL EFFECTIVENESS (%)	ACTUAL EFFECTIVENESS (%)	CONTINUED USE AT 1 YEAR (%)	USE OF METHOD BY U.S. WOMEN AT RISK OF UNINTENDED PREGNANCY (%)
No method	15	15		10
Fertility awareness	96	76	47	1.2
Withdrawal	96	78	46	4.4
Barrier methods				
Condoms	98	82	43	13.7
Diaphragm	94	82	57	2
Spermicides	82	72	43	1
Sterilization				
Female	99.5	99.5	100	22.6
Male	99.5	99.9	100	7.4
Intrauterine device				9.3
Copper T	99.4	99.8	85	
Progestin-containing	99.8	99.8	88	
Hormonal contraceptives				
Combined and progestin only	99.7	91	67	23.3
Transdermal patch	99.7	91	67	0.5
Vaginal ring	99.7	91	67	1.8
Implant				1.2
Depo-Provera	99.8	94	56	
Subdermal implant	99.5	99.5	84	
Emergency contraception	95	-	-	11

Sources: Data from J Trussell et al: Contraceptive Efficacy, in Contraceptive Technology, 20th revised ed, RA Hatcher et al (eds). New York, Ardent Media, 2011; CDC. NCHS National Survey of Family Growth, 2011–2013; J Jones et al: Current contraceptive use in the United States, 2006–2010, and changes in patterns of use since 1995. Natl Health Stat Report 60:1, 2012, and NE Birgisson et al: Preventing unintended pregnancy: The contraceptive CHOICE project in review. J Womens Health (Larchmt) 24:349, 2015.

of skin reactions and a lower efficacy in women weighing >90 kg. The transdermal mode of delivery is associated with a higher steady state comparable to that of a 40-µg ethinyl estradiol oral contraceptive. Hormonal contraceptives offer additional benefits such as regulation of menstrual cycles; suppression of ovarian cysts; and decrease in menorrhagia, dysmenorrhea, and hyperandrogenism symptoms; in addition, they reduce the risk of both endometrial (50% reduction) and ovarian cancer (40% reduction). Common side effects include nausea, breast tenderness, bloating, and intermenstrual bleeding. There may be a mild increase in BP in some patients, and it is recommended to check BP at follow-up visits. In large studies and meta-analyses, hormonal contraceptives are not associated with significant weight gain, mood changes, or effect on libido. Prior to administering hormonal contraceptives, a detailed patient history should be obtained to determine any absolute or relative contraindications to their use. Due to the low but slightly increased risk of deep-vein thrombosis (DVT) associated with estrogen-containing hormonal contraceptives (3–15 per 10,000 women-years), they are contraindicated in the immediate postpartum period, in smokers over the age of 35 years, and in women with a history of hereditary thrombophilias or DVT. The association between risk of DVT and different doses of estrogen (ethinyl estradiol <35 µg) or different routes of administration (transdermal patch) is weak. There is, however, some association between third- and fourth-generation progestins and risk of DVT. Routine screening for familial thrombotic disorders is not recommended prior to prescribing hormonal contraceptives. Although obesity is associated with decreased fertility, the vast majority of women with obesity do not experience infertility. The USMEC classifies obesity alone as risk category 2, where the benefits of taking hormonal contraceptives outweigh any theoretical risk.

**PROGESTIN-ONLY HORMONAL CONTRACEPTION** Different types of progestins are used for contraception in oral pills, injectable forms, subdermal implants, and IUDs and may be an option for women who have contraindications to the use of estrogen-containing contraceptives (e.g., migraine with aura, DVT, stroke, breast-feeding). The failure rate with *progestin-only pills* is 9 pregnancies per 100 women per year, whereas the failure rate of progestin IUDs is 0.1 pregnancies per 100

women per year. In addition to acting as a spermicidal, the *levonorgestrel IUD* also thickens the cervical mucus and thins the endometrium, thereby decreasing its receptivity. The common side effect is irregular bleeding, pain, and rarely expulsion. Breakthrough bleeding or unscheduled bleeding is commonly reported, as estrogen usually serves to stabilize the endometrial lining and prolonged exposure to progestin alone results in a thinner decidualized lining. Depending on the device used, the progestin IUD is effective for 3–7 years. The *injectable form of progesterone* (medroxyprogesterone acetate) is administered every 3 months with a failure rate of 3 pregnancies per 100 women per year. Its side effects include weight gain, irregular menses, amenorrhea, and mood changes, and there is a slow return to ovulation and fertility after discontinuation (6–9 months). The *subdermal implant* contains etonogestrel and is placed easily over the triceps muscle in the inner arm using local anesthesia. It lasts up to 5 years and has a failure rate of 0.05 pregnancies per 100 women per year. Findings from the Contraceptive Choice research project showed that continuation rates were higher for LARC (IUDs and implants) compared to short-acting methods. LARCs are the most effective reversible form of contraception with high continuation and satisfaction rates; hence, they are a good choice in adolescents and nulliparous women.

**Nonhormonal IUD** IUDs are a commonly used form of contraception worldwide and are available as hormonal and nonhormonal devices. The nonhormonal copper IUD works as a spermicidal and is effective for up to 12 years with a failure rate of <1 pregnancy per 100 women per year. Patients should be counseled regarding the increased risk of heavy vaginal bleeding and dysmenorrhea resulting in higher discontinuation rates compared to the levonorgestrel-containing IUDs. IUDs can be used in adolescents and adult women and are typically inserted and removed as an office procedure with use of mild analgesics. They can be inserted anytime during a menstrual cycle, referred to as interval insertion, and in the immediate postpartum and postabortion period.

**Barrier Contraception** The barrier forms of contraception include condoms (male, female) and diaphragm and cervical cap and

have lower effectiveness secondary to inconsistent and incorrect use. They offer several advantages including minimal side effects, lower cost, no requirement for a prescription, and protection from sexually transmitted infections. The failure rate for male and female condoms is 17–21 pregnancies per 100 women per year. Spermicides can be used in conjunction with barrier methods to improve effectiveness.

**Lactational Contraception** Lactation may serve as an effective form of contraception during the first 6 postpartum months if there is exclusive breast-feeding and menstrual cycles have not resumed. The contraceptive effect occurs due to suppression of gonadotropin-releasing hormone pulsatility associated with suckling. The failure rate under these circumstances can be as low as 0.5–1.5 pregnancies per 100 women per year.

**Fertility Awareness** The standard days method is typically used by women with regular menstrual cycles whereby they track their cycles to avoid intercourse from cycle days 8–19.

**Emergency Contraception** Also known as postcoital contraception, this method is used after an unprotected or inadequately protected act of intercourse. The probability of pregnancy independent of the time of the month is 8%, but the probability varies significantly in relation to proximity to ovulation and may be as high as 30%. Many women are not aware of the availability of emergency contraception and its appropriate use. As the probability of pregnancy is highest if there has been unprotected intercourse during the 3 days prior to ovulation, the timing of administration and type of emergency contraceptive used determine the efficacy. The emergency contraception options include the copper IUD and oral medications such as ulipristal acetate, levonorgestrel, and combined hormonal pills. The copper IUD prevents fertilization and implantation and is the most effective choice if inserted within 5 days of unprotected intercourse. It can also be offered to obese women in whom other hormonal forms of emergency contraceptive may be less effective. Ulipristal acetate, a progesterone receptor antagonist, blocks the ability of endogenous progesterone to act on its receptors and inhibits the LH surge, delaying or inhibiting ovulation, and may directly inhibit follicular rupture. It is administered as a 30-mg single dose up to 5 days after unprotected intercourse. Levonorgestrel administered as a single dose will prevent or delay ovulation and is associated with fewer side effects compared to combined hormonal pills. Overall, the failure rate for all hormonal emergency contraception is 1–3%, with ulipristal acetate being the most effective. Emergency contraception should be offered to all women who ask for it up to 5 days after unprotected intercourse and not delayed in order to obtain a pregnancy test or perform a clinical examination. Although body weight can affect the efficacy of emergency hormonal contraception, it should not be withheld from overweight and obese women.

### CONTRACEPTION COUNSELING

Patients should be provided information regarding the different methods of contraception, side effects, noncontraceptive benefits, efficacy, need for strict compliance, and impact on future fertility. In order to facilitate patient-centric care, the provider should discuss plans for future pregnancy and whether childbearing is complete. A detailed patient history should be reviewed to identify potential contraindications such as migraines with aura, smoking, and hypertension. Providers should refer to the most updated USMEC or WHO Medical Eligibility Criteria for Contraceptive Use guidelines when counseling patients with associated comorbidities. As part of the shared decision-making approach, the patient's choice should be the guiding factor, and the discussion should be nonjudgmental. Adolescents should be offered access to the full range of contraceptive options. In a low-risk patient, hormonal contraceptives can be prescribed from menarche to menopause; regular evaluation of side effects and assessment of changes in the patient's medical history, however, are required.

### FURTHER READING

Centers for Disease Control and Prevention: Reproductive health. Available at <https://www.cdc.gov/reproductivehealth/Infertility/#e>. Accessed December 23, 2020.

Cooper TG et al: World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 16:231, 2010.

Curtis KM, Peipert JF: Long-acting reversible contraception. *N Engl J Med* 376:461, 2017.

Curtis KM et al: U.S. medical eligibility criteria for contraceptive use, 2016. *MMWR Recomm Rep* 65(3):1, 2016.

Infertility Workup for the Women's Health Specialist: ACOG Committee Opinion, Number 781. *Obstet Gynecol* 133:e377, 2019.

Kulkarni AD et al: Fertility treatments and multiple births in the United States. *N Engl J Med* 369:2218, 2013.

Mascarenhas MN et al: National, regional, and global trends in infertility prevalence since 1990: A systematic analysis of 277 health surveys. *PLoS Med* 9:e1001356, 2012.

Slama R et al: Estimation of the frequency of involuntary infertility on a nation-wide basis. *Hum Reprod* 27:1489, 2012.

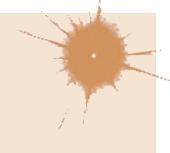
Steiner AZ et al: Association between biomarkers of ovarian reserve and infertility among older women of reproductive age. *JAMA* 318:1367, 2017.

World Health Organization: Infertility. <https://www.who.int/news-room/fact-sheets/detail/infertility>. Accessed December 23, 2020.

397

## Sexual Dysfunction

Kevin T. McVary



Male sexual dysfunction affects up to 31% of middle-aged and elderly men, whereas female sexual dysfunction, although studied less intensely, has a higher prevalence (43%) than male sexual dysfunction. Demographic changes, the popularity of newer treatments, and greater awareness of sexual dysfunction by patients and society have led to increased diagnosis and associated health care expenditures for the management of this common disorder. Sexual health and satisfaction with sex life are important aspects of quality of life for many, including those in poor health. Because many patients are reluctant to initiate discussion of their sex lives, physicians should address this topic directly to elicit a history of sexual dysfunction. Specifically addressing sexual health should be a routine part of the clinical encounter.

### MALE SEXUAL DYSFUNCTION

#### PHYSIOLOGY OF MALE SEXUAL RESPONSE

Normal male sexual function includes (1) sufficient libido, (2) the ability to achieve and maintain penile erection, (3) ejaculation, and (4) detumescence. *Libido* refers to sexual desire and is influenced by a variety of visual, olfactory, tactile, auditory, imaginative, and hormonal stimuli. Sex steroids, particularly testosterone, act to increase libido. Libido can be diminished by emotional context, systemic illness, hormonal disturbances, psychiatric disorders, and medications.

Penile tumescence leading to erection depends on an increased flow of blood into the lacunar network accompanied by complete relaxation of the arteries and corporal smooth muscle. The microarchitecture of the corpora is composed of a mass of smooth muscle (trabecula) that contains a network of endothelial-lined vessels (lacunar spaces). Subsequent compression of the trabecular smooth muscle against the fibroelastic tunica albuginea causes a passive closure of the emissary veins and accumulation of blood in the corpora. In the presence of a full erection and a competent valve mechanism, the corpora become non-compressible cylinders from which blood does not escape. This cascade of relaxation and venous occlusion culminates in a rigid erection.

The central nervous system (CNS) exerts an important influence by either stimulating or antagonizing spinal pathways that mediate erectile function and ejaculation. The erectile response is mediated

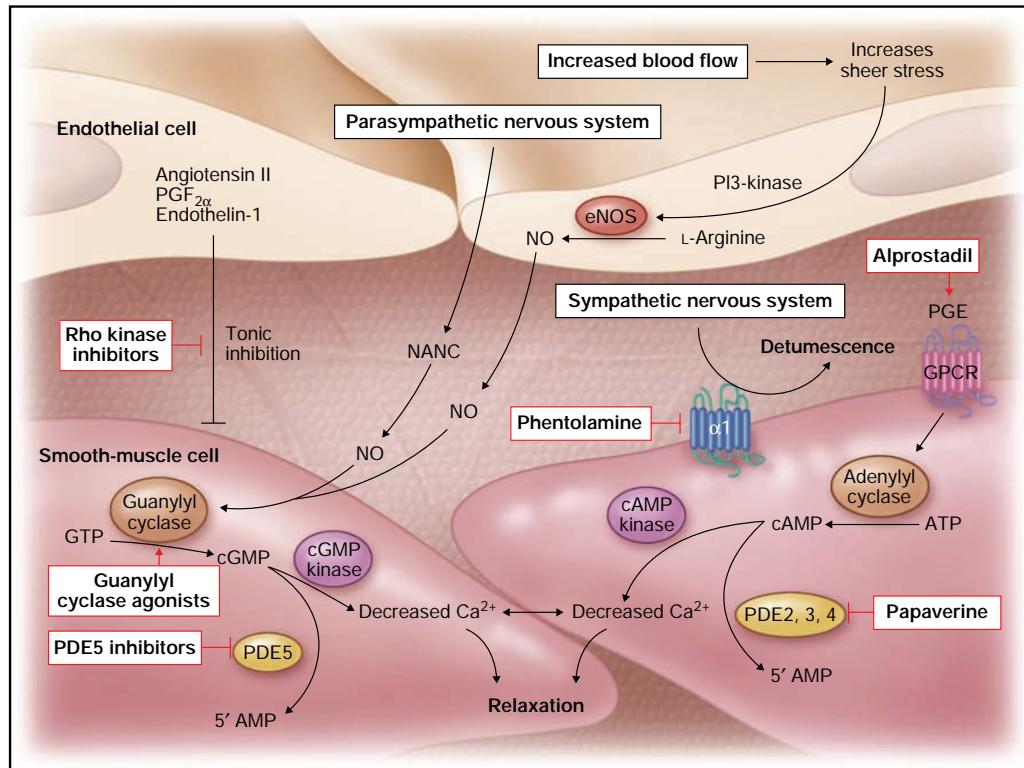
**3056** by a combination of central (psychogenic) innervation and peripheral (reflexogenic) innervation. Sensory nerves that originate from receptors in the penile skin and glans converge to form the dorsal nerve of the penis, which travels to the S2-S4 dorsal root ganglia via the pudendal nerve. Parasympathetic nerve fibers to the penis arise from neurons in the intermediolateral columns of the S2-S4 sacral spinal segments. Sympathetic innervation originates from the T-11 to the L-2 spinal segments and descends through the hypogastric plexus.

Neural input to smooth-muscle tone is crucial to the initiation and maintenance of an erection. There is also an intricate interaction between the corporal smooth-muscle cell and its overlying endothelial cell lining (Fig. 397-1). Nitric oxide, which induces vascular relaxation, promotes erection and is opposed by endothelin 1 (ET-1) and Rho kinase, which mediate vascular contraction. Nitric oxide is synthesized from L-arginine by nitric oxide synthase (NOS) and is released from the nonadrenergic, noncholinergic (NANC) autonomic nerve supply to act postjunctionally on smooth-muscle cells. Nitric oxide increases the production of cyclic 3',5'-guanosine monophosphate (cyclic GMP), which induces relaxation of smooth muscle (Fig. 397-2). Cyclic GMP is metabolized by phosphodiesterase type 5 (PDE-5). Inhibitors of PDE-5 such as the oral medications sildenafil, tadalafil, vardenafil, and avanafil maintain erections by reducing the breakdown of cyclic

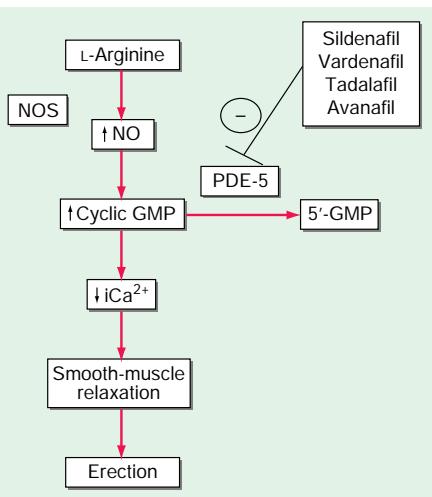
GMP. However, if nitric oxide is not produced at some level, PDE-5 inhibitors are ineffective, as these drugs facilitate, but do not initiate, the initial enzyme cascade. In addition to nitric oxide, vasoactive prostaglandins (PGE, PGF<sub>2α</sub>) are synthesized within the cavernosal tissue and increase cyclic AMP levels, also leading to relaxation of cavernosal smooth-muscle cells.

*Ejaculation* is stimulated by the sympathetic nervous system; this results in contraction of the epididymis, vas deferens, seminal vesicles, and prostate, causing seminal fluid to enter the urethra. Seminal fluid emission is followed by rhythmic contractions of the bulbocavernosus and ischiocavernosus muscles, leading to ejaculation. This is followed by expulsion, characterized by stereotypic rhythmic contractions of the striated perineal muscles, leading to forceful expulsion of semen with the bladder neck closed. This emission and expulsion are controlled by the autonomic (parasympathetic and sympathetic) and somatic spinal centers, respectively. The synchronization between autonomic and somatic spinal centers is orchestrated by interneurons that form a spinal ejaculation generator that is present in mammals including man.

*Premature ejaculation* usually is related to anxiety or a learned behavior and is amenable to behavioral therapy or treatment with medications such as selective serotonin reuptake inhibitors (SSRIs).



**FIGURE 397-1** Pathways that control erection and detumescence. Outflow from the parasympathetic nervous system leads to relaxation of the cavernous sinusoids in two ways, both of which increase the concentration of nitric oxide (NO) in smooth-muscle cells. First, NO is the neurotransmitter in nonadrenergic, noncholinergic (NANC) fibers; second, stimulation of endothelial nitric oxide synthase (eNOS) through cholinergic output causes increased production of NO. The NO produced in the endothelium then diffuses into the smooth-muscle cells and decreases its intracellular calcium concentration through a pathway mediated by cyclic guanosine monophosphate (cGMP), leading to relaxation. A separate mechanism that decreases the intracellular calcium level is mediated by cyclic adenosine monophosphate (cAMP). With increased cavernosal blood flow, as well as increased levels of vascular endothelial growth factor (VEGF), the endothelial release of NO is further sustained through the phosphatidylinositol 3 (PI3) kinase pathway. Active treatments (red boxes) include drugs that affect the cGMP pathway (phosphodiesterase type 5 [PDE-5] inhibitors and guanylyl cyclase agonists), the cAMP pathway (alprostadil), or both pathways (papaverine), along with neural-tone mediators (phentolamine and Rho kinase inhibitors). Agents that are being developed include guanylyl cyclase agonists (to bypass the need for endogenous NO) and Rho kinase inhibitors (to inhibit tonic contraction of smooth-muscle cells mediated through endothelin).  $\alpha 1$ ,  $\alpha$ -adrenergic receptor; GPCR, G protein-coupled receptor; GTP, guanosine triphosphate;  $i\text{Ca}^{2+}$ , intracellular calcium; NOS, nitric oxide synthase; PGE, prostaglandin E; PGF, prostaglandin F. (Reproduced with permission from KT McVary: Clinical practice. Erectile dysfunction. *N Engl J Med* 357:2472, 2007.)



**FIGURE 397-2** Biochemical pathways modified by phosphodiesterase type 5 (PDE-5) inhibitors. Sildenafil, vardenafil, tadalafil, and avanafil enhance erectile function by inhibiting PDE-5, thereby maintaining high levels of cyclic 3',5'-guanosine monophosphate (cyclic GMP). iCa<sup>2+</sup>, intracellular calcium; NO, nitric oxide; NOS, nitric oxide synthase.

**Retrograde ejaculation (RE)** results when the internal urethral sphincter does not close; it may occur in men with diabetes or after surgery involving the bladder neck. **Anejaculation**, the failure of a portion or the whole of the emission process often confused with RE, is commonly the result of selective alpha blockers used in male voiding dysfunction (e.g., tamsulosin, silodosin).

**Detumescence** is mediated by norepinephrine from the sympathetic nerves, endothelin from the vascular surface, and smooth-muscle contraction induced by postsynaptic  $\alpha$ -adrenergic receptors and activation of Rho kinase. These events increase venous outflow and restore the flaccid state. Venous leak can cause premature detumescence and is caused by insufficient relaxation of the corporal smooth muscle rather than a specific anatomic defect. **Priapism** refers to a persistent and painful erection and may be associated with sickle cell anemia, hypercoagulable states, spinal cord injury, or injection of vasodilator agents into the penis.

## ERECTILE DYSFUNCTION

**Epidemiology** Erectile dysfunction (ED) is not considered a normal part of the aging process. Nonetheless, it is associated with certain physiologic and psychological changes related to age. In the Massachusetts Male Aging Study (MMAS), a community-based survey of men aged 40–70, 52% of responders reported some degree of ED. Complete ED occurred in 10% of respondents, moderate ED in 25%, and minimal ED in 17%. The incidence of moderate or severe ED more than doubled between the ages of 40 and 70. In the National Health and Social Life Survey (NHSLS), which included a sample of men and women aged 18–59, 10% of men reported being unable to maintain an erection (corresponding to the proportion of men in the MMAS reporting severe ED). Incidence was highest among men in the age group 50–59 (21%) and men who were poor (14%), divorced (14%), and less educated (13%).

The incidence of ED is also higher among men with certain medical disorders, such as diabetes mellitus, obesity, lower urinary tract symptoms secondary to benign prostatic hyperplasia (LUTS/BPH), heart disease, hypertension, decreased high-density lipoprotein (HDL) levels, and diseases associated with general systemic inflammation (e.g., rheumatoid arthritis). Cardiovascular disease and ED share etiologies as well as pathophysiology (e.g., endothelial dysfunction), and the degree of ED appears to correlate with the severity of cardiovascular

disease. Consequently, ED represents a “sentinel symptom” in patients with occult cardiovascular and peripheral vascular disease.

Smoking is also a significant risk factor in the development of ED. Medications used in treating diabetes or cardiovascular disease are additional risk factors (see below). There is a higher incidence of ED among men who have undergone radiation or surgery for prostate cancer and in those with a lower spinal cord injury. Psychological causes of ED include depression, anger, stress from employment or relationships, anxiety, and other stress-related causes.

**Pathophysiology** ED may result from three basic mechanisms: (1) failure to initiate (psychogenic, endocrinologic, or neurogenic), (2) failure to fill (arteriogenic), and (3) failure to store adequate blood volume within the lacunar network (veno-occlusive dysfunction). These categories are not mutually exclusive, and multiple factors contribute to ED in many patients. For example, diminished filling pressure can lead secondarily to venous leak. Psychogenic factors frequently coexist with other etiologic factors and should be considered in all cases. Diabetic, atherosclerotic, and drug-related causes account for >80% of cases of ED in older men.

**Vasculogenic** The most common organic cause of ED is a disturbance of blood flow to and from the penis. Atherosclerotic or traumatic arterial disease can decrease flow to the lacunar spaces, resulting in decreased rigidity and an increased time to full erection. Excessive outflow through the veins despite adequate inflow also may contribute to ED. Structural alterations to the fibroelastic components of the corpora may cause a loss of compliance and inability to compress the tunical veins. This condition may result from aging, increased cross-linking of collagen fibers induced by nonenzymatic glycosylation, hypoxemia, or altered synthesis of collagen associated with hypercholesterolemia.

**Neurogenic** Disorders that affect the sacral spinal cord or the autonomic fibers to the penis preclude nervous system relaxation of penile smooth muscle, thus leading to ED. In patients with spinal cord injury, the degree of ED depends on the completeness and level of the lesion. Patients with incomplete lesions or injuries to the upper part of the spinal cord are more likely to retain erectile capabilities than are those with complete lesions or injuries to the lower part. Although 75% of patients with spinal cord injuries have some erectile capability, only 25% have erections sufficient for penetration. Other neurologic disorders commonly associated with ED include multiple sclerosis and peripheral neuropathy. The latter is often due to either diabetes or alcoholism. Pelvic surgery may cause ED through disruption of the autonomic nerve supply.

**Endocrinologic** Androgens increase libido, but their exact role in erectile function is unclear. Individuals with castrate levels of testosterone can achieve erections from visual or sexual stimuli. Nonetheless, normal levels of testosterone appear to be important for erectile function, in which the upregulation of nitric oxide synthase and the nitric oxide cascade is optimized (Fig. 397-1A). Androgen replacement therapy can improve depressed erectile function when it is secondary to hypogonadism; however, it is not useful for ED when endogenous testosterone levels are normal. Increased prolactin may decrease libido by suppressing gonadotropin-releasing hormone (GnRH) resulting in decreased testosterone levels. Treatment of hyperprolactinemia with dopamine agonists can restore libido and eugonadism.

**Diabetic** ED occurs in 35–75% of men with diabetes mellitus. Pathologic mechanisms are related primarily to diabetes-associated vascular and neurologic complications. Diabetic macrovascular complications are related mainly to age, whereas microvascular complications correlate with the duration of diabetes and the degree of glycemic control (Chap. 403). Individuals with diabetes also have reduced amounts of nitric oxide synthase in both endothelial and neural tissues.

**3058 Psychogenic** Two mechanisms contribute to the inhibition of erections in psychogenic ED. First, psychogenic stimuli to the sacral cord may inhibit reflexogenic responses, thereby blocking activation of vasodilator outflow to the penis. Second, excess sympathetic stimulation in an anxious man may increase penile smooth-muscle tone. The most common causes of psychogenic ED are performance anxiety, depression, relationship conflict, loss of attraction, sexual inhibition, conflicts over sexual preference, sexual abuse in childhood, and fear of pregnancy or sexually transmitted disease. Almost all patients with ED, even when it has a defined organic basis, develop a psychogenic component as a reaction to ED.

**Medication-Related** Medication-induced ED (**Table 397-1**) is estimated to occur in 25% of men seen in general medical clinics. The adverse effects related to drug therapy are additive, especially in older men. In addition to the drug itself, the underlying disease being treated is likely to contribute to sexual dysfunction (e.g., hypertension). Among the antihypertensive agents, the thiazide diuretics and beta blockers have been implicated most frequently. Calcium channel blockers and angiotensin-converting enzyme inhibitors are cited less frequently. These drugs may act directly at the corporal level (e.g., calcium channel blockers) or indirectly by reducing pelvic blood pressure, which is important in the development of penile rigidity.  $\alpha$ -Adrenergic blockers are less likely to cause ED. Estrogens, GnRH agonists, H<sub>2</sub> antagonists, and spironolactone cause ED by suppressing gonadotropin production or by blocking androgen action. Antidepressant and antipsychotic agents—particularly neuroleptics, tricyclics, and SSRIs—are associated with erectile, ejaculatory, orgasmic, and sexual desire difficulties. Among the SSRIs, paroxetine and escitalopram have been associated with the highest risk of sexual dysfunction. Bupropion, nefazodone, and mirtazapine appear less likely to cause sexual dysfunction. A number of molecular pathways have been implicated in antidepressant-induced sexual adverse events. Serotonin has been hypothesized to inhibit normal sexual response by decreasing dopamine-enhanced libido, arousal, and erection and by increasing prolactin release. SSRIs have also been shown to be potent inhibitors of nitric oxide synthase.

If there is a strong association between the institution of a drug and the onset of ED, alternative medications should be considered. Otherwise, it is often practical to treat the ED without attempting multiple changes in medications as it may be difficult to establish a causal role for a drug.

## APPROACH TO THE PATIENT

### Erectile Dysfunction

A good physician–patient relationship helps unravel the possible causes of ED, many of which require discussion of personal and sensitive topics. For this reason, a primary care provider is often ideally suited to initiate the evaluation. However, a significant percentage of men experience ED and remain undiagnosed unless specifically questioned about this issue. By far the two most common reasons for underreporting of ED are patient embarrassment and perceptions of physicians' inattention to the disorder. Once the topic is initiated by the physician, patients are more willing to discuss their potency issues. A complete medical and sexual history should be taken in an effort to assess whether the cause of ED is organic, psychogenic, or multifactorial (**Fig. 397-3**).

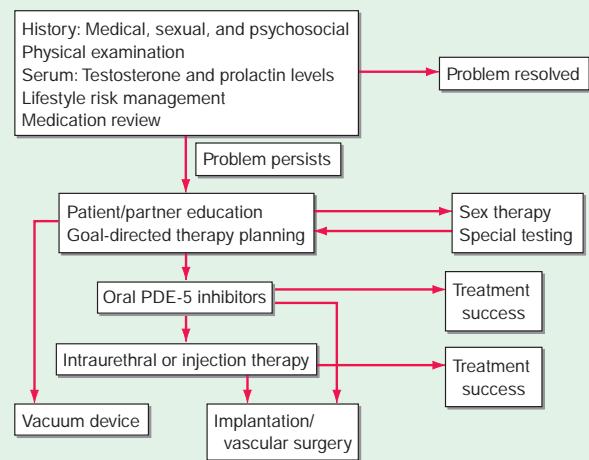
Both the patient and his sexual partner should be interviewed regarding sexual history. ED should be distinguished from other sexual problems, such as premature ejaculation. Lifestyle factors such as sexual orientation, the patient's distress from ED, performance anxiety, and details of sexual techniques should be addressed. Validated questionnaires are available to assess ED, including the International Index of Erectile Function (IIEF) and the more easily administered Sexual Health Inventory for Men (SHIM), a validated abridged version of the IIEF. These can assess the severity of ED,

**TABLE 397-1 Drugs Associated with Erectile Dysfunction**

CLASSIFICATION	DRUGS	POSSIBLE SUBSTITUTES
Diuretics	Thiazides Spironolactone	
Antihypertensives	Calcium channel blockers  Methyldopa Clonidine Reserpine Beta blockers Guanethidine	$\alpha$ -Adrenergic blockers Prazosin Terazosin Doxazosin ACE inhibitors
Cardiac/antihyperlipidemics	Digoxin  Gemfibrozil Clofibrate	
Antidepressants	Selective serotonin reuptake inhibitors  Tricyclic antidepressants Lithium Monoamine oxidase inhibitors	Bupropion Nefazodone Mirtazapine
Tranquilizers	Butyrophenones Phenothiazines	
H <sub>2</sub> antagonists	Ranitidine  Cimetidine	Proton pump inhibitors (PPI)  Omeprazole Esomeprazole Pantoprazole Rabeprazole
Hormones	Progesterone Estrogens Corticosteroids GnRH agonists 5 $\alpha$ -Reductase inhibitors Cyproterone acetate	
Cytotoxic agents	Cyclophosphamide Methotrexate Roferon-A	
Anticholinergics	Disopyramide Anticonvulsants	
Recreational	Ethanol Cocaine Marijuana	

Abbreviations: ACE, angiotensin-converting enzyme; GnRH, gonadotropin-releasing hormone.

measure treatment effectiveness, and guide future management. The initial evaluation of ED begins with a review of the patient's medical, surgical, sexual, and psychosocial histories. The history should note whether the patient has experienced pelvic trauma, surgery, or radiation. In light of the increasing recognition of the relationship between lower urinary tract symptoms (LUTS/BPH) and ED, it is advisable to evaluate for the presence of associated urinary symptoms. Questions should focus on the onset of symptoms, the presence and duration of partial erections, and the progression



**FIGURE 397-3** Algorithm for the evaluation and management of patients with erectile dysfunction. PDE, phosphodiesterase.

of ED. A history of nocturnal or early morning erections may be useful for distinguishing physiologic ED from psychogenic ED. Nocturnal erections occur during rapid eye movement (REM) sleep and require intact neurologic and circulatory systems. Organic causes of ED generally are characterized by a gradual and persistent change in rigidity or the inability to sustain nocturnal, coital, or self-stimulated erections. The patient should be questioned about the presence of penile curvature or pain with coitus. It is also important to address libido, as decreased sexual drive and ED are sometimes the earliest signs of endocrine abnormalities (e.g., increased prolactin, decreased testosterone levels). It is useful to ask whether the problem is confined to coitus with one partner or also involves other partners; ED not uncommonly arises with new or extramarital sexual relationships. Situational ED, as opposed to consistent ED, suggests psychogenic causes. For men being treated for ED, referral to a mental health professional should be considered to promote treatment adherence, reduce performance anxiety, and integrate treatments into a sexual relationship. Ejaculation is much less commonly affected than erection, but questions should be asked about whether ejaculation is normal, premature, delayed, or absent. Relevant risk factors should be identified, such as diabetes mellitus, coronary artery disease (CAD), and neurologic disorders. The patient's surgical history should be explored with an emphasis on bowel, bladder, prostate, and vascular procedures. A complete drug history, including tobacco, alcohol, marijuana, and illicit drug inquiries, is also important. Social changes that may precipitate ED are also crucial to the evaluation, including health worries, spousal death, divorce, relationship difficulties, and financial concerns.

Because ED commonly involves a host of endothelial cell risk factors, men with ED report higher rates of overt and silent myocardial infarction. Therefore, ED in an otherwise asymptomatic male warrants consideration of other vascular disorders, including CAD.

Men who suffer from ED are at high risk for concomitant LUTS from BPH and vice versa. Given that some treatments of one disorder will impact the other, the clinician should consider an assessment of LUTS in any man with ED.

The physical examination is an essential element in the assessment of ED. Signs of hypertension as well as evidence of thyroid, hepatic, hematologic, cardiovascular, or renal diseases should be sought. An assessment should be made of the endocrine and vascular systems, the external genitalia, and the prostate gland. The penis should be palpated carefully along the corpora to detect fibrotic plaques. Reduced testicular size and loss of secondary

sexual characteristics are suggestive of hypogonadism. Neurologic examination should include assessment of anal sphincter tone, investigation of the bulbocavernosus reflex, and testing for peripheral neuropathy.

Although hyperprolactinemia is uncommon, a serum prolactin level should be measured in hypogonadal men, as decreased libido and/or ED may be the presenting symptoms of a prolactinoma or another mass lesion of the sella ([Chap. 380](#)). The serum testosterone level should be measured, and if it is low, gonadotropins should be measured to determine whether hypogonadism is primary (testicular) or secondary (hypothalamic-pituitary) in origin ([Chap. 391](#)). If not performed recently, serum chemistries, complete blood count (CBC), hemoglobin A<sub>1c</sub>, and lipid profiles may be of value, as they can yield evidence of anemia, diabetes, hyperlipidemia, or other systemic diseases associated with ED.

Additional diagnostic testing is rarely necessary in the evaluation of ED. However, in selected patients, specialized testing may provide insight into pathologic mechanisms of ED and aid in the selection of treatment options. Optional specialized testing includes (1) studies of nocturnal penile tumescence and rigidity, (2) vascular testing (in-office injection of vasoactive substances, penile Doppler ultrasound, penile angiography, dynamic infusion cavernosography/cavernosometry), (3) neurologic testing (biothesiometry-graded vibratory perception, somatosensory-evoked potentials), and (4) psychological diagnostic tests. The information potentially gained from these procedures must be balanced against their invasiveness, cost, and impact on ultimate treatment outcome.

Clinicians should counsel men with ED who have comorbidities known to negatively affect erectile function that lifestyle modifications, including changes in diet and increased physical activity, improve overall health and may improve erectile function.

## TREATMENT

### Male Sexual Dysfunction

#### PATIENT EDUCATION

Patient and partner education is essential in the treatment of ED. In goal-directed therapy, education facilitates understanding of the disease, the results of the tests, and the selection of treatment. Discussion of treatment options helps clarify how treatment is best offered and stratify first- and second-line therapies. Patients with high-risk lifestyle issues such as obesity, smoking, alcohol abuse, and recreational drug use should be counseled on the role those factors play in the development of ED.

Therapies currently employed for the treatment of ED include oral phosphodiesterase type 5 inhibitor (PDE-5i) therapy (most commonly used), injection therapies, testosterone therapy, penile devices, and psychological therapy. In addition, limited data suggest that treatments for underlying risk factors and comorbidities—for example, weight loss, exercise, stress reduction, and smoking cessation—may improve erectile function. Decisions regarding therapy should take into account the preferences and expectations of patients and their partners.

#### ORAL AGENTS

Sildenafil, tadalafil, vardenafil, and avanafil are the only approved and effective oral agents for the treatment of ED. These four medications have markedly improved the management of ED because they are effective for the treatment of a broad range of causes, including psychogenic, diabetic, vasculogenic, post radical prostatectomy (nerve-sparing procedures), and post spinal cord injury. They belong to a class of medications that are selective and potent inhibitors of PDE-5, the predominant phosphodiesterase isoform found in the penis. They are administered in graduated doses and enhance erections after sexual stimulation ([Fig. 397-2](#)). The onset of action is ~30–120 min, depending on the medication used and other factors, such as recent food intake. Reduced initial

doses should be considered for patients who are elderly, are taking concomitant alpha blockers, have renal insufficiency, or are taking medications that inhibit the CYP3A4 metabolic pathway in the liver (e.g., erythromycin, cimetidine, ketoconazole, clarithromycin, diltiazem, itraconazole, ritonavir, verapamil, grapefruit, and possibly itraconazole and mibefradil), as they may increase the serum concentration of the PDE-5i or promote hypotension.

Initially, there were concerns about the cardiovascular safety of these drugs. It is known that these agents can act as mild vaso-dilators, and warnings exist about orthostatic hypotension with concomitant use of alpha blockers. The use of PDE-5is is not contraindicated in men who are also receiving alpha blockers, but they must be stabilized on the medication prior to initiating PDE-5i therapy. Earlier concerns that the use of PDE-5is would increase cardiovascular events have been mitigated by the results of several controlled trials showing no increase in myocardial ischemic events or overall mortality compared to the general population.

Several randomized trials have demonstrated the efficacy of this class of medications. There are no compelling data to support the superiority of one PDE-5i over another. Subtle differences between agents have variable clinical relevance (**Table 397-2**).

Patients may fail to respond to a PDE-5i for several reasons (**Table 397-3**). Some patients may not tolerate PDE-5i secondary to adverse events from vasodilation in nonpenile tissues expressing PDE-5 or from the inhibition of homologous nonpenile isozymes (i.e., PDE-6 found in the retina). Abnormal vision attributed to the effects of PDE-5i on retinal PDE-6 is of short duration, reported only with sildenafil, and not thought to be clinically significant. A more serious concern is the possibility that PDE-5is may cause nonarteritic anterior ischemic optic neuropathy (NAION); although data to support that association are limited, it is prudent to avoid the use of these agents in men with a prior history of NAION.

Testosterone supplementation combined with a PDE-5i may be beneficial in improving erectile function in hypogonadal men with ED who are unresponsive to PDE-5i alone. These drugs do not affect ejaculation, orgasm, or sexual drive. Side effects associated with PDE-5is include headaches (19%), facial flushing (9%), dyspepsia (6%), and nasal congestion (4%). Approximately 7% of men using sildenafil may experience transient altered color vision (blue halo effect), and 6% of men taking tadalafil may experience loin pain. PDE-5i is contraindicated in men receiving nitrate therapy for

**TABLE 397-3 Issues to Consider if Patients Report Failure of Phosphodiesterase Type 5 Inhibitor (PDE-5i) to Improve Erectile Dysfunction**

1. A trial of medication on at least 6 different days at the maximal dose should be performed before declaring patient nonresponsive to PDE-5i use.
2. Confirm that the patient did not partake in a high-fat meal prior to taking medication; pertains to sildenafil.
3. Failure to include physical and psychic stimulation at the time of foreplay to induce endogenous NO.
4. Took medications at an appropriate time frame *prior* to step 3: half-hour prior for avanafil, 1 h for sildenafil/vardenafil, or 2.5 h for tadalafil.
5. Unrecognized hypogonadism.

Abbreviation: NO, nitric oxide.

cardiovascular disease, including agents delivered by the oral, sublingual, transnasal, and topical routes. These agents can potentiate its hypotensive effect and may result in profound shock. Likewise, amyl/butyl nitrate “poppers” may have a fatal synergistic effect on blood pressure. PDE-5is also should be avoided in patients with congestive heart failure and cardiomyopathy because of the risk of vascular collapse. Because sexual activity leads to an increase in physiologic expenditure (5–6 metabolic equivalent tasks [METs]), physicians have been advised to exercise caution in prescribing any drug for sexual activity to those with active coronary disease, heart failure, borderline hypotension, or hypovolemia and to those on complex antihypertensive regimens.

Although the various forms of PDE-5is have a common mechanism of action, there are a few differences among the four agents (Table 397-2). Tadalafil is unique in its longer half-life, and avanafil appears to have the fastest onset of action. All four drugs are effective for patients with ED of all ages, severities, and etiologies. Although there are pharmacokinetic and pharmacodynamic differences among these agents, clinically relevant differences are not clear.

#### ANDROGEN THERAPY

Testosterone replacement is used to treat both primary and secondary causes of hypogonadism (**Chap. 391**). Men with ED and testosterone deficiency (TD) who are considering ED treatment with a PDE-5i should be informed that PDE-5is may be more effective if combined

**TABLE 397-2 PDE-5 Inhibitors<sup>a</sup>**

DRUG	ONSET OF ACTION	T <sub>1/2</sub>	DOSE	ADVERSE EFFECTS	CONTRAINdications
Sildenafil	T <sub>max</sub> 30–120 min Duration 4 h High-fat meal decreases absorption Alcohol use may affect efficacy	2–5 h	25–100 mg Starting dose 50 mg	Headache, flushing, dyspepsia, nasal congestion, altered vision	Nitrates Hypotension Cardiovascular risk factors Retinitis pigmentosa Change dose with some antiretrovirals Should be on stable dose of alpha blockers
Vardenafil	T <sub>max</sub> 30–120 min Duration 4–5 h High-fat meal decreases absorption ETOH may affect efficacy	4.5 h	5–10 mg	Headache, flushing, rhinitis, dyspepsia	Same as sildenafil May have minor prolongation of QT interval Concomitant use of class I anti-arrhythmic
Tadalafil	T <sub>max</sub> 30–60 min Duration 12–36 h Plasma concentration not affected by food or ETOH	17.5 h	10 or 20 mg; 2.5 or 5 mg for daily dose	Headache, dyspepsia, backpain, nasal congestion, myalgia	Same as sildenafil
Avanafil	T <sub>max</sub> 30 min Duration 2 h Plasma concentration not affected by food	3–5 h	50, 100, and 200 mg dose	Headache, flushing, nasal congestion, nasopharyngitis back pain	Same as sildenafil

<sup>a</sup>Sildenafil, vardenafil, tadalafil, and the newest option, avanafil, appear to be equally effective, but tadalafil has a longer duration of action and avanafil has a more rapid onset.

Abbreviation: ETOH, ethanol; PDE-5, phosphodiesterase type 5.

with testosterone therapy. Androgen supplementation in the setting of normal testosterone is not efficacious in the treatment of ED and is discouraged secondary to additional risk for toxicity without benefit. Methods of androgen replacement include transdermal patches and gels, including cutaneous nasal and axillary gels. Parenteral long-acting testosterone esters (enanthate and cypionate), long-acting subcutaneous pellets, and oral preparations (17 $\alpha$ -alkylated derivatives) are also available (Chap. 391). With the possible exception a newer oral testosterone undecanoate, oral androgen preparations have the potential for hepatotoxicity and should be avoided.

The increased scrutiny of testosterone caused the U.S. Food and Drug Administration (FDA) to issue a warning that there is a “weak signal” that testosterone replacement therapy increases the risk of thromboembolic events and may have addictive properties. Although testosterone therapy has known risks, such as water retention in heart failure patients and worsening sleep apnea, increasing evidence suggests that, when monitored appropriately, this therapy decreases the risk for metabolic syndrome, changes body composition by increasing lean muscle mass, and improves insulin sensitivity and average hemoglobin A<sub>1c</sub>. This evidence, combined with the fact that hypogonadism is a known risk factor for metabolic syndrome and cardiovascular disease, has led to the conclusion that testosterone therapy for age-related hypogonadism in fact improves overall health and decreases the risk of cardiovascular events. It is important to note that men with secondary hypogonadism who desire fertility should not be treated directly with testosterone, but with an alternative such as the selective estrogen receptor modulator (SERM) clomiphene citrate, which increases gonadotropin levels, stimulating testicular testosterone production.

Testosterone circulates in the body in two forms: free and unbound or that bound to proteins such as albumin or sex hormone-binding globulin (SHBG). SHBG has a very high affinity for testosterone, and thus, testosterone bound to SHBG does not bind to the androgen receptor and is not bioavailable. Bioavailable testosterone is any testosterone that is not bound to SHBG. Unfortunately, reliable assays to directly measure bioavailable testosterone or free testosterone are expensive, difficult to perform, and thus not offered by most laboratories. However, direct measurement of SHBG is inexpensive and reliable, allowing free and bioavailable testosterone to be calculated.

Men who receive testosterone should be reevaluated after 3–6 months and at least annually thereafter for testosterone levels, erectile function, and adverse effects, which may include gynecomastia, sleep apnea, development or exacerbation of LUTS or BPH, prostate cancer, lowering of HDL, erythrocytosis, elevations of liver function tests, and reduced fertility. Periodic reevaluation should include measurement of hemoglobin and prostate-specific antigen and digital rectal examination. Therapy should be discontinued in patients who do not respond within 6 months without an alternate explanation (e.g., elevated estradiol).

### VACUUM CONSTRICION DEVICES

Vacuum constriction devices (VCDs) are a well-established non-invasive therapy. They are a reasonable treatment alternative for select patients who cannot take PDE-5is or do not desire other interventions. VCDs draw venous blood into the penis and use a constriction ring to restrict venous return and maintain tumescence. Adverse events with VCD include pain, numbness, bruising, and altered ejaculation. Additionally, many patients complain that the devices are cumbersome and that the induced erections have a nonphysiologic appearance and feel.

### INTRAURETHRAL ALPROSTADIL

If a patient fails to respond to oral agents, a reasonable next choice is intraurethral or self-injection of vasoactive substances. Intraurethral prostaglandin E<sub>1</sub> (alprostadil), in the form of a semisolid pellet (doses of 125–1000 µg), is delivered with an applicator. Approximately 65% of men receiving intraurethral alprostadil respond with an erection when tested in the office, but <50% achieve successful coitus at

home. Intraurethral insertion is associated with a markedly reduced incidence of priapism in comparison to intracavernosal injection.

### INTRACAVERNOSAL SELF-INJECTION

Injection of synthetic formulations of alprostadil is effective in 70–80% of patients with ED, but discontinuation rates are high because of the invasive nature of administration. Doses range between 1 and 40 µg. Injection therapy is contraindicated in men with a history of hypersensitivity to the drug and men at risk for priapism (hypercoagulable states, sickle cell disease). Side effects include local adverse events, prolonged erections, pain, and fibrosis with chronic use. Various combinations of alprostadil, phentolamine, and/or papaverine sometimes are used.

### SURGERY

An important but less frequently used form of therapy for ED involves the surgical implantation of a semirigid or inflatable penile prosthesis. Because of the permanence of prosthetic devices, patients should first consider less invasive options for treatment. These surgical treatments are associated with a low rate of complications and are used for those who do not want the less spontaneous medical treatments, in PDE-5i-refractory ED, or in men who cannot tolerate such medications. Despite the requirement for surgery, penile prostheses are associated with very high rates of patient and partner satisfaction.

### SEX THERAPY

A course of sex therapy may be useful for addressing specific interpersonal factors that may affect sexual functioning. These approaches may be useful in patients who have psychogenic or social components to their ED, although data from randomized trials are scanty and inconsistent. It is preferable to include both partners in therapy if the patient is involved in an ongoing relationship.

## FEMALE SEXUAL DYSFUNCTION

Female sexual dysfunction (FSD) has traditionally included disorders of desire, arousal, pain, and muted orgasm. The associated risk factors for FSD are similar to those in males: cardiovascular disease, endocrine disorders, hypertension, neurologic disorders, and smoking (Table 397-4). Women with hypertension report significantly lower sexual satisfaction (especially younger women).

### EPIDEMIOLOGY

Epidemiologic data are limited, but the available estimates suggest that as many as 43% of women complain of at least one sexual problem. Despite the recent interest in organic causes of FSD, desire and arousal phase disorders (including lubrication complaints) remain the most

**TABLE 397-4 Risk Factors for Female Sexual Dysfunction**

Neurologic disease: stroke, spinal cord injury, parkinsonism
Trauma, genital surgery, radiation
Endocrinopathies: diabetes, hyperprolactinemia
Liver and/or renal failure
Cardiovascular disease, especially hypertension
Psychological factors and interpersonal relationship disorders: sexual abuse, life stressors
Medications
Antiandrogens: cimetidine, spironolactone
Antidepressants, alcohol, hypnotics, sedatives
Antiestrogens or GnRH antagonists
Antihistamines, sympathomimetic amines
Antihypertensives: diuretics, calcium channel blockers
Alkylating agents
Anticholinergics

Abbreviation: GnRH, gonadotropin-releasing hormone.

## PHYSIOLOGY OF THE FEMALE SEXUAL RESPONSE

The normal female sexual response requires the presence of estrogens. A role for androgens is also likely but less well established. In the CNS, estrogens and androgens work synergistically to enhance sexual arousal and response. A number of studies report enhanced libido in women during preovulatory phases of the menstrual cycle, suggesting that hormones involved in the ovulatory surge (e.g., estrogens) increase desire.

Sexual motivation is heavily influenced by context, including the environment and partner factors. Once sufficient sexual desire is reached, sexual arousal is mediated by the central and autonomic nervous systems. Cerebral sympathetic outflow is thought to increase desire, and peripheral parasympathetic activity results in clitoral vasocongestion and vaginal secretion (lubrication).

The neurotransmitters for clitoral corporal engorgement are similar to those in the male penile tissues, with a prominent role for neural, smooth-muscle, and endothelial released nitric oxide (NO). A fine network of vaginal nerves and arterioles promotes a vaginal transudate. The major transmitters of this complex vaginal response are not certain, but roles for NO and vasoactive intestinal polypeptide (VIP) are suspected. Investigators studying the normal female sexual response have challenged the long-held construct of a linear and unmitigated relationship between initial desire, arousal, vasocongestion, lubrication, and eventual orgasm. Caregivers should consider a paradigm of a positive emotional and physical outcome with one, many, or no orgasmic peak and release.

Although there are anatomic differences as well as variation in the density of vascular and neural beds in males and females, the primary effectors of sexual response are strikingly similar. Intact sensation is important for arousal. Thus, reduced levels of sexual functioning are more common in women with peripheral neuropathies (e.g., diabetes). Vaginal lubrication is a transudate of serum that results from the increased pelvic blood flow associated with arousal. Vascular insufficiency from a variety of causes may compromise adequate lubrication and result in dyspareunia. Cavernosal and arteriole smooth-muscle relaxation occurs via increased NO synthase (NOS) activity and produces engorgement in the clitoris and the surrounding vestibule. Orgasm requires an intact sympathetic outflow tract; hence, orgasmic disorders are common in female patients with spinal cord injuries.

## APPROACH TO THE PATIENT

### Female Sexual Dysfunction

Many women do not volunteer information about their sexual response. Open-ended questions in a supportive atmosphere are helpful in initiating a discussion of sexual integrity in women who are reluctant to discuss such issues. Once a complaint has been voiced, a comprehensive evaluation should be performed, including a medical history, a psychosocial history, a physical examination, and limited laboratory testing.

The history should include the usual medical, surgical, obstetric, psychological, gynecologic, sexual, and social information. Past experiences, intimacy, knowledge, and partner availability should also be ascertained. Medical disorders that may affect sexual health should be delineated. They include diabetes, cardiovascular disease, gynecologic conditions, obstetric history, depression, anxiety disorders, and neurologic disease. Medications should be reviewed as they may affect arousal, libido, and orgasm. The need for counseling and recognizing life stresses should be identified. The physical examination should assess the genitalia, including the clitoris. Pelvic floor examination may identify prolapse or other disorders. Laboratory studies are needed, especially if menopausal status is uncertain. Estradiol, follicle-stimulating hormone (FSH), and

luteinizing hormone (LH) are usually obtained, and dehydroepiandrosterone (DHEA) should be considered as it reflects adrenal androgen secretion. A CBC, liver function assessment, and lipid studies may be useful, if not otherwise obtained. Complicated diagnostic evaluation such as clitoral Doppler ultrasonography and biothesiometry require expensive equipment and are of uncertain utility. It is important for the patient to identify which symptoms are most distressing.

The evaluation of FSD previously occurred exclusively in a psychosocial context. However, inconsistencies between diagnostic categories based only on psychosocial considerations and the emerging recognition of organic etiologies have led to a new classification of FSD. This diagnostic scheme is based on four components that are not mutually exclusive: (1) *hypoactive sexual desire*—the persistent or recurrent lack of sexual thoughts and/or receptivity to sexual activity, which causes personal distress; hypoactive sexual desire may result from endocrine failure or may be associated with psychological or emotional disorders, (2) *sexual interest arousal disorder*—the persistent or recurrent inability to attain or maintain sexual excitement, which causes personal distress, (3) *orgasmic disorder*—the persistent or recurrent loss of orgasmic potential after sufficient sexual stimulation and arousal, which causes personal distress, and (4) *sexual pain disorder*—persistent or recurrent genital pain associated with noncoital sexual stimulation, which causes personal distress. This newer classification emphasizes “personal distress” as a requirement for dysfunction and provides clinicians with an organized framework for evaluation before or in conjunction with more traditional counseling methods.

## TREATMENT

### Female Sexual Dysfunction

#### GENERAL

An open discussion with the patient is important as couples may need to be educated about normal anatomy and physiologic responses, including the role of orgasm, in sexual encounters. Physiologic changes associated with aging and/or disease should be explained. Couples may need to be reminded that clitoral stimulation rather than coital intromission may be more beneficial.

Behavioral modification and nonpharmacologic therapies should be a first step. Patient and partner counseling may improve communication and relationship strains. Lifestyle changes involving known risk factors can be an important part of the treatment process. Emphasis on maximizing physical health and avoiding lifestyles (e.g., smoking, alcohol abuse) and medications likely to produce FSD is important (Table 397-3). The use of topical lubricants may address complaints of dyspareunia and dryness. Contributing medications such as antidepressants may need to be altered, including the use of medications with less impact on sexual function, dose reduction, medication switching, or drug holidays.

#### HORMONAL THERAPY

In postmenopausal women, estrogen replacement therapy may be helpful in treating vaginal atrophy, decreasing coital pain, and improving clitoral sensitivity (Chap. 395). Menopause and its transition represent significant risk factors for the development of vulvovaginal atrophy-related sexual dysfunction. Available vaginal estrogen preparations include conjugated equine estrogens, estradiol vaginal cream, a sustained-release intravaginal estradiol ring, or a low-dose estradiol tablet. Vaginal estrogen preparations with the lowest systemic absorption rate may be preferred in women with history of breast cancer and severe vaginal atrophy. Vaginal lubricants and moisturizers applied on a regular basis have an efficacy comparable to that of local estrogen therapy and should be offered to women wishing to avoid the use of vaginal estrogens. If a hormonal supplement is chosen, then estrogen replacement in

the form of local cream is the preferred method as it avoids systemic side effects. Androgen levels in women decline substantially before menopause. However, low levels of testosterone or DHEA are not effective predictors of a positive therapeutic outcome with androgen therapy. The widespread use of exogenous androgens is not supported by the literature except in select circumstances (premature ovarian failure or menopausal states) and in secondary arousal disorders.

Arophic vaginitis is very common in postmenopausal women and is most commonly treated with estrogen-based treatments. However, many women are hesitant to use estrogen-based treatments due to health concerns or are unable to use them due to a history of breast cancer or endometrial cancer. Hyaluronic acid vaginal gel has been found to be efficacious in treating atrophic vaginitis.

### ORAL AGENTS

Flibanserin, originally developed as an antidepressant, has been approved by the FDA as a treatment for low sexual desire in premenopausal women. Flibanserin, a postsynaptic agonist of serotonin receptor 1A and antagonist of serotonin receptor 2A, increases sexual desire and reduces resultant stress in women with hyposexual desire disorder (HSDD) with few adverse effects. Flibanserin has two principal pharmacologic actions in neural microcircuits: it acts as a full agonist at postsynaptic 5-HT<sub>1A</sub> receptors and an antagonist at postsynaptic 5-HT<sub>2A</sub> receptors. Exclusive binding at these receptors differentiates flibanserin from buspirone and bupropion. This action in the prefrontal cortex causes the downstream release of dopamine and norepinephrine and reduction of serotonin. Flibanserin acts selectively on pyramidal neurons that excite brainstem 5-HT neurons yet also selectively on pyramidal neurons that inhibit brainstem norepinephrine and dopamine neurons.

Flibanserin may boost sex drive in women who experience low sexual desire and who find the experience distressing. The drug should be discontinued if there is no improvement in sex drive after 8 weeks. Potentially serious side effects include low blood pressure, dizziness, and fainting, particularly if it is mixed with alcohol. Other common adverse events include dizziness, nausea, fatigue, sleepiness, and insomnia. Health care professionals and pharmacies dealing with flibanserin have to undergo a certification (risk evaluation and mitigation strategy [REMS]) process, and patients need to submit a written agreement to abstain from alcohol. The goal of the flibanserin REMS is to inform patients about the increased risk of hypotension and syncope due to an interaction with alcohol.

Bremelanotide, a melanocortin 4 receptor agonist, has recently been approved for HSDD. It demonstrates significant improvement in desire and a significant decrease in distress related to lack of desire. The most common adverse effects include nausea (39.9%), facial flushing (20.4%), and headache (11%). Bremelanotide's place in therapy is unknown, as the trials met statistical significance for change in sexual desire elements and distress related to sexual desire, yet the clinical benefit may only be modest. It is a subcutaneous injection given 45 min prior to sexual activity. Bremelanotide has no clinically significant interactions with ethanol. Prescribing guidelines recommend no more than one dose in 24 h and no more than eight doses per month. Individuals should discontinue use after 8 weeks without benefit.

The efficacy of PDE-5is in FDS has been a marked disappointment in light of the proposed role of NO-dependent physiology in the normal female sexual response. The use of PDE-5is for FSD should be discouraged pending proof that they are effective.

### CLITORAL VACUUM DEVICE

In patients with arousal and orgasmic difficulties, the option of using a clitoral vacuum device may be explored. This handheld battery-operated device has a small soft plastic cup that applies a vacuum over the stimulated clitoris. This causes increased cavernosal blood flow, engorgement, and vaginal lubrication.

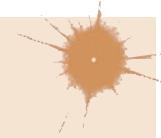
### FURTHER READING

- Bhasin S: A perspective on the evolving landscape in male reproductive medicine. *J Clin Endocrinol Metab* 101:827, 2016.
- Burnett AL et al: Erectile dysfunction: AUA Guideline. *J Urol* 200:633, 2018.
- Cappelleri JC, Rosen RC: The Sexual Health Inventory for Men (SHIM): A 5-year review of research and clinical experience. *Int J Impot Res* 17:307, 2005.
- Geerken MJM et al: Sexual dysfunction and bother due to erectile dysfunction in the healthy elderly male population: Prevalence from a systematic review. *Eur Urol Focus* 6:776, 2020.
- McVary KT: Clinical practice. Erectile dysfunction. *N Engl J Med* 357:2472, 2007.
- Wheeler LJ, Guntupalli SR: Female sexual dysfunction: Pharmacologic and therapeutic interventions. *Obstet Gynecol* 136:174, 2020.

**398**

## Women's Health

Emily Nosova, Andrea Dunai



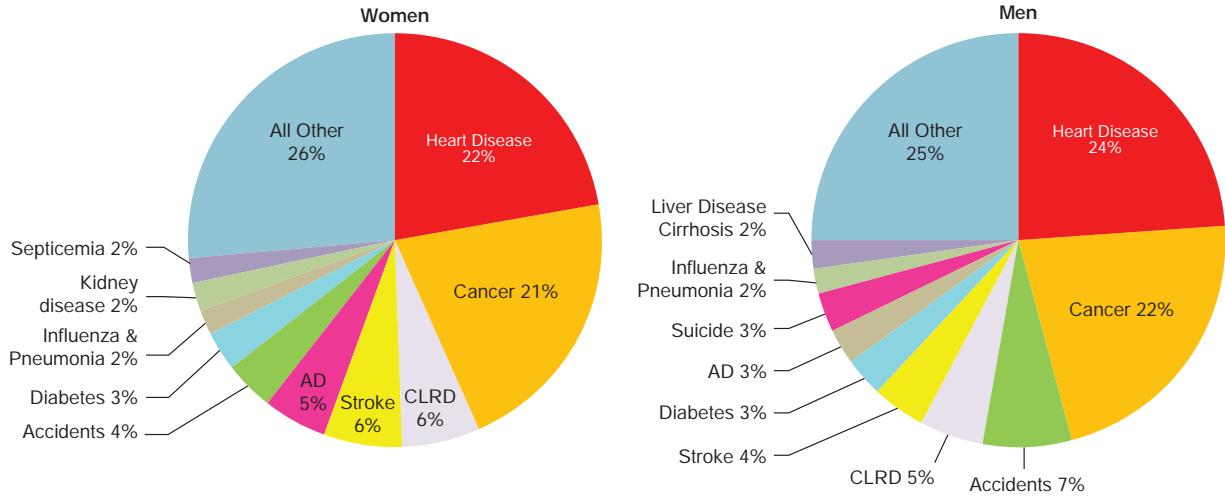
The clinical discipline of women's health is well established. Indeed, its emphasis on greater attention to patient education and medical decision-making is a paradigm for what has become known as patient-centered health care. Moreover, the recognition of sex differences in gene expression, disease processes, and health outcomes is an important example of precision medicine. Sex difference refers to the biologic differences conferred by sex chromosomes and hormones. In contrast, gender differences are related to psychosocial roles and cultural expectations. The study of sex differences continues to grow as a scientific discipline. In 2016, the National Institutes of Health recognized its importance by implementing the expectation that sex should be considered as biologic variable in study designs, analyses, and reporting in not only human but also vertebrate animal research. Strong scientific justification must be provided to limit research to only one sex.

### DISEASE RISK: REALITY AND PERCEPTION

The leading causes of death are the same in women and men: (1) heart disease and (2) cancer (Fig. 398-1). The leading cause of cancer death, lung cancer, is the same in both sexes. Breast cancer is the second leading cause of cancer death in women, but it causes ~70% fewer total deaths than does lung cancer. Men are more likely than women to die from suicide and accidents.

Maternal mortality continues to be higher in the United States than in other industrialized nations and is associated with substantial health disparities in maternal deaths. U.S. maternal mortality rates declined for the majority of the twentieth century given improvements in maternity care and safer surgical techniques; however, the rates began to rise again in 2000. The most current national data for maternal mortality were reported in 2018, over 10 years since the prior updated statistics. Over the past decade, the mortality rate has remained relatively stable. In 2018, the mortality rate was 17.4 deaths per 100,000 live births. The mortality rates for non-Hispanic black women were highest at 37.3 deaths per 100,000 live births (2.5 times the rate for non-Hispanic white women [14.9 deaths per 100,000 live births], 3.2 times the rate for Hispanic women [11.8 deaths], and 2.8 times the rate for Asian women [13.3 deaths]).

Women's risk for many diseases increases at menopause. The median age of menopause in Caucasian women from industrialized countries is between 50 and 52 years, where women spend one-third of their lives in the postmenopausal period. Menopause occurs at earlier ages in Hispanic and African-American women as well as in women of lower socioeconomic status. Estrogen levels fall abruptly at menopause, inducing a variety of physiologic and metabolic responses. Rates of



**FIGURE 398-1** Percent distribution of 10 leading causes of death in (A) women compared to (B) men in the United States in 2018. In both women and men, the first and second leading causes of death are the same, heart disease and cancer, respectively. Causes of death then diverge by sex. For example, accidents are the third leading cause of death in men but the sixth leading cause of death in women. Chronic lower respiratory disease (CLRD), stroke, and Alzheimer's disease (AD) cause a larger percentage of deaths in women than in men. Suicide is among the 10 leading causes of death in men but not in women. (Data from SL Murphy, J Xu, KD Kochanek, E Arias, B Tejada-Vera: Deaths: Final Data for 2018. *Natl Vital Stat Rep* 69:1, 2021.)

cardiovascular disease (CVD) increase and bone density decreases rapidly after menopause.

In the United States, women live on average 5.0 years longer than men, with a life expectancy at birth in 2018 of 81.2 years in women compared with 76.2 years in men of all races. Life expectancy was lower in African Americans of both sexes and higher in Hispanics of both sexes than their Caucasian counterparts. Accordingly, elderly women outnumber elderly men, so that age-related conditions, such as hypertension, have a female preponderance.

## SEX DIFFERENCES IN HEALTH AND DISEASE

### ALZHEIMER'S DISEASE

(See also Chap. 431.) Alzheimer's disease (AD) affects approximately twice as many women as men. Because the risk for AD increases with age, part of this sex difference is accounted for by the fact that women live longer than men. However, even in relatively younger groups (60–70 years of age), there is still a higher incidence of AD among women. Additional factors may contribute to the increased risk for AD in women, including sex differences in brain size, structure, and functional organization. Multimodal neuroimaging has demonstrated that certain biomarkers of the preclinical phase of AD, including a decline in neuronal mitochondrial function and impaired cerebral glucose metabolism, are evident earlier in women and are even distinguishable during the perimenopausal endocrine transition. There is emerging evidence for sex-specific differences in gene expression, not only for genes on the X and Y chromosomes but also for some autosomal genes. These genetic differences may translate into variable severity of AD, with women experiencing greater deficits in cognition. The A4 allele of the apolipoprotein E gene (*APO* 4), a cholesterol carrier integral for lipid transport in the brain, is a major risk factor for AD. Recent studies show that the *APO* 4 genotype is strongly linked to development of sporadic AD in women (Fig. 398-2). Women who carry either the *APO* 4 homo- or heterozygous isoform have an increased risk of progressing from healthy aging patterns to cognitive impairment or AD, whereas men who carry either isoform experience marginal impact on their memory or cognition.

Estrogens have pleiotropic genomic and nongenomic effects on the central nervous system, including neurotrophic actions in key areas involved in cognition and memory. Women with AD have lower endogenous estrogen levels than do women without AD. These

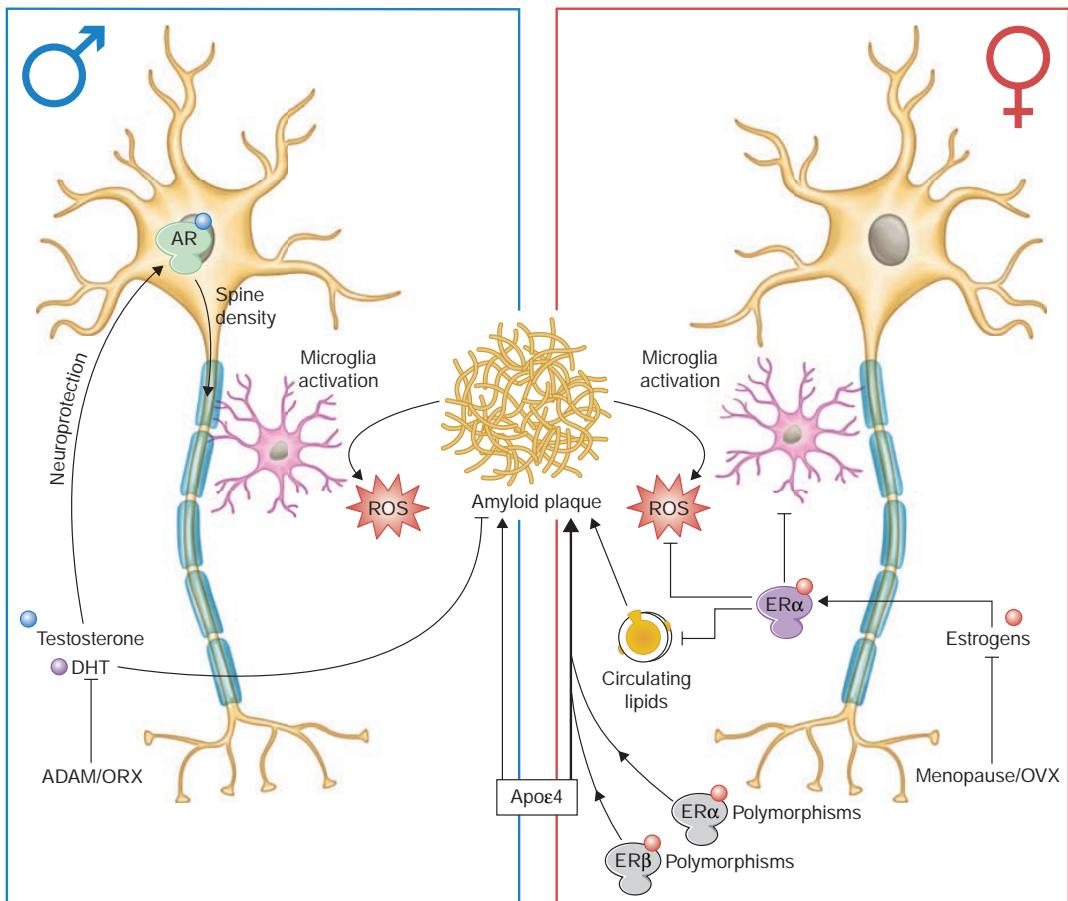
observations have led to the hypothesis that estrogen is neuroprotective. The Women's Health Initiative Memory Study (WHIMS), an ancillary study in the Women's Health Initiative (WHI) in women aged 65 years, found significantly increased risk for both dementia and mild cognitive impairment in women receiving estrogen alone or estrogen with progestin compared to placebo. However, the Kronos Early Estrogen Prevention Study (KEEPS), a randomized clinical trial of early hormone therapy (HT) initiation after menopause that compared conjugated equine estrogen (CEE), transdermal estradiol (both estrogen arms included cyclic oral micronized progesterone), and placebo, found no adverse effect of HT on cognitive function. In summary, there is no evidence from placebo-controlled trials that HT improves cognitive function.

While studies have shown a link between female sex and AD, other neurodegenerative disorders, including Parkinson's disease (PD) and amyotrophic lateral sclerosis (ALS), exhibit a stronger association with male sex. Men are 1.5 times more likely to develop PD than women across all age groups. A possible explanation for the male predilection may be the effect of Y-chromosome exclusive gene sex-determining region Y (SRY) on nigrostriatal dopaminergic (NSDA) neurons: the SRY upregulates neuronal numbers, synthesis of dopamine, and metabolism of neurons.

ALS is a highly variable disease in symptomatology and age at onset. Men are diagnosed with ALS earlier in life and exhibit a more severe disease course, even though survival patterns are similar between the two sexes. Despite this variability, epidemiologic data have demonstrated that, in females, ALS starts in the bulbar tract, whereas in males, ALS tends to begin in the motor neurons of the lumbar tract. Reasons for the sex differences observed in ALS remain unclear. However, there is some evidence to suggest that a more prolonged reproductive condition, defined as a longer duration exposure to estrogen via oral contraceptive use versus natural menopause, may exert a neuroprotective role on motor neurons among women diagnosed with ALS.

### CVD AND STROKE

(See also Chap. 273.) There are major sex differences in CVD, the leading cause of death in developed countries. However, there are also major gender differences because of perceptions by both women and their health care providers that women are at lower risk for CVD. As a result of these misconceptions, women are less likely to seek medical help when they experience symptoms of CVD. Health care providers are less likely to suspect CVD, so women receive fewer interventions for modifiable



**FIGURE 398-2** Sex differences and actions of sex steroid hormones on amyloid plaque deposition, neuroinflammation, and neuroprotection. The  $\epsilon 4$  allele of the apolipoprotein E gene (*APoE4*), a cholesterol carrier integral for lipid transport in the brain, has been identified as a major genetic risk factor for the sporadic development of Alzheimer's disease (AD). Women who carry either the homo- or heterozygous *APoE4* isoform have higher rates of amyloid plaque deposition. The *APoE4* variant has relatively limited effects in men with either the homo- or the heterozygous isoforms. AR, androgen receptor; DHT, dihydrotestosterone; ER, estrogen receptor; ROS, reactive oxygen species. (Reproduced with permission from E Vegeto et al: The role of sex and sex hormones in neurodegenerative diseases. *Endocr Rev* 41:273, 2020.)

risk factors as well as fewer acute interventions than do men. Women and their health care providers are also less aware that prodromal symptoms of cardiac disease differ in women compared to men. Women are less likely than men to present with chest pain and more likely to present with fatigue, shortness of breath, indigestion/nausea, and anxiety.

Sex steroids have major effects on the cardiovascular system and lipid metabolism. Estrogen increases high-density lipoprotein (HDL) and lowers low-density lipoprotein (LDL), whereas androgens have the opposite effect. Estrogen has direct vasodilatory effects on the vascular endothelium, enhances insulin sensitivity, and has antioxidant and anti-inflammatory properties. There is a striking increase in CVD after both natural and surgical menopause, suggesting that endogenous estrogens are cardioprotective. Women also have longer QT intervals on electrocardiograms, and this increases their susceptibility to certain arrhythmias.

CVD presents differently in women, who are usually 10–15 years older than their male counterparts and are more likely to have comorbidities such as hypertension, congestive heart failure, and diabetes mellitus (DM). In the Framingham study, angina was the most common initial symptom of CVD in women, whereas myocardial infarction (MI) was the most common initial presentation in men. Women more often have atypical symptoms such as fatigue, anxiety, nausea, indigestion, and upper back pain. Although awareness that heart disease is the leading cause of death in women has nearly doubled over the past 15 years, women remain less aware that its symptoms are often

atypical and are less likely to contact 9-1-1 when they experience such symptoms.

Deaths from CVD had decreased markedly in men since 1980, whereas CVD deaths only started to decrease substantially in women beginning in 2000. After 2010, death rates from CVD among both sexes stabilized and even began to increase slightly in men. Women with MI are more likely to present with cardiac arrest or cardiogenic shock, whereas men are more likely to present with ventricular tachycardia. Further, younger women with MI are more likely to die than are men of similar age. However, this mortality gap has decreased in recent years because younger women have experienced greater improvements in survival after MI than men. The improvement in survival is due largely to a reduction in comorbidities, suggesting a greater attention to modifiable risk factors in women.

Sex differences account for more variable short-term outcomes observed among women with CVD who receive therapeutic intervention, as compared to men. Women undergoing CABG surgery have more advanced disease, a higher perioperative mortality rate, less relief of angina, and less graft patency; however, 5- and 10-year survival rates are similar. Women undergoing percutaneous transluminal coronary angioplasty have lower rates of initial angiographic and clinical success than men, but they also have a lower rate of restenosis and a better long-term outcome. Women may benefit less and have more frequent serious bleeding complications from thrombolytic therapy compared with men. Factors such as older age, more comorbid conditions,

smaller body size, and more severe CVD in women at the time of events or procedures account in part for the observed sex differences.

Elevated cholesterol levels, hypertension, smoking, obesity, low HDL cholesterol levels, DM, and lack of physical activity are important risk factors for CVD in both men and women. Total triglyceride levels are an independent risk factor for CVD in women but not in men. Low HDL cholesterol and DM are more important risk factors for CVD in women than in men. Several disorders affect women exclusively, such as pregnancy-associated hypertension, preeclampsia, gestational DM, and polycystic ovary syndrome, or predominantly, such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). Cholesterol-lowering drugs are equally effective in men and women for primary and secondary prevention of CVD. In contrast to men, randomized trials showed that aspirin was not effective in the primary prevention of CVD in women; it did significantly reduce the risk of ischemic stroke.

Recent studies demonstrate strong associations between certain adverse pregnancy outcomes (APO) and the development of CVD in postmenopausal women, a risk that has been previously underappreciated. In an analysis of WHI data, nearly 29% of women surveyed reported at least one APO, defined as a hypertensive disorder during pregnancy, low or high neonatal birth weight, gestational diabetes, or preterm delivery of 3 weeks or greater. Each outcome, when analyzed independently, was significantly associated with the development of CVD after menopause. When multivariate modeling incorporated all of the APO variables and accounted for body mass index (BMI), socioeconomic status, and parity, low birth weight and hypertensive disorders remained significantly associated with later-onset CVD. A recent meta-analysis examined the association of female reproductive factors and the development of future CVD. The greatest risk for CVD (at least 2-fold) was conferred by a history of stillbirth, preterm birth, or preeclampsia, followed by a 1.5- to 1.9-fold risk with gestational diabetes and hypertension, premature ovarian insufficiency, and placental abruption; the lowest risk (<1.5-fold) was associated with early menarche, early menopause, parity, and polycystic ovary syndrome. Given these strong associations, targeted counseling and increased surveillance of CVD risk factors is warranted for women in high-risk groups.

The sex differences in CVD prevalence, beneficial biologic effects of estrogen on the cardiovascular system, and reduced risk for CVD in observational studies led to the hypothesis that HT was cardioprotective. However, the WHI, which studied >16,000 women on CEE plus medroxyprogesterone acetate (MPA) or placebo and >10,000 women with hysterectomy on CEE alone or placebo, did not demonstrate a benefit of HT for the primary or secondary prevention of CVD. In addition, CEE plus MPA was associated with an increased risk for CVD, particularly in the first year of therapy, whereas CEE alone neither increased nor decreased CVD risk. Both HT groups were associated with an increased risk for ischemic stroke. In a subgroup analysis of the WHI estrogen-alone trial, a relatively younger age (50–59 years) combined with a history of bilateral salpingo-oophorectomy (BSO) was associated with a >30% CEE treatment-associated reduction in all-cause mortality, whereas CEE-treated older women with prior BSO did not see a significant reduction in any other outcomes, including incidence of coronary artery disease, invasive breast cancer, all-cause mortality, and a composite index of the aforementioned outcomes plus stroke, hip fracture, pulmonary embolism, and colorectal cancer. These results suggest that postmenopausal women younger than 60 with prior BSO may have mortality benefit from HT, while women older than 60 with BSO may suffer consequences associated with HT. More recent data from KEEPS indicate that even if estrogen therapy is initiated shortly after the menopausal transition, it does not reduce atherosclerotic progression or impact CVD outcomes. Additionally, HT and placebo groups have similar outcomes with respect to venous thromboembolism and breast cancer. Although HT does not slow CVD development as previously thought, findings from KEEPS suggest that treated women experience significant improvements in vaso-motor symptoms, mood, sexual function, and bone density, especially when therapy is started sooner after menopause onset.

**HT is discussed further in Chap. 395.**

## DIABETES MELLITUS

**(See also Chap. 403.)** Women are more sensitive to insulin than men. Despite this, the prevalence of type 2 DM is similar in men and women. There is a sex difference in the relationship between endogenous androgen levels and DM risk. Higher bioavailable testosterone levels are associated with increased risk in women, whereas lower bioavailable testosterone levels are associated with increased risk in men. This observation has been confirmed in a recent Mendelian randomization that found that genetically determined higher testosterone increases risk for DM in women but reduces risk in men. Polycystic ovary syndrome, preeclampsia, pregnancy-associated hypertension, and gestational DM—common conditions in premenopausal women—are associated with a significantly increased risk for type 2 DM. Among individuals with DM, women have a greater risk for MI than do men. Women with DM have a sixfold greater risk of dying of CVD compared to women without DM.

Premenopausal women with DM lose the cardioprotective effect of female sex and have rates of CVD identical to those in males. These women have impaired endothelial function and reduced coronary vasodilatory responses, which may predispose to cardiovascular complications. Women with DM are more likely to have left ventricular hypertrophy. Women with DM receive less aggressive treatment for modifiable CVD risk factors than men with DM.

## HYPERTENSION

**(See also Chap. 277.)** After age 60, hypertension is more common in U.S. women than in men, largely because of the high prevalence of hypertension in older age groups and the longer survival of women. Isolated systolic hypertension is present in 30% of women >60 years old. Sex hormones affect blood pressure. Both normotensive and hypertensive women have higher blood pressure levels during the follicular phase than during the luteal phase. In the Nurses' Health Study, the relative risk of hypertension was 1.8 in current users of oral contraceptives, but this risk is lower with the newer low-dose contraceptive preparations. HT is not associated with hypertension. Among secondary causes of hypertension, there is a female preponderance of renal artery fibromuscular dysplasia.

The benefits of treatment for hypertension have been dramatic in both women and men. A meta-analysis of the effects of hypertension treatment, the Individual Data Analysis of Antihypertensive Intervention Trial, found a reduction of risk for stroke and major cardiovascular events in women. The effectiveness of various antihypertensive drugs appears to be comparable in women and men; however, women may experience more side effects, such as cough with angiotensin-converting enzyme inhibitors.

## AUTOIMMUNE DISORDERS

**(See also Chap. 355.)** Most autoimmune disorders occur more commonly in women than in men; they include autoimmune thyroid and liver diseases, SLE, RA, scleroderma, multiple sclerosis (MS), and idiopathic thrombocytopenic purpura. However, there is no sex difference in the incidence of type 1 DM, and ankylosing spondylitis occurs more commonly in men. Sex differences in both immune responses and adverse reactions to vaccines have been reported. For example, there is a female preponderance of postvaccination arthritis.

Adaptive immune responses are more robust in women than in men; this may be explained by the stimulatory actions of estrogens and the inhibitory actions of androgens on the cellular mediators of immunity. Consistent with an important role for sex hormones, there is variation in immune responses during the menstrual cycle, and the activity of certain autoimmune disorders is altered by castration or pregnancy (e.g., RA and MS may remit during pregnancy). Nevertheless, the majority of studies show that exogenous estrogens and progestins in the form of HT or oral contraceptives do not alter autoimmune disease incidence or activity. Exposure to fetal antigens, including circulating fetal cells that persist in certain tissues, has been speculated to increase the risk of autoimmune responses. There is clearly an important genetic component to autoimmunity, as indicated by the familial clustering and HLA association of many such disorders.

X chromosome genes also contribute to sex differences in immunity. Indeed, nonrandom X chromosome inactivation may be a risk factor for autoimmune diseases.

## HIV INFECTION

**(See also Chap. 202.)** Women accounted for almost 18% of the ~36,400 new HIV diagnoses in the United States in 2018. Annual HIV diagnoses remained stable among women from 2014 to 2018. In 2018, the infectivity rate for black/African-American females was 13 times the rate for white females and 4 times higher than Hispanic/Latino females. Nevertheless, AIDS remains an important cause of death in younger women, particularly African-American women aged 25–44 years. Heterosexual contact with an at-risk partner is the fastest-growing transmission category, and women are more susceptible to HIV infection during vaginal sex than men. This increased susceptibility is accounted for in part by an increased prevalence of sexually transmitted diseases, i.e., gonorrhea and syphilis, in women.

Some studies have suggested that hormonal contraceptives may increase the risk of HIV transmission. Progesterone has been shown to increase susceptibility to infection in nonhuman primate models of HIV. Women are also more likely to be infected by multiple variants of the virus than men. Women with HIV have more rapid decreases in their CD4 cell counts than do men. Compared with men, HIV-infected women more frequently develop candidiasis, but Kaposi's sarcoma is less common than it is in men. Women have more adverse reactions, such as lipodystrophy, dyslipidemia, and rash, with antiretroviral therapy than do men. This observation is explained in part by sex differences in the pharmacokinetics of certain antiretroviral drugs, resulting in higher plasma concentrations in women.

## OBESITY

**(See also Chap. 402.)** The prevalence of both obesity ( $BMI \geq 30 \text{ kg/m}^2$ ) and abdominal obesity (waist circumference  $\geq 88 \text{ cm}$  in women) are similar in U.S. women and men. In 2018, the age-adjusted prevalence of obesity in U.S. adults was 42.4%, and there were no significant differences between women and men, even across different age groups. In 2018, the prevalence of obesity was 41.9% for women and 43.0% for men. The prevalence of obesity was highest among non-Hispanic black women (56.9%) as compared with non-Hispanic white (39.8%), Hispanic (43.7%), and non-Hispanic Asian women (17.2%). Non-Hispanic black women had a higher prevalence of obesity compared to non-Hispanic black men. There were no significant differences in prevalence between men and women among non-Hispanic white, non-Hispanic Asian, or Hispanic adults. More than 80% of patients who undergo bariatric surgery are women. Pregnancy and menopause are risk factors for obesity.

There are major sex differences in body fat distribution. Women characteristically have a gluteal and femoral or gynoid pattern of fat distribution, whereas men typically have a central or android pattern. Women have more subcutaneous fat than men. In women, endogenous androgen levels are positively associated with abdominal obesity, and androgen administration increases visceral fat. In contrast, there is an inverse relationship between endogenous androgen levels and abdominal obesity in men. Further, androgen administration decreases visceral fat in these obese men. The reasons for these sex differences in the relationship between visceral fat and androgens are unknown; however, emerging evidence suggests that there is a contribution of genetic variation. Studies in humans also suggest that sex steroids play a role in modulating food intake and energy expenditure.

In men and women, abdominal obesity characterized by increased visceral fat is associated with an increased risk for CVD and DM. Obesity increases a woman's risk for certain cancers, in particular postmenopausal breast and endometrial cancer, in part because adipose tissue provides an extragonadal source of estrogen through aromatization of circulating adrenal and ovarian androgens, especially the conversion of androstenedione to estrone. Obesity increases the risk of infertility, miscarriage, and complications of pregnancy.

## OSTEOPOROSIS

**(See also Chap. 411.)** Osteoporosis is about five times more common in postmenopausal women than in age-matched men, and osteoporotic

hip fractures are a major cause of morbidity in elderly women. Men accumulate more bone mass and lose bone more slowly than do women. Sex differences in bone mass are found as early as infancy. Calcium intake, vitamin D, and estrogen all play important roles in bone formation and bone loss. Particularly during adolescence, calcium intake is an important determinant of peak bone mass. Vitamin D deficiency is surprisingly common in elderly women, occurring in >40% of women living in northern latitudes. Receptors for estrogens and androgens have been identified in bone. Estrogen deficiency is associated with increased osteoclast activity and a decreased number of bone-forming units, leading to net bone loss. The aromatase enzyme, which converts androgens to estrogens, is also present in bone. Estrogen is an important determinant of bone mass in men (derived from the aromatization of androgens) as well as in women.

## PHARMACOLOGY

On average, women have lower body weights, smaller organs, a higher percentage of body fat, and lower total-body water than men. There are also important sex differences in drug action and metabolism that are not accounted for by these differences in body size and composition. Sex steroids alter the binding and metabolism of a number of drugs. Further, menstrual cycle phase and pregnancy can alter drug action. Women also take more medications than men, including over-the-counter formulations and supplements. The greater use of medications combined with these biologic differences may account for the reported higher frequency of adverse drug reactions in women than in men.

Two-thirds of cases of drug-induced torsades des pointes, a rare, life-threatening ventricular arrhythmia, occur in women because they have a longer, more vulnerable QT interval. These drugs, which include certain antihistamines, antibiotics, antiarrhythmics, and antipsychotics, can prolong cardiac repolarization by blocking cardiac voltage-gated potassium channels.

## PSYCHOLOGICAL DISORDERS

**(See also Chap. 452.)** Depression, anxiety, and affective and eating disorders (bulimia and anorexia nervosa) are more common in women than in men. Epidemiologic studies from both developed and developing nations consistently find major depression to be twice as common in women as in men, with the sex difference becoming evident in early adolescence. Depression occurs in 10% of women during pregnancy and in 10–15% of women during the postpartum period. There is a high likelihood of recurrence of postpartum depression with subsequent pregnancies. The incidence of major depression diminishes after the age of 45 years and does not increase with the onset of menopause. Depression in women appears to have a worse prognosis than does depression in men; episodes last longer, and there is a lower rate of spontaneous remission. Schizophrenia and bipolar disorders occur at equal rates in men and women, although there may be sex differences in symptoms.

Both biologic and social factors account for the greater prevalence of depressive disorders in women. Men have higher levels of the neurotransmitter serotonin. Sex steroids also affect mood, and fluctuations during the menstrual cycle have been linked to symptoms of premenstrual syndrome. Sex hormones differentially affect the hypothalamic-pituitary-adrenal responses to stress. Testosterone appears to blunt cortisol responses to corticotropin-releasing hormone. Both low and high levels of estrogen can activate the hypothalamic-pituitary-adrenal axis.

## COVID-19 INFECTION

**(See also Chap. 199.)** Soon after the discovery of COVID-19, which was identified in November 2019 in Wuhan, China, as being caused by the novel coronavirus SARS-CoV-2, it was evident that there were appreciable sex differences in severity and outcomes. Indeed, initial observational data from the winter of 2019 and spring of 2020 demonstrated a higher overall incidence of infectious cases, hospitalizations, intensive care unit admissions, and case-fatality rates among men as compared to women. More pronounced sex differences were observed with advanced age, with a higher overall incidence in older male age groups. These sex differences have persisted among different racial, ethnic, and socioeconomic groups and across all continents, as SARS-CoV-2 became a global pandemic.

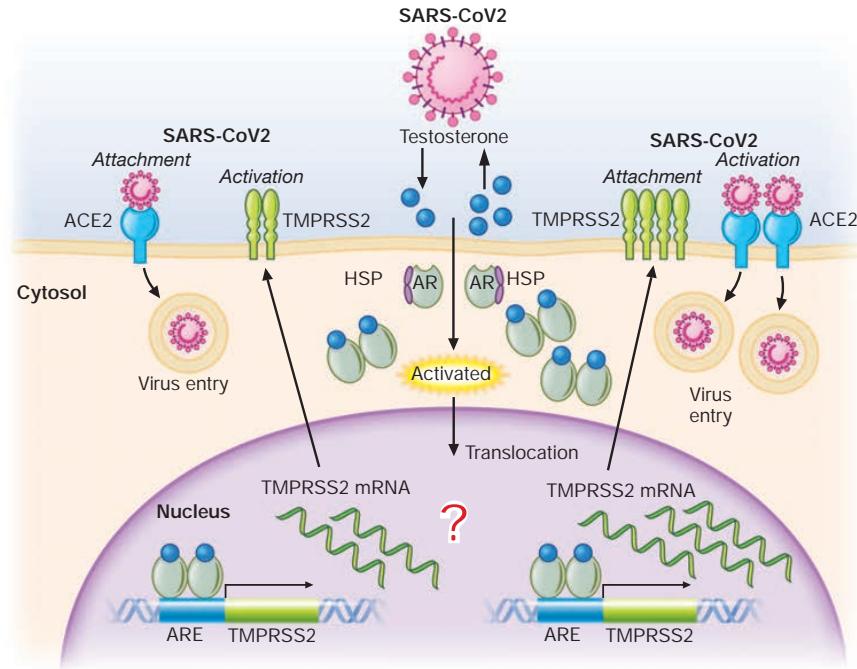
There are several potential mechanisms for these sex-specific effects of SARS-CoV-2 infection. The virus's entry point into cells is the membrane-bound angiotensin-converting enzyme 2 (ACE2) receptor, and it also harnesses the protease TMPRSS2, a cellular serine protease (Fig. 398-3). Circulating levels of ACE2, which is expressed in a variety of tissues including the lung, heart, and kidneys, have been reported to be relatively higher in men who have diabetes and/or kidney disease, as well as in healthy men; however, not all studies have reported similar sex differences. One hypothesis is that upregulation of the ACE2 receptor in men may provide greater opportunity for cellular entry, viral replication, and development of symptoms and deleterious sequelae.

ACE2 plays a critical role in bronchial transient secretory cells/type II alveolar cells as well as the renin-angiotensin-aldosterone system (RAAS). In the RAAS, ACE2 opposes angiotensin II's vasoconstrictive actions by converting angiotensin II to the vasodilatory angiotensin 1-7 in critical tissues, including cardiac myocytes, cardiac fibroblasts, and coronary endothelial cells. Importantly, recent evidence has shown the impact of sex and sex hormones on the RAAS and ACE2: estrogen downregulates angiotensin II receptor type 1 and regulates renin activity, as well as modulates local RAAS in the atrial myocardium. Furthermore, it has been shown that ovariectomized females have increased ACE2 activity and expression in their kidney and adipose tissue and that estradiol replacement reduces ACE2 expression. On the contrary, orchectomized males have decreased ACE2 activity. Estrogen appears to reduce ACE2 expression in the heart and kidney in both rodent and human studies.

TMPRSS2, a vital contributor to SARS-CoV-2 cellular invasion, is a protein that is most abundantly expressed in prostate epithelial tissue, including high-grade prostate cancers and metastases. Accordingly, the protein's involvement in viral priming is thought to be an important reason for the higher case-fatality rate observed in men; however, the association has not yet been proven. TMPRSS2 is also expressed in airway epithelia, where its physiologic function is not entirely clear. Transcription of the cellular protein is regulated by androgenic ligands and androgen receptor binding element; it is unknown whether estrogen plays a role in its regulation. Emerging *in vitro* studies have demonstrated that a TMPRSS2 inhibitor blocks viral entry into cells. These data may serve as an important foundation for sex-specific and personalized therapeutic approaches in the future.

### SUBSTANCE ABUSE AND TOBACCO

(See also Chaps. 453 and 454.) Substance abuse is more common in men than in women. However, one-third of Americans who suffer from alcoholism are women. Women are less likely to be diagnosed with alcoholism than men. A greater proportion of men than women seek help for alcohol and drug abuse. Men are more likely to go to an alcohol or drug treatment facility, whereas women tend to approach a primary care physician or mental health professional for help under the guise of a psychosocial problem. Blood alcohol levels are higher in women than in men after drinking equivalent amounts of alcohol, adjusted for body weight. This greater bioavailability of alcohol in women is due to both the smaller volume of distribution and the slower gastric metabolism of alcohol secondary to lower activity of gastric alcohol dehydrogenase than is the case in men. Women with



**FIGURE 398-3** Proposed sex hormone differences in TMPRSS2-mediated SARS-CoV-2 host cell entry. The virus entry point into cells is the membrane-bound angiotensin-converting enzyme 2 (ACE2) receptor. The cell-membrane protease, TMPRSS2, is also vital for host cell entry. Circulating levels of ACE2, expressed abundantly in the lung, heart, and kidney tissues, have been reported to be relatively higher in men. Upregulation of the ACE2 receptor in men may provide greater opportunity for cellular entry, viral replication, symptom development, and multiorgan involvement. (Reproduced with permission from C Gebhard et al: Impact of sex and gender on COVID-19 outcomes in Europe. *Biol Sex Differ* 11:29, 2020.)

alcoholism have a higher mortality rate than do women and men without alcoholism. Women also appear to develop alcoholic liver disease and other alcohol-related diseases with shorter drinking histories and lower levels of alcohol consumption. Alcohol abuse also poses special risks to a woman, adversely affecting fertility and the health of the baby (fetal alcohol syndrome). Even moderate alcohol use increases the risk of breast cancer, hypertension, and stroke in women.

More men than women smoke tobacco, but this sex difference continues to decrease. Women have a much larger burden of smoking-related disease. Smoking markedly increases the risk of CVD in premenopausal women and is also associated with a decrease in the age of menopause. Women who smoke are more likely to develop chronic obstructive pulmonary disease and lung cancer than men and at lower levels of tobacco exposure. Postmenopausal women who smoke have lower bone density than women who never smoked. Smoking during pregnancy increases the risk of preterm deliveries and low birth weight infants.

### VIOLENCE AGAINST WOMEN

More than one in four women in the United States have experienced rape, physical violence, and/or stalking by an intimate partner. Adult women are much more likely to be raped by a spouse, ex-spouse, or acquaintance than by a stranger. Intimate partner violence (IPV) is a leading cause of death among young women. Rates of reported IPV in the United States increased dramatically amid stay-at-home orders during the COVID-19 pandemic. IPV is an important risk factor for depression, substance abuse, and suicide in women. Screening instruments can accurately identify women experiencing IPV and should be administered in settings that ensure adequate privacy and safety.

### SUMMARY

Women's health is now a mature discipline, and the importance of sex differences in biologic processes is well recognized. Nevertheless, ongoing misperceptions about disease risk, not only among women but

also among their health care providers, result in inadequate attention to modifiable risk factors. Research into the fundamental mechanisms of sex differences will provide important biologic insights. Furthermore, those insights will have an impact on both women's and men's health.

## FURTHER READING

- Bunders MJ, Altfeld M: Implications of sex differences in immunity for SARS-CoV-2 pathogenesis and design of therapeutic interventions. *Immunity* 53:3, 2020.
- Manson JE et al: Menopausal estrogen-alone therapy and health outcomes in women with and without bilateral oophorectomy: A randomized trial. *Ann Intern Med* 171:6, 2019.
- National Institutes of Health: Sex/gender influences in health and disease. <https://orwh.od.nih.gov/sex-gender-sexgender-influences-health-and-disease>.
- Vegeiro E et al: The role of sex and sex hormones in neurodegenerative diseases. *Endocr Rev* 41:273, 2020.
- Zhao D et al: Endogenous sex hormones and incident cardiovascular disease in post-menopausal women. *J Am Coll Cardiol* 71:22, 2018.

increase in mortality rates among people aged 25–64 was the highest in the Ohio Valley and Appalachia, and in the New England states of New Hampshire, Maine, and Vermont. The rising mortality rates in young and middle-aged men have been attributed to an increase in deaths due to drug overdose, alcohol-related liver disease, and suicide. The rising rates of “deaths of despair” among young and middle-aged men, especially white non-Hispanics, have been associated with deterioration in economic and social well-being, reduced rates of marriage and labor force participation, and poor physical and mental health.

The biologic bases of sex differences in disease susceptibility, progression, and manifestation remain incompletely understood and are likely multifactorial. Undoubtedly, sex-specific differences in the genetic architecture and circulating sex hormones influence disease phenotype; additionally, epigenetic effects of sex hormones during fetal life, early childhood, and pubertal development may epigenetically imprint sexual and nonsexual behaviors, body composition, and disease susceptibility. The circulating and tissue concentrations of sex hormones differ substantially in men and women, and these hormonal differences may affect gene expression in cells of males and females in all parts of the body. The presence of only one X chromosome in men renders them more susceptible to X-linked disorders than women. Due to the X inactivation of one randomly chosen X chromosome, women's bodies contain two epigenetically different cell populations. The genes that do not undergo X inactivation exhibit dosage differences between male and female cells. Expression of the Y chromosome genes in men may affect the function of somatic cells containing the Y chromosome. The differences in the imprinting of maternally and paternally derived genes may also contribute to sex differences in the expression of disease. Reproductive load and physiologic changes during pregnancy, including profound hormonal and metabolic shifts and microchimerism (transfer of cells from the mother to the fetus and from the fetus to the mother), may affect disease susceptibility and disease severity in women. Sociocultural norms of child-rearing practices, societal expectations of gender roles, and the long-term economic impact of these practices and gender roles influence health behaviors and disease risk. Furthermore, the trajectories of age-related changes in sex hormones during the reproductive and postreproductive years vary substantially between men and women and influence the sex-specific patterns of the temporal evolution of age-related conditions such as osteoporosis, breast cancer, and autoimmune disease.

In a reflection of the growing attention on issues related to men's health, men's health clinics have mushroomed all over the country. Although the major threats to men's health have not changed—heart disease, cancer, and unintentional injury continue to dominate the list of major medical causes of morbidity and mortality in men—the men who attend men's health clinics do so largely for sexual, reproductive, and urologic health concerns involving common conditions, such as androgen deficiency syndromes, age-related decline in testosterone levels, sexual dysfunction, muscle dysmorphia and anabolic-androgenic steroid (AAS) use, lower urinary tract symptoms (LUTS), and medical complications of prostate cancer therapy, which are the subjects of this chapter. Additionally, we are witnessing the emergence of new categories of body image disorders in men that had not been recognized until the 1980s, such as the body dysmorphia syndrome and the use of performance-enhancing drugs to increase muscularity and lean appearance. Although menopause has been the subject of intense investigation for more than five decades, these issues that are specific to men's health are just beginning to gain the attention that they deserve because of their high prevalence and impact on overall health, well-being, and quality of life.

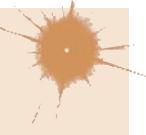
## AGING-RELATED CHANGES IN MALE REPRODUCTIVE FUNCTION

A number of cross-sectional and longitudinal studies (e.g., the Baltimore Longitudinal Study of Aging, the Framingham Heart Study [FHS], the Massachusetts Male Aging Study, and the European Male Aging Study [EMAS]) have established that testosterone concentrations decrease with advancing age. This age-related decline starts in the third decade of life and progresses slowly (**Fig. 399-1**); the rate

## 399

## Men's Health

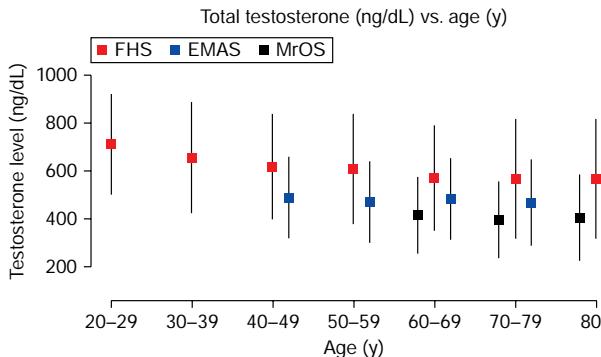
Shalender Bhansali



The emergence of men's health as a distinct discipline within internal medicine is founded on the wide consensus that men and women differ across their life span in their susceptibility to disease, in the clinical manifestations of the disease, and in their response to treatment. Furthermore, men and women weigh the health consequences of illness differently and have different motivation for seeking care. Men and women experience different types of disparities in access to health care services and in the manner in which health care is delivered to them because of a complex array of socioeconomic and cultural factors. Attitudinal and institutional barriers to accessing care, fear and embarrassment due to the perception that it is not manly to seek medical help, and reticence on the part of patients and physicians in discussing issues related to sexuality, drug use, and aging have heightened the need for programs tailored to address the specific health needs of men.

The sex differences in disease prevalence, susceptibility, and clinical manifestations of the disease were discussed in **Chap. 398** (Women's Health) and will not be discussed here. It is notable that the two leading causes of death in both men and women—heart disease and cancer—are the same. However, men have higher prevalence of neurodevelopmental and degenerative disorders, substance use disorders, including the use of performance-enhancing drugs and alcohol dependence, diabetes, and cardiovascular disease, and women have a higher prevalence of autoimmune disorders, depression, rheumatologic disorders, and osteoporosis. Men are substantially more likely to die from accidents, suicides, and homicides than women. Among men 15–34 years of age, unintentional injuries, homicides, and suicides account for over three-fourths of all deaths. Among men 35–64 years of age, heart disease, cancer, and unintentional injuries are the leading causes of death. Among men 65 years of age, heart disease, cancer, lower respiratory tract infections, and stroke are the major causes of death.

From 1999 to 2010, the mortality rates in the United States decreased for men and women of all age groups, largely due to reduced death rates from heart attacks, cancer, motor vehicle injuries, and HIV infection. However, during the past decade, troubling disparities in sex-specific mortality rates have emerged among middle-aged men in the United States. From 2010 to 2017, the death rates have risen and life expectancy has decreased for young and middle-aged men. The



**FIGURE 399-1** Age-related decline in total testosterone levels. Total testosterone levels measured using liquid chromatography-tandem mass spectrometry in men of the Framingham Heart Study (FHS), the European Male Aging Study (EMAS), and the Osteoporotic Fractures in Men Study (MrOS). (Reproduced with permission from S Bhasin et al: Reference ranges for testosterone in men generated using liquid chromatography tandem mass spectrometry in a community-based sample of healthy nonobese young men in the Framingham Heart Study and applied to three geographically distinct cohorts. *J Clin Endocrinol Metab* 96:2430, 2011.)

of decline in testosterone concentrations is greater in obese men, in men with chronic illness, and in those taking medications than in healthy older men. Because sex hormone-binding globulin (SHBG) concentrations are higher in older men than in younger men, free or bioavailable testosterone concentrations decline with aging to a greater extent than total testosterone concentrations. The age-related decline in testosterone is due to defects at all levels of the hypothalamic-pituitary-testicular (HPT) axis: pulsatile gonadotropin-releasing hormone (GnRH) secretion is attenuated, luteinizing hormone (LH) response to GnRH is reduced, and testicular response to LH is impaired. However, the gradual rise of LH with aging suggests that testis dysfunction is the main cause of declining androgen levels. The magnitude and trajectory of age-related decline in testosterone levels are affected by adiposity and weight change, comorbid conditions, and genetic factors. In the EMAS, 2.1% of community-dwelling men aged 40–70 years had total testosterone levels <317 ng/dL and a free testosterone level of <64 pg/mL, as well as sexual symptoms.

In epidemiologic surveys, low total and bioavailable testosterone concentrations in middle-aged and older men have been associated with decreased sexual desire, poor erections, and diminished early morning erections; lower appendicular skeletal muscle mass, muscle strength, and self-reported physical function; increased risk of mobility limitation and falls; higher visceral fat mass, insulin resistance, and type 2 diabetes; reduced telomere length and increased all-cause and cardiovascular mortality; lower areal and volumetric bone mineral density and bone quality; and higher rates of bone fractures (**Table 399-1**). Hypogonadal men often report low mood. However, testosterone levels have not been consistently associated with major depressive disorder; rather, low testosterone levels are more robustly associated with late-onset, low-grade, persistent depressive disorder previously referred to as dysthymia. An analysis of signs and symptoms in older men in the EMAS revealed a syndromic association of sexual symptoms with total testosterone levels <320 ng/dL and free testosterone levels <64 pg/mL in community-dwelling older men. Neither testosterone nor dihydrotestosterone levels are associated with the risk of prostate cancer or LUTS.

Mendelian randomization studies using data from the United Kingdom Biobank Study found a sexual dimorphic relation between genetically determined testosterone levels and the risk of type 2 diabetes; in men, lower genetically determined testosterone levels were associated with higher risk of type 2 diabetes, but in women, higher genetically determined testosterone levels were associated with higher risk of type 2 diabetes. Higher genetically determined testosterone levels were also associated with increased risk of prostate cancer in men in this study.

**TABLE 399-1** Association of Testosterone Levels with Outcomes in Older Men

1. Positively associated with:
  - Muscle mass and muscle strength
  - Self-reported and performance-based measures of physical function
  - Sexual desire
  - Bone mineral density, bone geometry and quality, and volumetric bone mineral density
2. Negatively associated with risk of:
  - Coronary artery disease
  - Type 2 diabetes mellitus
  - Metabolic syndrome
  - All-cause mortality
  - Falls and fracture risk
  - Dementia and Alzheimer's disease
  - Frailty
  - Late-onset low-grade persistent depressive disorder (dysthymia)
3. Not associated with:
  - Lower urinary tract symptoms
  - Erectile dysfunction
  - Major depressive disorder

Among the small number of randomized trials that have evaluated the efficacy of testosterone treatment in older men, the Testosterone Trials (TTrials)—a set of seven coordinated placebo-controlled trials of testosterone replacement conducted in 788 community-dwelling older men aged 65 years or older, who had an average of two morning, fasting total testosterone levels, measured using liquid chromatography-tandem mass spectrometry (LC-MS/MS), <275 ng/dL—have provided the most comprehensive data on the efficacy of testosterone treatment. The eligible men in the TTrials were required to have one or more of the following: low sexual desire, mobility limitation, and/or fatigue, and they were allocated using minimization to receive either placebo gel or testosterone gel for 1 year. Testosterone treatment was associated with greater improvement in overall sexual activity, sexual desire, erectile function, and satisfaction with sexual experience than placebo (**Table 399-2**). Testosterone treatment improved volumetric as well as areal bone density and estimated bone strength more than placebo; the improvements in volumetric bone density in the spine were greater than in the hip and greater in the trabecular than the peripheral bone. Testosterone treatment also corrected anemia in a greater proportion of older men with unexplained anemia of aging than placebo. No trials have been large enough or long enough to determine the long-term benefits of testosterone treatment in older men on clinically important outcomes such as disability, fractures, progression from prediabetes to diabetes, remission of dysthymia, and progression to Alzheimer's disease (AD) in men at risk of AD.

Testosterone therapy of healthy older men with low or low-normal testosterone levels is associated with greater increments in lean body mass, grip strength, and some measures of physical function than those associated with placebo (**Fig. 399-2**). In the TTrials, testosterone therapy of older hypogonadal men with self-reported mobility limitation consistently improved self-reported walking ability and modestly improved 6-min walking distance across all TTrials participants but did not affect falls. Testosterone treatment of hypogonadal men without a depressive disorder has been associated with a small but significantly greater improvement in depressive symptoms compared to placebo; however, testosterone treatment alone or as an adjunct to anti-depressant pharmacologic therapy has not been found to be efficacious in major depressive disorder. Two small, randomized trials in men with late-onset, low-grade persistent depressive disorder (dysthymia) have reported improvements in depressive symptoms in dysthymic men with low testosterone levels. In a large randomized trial (T4DM Trial) in men aged 50–74 years without hypogonadism, who had a waist circumference of 95 cm and impaired glucose tolerance or newly diagnosed type 2 diabetes, testosterone treatment in conjunction with

**TABLE 399-2 The Main Findings of the Testosterone Trials (TTrials)**

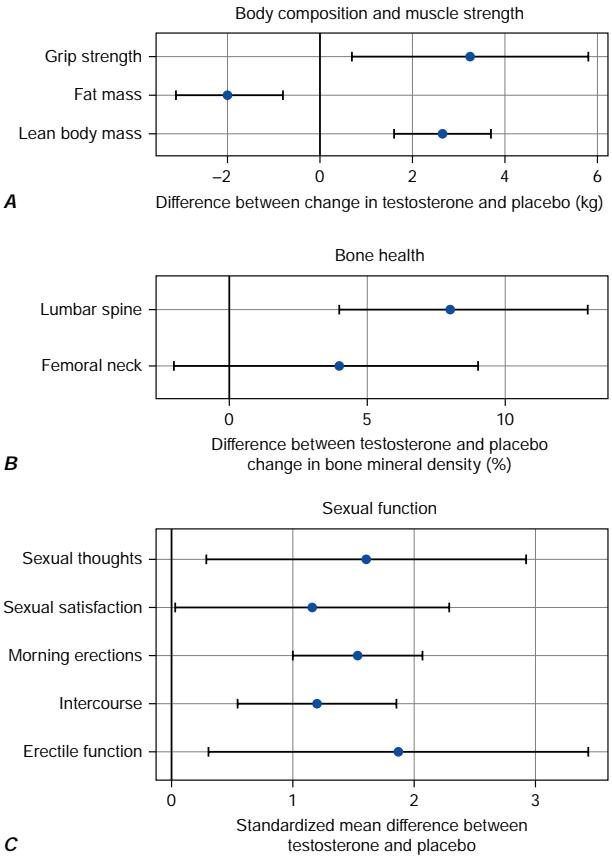
TRIAL	PRIMARY OUTCOME(S)	MAIN FINDINGS
Sexual Function Trial	Sexual activity	Testosterone treatment improved sexual activity, sexual desire, and erectile function.
Physical Function Trial	Distance walked over 6 min and self-reported physical function	Testosterone treatment consistently improved self-reported walking ability and modestly improved 6-min walk test distance across all TTrials participants but did not affect falls.
Vitality Trial	Energy measured using the Functional Assessment of Chronic Illness Therapy (FACT-F)	Testosterone did not improve energy but modestly improved mood and depressive symptoms.
Anemia Trial	The proportion of men with unexplained anemia who increased their hemoglobin 1.0 g/dL and experienced correction of anemia	Testosterone treatment, compared to placebo, was associated with a greater proportion of men with unexplained anemia increasing their hemoglobin by 1.0 g/dL and correcting their anemia.
Cognitive Function Trial	Delayed paragraph recall (Wechsler-Memory scale, a measure of memory)	Testosterone treatment did not improve delayed paragraph recall, visual memory, spatial ability, subjective memory complaints, or global cognitive function.
Bone Trial	Volumetric bone mineral density (vBMD) assessed using quantitative computed tomography	Testosterone treatment increased vBMD of the trabecular as well as peripheral bone in the spine and hip and increased estimated bone strength in the spine and hip more than placebo.
Cardiovascular Trial	Noncalcified coronary artery plaque volume determined by computed tomographic angiography	Testosterone treatment was associated with greater increase in the volume of noncalcified plaque in the coronary arteries than placebo.

Note: The TTrials were a set of seven coordinated placebo-controlled trials whose primary goal was to determine whether testosterone treatment for 1 year of men aged 65 years or older with an average of two morning, fasting total testosterone levels <275 ng/dL plus one or more of three conditions (low sexual desire, mobility limitation, and/or low vitality) was more efficacious than placebo in improving sexual function, mobility, and/or vitality. The other four linked trials evaluated the effects of testosterone treatment on volumetric bone mineral density, anemia, cognitive function, and coronary artery plaque volume.

a lifestyle program for 2 years reduced the proportion of participants with type 2 diabetes more than placebo plus lifestyle program. Testosterone therapy has not been shown to improve fracture risk, cognitive function, or response to phosphodiesterase inhibitors in older men.

Neither the long-term risks nor the clinical benefits of testosterone therapy in older men have been demonstrated in adequately powered trials. Erythrocytosis is the most frequent adverse event associated with testosterone treatment. While there is no evidence that testosterone causes prostate cancer, there is concern that testosterone therapy might cause subclinical prostate cancers to grow. Testosterone therapy is associated with increased risk of detection of prostate events. Testosterone does not worsen LUTS in older men who do not have severe LUTS prior to treatment.

There is no clear evidence that testosterone treatment increases the risk of major adverse cardiovascular events (MACE). No randomized trial to date has been long enough or large enough to determine whether testosterone increases the risk of MACE. In two placebo-controlled trials, the rates of atherosclerosis progression did not differ significantly between the testosterone and placebo groups. In the Cardiovascular Trial of the TTrials, testosterone treatment was associated with a greater increase in the volume of the noncalcified plaque, compared to placebo. A large randomized trial to determine the effects of testosterone replacement therapy on MACE in middle-aged and older hypogonadal men aged 45–85 years (TRAVERSE Trial, ClinicalTrials.gov identifier: NCT03518034) is in progress. The number of venous



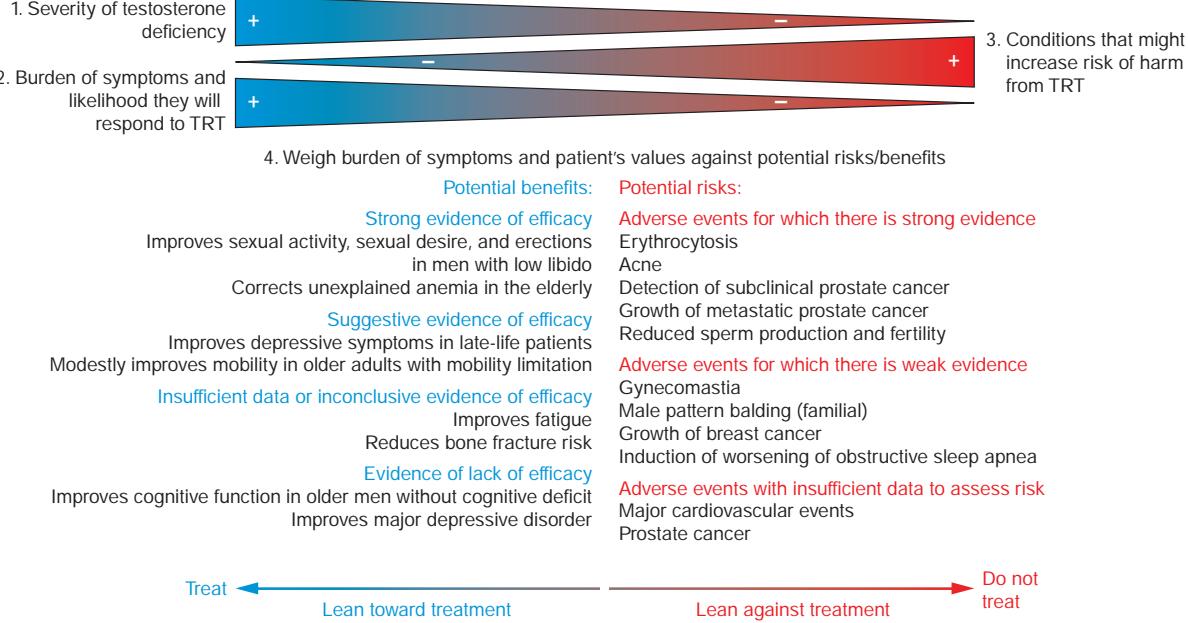
**FIGURE 399-2** The effects of testosterone therapy on body composition, muscle strength, bone mineral density (BMD), and sexual function in intervention trials. The point estimates and the associated 95% confidence intervals are shown. **A.** The effects of testosterone therapy on lean body mass, grip strength, and fat mass in a meta-analysis of randomized trials. **B.** The effects of testosterone therapy on lumbar and femoral BMD in a meta-analysis of randomized trials. **C.** The effects of testosterone therapy on measures of sexual function in men with baseline testosterone <10 nmol/L (290 ng/dL). (**A**, Data from S Bhasin et al: Drug insight: Testosterone and selective androgen receptor modulators as anabolic therapies for chronic illness and aging. *Nat Clin Pract Endocrinol Metab* 2:146, 2006; **B**, Data from MJ Tracz et al: Testosterone use in men and its effects on bone health. A systematic review and meta-analysis of randomized placebo-controlled trials. *J Clin Endocrinol Metab* 91:2011, 2006. **C**, Data from AM Isidori et al: Effects of testosterone on sexual function in men: results of a meta-analysis. *Clin Endocrinol (Oxf)* 63:381, 2005.)

thromboembolic events in randomized testosterone trials has been too small to draw meaningful inferences. The risk for venous thromboembolic events may be increased in men with hypercoagulable states.

## APPROACH TO THE PATIENT

### Older Men with Age-Related Decline in Testosterone

Population screening of all older men for low testosterone levels is not recommended; testing should be restricted to men who have symptoms or physical features attributable to androgen deficiency. Testosterone treatment of older men with symptomatic testosterone deficiency offers some clinical benefits (e.g., improvement of sexual symptoms in men with low libido, correction of anemia), but because of the lack of evidence of long-term safety and limited evidence of long-term efficacy, an expert panel of the Endocrine Society recommended against testosterone treatment of *all* older men with low testosterone levels. Instead, the expert panel recommended that “testosterone therapy should be offered on an



**FIGURE 399-3** An individualized, patient-centric approach to shared treatment decision-making in older men with testosterone deficiency. Testosterone treatment is not indicated in all older men with low testosterone levels. The decision to treat should be individualized based on considerations of the severity of testosterone deficiency, the burden of symptoms and conditions associated with testosterone deficiency, the presence of conditions that might increase the patient's risk of harm from testosterone treatment, and the patient's values and willingness to accept the uncertainty of the long-term benefits and risks and the burden of treatment and monitoring. TRT, testosterone replacement therapy. (Reproduced with permission from S Bhasin. Testosterone replacement in aging men: an evidence-based patient-centric perspective. *J Clin Invest* 131:e146607, 2021.)

individualized basis . . . in men >65 years who have symptoms or conditions suggestive of testosterone deficiency (e.g., low libido or unexplained anemia) and consistently low testosterone.” The decision to offer testosterone treatment to older men with low testosterone levels should be a shared decision, guided by an individualized assessment of potential benefits and risks, and careful weighing of the burden of symptoms/conditions against the potential benefits and risks (Fig. 399-3). Evaluate whether the patient has clear evidence of testosterone deficiency. The diagnosis of testosterone deficiency should be made on the basis of two or more early morning, fasting testosterone levels below the lower limit of normal for healthy young men plus the presence of symptoms. Weigh the burden of symptoms/conditions against the known benefits and the uncertainty of long-term harm. Ascertain whether the patient has any conditions that might increase the risk of harm, such as prostate cancer, severe LUTS, erythrocytosis, or deep-venous thrombosis. Older men considering testosterone supplementation should undergo baseline evaluation of risk factors for prostate cancer. The initiation of prostate screening and monitoring should be a shared decision because prostate cancer screening has some risks. A shared decision to treat should be accompanied by a standardized monitoring plan to optimize the benefit-to-risk ratio.

## AGE-RELATED CHANGES IN FECUNDITY

Although testicular morphology, semen production, and fertility are maintained up to a very old age in men, advanced paternal age is a risk factor for reduced fertility. Compared to men aged 21–25 years, men >50 years old have lower sperm motility and sperm morphology, a higher frequency of sperm tail defects, and lower fecundity. The fecundity is reduced when both parents are >40 years old. Increased workforce participation and changing career expectations of women, a higher age at reproductive union, and the availability of contraceptives that enable couples to separate their sexual and procreative lives have underpinned powerful secular trends toward postponement of childbearing to an older age. The median age at first childbirth has been

increasing steadily across the world; postponement of childbirth to an older age increases the risk of involuntary childlessness because of the adverse effects of advanced maternal and paternal age on fecundity, increased risk of comorbidities that may indirectly affect fecundity, and the age-related changes in reproductive behaviors. Increased paternal age is associated with increased risk of germline mutations in the *FGFR2*, *FGFR3*, and *RET* genes and the associated autosomal dominant diseases, such as achondroplasia, Pfeiffer's syndrome, Crouzon's syndrome, Apert's syndrome, multiple endocrine neoplasia (MEN) 2A, and MEN 2B. Advanced paternal age also increases the risk of Klinefelter's syndrome, trisomy 13 and 18, neurodevelopmental disorders such as schizophrenia, autism, bipolar disorders, and cardiac malformations such as ventricular septal defects, atrial septal defects, and patent ductus arteriosus.

**Sexual Dysfunction** Various forms of sexual dysfunction are a major motivating factor for men seeking care at men's health clinics. The landmark descriptions of the human sexual response cycle by Masters and Johnson demonstrating that men and women display predictable physiologic responses after sexual stimulation provided the basis for rational classification of human sexual disorders. Accordingly, sexual disorders have been classified into four categories depending on phase of sexual response cycle in which the abnormality exists:

1. Hypoactive sexual desire disorder
2. Erectile dysfunction
3. Ejaculatory and orgasmic disorders
4. Disorders of pain

Classification of the patient's disorder into these categories is important as the etiologic factors, diagnostic tests, and therapeutic strategies vary for each class of sexual disorder. Historically, the classification and nomenclature for sexual disorders were based on the *Diagnostic and Statistical Manual of Mental Disorders* (DSM), based on the erroneous belief that sexual disorders in men are largely psychogenic in origin. However, the recognition of erectile dysfunction as a manifestation of systemic disease and the availability of easy-to-use oral selective phosphodiesterase-5 (PDE5) inhibitors have placed sexual disorders in men

within the purview of the primary care provider. These disorders have been discussed in [Chap. 397](#) (Sexual Dysfunction).

### MUSCLE DYSMORPHIA SYNDROME IN MEN—A FORM OF BODY IMAGE DISORDER

Muscle dysmorphia is a form of body image disorder characterized by a pathologic preoccupation with muscularity and leanness. The men with muscle dysmorphia express a strong desire to be more muscular and lean. These men describe shame and embarrassment about their body size and shape and often report aversive symptoms such as dissatisfaction with appearance, preoccupation with bodybuilding and muscularity, and functional impairment. Patients with muscle dysmorphia also report higher rates of mood and anxiety disorders and obsessive and compulsive behaviors than individuals with no history of muscle dysmorphia. These men often experience impairment of social and occupational functioning.

Patients with muscle dysmorphia syndrome—nearly all men—are almost always engaged in weightlifting and body building and are more likely to use performance-enhancing drugs, especially AASs, than men in the general population or even weightlifters without body dysmorphia. Muscle dysmorphia disorder exposes men to an increased risk of disease due to the combined interactive effects of the intensity of physical exercise, the use of performance-enhancing drugs, and other lifestyle factors associated with weightlifting and the use of performance-enhancing drugs. These patients are also at increased risk of functioning poorly in their occupation and social life than men without this disorder. No randomized trials of any treatment modalities have been conducted; anecdotally, behavioral and cognitive therapies have been tried with varying degrees of success.

**AAS Abuse by Athletes and Recreational Bodybuilders** The illicit use of AASs to enhance athletic performance first surfaced in the 1950s among powerlifters and spread rapidly to other sports, professional as well as high school athletes, and recreational bodybuilders. In the early 1980s, the use of AAS spread beyond the athletic community into the general population. As many as 3 million Americans—most of them men—have likely used these compounds. Most AAS users are not athletes, but rather recreational weightlifters who use these drugs to look lean and more muscular.

The most commonly used AASs include testosterone esters, nandrolone, stanozolol, methandienone, and methenolol. AAS users generally use increasing doses of multiple steroids in a practice known as stacking.

The adverse effects of long-term AAS abuse remain poorly understood. Most of the information about the adverse effects of AAS has emerged from case reports, uncontrolled studies, or clinical trials that used replacement doses of testosterone ([Table 399-3](#)). The adverse event data from clinical trials using physiologic replacement doses of testosterone have been extrapolated unjustifiably to AAS users who may administer 10–100 times the replacement doses of testosterone over many years and to support the claim that AAS use is safe. A substantial fraction of AAS users also use other drugs that are perceived to be muscle-building or performance-enhancing, such as growth hormone; erythropoiesis-stimulating agents; insulin; stimulants such as amphetamine, clenbuterol, cocaine, ephedrine, and thyroxine; and drugs perceived to reduce adverse effects such as human chorionic gonadotropin (hCG), aromatase inhibitors, or estrogen antagonists. Men who abuse AAS are more likely to engage in other high-risk behaviors than nonusers. The adverse events associated with AAS use may be due to AAS themselves, concomitant use of other drugs, high-risk behaviors, and host characteristics that may render these individuals more susceptible to AAS use or to other high-risk behaviors.

The high rates of mortality and morbidities observed in AAS users are alarming. One Finnish study reported 4.6 times the risk of death among elite power lifters than in age-matched men from the general population. The causes of death among power lifters included suicides, myocardial infarction, hepatic coma, and non-Hodgkin's lymphoma. A retrospective review of patient records in Sweden also reported higher standardized mortality ratios for AAS users than for nonusers.

**TABLE 399-3 Potential Adverse Effects Associated with the Use of Anabolic-Androgenic Steroids**

ORGAN SYSTEM	EFFECT
Cardiovascular	Dyslipidemia Atherosclerotic disease Sudden death Myocardial fibrosis, cardiomyopathy Cardiac conduction abnormalities Hypertension
Neuroendocrine	HPT suppression: hypogonadism on AAS withdrawal Gynecomastia
Females	Virilizing effects
Neuropsychiatric	Major mood disorders (mania, hypomania, depression) Aggression, violence AAS dependence Neuronal apoptosis: cognitive deficits
Hematologic	Polycythemia Hypercoagulability and thrombosis
Hepatic	Inflammatory and cholestatic effects Peliosis hepatitis (rare) Neoplasms (rare)
Musculoskeletal	Premature epiphyseal closure (in adolescents) Tendon rupture
Kidney	Renal failure secondary to rhabdomyolysis Focal segmental glomerulosclerosis
Dermatologic	Acne Striae

*Abbreviations:* AAS, anabolic-androgenic steroids; HPT axis, hypothalamic-pituitary-testicular axis.

*Source:* Data from HG Pope Jr et al: Adverse health consequences of performance-enhancing drugs: an Endocrine Society scientific statement. *Endocr Rev* 35:341, 2014.

Studies indicate that 32% of deaths among AAS users were suicidal, 26% homicidal, and 35% accidental. The median age of death among AAS users—24 years—is even lower than that for heroin or amphetamine users.

Numerous reports of cardiac death among young AAS users raise concerns about the adverse cardiovascular effects of AAS. High doses of AAS may induce proatherogenic dyslipidemia, increase thrombosis risk via effects on clotting factors and platelets, induce vasospasm through their effects on vascular nitric oxide, and induce myocardial hypertrophy and fibrosis.

Replacement doses of testosterone, when administered parenterally, are associated with only a small decrease in high-density lipoprotein (HDL) cholesterol and little or no effect on total cholesterol, low-density lipoprotein (LDL) cholesterol, and triglyceride levels. In contrast, supraphysiologic doses of testosterone and orally administered, 17-*-alkylated, nonaromatizable AAS are associated with marked reductions in HDL cholesterol and increases in LDL cholesterol.*

Recent studies of AAS users using tissue Doppler and strain imaging and magnetic resonance imaging (MRI) have reported diastolic and systolic dysfunction, including significantly lower early and late diastolic tissue velocities, reduced E/A ratio, and reduced peak systolic strain in AAS users than in nonusers. Power athletes using AAS often have short QT intervals but increased QT dispersion, which may predispose them to ventricular arrhythmias. Long-term AAS use may be associated with myocardial hypertrophy and fibrosis. Myocardial tissue of power lifters using AAS has been shown to be infiltrated with fibrous tissue and fat droplets. AAS users demonstrate higher coronary artery plaque volume than nonusers, and lifetime AAS dose is associated with coronary atherosclerotic burden.

Long-term AAS use suppresses LH and follicle-stimulating hormone (FSH) secretion and inhibits endogenous testosterone production and spermatogenesis. Men who have used AAS for more than a few months experience marked suppression of the HPT axis after stopping AAS

that may be associated with sexual dysfunction, fatigue, infertility, and depressive symptoms. In some AAS users, HPT suppression may last more than a year, and in a few individuals, complete recovery may not occur. The symptoms of androgen deficiency during AAS withdrawal may cause some men to revert back to using AAS, leading to continued use and AAS dependence. As many as 30% of AAS users develop a syndrome of AAS dependence, characterized by long-term AAS use despite adverse medical and psychiatric effects. In some men's health clinics, as many as 25% of young men receiving testosterone replacement therapy have anabolic steroid withdrawal hypogonadism.

Supraphysiologic doses of testosterone may also impair insulin sensitivity. Orally administered androgens have been associated with insulin resistance and diabetes.

Unsafe injection practices, high-risk behaviors, and increased rates of incarceration render AAS users at increased risk of HIV and hepatitis B and C. In one survey, nearly 1 in 10 gay men had injected AAS or other substances, and AAS users were more likely to report high-risk unprotected anal sex than other men.

Some AAS users develop hypomanic and manic symptoms during AAS exposure (irritability, aggressiveness, reckless behavior, and occasional psychotic symptoms, sometimes associated with violence) and major depression (sometimes associated with suicidality) during AAS withdrawal. Users may also develop other forms of illicit drug use, which may be potentiated or exacerbated by AAS.

AAS use has been associated with difficulties with spatial as well as working memory, problem solving, and attention, and structural and functional changes in many brain regions involved in inhibitory control and emotional regulation. A structural MRI study of users of high doses of AAS reported smaller cortical, gray matter, putamen, and corpus callosum volumes. Both low and high androgen levels have been associated with increased A<sub>1</sub> and tau-P levels and A<sub>2</sub> toxicity. These data have raised concern that long-term AAS use may increase the risk of Alzheimer's disease and related dementias.

Elevated liver enzymes, cholestatic jaundice, hepatic neoplasms, and peliosis hepatitis have been reported with oral, 17 $\alpha$ -alkylated AAS. AAS use may cause muscle hypertrophy without compensatory adaptations in tendons, ligaments, and joints, thus increasing the risk of tendon and joint injuries. AAS use is associated with acne, baldness, and increased body hair.

## APPROACH TO THE PATIENT

### Detection of AAS Use

AAS users generally mistrust physicians and seek medical help infrequently; when they do seek medical help, it is often for the treatment of AAS withdrawal syndrome, infertility, gynecomastia, or other medical or psychiatric complications of AAS use. The suspicion of AAS use should be raised by the increased hemoglobin and hematocrit levels; suppressed LH, FSH, and testosterone levels; low HDL cholesterol; and low testicular volume and sperm density in a person who looks highly muscular (**Table 399-4**). A combination of these findings along with self-report of their use by the patient—which usually can be elicited by a tactful interview—are often sufficient to establish a diagnosis in clinical practice.

Accredited laboratories use gas chromatography and mass spectrometry or liquid chromatography and mass spectrometry to detect AAS abuse. In recent years, the availability of high-resolution mass spectrometry and tandem mass spectrometry has further improved the sensitivity of detecting AAS abuse. Illicit testosterone use is detected generally by the application of the measurement of the urinary testosterone-to-epitestosterone ratio and further confirmed by the use of the  $^{13}\text{C}:\text{^{12}\text{C}}$  ratio in testosterone by the use of isotope ratio combustion mass spectrometry. Exogenous testosterone administration increases urinary testosterone glucuronide excretion and consequently the testosterone-to-epitestosterone ratio. Ratios  $>4$  suggest exogenous testosterone use but can also reflect genetic variation. Genetic variations in the uridine diphospho-glucuronyl

**TABLE 399-4 Detection of the Use of Anabolic-Androgenic Steroids**

Clinical indicators that should raise suspicion of anabolic-androgenic steroid use
1. Very muscular phenotype
2. Reduced testicular volume (<15 mL)
Laboratory indicators
1. Suppressed LH and FSH levels
2. Increased hematocrit
Detection of anabolic-androgenic steroids
1. LC-MS/MS analysis of urine
Detection of exogenous testosterone use
1. Urinary testosterone-to-epitestosterone ratio
2. Isotope ratio mass spectrometry analysis to detect differences in $^{13}\text{C}:\text{^{12}\text{C}}$ ratio in exogenous and endogenous testosterone

*Note:* In clinical settings, the use of anabolic-androgenic steroids can often be ascertained simply by direct questioning. Reduced testicular volume, suppressed LH and FSH, and increased hematocrit in an unusually muscular man should raise suspicion of anabolic-androgenic steroid use. Although rarely needed in clinical practice, recent use of anabolic-androgenic steroids can be confirmed by LC-MS/MS analysis of urine. Exogenous testosterone use can be detected using the urinary testosterone-to-epitestosterone ratio and isotope ratio mass spectrometry analysis to detect differences in  $^{13}\text{C}:\text{^{12}\text{C}}$  ratio in exogenous and endogenous testosterone.

*Abbreviations:* FSH, follicle-stimulating hormone; LC-MS/MS, liquid chromatography-tandem mass spectrometry; LH, luteinizing hormone.

transferase 2B17 (*UGT2B17*), the major enzyme for testosterone glucuronidation, affect the testosterone-to-epitestosterone ratio. Synthetic testosterone has a lower  $^{13}\text{C}:\text{^{12}\text{C}}$  ratio than endogenously produced testosterone, and these differences in the  $^{13}\text{C}:\text{^{12}\text{C}}$  ratio can be detected by isotope ratio combustion mass spectrometry and used to confirm exogenous testosterone use in individuals with a high testosterone-to-epitestosterone ratio.

## TREATMENT

### Integrated Management of Patients with AAS Use

The nonathlete weightlifters who abuse AAS frequently do not seek medical treatment and often mistrust physicians. They also do not view these drugs and the associated lifestyle as deleterious to their health. In turn, many internists erroneously view AAS abuse as largely a problem of cheating in competitive sports, while, in fact, most AAS users are not athletes at all. Also, physicians often have a poor understanding of the factors motivating the use of these performance-enhancing drugs, the long-term health effects of AAS, and the associated psychopathologies that may affect treatment choices.

In addition to treating the underlying body dysmorphism disorder that motivates the use of these drugs, the treatment should be directed at the symptoms or the condition for which the patient seeks therapy, such as infertility, sexual dysfunction, gynecomastia, or depressive symptoms. Accordingly, therapy may include some combination of cognitive and behavioral therapy for muscle dysmorphia syndrome, antidepressant therapy for depression, selective PDE5 inhibitors for erectile dysfunction, and/or use of selective estrogen receptor modulators or aromatase inhibitors to reactivate HPT axis or hCG to restore testosterone levels.

As discussed above, AASs suppress the male hypothalamic-pituitary-gonadal axis, and men with long-term AAS use may experience symptoms of profound androgen deficiency such as sexual dysfunction, fatigue, and depressive symptoms during AAS withdrawal. Some of these patients may resume AAS use or start using other drugs to combat the distressing withdrawal symptoms. There are no randomized trials of any therapies for AAS withdrawal. Case reports and clinical experience suggest that administration of selective estrogen receptor modulators, CYP19 aromatase inhibitors, or hCG may restore circulating testosterone levels. Clomiphene citrate, a partial estrogen receptor agonist, administered in a dose

of 25–50 mg on alternate days, can increase LH and FSH levels and restore testosterone levels in a vast majority of men with AAS withdrawal syndrome. However, the recovery of sexual function during clomiphene administration is variable despite improvements in testosterone levels. Anecdotally, other aromatase inhibitors such as anastrozole have also been used. hCG, administered by intramuscular injections of 750–1500 IU three times each week, can raise testosterone levels into the normal range. Some patients may not respond to either clomiphene or hCG therapy, raising the possibility of irreversible long-term toxic effects of AAS on Leydig cell function.

Adjunctive cognitive and behavioral therapy or antidepressants to treat depression inadequately responsive to endocrine therapies alone may be needed. Emerging human and animal evidence suggests AAS and opioids likely promote dependence via common mechanisms. The opioid antagonist naltrexone blocks AAS dependence in animals. Therefore, treatments for human opioid dependence might also benefit AAS dependence. Many patients who abuse AAS suffer from body-image disorder and require psychiatric treatment for this underlying disorder.

### LUTS IN MEN

LUTS in men include storage symptoms (urgency, daytime as well as nighttime frequency, and urgency incontinence), voiding disturbances (slow or intermittent stream, difficulty in initiating micturition, straining to void, pain or discomfort during the passage of urine, and terminal dribbling), or postmicturition symptoms (a sense of incomplete voiding after passing urine and postmicturition dribble). The overactive bladder syndrome refers to urgency with or without urgency incontinence, usually with urinary frequency and nocturia, and is often due to detrusor muscle overactivity. A presumptive diagnosis of benign prostatic hyperplasia should be made only in men with LUTS, who have demonstrable evidence of prostate enlargement and obstruction based on the size of the prostate. LUTS have historically been attributed to benign prostatic hyperplasia, although it has become apparent that the pathophysiologic mechanisms of LUTS are complex and multifactorial and may include structural or functional abnormalities of the bladder, bladder neck, prostate, distal sphincter mechanism, and urethra, as well as abnormalities in the neural control of the lower urinary tract. Diuretics, antihistamines, antidepressants, and other medications that have anticholinergic properties can cause or exacerbate LUTS in older men. The intensity of LUTS tends to fluctuate over time.

LUTS is highly prevalent in older men, affecting nearly 50% of men >65 and 70% of men >80 years old. The LUTS adversely affects quality of life because of its impact on sleep, ability to perform activities of daily living, and depressive symptoms. LUTS is often associated with erectile dysfunction.

### APPROACH TO THE PATIENT

#### Lower Urinary Tract Symptoms

Medical evaluation should include assessment of potential causes of symptoms; medications including herbal and over-the-counter products that might contribute to symptoms; the symptom severity and bother using an International Prostate Symptom Score; and in some patients, a frequency-volume chart. The impact of LUTS on sleep, activities of daily living, and quality of life should be evaluated. Evaluation should also include digital prostate examination, neurologic examination focused on perineum and lower extremities, urinalysis, fasting blood glucose, electrolytes, creatinine, and prostate-specific antigen (PSA). Urodynamic studies are not required in most patients but are recommended when invasive surgical therapies are being considered. A urologic referral may be appropriate if the patient has hydronephrosis, renal insufficiency, recurrent urinary tract infections, hematuria, or history of acute urinary retention.

## TREATMENT

### Patients with LUTS

Considerations of the severity of symptoms; the impact of symptoms on sleep, activities of daily living, and quality of life; the natural history of the disease; and potential adverse effects of the intervention should guide the decision to intervene. In men with mild to moderately severe LUTS, the symptoms typically progress slowly over many years and may remain stable or even improve in some men. The men who have mild symptoms can usually be reassured and followed. Several simple steps such as reducing caffeine and alcohol intake, especially late in the day, taking the diuretic medication early in the day, avoiding excessive water intake close to bedtime, bladder training, pelvic floor exercises including biofeedback to promote pelvic floor relaxation, and timed voiding regimens or double voiding to ensure complete emptying of the bladder may be helpful in reducing the severity of symptoms. Men with mild to moderate bothersome LUTS can be treated effectively using  $\alpha_1$ -adrenergic antagonists, steroid 5'-reductase inhibitors, PDE5 inhibitors, or anticholinergic agents alone or in combination. Selective  $\alpha_1$ -adrenergic antagonists are typically the first line of therapy; their side effects include hypotension, dizziness, nasal congestion, retrograde or delayed ejaculation, and rarely floppy iris syndrome. In men with probable benign prostate obstruction with gland enlargement and LUTS, therapy with steroid 5'-reductase inhibitors, finasteride, or dutasteride for 1 or more years improves urinary symptoms and flow rate and reduces prostatic volume. Long-term treatment with 5'-reductase inhibitors can reduce the risk of acute urinary retention and need for prostate surgery. Combined administration of a steroid 5'-reductase inhibitor and an  $\alpha_1$ -adrenergic blocker can rapidly improve urinary symptoms and reduce the relative risk of acute urinary retention and surgery. PDE5 inhibitors, when administered chronically alone or in combination with  $\alpha_1$ -adrenergic blockers, are effective in improving LUTS and erectile dysfunction through their effects on nitric oxide-cyclic guanosine monophosphate (cGMP) in the bladder, urethra, and prostate. PDE5 inhibitors do not improve uroflow parameters, and their hypotensive effect may be potentiated by  $\alpha_1$ -adrenergic blockers. Anticholinergic drugs are used for the treatment of overactive bladder in men with prominent irritative symptoms, such as frequency, urgency, and incontinence, and no evidence of elevated postvoid residual urine. Containment products, such as pads, can help improve social life in men who have severe storage symptoms, including incontinence. Surgery is indicated when medical therapy fails, symptoms progress in spite of medical therapy, or the patient develops acute urinary retention, hydronephrosis, renal insufficiency, or recurrent urinary tract infections, or if the patient has postvoid residual urine volume >25% of the urinary bladder volume.

### MEDICAL COMPLICATIONS OF PROSTATE CANCER THERAPY

Prostate cancer is the most common malignancy in American men, accounting for 19% of all diagnosed cancers and ~8% of all cancer deaths; its incidence is on the rise, partly due to increased screening with PSA. The American Cancer Society estimates that, in 2021, 248,530 new cases of prostate cancer will be diagnosed in the United States and 34,130 men will die from this disease. The majority of these men have low-grade, organ-confined prostate cancer and excellent prospects of long-term survival. Substantial improvement in survival in men with prostate cancer has focused attention on the high prevalence of sexual dysfunction, physical dysfunction, and low vitality in the men, which are important contributors to poor quality of life among patients treated for prostate cancer. The pathophysiology of these symptoms after radical prostatectomy is multifactorial, but denervation and androgen deficiency are important contributors to these symptoms.

Androgen deficiency is common in men with prostate cancer. Testosterone levels decline with age, and men with prostate cancer are at risk of having low testosterone levels simply by virtue of their age.

**3076** However, total and free testosterone levels are even lower in men with prostate cancer who have undergone prostatectomy, when compared to noncancer age-matched controls. This age-related androgen deficiency in men with prostate cancer is associated with fatigue, sexual dysfunction, mobility limitation, and decreased physical function. Even with bilateral nerve-sparing procedure, >50% of men develop sexual dysfunction after surgery. Although there is some recovery of sexual function with passage of time, 40–50% of men undergoing radical prostatectomy find their sexual performance to be a moderate to large problem 18 months after surgery. Sexual problems are a source of psychosocial distress in men with localized prostate cancer. The men with locally advanced or metastatic prostate cancer who undergo androgen deprivation therapy (ADT) encounter even more distressing symptoms because of the profound androgen deficiency. In addition to fatigue, sexual dysfunction, and hot flushes, these men are at increased risk for diabetes, metabolic syndrome, coronary heart disease, and frailty.

**Testosterone Therapy in Men with a History of Prostate Cancer** A history of prostate cancer has historically been considered a contraindication for testosterone therapy. This guidance is based on observations that testosterone promotes the growth of metastatic prostate cancer. Metastatic prostate cancer generally regresses after orchidectomy and ADT. Androgen receptor signaling plays a central role in maintaining growth of normal prostate and prostate cancer. PSA levels are lower in hypogonadal men and increase after testosterone therapy. Prostate volume is lower in hypogonadal men and increases after testosterone therapy to levels seen in age-matched controls.

However, the role of testosterone in prostate cancer is complex. Epidemiologic studies and their meta-analyses have not revealed a consistent relationship between serum testosterone and prostate cancer. Others have reported that low testosterone levels are associated with high-grade cancers. In a landmark randomized trial, testosterone therapy of older men with low testosterone did not affect intraprostatic androgen levels or the expression of androgen-dependent prostatic genes. The suppression of circulating testosterone levels by a GnRH antagonist also does not affect intraprostatic androgen concentrations. Open-label trials and retrospective analyses of testosterone therapy in men with prostate cancer, who have undergone radical prostatectomy and have undetectable PSA levels after radical prostatectomy, have found very low rates of PSA recurrence. Even in men with high-grade prostatic intraepithelial neoplasia (HGPIN)—a group at high risk of developing prostate cancer—testosterone therapy for 1 year did not increase PSA or rates of prostate cancer.

A majority of men diagnosed with prostate cancer today have localized disease that can be potentially cured by radical prostatectomy. The men with organ-confined prostate cancer (pT2, N0, M0) and Gleason score <6 are at a very low risk of disease recurrence after radical prostatectomy with 0.5% biochemical recurrence rate and 0.2% local recurrence rate at >10–15 years. Similarly, preoperative PSA <10 ng/mL is associated with lower risk of disease recurrence than PSA >10 ng/mL. After radical prostatectomy, in the absence of residual cancer, PSA becomes undetectable within a month. An undetectable PSA after radical prostatectomy is a good indicator of biochemical recurrence-free survival at 5 years. Therefore, men with organ-confined prostate cancer (pT2), Gleason score <6, and a preoperative PSA of <10 ng/mL, who have had undetectable PSA levels (<0.1 ng/mL) for >2 years after radical prostatectomy, have very low risk of disease recurrence (<0.5% at 10 years) and may be considered for testosterone therapy on an individualized basis. If testosterone therapy is instituted, it should be associated with careful monitoring of PSA levels and close consultation with a urologist.

### MEDICAL COMPLICATIONS OF ADT

In patients with prostate cancer and distant metastases, ADT improves survival. In patients with locally advanced disease, ADT in combination with external beam radiation or as an adjuvant therapy (post-prostatectomy and pelvic lymphadenectomy) also has been shown to improve survival. However, ADT is being increasingly used as primary therapy in men with localized disease and in men encountering biochemical recurrence without clear evidence of survival advantage. The

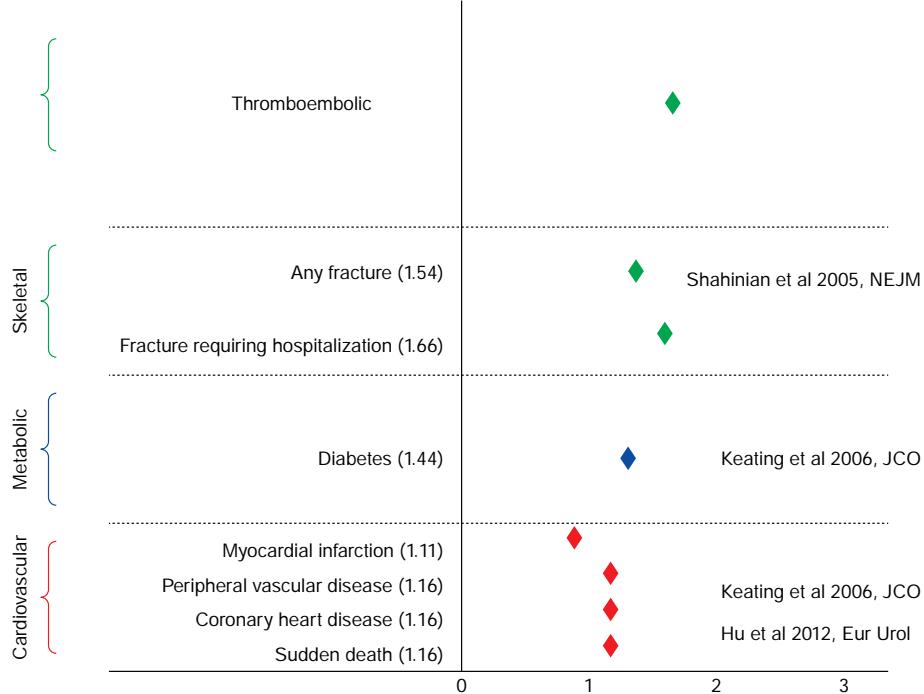
overall use of ADT in men with prostate cancer has increased in the past two decades, and its use in men with localized disease and biochemical recurrence accounts for a substantial fraction of this increase. Since most men with prostate cancer die of conditions other than their primary malignancy, recognition and management of these adverse effects is paramount.

Profound hypogonadism resulting from ADT is associated with sexual dysfunction, vasomotor symptoms, gynecomastia, decreased muscle mass and strength, frailty, increased fat mass, anemia, fatigue, bone loss, loss of body hair, depressive symptoms, and reduced quality of life. Diabetes and cardiovascular disease have recently been added to the list of these complications (Fig. 399-4). Treatment with GnRH agonists in men with prostate cancer is associated with rapid induction of insulin resistance, hyperinsulinemia, and a significant increase in the risk of incident diabetes. Metabolic syndrome is prevalent in >50% of men undergoing long-term ADT when compared to age-matched men with prostate cancer not undergoing ADT (22%) and their age-matched eugonadal counterparts (20%). Some but not all studies have reported an increased risk of cardiovascular events, death due to cardiovascular events, and peripheral vascular disease in men undergoing ADT. Some reports suggest that men receiving ADT are at an increased risk of thromboembolic events and cognitive dysfunction. The rates of acute kidney injury are higher in men currently receiving ADT than in men not receiving ADT; the increased risk appears to be particularly associated with the use of combined regimens of a GnRH agonist plus an antiandrogen. ADT also is associated with substantially increased risk of osteoporosis and bone fractures.

## APPROACH TO THE PATIENT

### Men Receiving ADT

The benefits of ADT in treating nonmetastatic prostate cancer should be carefully weighed against the risks of ADT-induced adverse events (Table 399-5). If ADT is medically indicated, consider whether intermittent ADT is a feasible option. Men being considered for ADT should undergo assessment of cardiovascular, diabetes, and fracture risk; this assessment may include measurement of blood glucose, plasma lipids, and bone mineral density by dual energy x-ray absorptiometry. Institute measures to prevent bone loss, including physical activity, adequate calcium and vitamin D intake, and pharmacologic therapy in men with a previous minimal trauma fracture and those with 10-year risk of a major osteoporotic fracture >20%, unless contraindicated. Bisphosphonates and denosumab have been shown to reduce fracture risk in men undergoing ADT, and zoledronic acid and denosumab have been approved by the U.S. Food and Drug Administration for the prevention of metastasis-related skeletal-related events in this population. Men with prostate cancer who are receiving ADT should be monitored for weight gain and diabetes. Encourage lifestyle interventions, including physical activity and exercise, and attention to weight, blood pressure, lipid profile, blood glucose, and smoking cessation to reduce the risk of cardiometabolic complications. In randomized trials, medroxyprogesterone, cyproterone acetate, and the serotonin reuptake inhibitor venlafaxine have been shown to be more efficacious than placebo in alleviating hot flushes. The side effects of these medications—increased appetite and weight gain with medroxyprogesterone, gynecomastia with estrogenic compounds, and dry mouth with venlafaxine—should be weighed against their relative efficacy. Acupuncture, soy products, vitamin E, herbal medicines, and transdermal estradiol have been used empirically for the treatment of vasomotor symptoms without clear evidence of efficacy. Gynecomastia can be prevented by the use of an antiestrogen, an aromatase inhibitor, or local radiation therapy; these therapies are effective in alleviating pain and tenderness but are less effective in reducing established gynecomastia. For long-standing gynecomastia that persists after cessation of ADT and is bothersome, mammoplasty is an effective treatment option.



**FIGURE 399-4** Adverse cardiometabolic and skeletal effects of androgen deprivation therapy (ADT) in men receiving ADT for prostate cancer. Administration of ADT has been associated with increased risk of thromboembolic events, fractures, and diabetes. Some, but not all, studies have reported increased risk of cardiovascular events in men receiving ADT. (Data from VB Shahinian et al: *N Engl J Med* 352:154, 2005; NL Keating et al: *J Clin Oncol* 24:4448, 2006; JC Hu et al: *Eur Urol* 61:1119, 2012.)

## PREVENTION OF SEXUALLY TRANSMITTED DISEASES

Adolescent boys and young men aged 15–24 years; men who have sex with men, have multiple sex partners, have unprotected sex without condom, or have sex with sex workers; men who use illicit drugs; men who have a history of previous sexually transmitted infection (STI); and transgender men are at increased risk for STIs. STIs increase the risk of oropharyngeal and anogenital cancers, liver disease, pelvic pain, infertility, inadvertent transmission of infection to others, and emergency department visits and are a preventable cause of excess morbidity and mortality. HIV, hepatitis B and C infections, and syphilis can have additional disease-specific complications. The prevention and treatment of STIs are discussed in Chap. 136. Additionally, the Centers for Disease Control and Prevention (CDC) and U.S. Preventive Services Task Force (USPSTF) have published guidelines on the prevention, treatment, and

pre- and postexposure prophylaxis of STIs. The approach to the prevention of STIs includes a structured risk assessment; counseling about safe sex practices including condom use; immunization of individuals at risk; diagnosis and treatment of infected individuals whether or not they are symptomatic; detection and treatment of sexual partners; and targeted sex education of adolescents and young men who are at high risk for STIs. The USPSTF recommends screening for HIV in all men aged 15–65 years, regardless of risk, and for hepatitis B virus and syphilis in men at increased risk. Because more than half of STIs occur in persons aged 15–24 years, the USPSTF also recommends behavioral counseling for all sexually active adolescents and adult men at increased risk of STIs to encourage condom use and other protective behaviors, including consideration of abstinence, reducing the number of sex partners, and avoidance of unsafe sex practices. Consistent and correct condom use is the most important method of preventing STIs. Effective immunizations are available against hepatitis B, human papillomavirus (HPV), and *Neisseria meningitidis*. The CDC's Advisory Committee on Immunization Practices (ACIP) recommends universal hepatitis B immunization for all unvaccinated adults presenting to an STI clinic, all HIV-infected adults, and health workers. Although ACIP recommends HPV vaccination in males aged 9–21 years and in men aged 9–26 if they have sex with men or have an immunocompromising condition, recent data suggest that the prevalence of HPV and its complications continue to increase until middle age, and some experts recommend extending the age limit for HPV vaccination. Meningococcal vaccination is indicated for men who have sex with men from an area of outbreak and for all HIV-infected men.

Because men seeking care in men's health clinics often do so for sexual and urogenital problems, these visits offer opportunities for counseling, screening, and treatment of STIs and institution of immunization and other preventive measures for STIs.

## SEX DIFFERENCES IN COVID-19 DISEASE OUTCOMES

The COVID-19 pandemic has highlighted sex differences in the susceptibility to respiratory viral infections. Men infected with SARS-CoV-2

**TABLE 399-5** Checklist for Men Undergoing Androgen Deprivation Therapy (ADT)

1. Weigh the risks and benefits of ADT and whether intermittent ADT is a feasible and safe option.
2. Perform a baseline assessment including fasting glucose, plasma lipids, blood pressure, bone mineral density, and FRAX score.
3. Optimize calcium and vitamin D intake, encourage structured physical activity and exercise, and consider pharmacologic therapy in men with a previous minimal trauma fracture and those with a 10-year risk of a major osteoporotic fracture >20%, unless contraindicated.
4. Monitor body weight, fasting glucose, plasma lipids, blood pressure, and bone mineral density, and encourage smoking cessation and physical activity.
5. In men who are receiving ADT and who experience bothersome hot flushes, as indicated by sleep disturbance or interference with work or activities of daily living, consider initial therapy with venlafaxine. If ineffective, add medroxyprogesterone acetate.
6. In men who experience painful breast enlargement, consider therapy with an estrogen receptor antagonist, such as tamoxifen.

**3078** virus are more likely to have a more severe disease, require mechanical ventilation, have disease complications, and die of the disease than women. Somewhat similar sex differences in morbidity and mortality have been reported for influenza infection. In the United States, the incidence and rates of hospitalization for influenza are higher in men than in women across all age groups. However, the sex-specific mortality rates associated with influenza vary substantially across countries and age groups. The sex differences in susceptibility to SARS-CoV-2 infection and morbidity have been attributed to behavioral factors, such as higher rates of smoking and alcohol use in men; biologic factors, such as higher rates of comorbid conditions in men than in women; sex differences in immune responses, including a poor T lymphocyte response to SARS-CoV-2 infection; and lower expression levels in men of X-linked genes that are involved in the innate detection of RNA viruses and that escape X inactivation in women, resulting in higher expression levels in women. Additionally, the expression of angiotensin-converting enzyme 2 (ACE2) and the cell surface transmembrane protease serine 2 (TMPRSS2), the two host proteins that facilitate the entry of SARS-CoV-2 into the alveolar cells, is regulated by androgens in subsets of lung epithelial cells, and it is possible that higher testosterone levels in men may contribute to increased susceptibility to the infection.

### FURTHER READING

- Abrams P et al: Evaluation and treatment of lower urinary tract symptoms in older men. *J Urol* 189:S93, 2013.
- Baggish A et al: Cardiovascular toxicity of illicit anabolic-androgenic steroid use. *Circulation* 135:1991, 2017.
- Basaria S: Cardiovascular disease associated with androgen deprivation therapy: Time to give it due respect. *J Clin Oncol* 33:1232, 2015.
- Bhasin S: Testosterone replacement in aging men: An evidence-based patient-centric perspective. *J Clin Invest* 131:e146607, 2021.
- Bhasin S et al: Testosterone therapy in men with hypogonadism: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 103:1715, 2018.
- Bhasin S et al: The implications of reproductive aging for the health, vitality and economic welfare of human societies. *J Clin Endocrinol Metab* 104:3821, 2019.
- Case A, Deaton A: Mortality and morbidity in the 21st century. *Brookings Papers on Economic Activity*, Spring 2017.
- Centers for Disease Control and Prevention: Mortality tables. <http://www.cdc.gov/nchs/deaths.htm> (HHS, CDC, NCHS).
- Centers for Disease Control and Prevention: Sexually transmitted diseases treatment guidelines. *MMWR Recomm Rep* 64:51, 2015.
- Choi PY et al: Muscle dysmorphia: A new syndrome in weightlifters. *Br J Sports Med* 36:375, 2002.
- Dos Santos MR, Bhasin S: Benefits and risks of testosterone treatment in men with age-related decline in testosterone. *Annu Rev Med* 72:75, 2021.
- Jones C et al: Management of lower urinary tract symptoms in men: Summary of NICE guidance. *BMJ* 340:c2354, 2010.
- López AM et al: Fracture risk in patients with prostate cancer on androgen deprivation therapy. *Osteoporos Int* 16:707, 2005.
- Pope HG Jr et al: Adverse health consequences of performance-enhancing drugs: An endocrine society scientific statement. *Endocr Rev* 35:341, 2014.
- Ruth KS et al: Using human genetics to understand the disease impacts of testosterone in men and women. *Nat Med* 26:252, 2020.
- Smith MR et al: Denosumab in men receiving androgen deprivation therapy for prostate cancer. *N Engl J Med* 361:745, 2009.
- Snyder PJ et al: Effects of testosterone treatment in older men. *N Engl J Med* 374:611, 2016.
- U.S. Preventive Health Services Task Force. Final recommendation statement sexually transmitted infections: Behavioral counseling. <https://www.uspreventiveservicestaskforce.org/Page/Document/RecommendationStatementFinal/sexually-transmitted-infections-behavioral-counseling1>. Accessed June 21, 2017.
- Wittert G et al: Testosterone treatment to prevent or revert type 2 diabetes in men enrolled in a lifestyle programme (T4DM): A

randomised, double-blind, placebo-controlled, 2-year, phase 3b trial. *Lancet Diabetes Endocrinol* 9:32, 2021.

Woolf SH, Schoomaker H: Life expectancy and mortality rates in the United States, 1959–2017. *JAMA* 322:1996, 2019.

400

## Lesbian, Gay, Bisexual, and Transgender (LGBT) Health

Baligh R. Yehia, Zachary B. R. McClain



### UNDERSTANDING LGBT HEALTH DISPARITIES

The acceptance of the lesbian, gay, bisexual, and transgender (LGBT) community has greatly increased over the past decade in certain communities and parts of the world. However, numerous studies highlight health disparities involving the care of LGBT people. Lesbian and bisexual women are less likely to receive recommended preventive screenings such as breast, cervical, and colorectal cancer screenings. Among men who have sex with men, rates of human papillomavirus-associated anal cancers are 17 times higher than those of heterosexual men. In addition, gay and bisexual men accounted for 70% of all new HIV diagnoses in the United States in 2018, and they disproportionately contract sexually transmitted infections. In 2018, men who have sex with men accounted for 64% of primary and secondary syphilis infections in the United States where the sex of the sexual partner was known. Transgender individuals have a higher prevalence of HIV infection and suicide compared with other groups.

Research has found that LGBT individuals are more likely to experience depression, anxiety, and alcohol and drug use than their counterparts. Most concerning are the rates of suicide attempts and ideation among the LGBT community, particularly youth. Lesbian, gay, and bisexual (LGB) youth are four times more likely to attempt suicide than their heterosexual peers, and 61% of gender variant youth reported suicidal ideation at some point in their life. Additionally, the recent U.S. Transgender Survey found that 40% of transgender young adults and adults reported attempting suicide at some point in their lives.

In addition, U.S. studies indicate that substance abuse is twice as common in LGBT youth compared with their counterparts. These findings are mirrored among LGBT adults: the prevalence of substance abuse disorders is 20–30% compared with ~9% in the general population.

These health issues are compounded by structural barriers to health care, including decreased access to medical care, lack of awareness to the unique health needs of LGBT individuals, and stigma and discrimination toward the LGBT community. Many LGB individuals perceive the health care setting and providers as threatening, which may lead to avoiding needed medical care or withholding important medical information. A large U.S. survey identified that 8% of LGB and 27% of transgender individuals were refused needed health care, and almost 11% of LGB and 21% of transgender people reported being subjected to harsh or abusive language by health care professionals. Apart from health care settings, more than two-thirds of LGB people report discrimination in their personal lives, and 90% of transgender individuals report harassment, mistreatment, or discrimination at work. Chronic exposure to high levels of stress from real or anticipated discrimination, referred to as “minority stress,” may be an important factor contributing to the poor health outcomes experienced by LGBT populations.

While some research on LGBT health has been conducted, there remains a great opportunity to better understand the needs and experiences of LGBT individuals. Moreover, many LGBT individuals experience health disparities across their life cycle (e.g., LGBT youth

are at greater risk of suicide and homelessness, whereas elderly LGBT individuals face barriers to health because of isolation and fewer family supports), necessitating a longitudinal approach to examining LGBT health issues. There are more limited data on the health of LGBT individuals outside the United States and Europe. However, studies demonstrate that problems are greatest when people cannot be open about their sexual orientation and gender identity. Encouraging greater LGBT acceptance and access to health care will be critical to improving outcomes and experiences for LGBT communities.

## CREATING POSITIVE HEALTH EXPERIENCES FOR LGBT PATIENTS

**Understanding Gender Identity and Sexual Orientation**  
Addressing health disparities and creating positive health care experiences require an understanding of the diversity of cultural expression and lives of LGBT persons. First, providers must be able to distinguish *gender identity* from *sexual orientation*. Gender identity is a person's internal sense of their gender. It should not be confused with sex assigned at birth, which is based on anatomy and biology. Gender

identity expands beyond the binary male and female and includes persons who think of their gender as containing elements of both or neither. Many individuals who do not identify with the gender that correlates with their sex assigned at birth often use the terms *transgender* or *trans-male* or *trans-female* to identify themselves. Sexual orientation refers to how one thinks of their physical or emotional attraction to others. Sexual orientation has three dimensions: attraction, behavior, and identity. *Attraction* refers to one's desire to be with someone, regardless of one's behavior or stated identity. For example, a woman may be attracted to another woman, but this attraction may never be acted upon and may not form part of her sexual identity. *Behavior* refers to a person's sexual and romantic partners. Although sexual identity often aligns with behavior, some individuals who identify as heterosexual may have same-gender partners and some individuals who identify as lesbian or gay may have different-gender partners. Lastly, *identity* refers to how a person defines their own sexuality. Common terms for sexual identity include *gay*, *lesbian*, *bisexual*, *straight*, *heterosexual*, *homosexual*, and *asexual* (Table 400-1). As individuals go through the process of understanding their sexuality and self-identity over time, they may change how they define their sexual identity.

**TABLE 400-1** Common LGBT Terminology and Definitions

TERM	DEFINITION
<b>Agender</b>	Identifying as having no gender.
<b>Asexual</b>	Experiencing little or no sexual attraction to others.
<b>Assigned sex at birth</b>	The sex (male or female) assigned to a child at birth, most often based on the child's external anatomy. Also referred to as birth sex, natal sex, biological sex, or sex.
<b>Bisexual</b>	A sexual orientation that describes a person who is emotionally and sexually attracted to people of their own gender and people of other genders.
<b>Cisgender</b>	A person whose gender identity and assigned sex at birth correspond (i.e., a person who is not transgender).
<b>Gay</b>	A sexual orientation that describes a person who is emotionally and sexually attracted to people of their own gender. It can be used regardless of gender identity, but it is more commonly used to describe men.
<b>Gender dysphoria</b>	Distress experienced by some individuals whose gender identity does not correspond with their assigned sex at birth. Manifests as clinically significant distress or impairment in social, occupational, or other important areas of functioning.
<b>Gender expression</b>	The way a person acts, dresses, speaks, and behaves (i.e., feminine, masculine, androgynous). Gender expression does not necessarily correspond to assigned sex at birth or gender identity.
<b>Gender identity</b>	A person's internal sense of being a man/male, woman/female, both, neither, or another gender.
<b>Gender nonconforming</b>	Expressing a gender that differs from a given society's norms for males and females.
<b>Heterosexual</b>	A sexual orientation that describes women who are emotionally and sexually attracted to men, and men who are emotionally and sexually attracted to women.
<b>Intersex (disorders of sexual development)</b>	A group of rare conditions where the reproductive organs and genitals do not develop as expected.
<b>Lesbian</b>	A sexual orientation that describes a woman who is emotionally and sexually attracted to other women.
<b>Men who have sex with men (MSM)/women who have sex with women (WSW)</b>	Categories used in research and public health to describe those who engage in same-sex sexual behavior, regardless of their sexual orientation. Individuals rarely use the terms MSM or WSW to describe themselves.
<b>Pangender</b>	Describes a person whose gender identity comprises many genders.
<b>Pansexual</b>	A sexual orientation that describes a person who is emotionally and sexually attracted to people regardless of gender.
<b>Queer</b>	An umbrella term used by some to describe people who think of their sexual orientation or gender identity as outside of societal norms. Some people view the term as more fluid and inclusive than traditional categories for sexual orientation and gender identity. Due to its history as a derogatory term, it is not embraced or used by all members of the LGBT community.
<b>Questioning</b>	Describes an individual who is unsure about or is exploring their own sexual orientation and/or gender identity.
<b>Same-sex attraction</b>	Describes the experience of a person who is emotionally and/or sexually attracted to people of the same gender. Use of this term is not indicative of a person's sexual behavior.
<b>Sexual orientation</b>	Describes how a person characterizes their physical and emotional attraction to others. Sexual orientation is distinct from sex, gender identity, and gender expression.
<b>Trans man/transgender man/female-to-male (FTM)</b>	A transgender person whose gender identity is male may use these terms to describe themselves. Some will just use the term <i>man</i> .
<b>Trans woman/transgender woman/male-to-female (MTF)</b>	A transgender person whose gender identity is female may use these terms to describe themselves. Some will just use the term <i>woman</i> .
<b>Transgender</b>	Describes a person whose gender identity and assigned sex at birth do not correspond. Also used as an umbrella term to include gender identities outside of male and female.
<b>Transition/affirmation</b>	For transgender persons, the process of coming to recognize, accept, and express one's gender identity. Most often, this refers to the period when a person makes social, legal, and/or medical changes, such as changing their clothing, name, and sex designation, as well as using medical interventions.

*Note:* It is important to note that definitions vary across communities, that they change over time, and that not all LGBT people agree with all these definitions.

**3080** The creation of a welcoming environment requires not making any assumptions about an individual's gender identity or sexual orientation. Both front-line staff and clinicians should be cognizant of patient communication. For example:

- Instead of saying "How may I help you, sir?", say "How may I help you?"
- Instead of saying "She is here for her appointment," say "The patient is here in the waiting room."
- Instead of saying "Do you have a wife?", say "Are you in a relationship?"
- Instead of saying "What are your mother's and father's names?", say "What are your parents' names?"

### Developing Comfort and Competency in Sexual Health

Developing comfort discussing sexual health and intimacy is critical to providing appropriate care. After inquiring about healthy relationships and relationship status, a good starting place is to ask if a patient is sexually active and, if so, with whom, how often, and what types of physical interactions and types of sex they have with their partner(s). These discussions can allow clinicians to focus subsequent conversations on issues most relevant to a patient's health. For example, a gay man with multiple sexual partners who engages in receptive anal sex without condoms is at high risk for HIV and sexually transmitted infections (STIs). It will be important to recommend more frequent screenings for STIs and discuss use of preexposure prophylaxis (PrEP) and condoms to prevent HIV and STIs. Additionally, if you are seeing a transgender man, it will be important to know if he still has natal female genitalia to ensure appropriate cancer screening, desire for having biologic children, and contraceptive needs. Notably, many if not most transgender people have not had gender affirmation surgery and retain their natal sex organs.

### Creating a Welcoming and Safe Health Care Environment

Hospitals and clinics can take a number of steps to create a welcoming and safe space for LGBT patients. This starts by establishing and communicating a nondiscrimination policy that clearly includes gender identity, gender expression, and sexual orientation protections. Additionally, hospitals and clinics can develop and implement an equal visitation policy to ensure equal visitation for LGBT patients from same-sex partners, parents, and other family and friends. Staff training in LGBT patient-centered care also is a key component of creating inclusive health environments. This includes covering LGBT cultural competency, caring for LGBT patients, creating an inclusive environment for LGBT patients and staff, and other topics important for LGBT health.

As hospitals and clinics continue to adopt electronic health records, collecting sexual orientation and gender identity information becomes increasingly important to delivering personalized care to LGBT individuals. It allows providers to monitor quality of care and track population-based outcomes. This information can be captured by three questions:

- How do you think of yourself? As straight or heterosexual; lesbian, gay, or homosexual; bisexual; something else; don't know; choose not to disclose.
- What is your current gender identity? Male; female; transgender male/trans man/female-to-male (FTM); transgender female/trans woman/male-to-female (MTF); genderqueer, neither exclusively male nor female; additional category, please specify; choose not to disclose.
- What sex were you assigned at birth on your original birth certificate? Male; female; choose not to disclose.

The physical environment of a hospital or clinic is important, but the majority of clinical spaces do not signal that they are safe spaces for LGBT patients. Most health care posters, pamphlets, and materials feature heterosexual individuals or couples; adding LGBT-friendly images and text can help signal that the hospital or clinic is a safe space for sexual and gender minorities. In addition, easily identifying LGBT-competent providers by using websites, buttons, and pins can help patients select providers and feel at ease when attending

appointments. Lastly, designating all-gender bathrooms is important to creating welcoming spaces, particularly for transgender and gender-nonconforming individuals.

### FUTURE DIRECTION IN LGBT HEALTH

While social and cultural acceptance of the LGBT community has improved in certain parts of the world, many LGBT individuals continue to experience discrimination, stigmatization, and violence. Inequitable health care policies and practices, lack of awareness of LGBT health issues, and limited understanding of the unique health needs of LGBT individuals contribute to decreased access to care and disparities in health outcomes for LGBT individuals. Addressing these barriers will require improved data collection on the LGBT population; understanding of the intersectionality of gender identity, sexual orientation, race/ethnicity, and other sociocultural determinants of health; and outcomes-focused research across the life course. In striving to deliver high-quality care experiences for all patients, hospitals, clinics, and providers need to focus on meeting the needs of the LGBT community.

### FURTHER READING

- Centers for Disease Control and Prevention: Lesbian, Gay, Bisexual, and Transgender Health. 2018. Available at [www.cdc.gov/lgbthealth/](http://www.cdc.gov/lgbthealth/). Accessed December 10, 2019.
- Fenway Health: Glossary of LGBT Terms for Health Care Teams. March 2018. Available at [www.lgbthealtheducation.org/wp-content/uploads/LGBT-Glossary\\_March2016.pdf](http://www.lgbthealtheducation.org/wp-content/uploads/LGBT-Glossary_March2016.pdf). Accessed December 10, 2019.
- Institute of Medicine: The Health of Lesbian, Gay, Bisexual, and Transgender (LGBT) People: Building a Foundation for Better Understanding. 2011. Available at [www.nap.edu/catalog.php?record\\_id=13128](http://www.nap.edu/catalog.php?record_id=13128). Accessed December 10, 2019.
- Institute of Medicine: Collecting Sexual Orientation and Gender Identity Data in Electronic Health Records: Workshop Summary. 2013. Available at [www.nap.edu/catalog/18260/collecting-sexual-orientation-and-gender-identity-data-in-electronic-health-records](http://www.nap.edu/catalog/18260/collecting-sexual-orientation-and-gender-identity-data-in-electronic-health-records). Accessed December 10, 2019.
- Joint Commission: Advancing Effective Communication, Cultural Competence, Patient- and Family-Centered Care for the Lesbian, Gay, Bisexual, and Transgender Community: A Field Guide. 2011. Available at [www.jointcommission.org/lgbt](http://www.jointcommission.org/lgbt). Accessed December 10, 2019.
- National Center for Transgender Equality: The Report of the 2015 U.S. Transgender Survey. 2015. Available at [www.ustranssurvey.org/reports/USTS](http://www.ustranssurvey.org/reports/USTS). Accessed December 10, 2019.
- Safer JD, Tangpricha V: Care of transgender persons. N Engl J Med 381:2451, 2019.
- Substance Abuse and Mental Health Services Administration: Top Health Issues for LGBT Populations Information & Resource Kit. 2012. Available at [store.samhsa.gov/product/top-health-issues-lgbt-populations/sma12-4684](http://store.samhsa.gov/product/top-health-issues-lgbt-populations/sma12-4684). Accessed December 10, 2019.

## Section 3 Obesity, Diabetes Mellitus, and Metabolic Syndrome

**401**

### Pathobiology of Obesity

Stephen O'Rahilly, I. Sadaf Farooqi



Adipose tissue evolved as a solution to the challenge of the intermittent availability of food. At times when food is plentiful, excess calories are converted to triglycerides and efficiently stored in the unilocular lipid droplets that occupy most of the volume of fat cells. When needed, the triglyceride is rapidly broken down to free fatty acids and glycerol, which provide an energy source to other sites throughout the body.

However, in environments where food is abundant and when individuals tend to be sedentary, the chronic excess of energy intake over expenditure leads to obesity. The risks of becoming obese under those circumstances and of developing the illnesses associated with obesity vary greatly between individuals, with that variation having a strong genetic basis.

## DEFINITION OF OBESITY AND OVERWEIGHT

Obesity is defined as a state of excess adipose tissue mass that adversely affects health. The direct measurement of fat mass is not something that is readily undertaken in routine clinical practice, so a proxy measure, the body mass index (BMI), is generally used. This is calculated as weight/height<sup>2</sup> (in kg/m<sup>2</sup>) (Fig. 401-1). BMI-based definitions of obesity and overweight have been established based on associations with certain morbidities and excess mortality. These definitions have been based largely on studies of predominantly white, Western populations, and there is growing evidence that the relationship between BMI and adverse outcomes may be different in people from other ethnic groups, usually in the direction of worse health outcomes being seen at lower levels of BMI. The World Health Organization (WHO) defines a BMI of 30 kg/m<sup>2</sup> as the cutoff point for obesity, while individuals with values between 25 and 30 kg/m<sup>2</sup> are classified as overweight. For individuals with a very muscular body habitus, the BMI may overestimate the amount of body fat. For any given BMI, women will generally have a higher percentage of body fat than men.

The extent to which different adipose depots expand in response to chronic overnutrition varies markedly between people. In general, females store more fat in subcutaneous tissues, especially on buttocks, thighs, and upper arms, whereas men are more prone to store fat in intraabdominal and truncal subcutaneous sites. A simple measure of fat distribution is provided by a measurement of the waist-to-hip ratio. Independent of how obese a person is, a waist-to-hip ratio >0.9 in women and >1.0 in men is associated with adverse health outcomes such as type 2 diabetes and dyslipidemia.

## EPIDEMIOLOGY

The annual National Health and Nutrition Examination Survey (NHANES) provides an ongoing record of the prevalence of obesity in the United States. In 2017–2018, 42.4% of U.S. adults aged 20 years old were obese with no significant differences in prevalence by age group. Non-Hispanic black people had the highest prevalence of obesity at 49.6%, followed by Hispanic (44.8%), non-Hispanic white (42.2%), and non-Hispanic Asian (17.4%) people. In the United States, Asians represent a highly heterogeneous group encompassing both East and South Asia as well as a substantial Filipino community. The risks of obesity and its complications may differ greatly between people from different parts of Asia; in general, the prevalence of obesity is somewhat higher in women than in men, with black women having the

highest prevalence at 56.9%. There has been a marked increase in the prevalence of obesity over time. For example, between 1976 and 1980, the NHANES survey reported a prevalence of 14.5%, indicating a near threefold increase over the past 40 years.

This trend is seen globally. According to the WHO, obesity has nearly tripled worldwide since 1975. In 2016, >1.9 billion adults aged 18 years old were overweight. Of these, >650 million were obese; 39% of adults aged 18 years old were overweight in 2016, and 13% were obese. Most of the world's population lives in countries where overweight and obesity kills more people than underweight.

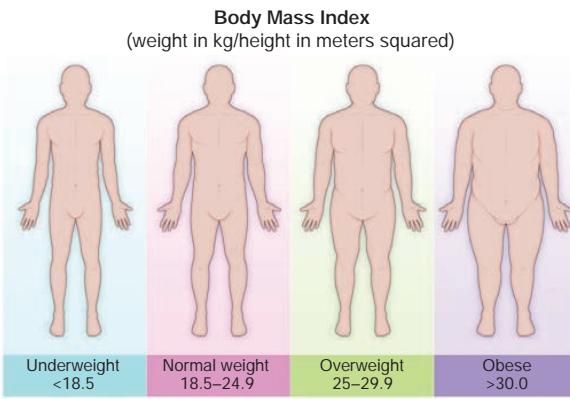
During this time, one of the most striking changes has been in the prevalence of obesity in children. In children, the relationship between BMI and body fat varies considerably with age and with pubertal maturation; however, when adjusted for age and sex, BMI is a reasonable proxy for fat mass. Using age- and sex-specific BMI cutoffs (overweight 91st percentile; obesity 99th percentile), in 2019, the WHO estimated that 38 million children under the age of 5 were overweight or obese, and in 2016, they reported that 340 million children and adolescents aged 5–19 were overweight or obese.

## PHYSIOLOGIC REGULATION OF ENERGY BALANCE

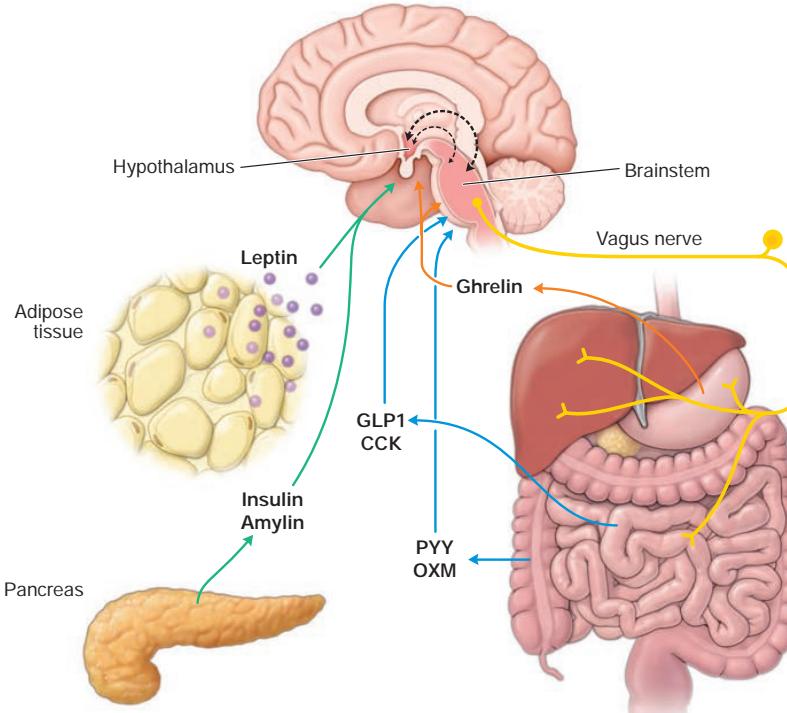
Discussions about obesity so frequently focus on the issues of personal choice or the obesogenic environment that it can be easy to forget that the amount of stored energy in our bodies is subject to homeostatic control by fundamental physiologic processes essential to our survival. In the 1940s, it was demonstrated that rodents defend their level of body fat; once returned to ad libitum diets after a short period of enforced caloric restriction or excess, animals either overconsumed or underconsumed calories until they returned to their previous status. Since that time, research has progressively dissected the signals that sense nutrient stores and the contents of our diets and how this information is integrated to control hunger, satiety, and the expenditure of energy. The key locus for the integration of these signals is the hypothalamus, an area of the brain at least partially outside the blood-brain barrier that facilitates its ability to receive hormonal signals and combine these with sensory, cognitive, and other neural inputs.

The hypothalamus receives two broad types of hormonal signals relevant to energy balance (Fig. 401-2). The circulating concentration of leptin, a peptide hormone produced by fat cells, increases as fat stores increase and declines as fat stores are depleted. Importantly, under conditions of caloric restriction, circulating leptin levels fall faster than the disappearance of fat. Humans born without functional leptin or leptin receptors, although normal weight at birth, become severely obese from an early age, largely as a result of an intense insatiable appetite. Clearly, a reduction of leptin below normal level is a powerful stimulus to food intake and largely explains the rebound overeating and weight regain that occurs after a period of starvation or dieting. The hypothalamus also receives hormonal signals that are more immediately related to the amount and type of food that has been ingested. Peripheral hormones such as cholecystokinin (CCK) from the stomach, glucagon-like peptide 1 (GLP-1) and gastric inhibitory polypeptide (GIP) from enteroendocrine cells of the small intestine, and peptide YY (PYY) and oxyntomodulin from the large intestine are secreted in response to eating a meal and/or the presence of nutrients in the intestinal lumen. Their release together with neural signals from the vagus nerve and the enteric nervous system contributes to satiety, often indirectly acting on the hypothalamus via projections from the brainstem. Insulin, produced by the pancreas in response to carbohydrate and protein-rich meals, also has effects on the hypothalamic neurons controlling energy balance.

The propeptide pro-opiomelanocortin (POMC) is expressed in a highly restricted population of hypothalamic neurons that project widely throughout the brain (Fig. 401-3). These neurons are responsive to the endocrine signals described above and are critical to the regulation of energy balance. The POMC-derived peptides - and -melanocyte-stimulating hormone (MSH) act on the melanocortin 4 receptor (MC4R) to regulate both food intake and aspects of energy expenditure that are influenced by the sympathetic nervous system.



**FIGURE 401-1 Definitions of overweight and obesity.** The World Health Organization defines obesity based on body mass index (BMI), which is calculated as weight in kilograms divided by the height in meters squared.



**FIGURE 401-2** The homeostatic regulation of body weight. In most people, body weight remains stable over long periods of time despite fluctuations in the amount of food we eat and the amount of activity we undertake. This homeostatic regulation of body weight is controlled by the neurons in the hypothalamus, which receive hormonal signals from adipose tissue such as leptin and neural and hormonal signals from the gut in response to meals. Glucagon-like peptide 1 (GLP1) and cholecystokinin (CCK) from enteroendocrine cells of the small intestine and peptide YY (PYY) and oxyntomodulin (OXM) from the large intestine are secreted in response to eating a meal and/or the presence of nutrients in the intestinal lumen. Their release, together with neural signals from the vagus nerve and the enteric nervous system, contributes to satiety, acting on the hypothalamus via projections from the brainstem. Insulin, produced by the pancreas in response to carbohydrate- and protein-rich meals and potentiated by the action of some of the gut hormones, also has effects on the hypothalamic neurons controlling energy balance. The release of the hormone ghrelin from the stomach increases in the unfed state and induces appetite by acting on hypothalamic neurons as well as on receptors in the brainstem.

-MSH, acting mostly through the MC3 receptor, appears to play more of a role in controlling linear growth and the disposition of nutrients into lean versus fat tissues. Signaling through both these melanocortin receptors is also subject to negative control by a different population of neurons, which make and release agouti-related peptide (AGRP), neuropeptide Y (NPY), and the inhibitory neurotransmitter -aminobutyric acid (GABA). AGRP actively switches off melanocortin receptors. Leptin, which suppress food intake, simultaneously stimulates POMC neurons and inhibits NPY/AGRP neurons. Human energy balance is highly sensitive to signaling through this system as people who have a genetic defect in just one of the two copies of the *MC4R* gene are very prone to overeat (hyperphagia) and to gain weight.

### THE PHYSIOLOGY OF NUTRIENT STORAGE IN ADIPOSE TISSUE

When energy intake exceeds energy expenditure, a small amount of that excess energy is stored as glycogen in liver and skeletal muscle. But if the imbalance is greater, then our bodies are designed to store that excess energy in a more efficient way as triacylglycerol (triglyceride). This fat is more efficient because, unlike glycogen, it does not need accompanying water, and when metabolized, it generates almost three times more energy per gram than does carbohydrate. Adipocytes (fat cells) have evolved to contain a highly specialized organelle, the unilamellar fat droplet, which holds the triglyceride within a single-layer of phospholipid that contains all the components needed for enzymes that make and breakdown triglycerides in a manner that is rapidly

responsive to metabolic requirements. No other type of cell is specifically designed to store fat safely in this manner, and many of the adverse consequences of obesity are likely caused not by having too much fat in adipocytes but by “nonprofessional” cells being forced to take up and store fat. Some new fat cells can be made in adulthood when ~10% of our fat cell population turns over every year.

### THE CAUSES OF OBESITY: AN INTERACTION OF GENES AND ENVIRONMENT

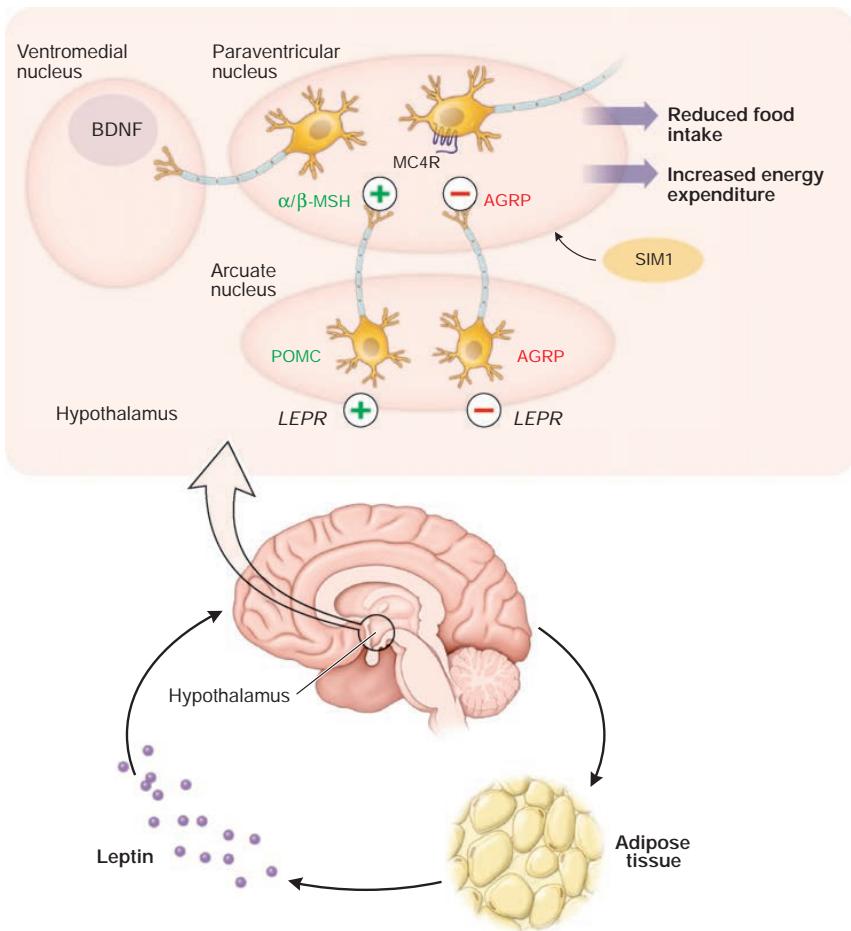
For a person to become obese, energy intake must exceed energy expenditure in a manner that is sufficiently sustained to result in the accumulation of a large excess of triglyceride in adipose tissue. As obesity is a cumulative pathology, if energy intake exceeds energy expenditure by even a small amount (as little as 7 kcal/d), this is sufficient to develop obesity over a matter of years or decades. Even where obesity is common, there are many people who are not overweight. Economic and social factors are likely to play a role as there are more normal-weight people in wealthier and more socially advantaged groups, at least in Western societies. It is also true, however, that because of discrimination, obese people may become socially and economically disadvantaged, which complicates interpretation of that data. We can, however, state with considerable certainty that genetic factors play a major role in predisposing people to a range of adiposity. We know this from a large number of studies comparing identical and nonidentical twins. It is particularly telling that the degree of adiposity in adult life of identical twins brought up in different families is very similar between the twins but is not at all correlated with that of the adoptive siblings with whom they were raised.

### THE RELATIVE ROLES OF EXCESS INTAKE AND LOWER ENERGY EXPENDITURE IN CONFERRING BIOLOGIC PREDISPOSITION

Do these heritable factors influence energy intake, energy expenditure, or both? It is clear that by the time a person is obese the amount of energy they expend in the resting state is more, not less, than a nonobese person. However, if an obese person loses weight by dieting, there is some evidence that they tend to be more “energy efficient” than a person who has never been obese, particularly in terms of how many calories they burn during a defined bout of muscular activity. However, the effects are subtle. It seems very likely that there are some individuals who are predominantly predisposed to develop obesity by virtue of a lower metabolic rate, but thus far, apart from severe hypothyroidism, concrete examples are scarce. In contrast, a much more consistent and compelling body of evidence supports the idea that the genetic predisposition to obesity is largely mediated through the brain's control of food intake. When studied in controlled settings, individuals who carry genetic variants that predispose to obesity tend to eat more and be less readily sated. This is very readily demonstrable when the mutation has a major effect on obesity predisposition, but similar data are now emerging in the case of common genetic variants with smaller effects.

### ENVIRONMENTAL FACTORS PREDISPOSING TO OBESITY

Obesity cannot exist in the absence of sufficient food to lay down and maintain excess fat stores. That fact not infrequently leads to the belief



**FIGURE 401-3 Hypothalamic pathways regulating body weight.** Neurons in the hypothalamus regulate energy intake and expenditure in response to leptin and other hormones. In the fed state, leptin stimulates primary neurons in the arcuate nucleus of the hypothalamus that express pro-opiomelanocortin (POMC). The POMC-derived peptides  $\alpha$ - and  $\beta$ -melanocyte-stimulating hormone (MSH) act on the melanocortin 4 receptor (MC4R) expressed on neurons in the paraventricular nucleus to reduce energy intake and increase energy expenditure. At the same time, leptin inhibits neurons expressing agouti-related peptide (AGRP), which switches off melanocortin receptors. When these and other key molecules, such as brain-derived neurotrophic factor (BDNF) and single minded-1 (SIM1), are disrupted by inherited mutations, affected individuals have hyperphagia and severe obesity.

that the principal cause of obesity must be either the obese person's ignorance of the role of excess caloric intake or their conscious choice to prioritize the immediate pleasures of eating over the long-term health harms associated with obesity. Taken to extremes, these views can engender serious social, economic, and medical discrimination against obese people. It is clear that genetic factors, however important they are in an individual's predisposition to obesity, cannot explain the marked increase in obesity prevalence that has occurred in the past few decades. We have to look to an environment that has become increasingly obesogenic to explain that phenomenon. In most developed and developing countries, energy-dense and highly palatable food and beverages have been aggressively marketed, made cheaper than ever before, provided in larger portions, and made available ubiquitously and continuously. This has been combined with the reduction in physical activity in work and domestic life due to mechanization and the change in the nature of employment. Even the control of our external temperature by artificial heating and cooling has meant less energy expended on thermoregulation. Taken together, these are likely to be the major factors driving the recent increase in obesity. It is important to remember, however, that a substantial proportion of the population remains normal weight under these circumstances and a large part of that is attributable to their genetic good fortune.

There is much current investigation into other environmental factors that might influence the development of obesity. Heated debates continue about the optimal balance of macronutrients in the diet to maintain normal weight and good health. Much of this revolves around the potential benefits of reducing the relative proportion of carbohydrates in the diet (**Chap. 402**). There seems to be reasonable consensus that, in the short term, diets that are rich in protein and fat and lower in carbohydrates more readily result in quick weight loss. This may be because the appetite-suppressing gut hormones discussed above increase more in response to protein than to carbohydrate, thus inducing earlier satiation. However, longer-term studies to date are less compelling, and the long-term increases in protein and fat intake are not without at least theoretical risks. A growing body of evidence suggests that exposures early in life, either in utero or in early postnatal life, might "program" individuals to develop obesity and/or cardiometabolic disease through effects that are often attributed to "epigenetics" (**Chap. 483**). This is an attractive idea, and if true, it would mean that time-limited and affordable interventions early in life might have lifelong benefits. Inevitably, it will take time to see if the promise of such interventions will be fulfilled. Much excitement has been generated by the increasing recognition of the diversity of our intestinal microbiome, which clearly has relevance to gastrointestinal health (**Chap. 471**). At present, it is premature to ascribe any significant role to the human microbiome in obesity or its adverse consequences.

### WHY DOESN'T LEPTIN PREVENT OBESITY?

Leptin is known to suppress food intake, and its levels rise as fat stores expand. So why does this not prevent us from becoming

obese? The most plausible explanation lies in the evolutionary history of leptin and the fact that it appears to defend strongly against the loss of body fat stores, with a fall in circulating leptin below a person's habitual level being a powerful stimulus to food intake, whereas the response to rises in leptin above the normal level is less pronounced. At higher levels of leptin, administering extra amounts of the hormone may have no discernible effect at all—a phenomenon that has come to be called leptin resistance. It is important to remember that even though a person appears to be leptin resistant, some leptin action is occurring; otherwise, the person would become as insatiably hungry and progressively obese as someone with congenital leptin deficiency (see below). It also seems likely that a subgroup of people may have relatively low leptin levels, which plays a role in the etiology of their obesity. There are likely other hormonal signals produced in severe obesity that, unlike leptin, continue to exert a suppressive effect on food intake and help to ensure that the expansion of adipose tissue does not become indefinitely cumulative.

### SINGLE-GENE DISORDERS LEADING TO OBESITY

The assessment of severely obese children and, indeed, adults should be directed at screening for potentially treatable endocrine and neurologic conditions and identifying genetic conditions so that appropriate

**TABLE 401-1 Classical Genetic Obesity Syndromes**

SYNDROME	INHERITANCE	ADDITIONAL CLINICAL FEATURES
Prader-Willi	Autosomal dominant	Hypotonia, failure to thrive in infancy, developmental delay, short stature, hypogonadotropic hypogonadism, sleep disturbance, obsessive behavior
Albright's hereditary osteodystrophy	Autosomal dominant	Short stature in some, skeletal defects, developmental delay, shortened metacarpals; hormone resistance when mutation on maternally inherited allele
Bardet-Biedl	Autosomal recessive	Syndactyly/brachydactyly/polydactyly, developmental delay, retinal dystrophy or pigmentary retinopathy, hypogonadism, renal abnormalities
Cohen's	Autosomal recessive	Facial dysmorphism, microcephaly, hypotonia, developmental delay, retinopathy
Carpenter's	Autosomal recessive	Acrocephaly, brachydactyly, developmental delay, congenital heart defects; growth retardation, hypogonadism
Alström's	Autosomal recessive	Progressive cone-rod dystrophy, sensorineural hearing loss, hyperinsulinemia, early type 2 diabetes mellitus, dilated cardiomyopathy, pulmonary, hepatic and renal fibrosis
Tubby	Autosomal recessive	Progressive cone-rod dystrophy, hearing loss

genetic counseling and, in some cases, treatment can be started. Clinically, it remains useful to categorize the genetic obesity syndromes as those with dysmorphism and/or developmental delay and those without these features (**Tables 401-1 and 401-2**). Although individually these monogenic disorders are rare, cumulatively, up to 20% of children with severe obesity have rare chromosomal abnormalities and/or highly penetrant genetic mutations that drive their obesity. This figure is likely to increase with wider accessibility to genetic testing and as new genes are identified. A genetic diagnosis can inform management (many such patients find it very difficult to lose weight through diet and exercise) and can inform clinical decision-making regarding the use of bariatric surgery (feasible in some; high risk in others) (**Chap. 402**). There are a number of drugs in clinical trials targeted specifically at patients with genetic obesity syndromes. Specifically, setmelanotide, a MC4R agonist, has been used effectively in phase 2/3 clinical trials in children who are genetically deficient in POMC or the leptin receptor. It is also being explored for the treatment of other genetic obesity syndromes affecting the melanocortin pathway.

### CLASSICAL SYNDROMIC DISORDERS

A number of syndromes were identified by clinicians long before their exact genetic cause was known. In these syndromes, obesity is associated with a stereotyped set of other anomalies, often neurodevelopmental in type. The precise genetic basis for the majority of these syndromes is now known. Prader-Willi syndrome (PWS) is the most common syndromic cause of obesity, with an estimated prevalence of ~1 in 25,000. It is an autosomal dominant disorder caused by deletion of an imprinted region on the paternal chromosome 15 (**Chap. 46**). The characteristic clinical features are hypotonia, feeding difficulties in infancy, developmental delay, hypogonadotropic hypogonadism, hyperphagia (increased food intake), and obesity. Children with PWS are short with reduced lean body mass and increased fat mass, features resembling those seen in growth hormone (GH) deficiency; GH treatment decreases body fat and increases linear growth and muscle mass and is now standard of care in this condition. Low levels of brain expression of the neuropeptide oxytocin and the nerve growth factor Brain-derived neurotrophic factor (BDNF) in PWS patients have suggested new therapeutic opportunities for these patients.

Inherited or de novo (not found in either parent) mutations in another imprinted gene, *GNAS1*, which encodes Gs protein, cause

**TABLE 401-2 Obesity Syndromes due to Mutations in Genes Controlling Energy Homeostasis Pathways**

GENE AFFECTED	INHERITANCE	ADDITIONAL CLINICAL FEATURES
Leptin	Autosomal recessive	Severe hyperphagia, frequent infections, hypogonadotropic hypogonadism, mild hypothyroidism
Leptin receptor	Autosomal recessive	Severe hyperphagia, frequent infections, hypogonadotropic hypogonadism, mild hypothyroidism
Proopiomelanocortin	Autosomal recessive	Hyperphagia, cholestatic jaundice or adrenal crisis due to ACTH deficiency, pale skin and red hair
Prohormone convertase 1	Autosomal recessive	Small-bowel enteropathy, postprandial hypoglycemia, hypothyroidism, ACTH deficiency, hypogonadism, central diabetes insipidus
Carboxypeptidase E	Autosomal recessive	
Melanocortin 4 receptor	Autosomal dominant	Hyperphagia, accelerated linear growth
Single-minded 1	Autosomal dominant	Hyperphagia, accelerated linear growth, speech and language delay, autistic traits
BDNF	Autosomal dominant	Hyperphagia, developmental delay, hyperactivity, behavioral problems including aggression
TrkB	Autosomal dominant	Hyperphagia, speech and language delay, variable developmental delay, hyperactivity, behavioral problems including aggression
SH2B1	Autosomal dominant	Hyperphagia, disproportionate hyperinsulinemia, early type 2 diabetes mellitus, behavioral problems including aggression

*Abbreviations:* ACTH, adrenocorticotropic hormone; BDNF, brain-derived neurotrophic factor; SH2B1, Src-homology-2 (SH2) B-adaptor protein-1 (SH2B1); TrkB, tropomyosin receptor kinase B.

a syndrome known as Albright's hereditary osteodystrophy (AHO) (**Chap. 412**). Maternal transmission of *GNAS1* mutations leads to AHO (characterized by short stature, obesity, and skeletal defects) plus resistance to several hormones (e.g., parathyroid hormone), whereas paternal transmission leads only to the AHO phenotype. The clinical spectrum is very broad, and some patients may present with obesity alone.

Bardet-Biedl syndrome is a rare autosomal recessive disease characterized by obesity, developmental delay, polydactyly, retinal dystrophy or pigmentary retinopathy, hypogonadism, and renal abnormalities. The same clinical features can arise from mutations in >20 genes, which disrupt signaling in primary cilia. Overlapping clinical features are seen in a number of other genetic obesity syndromes (Table 401-1).

### DISORDERS OF LEPTIN-MELANOCORTIN SIGNALING

Homozygous mutations that disrupt the production or action of leptin are rare but result in a disorder that is treatable. Children with homozygous loss-of-function leptin mutations have rapid weight gain in the first few months of life, resulting in severe obesity due to an intense drive to eat (hyperphagia) and impaired satiety with food-seeking behavior soon after the end of a meal. Congenital leptin deficiency can be treated with subcutaneous injections of recombinant leptin, which reduce hunger, increase satiety, and lead to weight loss. Similar clinical features are seen in patients with homozygous mutations in the leptin receptor gene, but they are not responsive to leptin treatment (Table 401-2). Normal pubertal development rarely occurs in adults with leptin or leptin receptor deficiency, with biochemical evidence of hypogonadotropic hypogonadism. However, there is some evidence

for the delayed but spontaneous onset of menses in a small number of leptin- and leptin receptor-deficient adults. Leptin treatment permits progression of pubertal development, suggesting that leptin is a permissive factor for the development of puberty.

Homozygous or compound heterozygous mutations in POMC lead to hyperphagia and early-onset obesity. As adrenocorticotropic hormone (ACTH) is produced in the pituitary gland by cleavage from POMC, patients also present with isolated ACTH deficiency (neonatal hypoglycemia and cholestatic jaundice). In the skin, POMC-derived melanocortin peptides act on melanocortin 1 receptors to induce pigmentation. For this reason, the lack of POMC-derived peptides in obese patients with POMC deficiency results in hypopigmentation of skin and hair, which is more noticeable in people of Caucasian ancestry who often have red hair. Prohormone convertase 1 (PCSK1) is an enzyme involved in the cleavage of POMC into ACTH, which is then further cleaved to make -MSH by carboxypeptidase E. Impaired processing of POMC contributes to the hyperphagic severe early-onset obesity and ACTH deficiency in people lacking PCSK1 who also have hypogonadotropic hypogonadism, postprandial hypoglycemia (due to impaired processing of proinsulin to insulin), and a neonatal enteropathy in early childhood. Heterozygous mutations that impair the function of MC4R are found in 5–6% of patients with severe early-onset obesity and at a frequency of ~1 in 300 in the general population, making this the most common gene in which variants contribute to obesity. MC4R mutations are inherited in a co-dominant manner, with variable penetrance and expression in heterozygous carriers; homozygous carriers are severely obese. Patients are often hyperphagic from early childhood and hyperinsulinemic and have increased lean mass and increased linear growth.

### GENETIC SUBTYPES OF OBESITY ASSOCIATED WITH NEUROBEHAVIORAL ABNORMALITIES

Both PWS patients and patients with mutations in *SIM1* (a gene that acts downstream of MC4R) exhibit a spectrum of behavioral abnormalities that overlap with autism-like features that could be related to reduced oxytocin signaling (Table 401-2). Mutations affecting BDNF and its receptor tropomyosin receptor kinase B (TrkB) cause speech and language delay, hyperphagia, and severe obesity, as well as hyperactivity, autistic traits, and impaired short-term memory. Interestingly, a common variant in BDNF (V66M), found in heterozygous form in ~20% of the population, is associated with a number of traits and neuropsychiatric disorders including anxiety and depression. Chromosomal deletion and mutations affecting Src-homology-2 (SH2) B-adaptor protein-1 (*SH2B1*) are associated with dominantly inherited, severe, early-onset obesity, disproportionate insulin resistance, early-onset type 2 diabetes, and behavioral problems including aggressive behavior.

### OBESITY SECONDARY TO OTHER DISORDERS

**Endocrine Disorders** Patients with hypothyroidism may gain weight and become obese, although it is rarely the sole cause of severe obesity. It is nonetheless prudent always to measure thyroid function in a patient presenting with obesity. Measurement of thyroid-stimulating hormone (TSH) will detect significant primary disease of the thyroid, but for rare secondary hypothyroidism, additional measurement of free thyroxine levels is needed (Chap. 383). Weight gain can also be a presenting feature of Cushing's syndrome. Clinically, the presence of spontaneous bruising, livid striae, myopathy, and marked centripetal distribution of body fat help to distinguish true endogenous hypercortisolism from common obesity. This condition is usually reasonably straightforward to diagnose based on tests that approximate cortisol production rates (24-h urine free cortisol) or the suppression of serum cortisol by dexamethasone (Chap. 386). Occasionally, in severely obese patients, effects of adiposity on glucocorticoid metabolism can make it difficult to interpret results, and more sophisticated tests, including those measuring diurnal rhythm of cortisol, may be necessary to establish or exclude the diagnosis with confidence. Weight gain can also be a presenting feature of patients with insulinoma, driven largely by the need to eat more frequently than normal to avoid hypoglycemia.

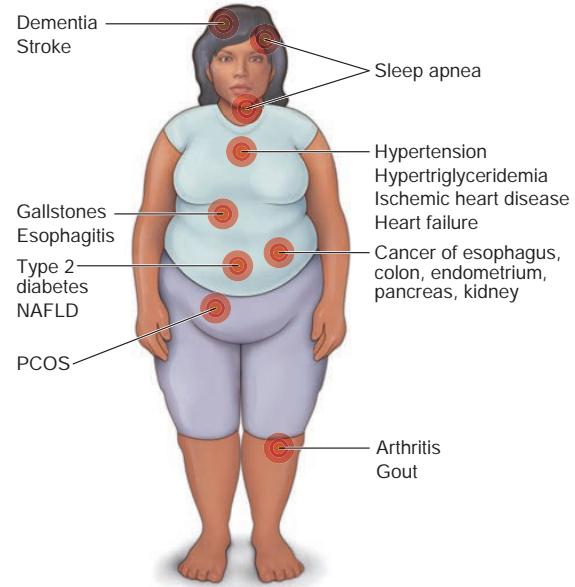
**Hypothalamic Damage** The hypothalamic regions that control energy balance can be disrupted by tumors (such as craniopharyngiomas), inflammatory masses, or after a severe head injury (Chap. 379). In such cases, there is often some accompanying evidence of disruption of the hormonal functions of the anterior or posterior pituitary, although it may be subtle and the history of hyperphagia and weight gain is often short. It is worth noting that in common obesity, GH levels in response to provocative testing may be somewhat lower than normal, but this does not necessarily suggest the presence of a structural lesion.

### ADVERSE CONSEQUENCES OF OBESITY

**Mechanistic Considerations** Obesity is associated with a wide range of pathologies that can adversely impact morbidity and mortality (Chap. 408). Some of these consequences are related, at least in part, to the direct mechanical or gravitational effects of the expanded fat mass itself (Fig. 401-4). However, the principal mechanisms behind many of the complications of obesity are less likely to be due to the expanded fat mass itself but more closely related to the chronic state of overnutrition itself and its effects on tissues throughout the body.

As people become obese, one of the first and most prominent biochemical abnormalities that develops is the need for increased circulating concentrations of insulin to maintain glucose homeostasis. This state of insulin resistance generally worsens with a greater degree of obesity, but there is a high degree of interindividual variability. It is more prominent when fat is distributed more centrally. Insulin resistance/hyperinsulinemia is likely to play a major role in the predisposition to metabolic endocrine and cardiovascular diseases seen more frequently in obesity and may even play a role in the predisposition of obese people to develop certain cancers.

The main sites of insulin action in the body are the liver and skeletal muscle. Thus, for insulin resistance to be discernible at the level of the whole body, the action of insulin must be disturbed in one or both of these tissues. It seems unlikely that an expanded fat cell mass would do that directly. How then does obesity lead to a state of insulin resistance?



**FIGURE 401-4** Obesity-related complications. The expanded fat mass that characterizes obesity predisposes to certain obesity-related complications (e.g., osteoarthritis of knees, reflux esophagitis, and obstructive sleep apnea) directly through its mass and/or volume. However, in the case of the metabolic, endocrine, and cardiovascular complications, the link is less clear. Further research is needed to establish whether some features of the expanded fat mass influence the development of these complications or whether other aspects of the chronically overnourished state, such as excess fat outside the fat depot, are more relevant. NAFLD, nonalcoholic fatty liver disease; PCOS, polycystic ovarian syndrome.

One hypothesis suggests a leading role for the inflammation that occurs in the adipose tissue in obesity (**Fig. 401-5**). This undoubtedly happens, as there are more macrophages in obese than nonobese adipose tissues, and this is associated with higher levels of inflammatory markers in the circulation of obese people. The majority of macrophages in obese adipose tissue are found in clusters around dead or dying adipocytes, so it appears that these cells are clearing debris after cell death. Studies in animal models provide strong support for the notion that this inflammatory state is mechanistically linked to insulin resistance, but evidence from humans for this is not as strong.

An alternative hypothesis is that as individuals become more obese they become less able to safely store nutrients in their adipose tissue and begin to redirect macronutrients to other tissues that are not designed for fat storage and may be damaged by the nutrient excess. This certainly happens to people who are born with a lack of adipose tissue (lipodystrophy) who, early in life, develop severe versions all the metabolic complications that are seen in obesity as they have no safe depot in which to store excess nutrients. There are stronger human data from both genetic and pharmacologic studies for the existence of the latter mechanism. How ectopic fat leads to insulin resistance and other damaging effects is still a puzzle, but it is very likely a major driver of pathology associated with obesity.

#### Metabolic Complications • DYSLIPIDEMIA

The insulin resistance of obesity is frequently associated with dyslipidemia characterized by high circulating triglycerides and low high-density lipoprotein cholesterol (**Chap. 407**). Occasionally, the hypertriglyceridemia may be severe enough to put the patient at risk of pancreatitis. Although there is a relationship between obesity and raised circulating levels of low-density lipoprotein cholesterol (which is the major risk factor for coronary artery disease), genetic factors independent of obesity and the type of dietary fat consumed probably have an even greater impact.

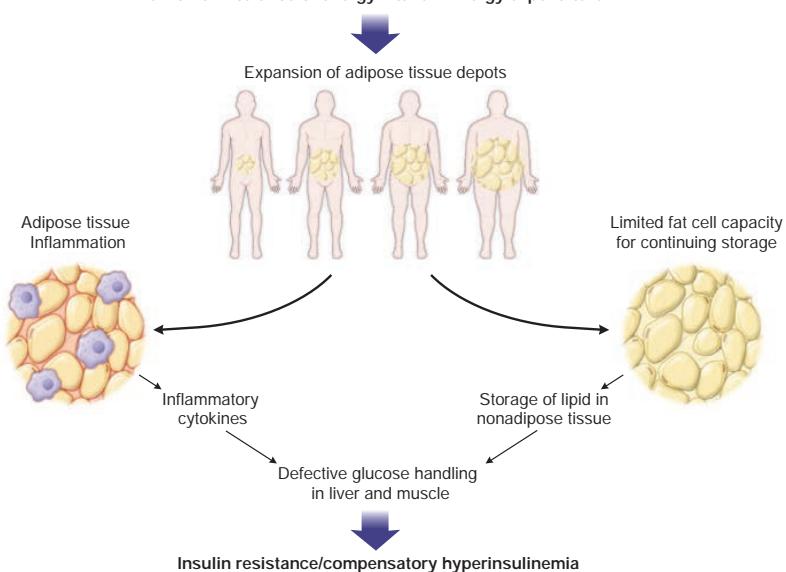
**FATTY LIVER DISEASE** Obesity is strongly associated with the presence of ectopic fat in hepatocytes. This can progress to nonalcoholic steatohepatitis (NASH), which can progress to the fibrosis, which is a precursor to cirrhosis (**Chap. 343**). The reported incidence of NASH-related cirrhosis and of hepatocellular carcinoma has increased markedly in step with the increase in the prevalence of obesity in adolescents and adults.

**TYPE 2 DIABETES** The insulin resistance characteristic of the over-nourished state strongly predisposes to the development of type 2 diabetes in people who, largely for genetic reasons, are less able to maintain the high levels of insulin secretion over many decades. Impaired glucose tolerance and type 2 diabetes are among the most common complications of obesity (**Chap. 403**).

**Endocrine Complications** In females, the insulin resistance/hyperinsulinemia frequently found in obesity strongly predisposes to the development of polycystic ovaries, characterized by irregular menstruation, anovulatory infertility, and hirsutism due to hyperandrogenism. In males, obesity is more often associated with a degree of central hypogonadism, where low circulating testosterone is associated with levels of luteinizing hormone and follicle-stimulating hormone that do not rise appropriately to compensate for the testosterone-deficient state.

**Dermatologic Complications** Obesity can result in problems with excessive skin folds that can cause discomfort through mechanical

Chronic imbalance of energy intake > Energy expenditure



**FIGURE 401-5** How does obesity cause metabolic disease? Insulin resistance is one of the earliest complications of obesity and underlies and precedes many of its adverse health consequences. The disposal and production of glucose by the most important tissues, muscle and liver, respectively, become less sensitive to insulin, and this results in a compensatory increase in insulin secretion from the pancreas. There are two main theories for the association of obesity with insulin resistance. In the first, products of macrophages and other inflammatory cells that are more abundant in obese adipose tissue can, through paracrine or endocrine routes, disturb insulin's action in muscle and liver cells. In the second, as adipose storage deposits fill up, they become less able to take on excessive calories, which end up being stored as ectopic lipid in tissues such as muscle and liver, which are not primarily designed to store nutrients of this type. The evidence in humans is stronger for the latter hypothesis.

irritation and can also become infected with fungi. Insulin resistance/hyperinsulinemia is associated with acanthosis nigricans, where areas such as axilla, groin, and the back of neck develop velvety hyperpigmentation. Hidradenitis suppurativa is a potentially disabling skin condition strongly associated with obesity. It is characterized by recurrent boils often with chronically draining sinus tracts affecting skin areas containing apocrine sweat glands.

**Cardiovascular Complications** Obese people, even if they do not have diabetes, have increased morbidity and mortality from atherosclerotic vascular disease, including coronary artery disease and stroke. The factors that result in this are complex and involve increased prevalence of hypertension, dyslipidemia, and insulin resistance/hyperinsulinemia. The rare condition of thrombotic thrombocytopenic purpura, which causes microvascular platelet thrombosis, thrombocytopenia, and hemolytic anemia due to the presence of abnormally large von Willebrand factor multimers, is strongly associated with obesity.

Independent of occlusive arterial disease, obese people are also at increased risk of heart failure, particularly characterized primarily by diastolic dysfunction, and of atrial fibrillation, the most common arrhythmia.

**Respiratory Complications** Exertional dyspnea is common in obesity, contributed to by the increased work required to move a greater mass as well as impacts of pressure on the diaphragm and thoracic cage on chest wall compliance. Enlargement of soft tissue of the mouth and throat and adipose depots around the airways contribute to the high prevalence of sleep apnea, although other factors may contribute to some forms, where central nocturnal hypoventilation also occurs.

**Gastrointestinal Disorders** Reflux esophagitis is the most common gastrointestinal complication of obesity, particularly occurring in those with high intraabdominal pressure. Gallstones are also more common in obese people, bringing increased risks of biliary colic, cholecystitis, pancreatitis, and gallbladder cancer.

**Rheumatologic Disorders** Osteoarthritis of the knee and gout are the two most common rheumatologic conditions clearly associated with obesity. Interestingly, despite obesity being described as a proinflammatory state, there is no evidence for an increase in rheumatoid arthritis or the seronegative arthritides among people who are obese.

**Cancers** Obesity is a risk factor for a number of common cancers. Indeed, it has recently been calculated that, at least in some countries, obesity has overtaken smoking as the greatest risk factor for developing cancer. Recent research has found that as the BMI increases by 5 kg/m<sup>2</sup>, cancer mortality increases by 10%. The largest effects are on colorectal, kidney, and pancreatic cancer, adenocarcinoma of the esophagus, and, in women, endometrial carcinoma. The recent rapid increase in the prevalence of esophageal adenocarcinoma is likely related to the marked recent increase in reflux esophagitis due to the raised intraabdominal pressure (with or without hiatus hernia) characteristic of central obesity.

**Response to Infection** The fact that obesity can influence the outcome of some infections has become very apparent with the COVID-19 pandemic. Obese patients have a substantially worse outcome if infected by SARS-CoV-2 through mechanisms that are as yet unclear. Obese patients also appear to be more susceptible to bacterial wound infections and postoperative sepsis.

**Disorders of the Central Nervous System** There is increasing evidence that obesity is a risk factor for dementia in later life, although how that risk is mediated is not clear. Idiopathic intracranial hypertension is a rare disorder that is strongly associated with obesity.

## CONCLUSION

Obesity is a medical disorder that has been greatly increasing in prevalence due to environmental factors that are ubiquitous in developed and developing countries. However, it is important to bear in mind that it is a highly heterogeneous condition, which in some people is attributable entirely to genetic causes, and that underlying genetic variation strongly influences the risk of obesity in all people. It is a serious condition leading to multiple adverse health outcomes and considerable human suffering. As our understanding of its pathogenesis increases, our duty to treat obese patients with understanding and compassion and to develop new and better options for its treatment and prevention is worthy of emphasis.

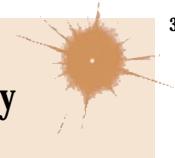
## FURTHER READING

- Casazza K et al: Myths, presumptions, and facts about obesity. *N Engl J Med* 368:446, 2013.
- Farooqi IS, O'Rahilly S: The genetics of obesity in humans. In: Feingold KR et al (eds). *Endotext*. South Dartmouth, MA, 2000.
- Friedman JM: Leptin and the endocrine control of energy balance. *Nat Metab* 1:754, 2019.
- Heymsfield SB, Wadden TA: Mechanisms, pathophysiology, and management of obesity. *N Engl J Med* 376:1492, 2017.
- Leibel RL et al: Changes in energy expenditure resulting from altered body weight. *N Engl J Med* 332:621, 1995.
- NCD Risk Factor Collaboration (NCD-RISC): Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: A pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet* 390:2627, 2017.
- O'Rahilly S: Harveian Oration 2016: Some observations on the causes and consequences of obesity. *Clin Med (Lond)* 16:551, 2016.

402

# Evaluation and Management of Obesity

Robert F. Kushner



More than 70% of U.S. adults are considered to be overweight or have obesity, and the prevalence of obesity is increasing rapidly in most of the industrialized world. Children and adolescents also are becoming more obese, indicating that the current trends will accelerate over time. Obesity is associated with an increased risk of multiple health problems, including hypertension, type 2 diabetes, dyslipidemia, obstructive sleep apnea, nonalcoholic fatty liver disease, degenerative joint disease, and some malignancies. Thus, it is important for health care providers to identify, evaluate, and treat patients for obesity and associated comorbid conditions.

## EVALUATION

Health care providers should screen all adult patients for obesity and offer intensive counseling and behavioral interventions to promote sustained weight loss. The four main steps in the evaluation of obesity, as described below, are (1) a focused obesity-related history that includes lifestyle questions about diet, physical activity, sleep, and stress; (2) a physical examination to determine the degree and type of obesity; (3) assessment of comorbid conditions; and (4) assessment of the patient's readiness to adopt lifestyle changes.

**The Obesity-Focused History** The first step in taking an obesity-focused history is to approach the topic in a sensitive manner. The reason for this concern is that the word *obesity* is a highly charged, emotive term. It has a significant pejorative meaning for many patients, leaving them feeling judged and blamed when labeled as such. This is not the case when patients are told that they have other chronic diseases such as diabetes or hypertension. Patients prefer that clinicians use more neutral words or terms such as *weight*, *excess weight*, *body mass index (BMI)*, or *unhealthy weight*, versus more stigmatizing terms such as *obesity*, *morbid obesity*, or *fatness*.

Information from the history should address the following seven questions:

- What factors contribute to the patient's obesity?
- How is the obesity affecting the patient's health?
- What is the patient's level of risk from obesity?
- What does the patient find difficult about managing weight?
- What are the patient's goals and expectations?
- Is the patient motivated to begin a weight management program?
- What kind of help does the patient need?

Although the vast majority of cases of obesity are promoted by behavioral factors that affect diet and physical activity patterns, the history may suggest secondary causes that merit further evaluation. Disorders to consider include polycystic ovarian syndrome, hypothyroidism, Cushing's syndrome, and hypothalamic disease. Drug-induced weight gain also should be considered. Common causes include medications for diabetes (insulin, sulfonylureas, thiazolidinediones), steroid hormones, antipsychotic agents (clozapine, olanzapine, risperidone), mood stabilizers (lithium), antidepressants (tricyclics, monoamine oxidase inhibitors, paroxetine, mirtazapine), and antiepileptic drugs (valproate, gabapentin, carbamazepine). Other medications, such as nonsteroidal anti-inflammatory drugs and calcium channel blockers, may cause peripheral edema but do not increase body fat.

The patient's current diet and physical activity patterns may reveal factors that contribute to the development of obesity and may identify behaviors to target for treatment. Physical fitness, in particular, is an important predictor of all-cause mortality rate independent of BMI and body composition and highlights the importance of taking a physical activity and exercise history during examination as well as emphasizing physical activity as a treatment approach.

**TABLE 402-1 Classification of Weight Status and Disease Risk**

CLASSIFICATION	BODY MASS INDEX (kg/m <sup>2</sup> )	OBESITY CLASS	DISEASE RISK
Underweight	<18.5	—	—
Healthy weight	18.5–24.9	—	—
Overweight	25.0–29.9	—	Increased
Obesity	30.0–34.9	I	High
Obesity	35.0–39.9	II	Very high
Extreme obesity	40	III	Extremely high

Source: Adapted with permission from WHO Consultation on Obesity (1997): Geneva, Switzerland, World Health Organization. Division of Noncommunicable Diseases & World Health Organization. Programme of Nutrition, Family and Reproductive Health (1998). Obesity: preventing and managing the global epidemic: report of a WHO Consultation on Obesity, Geneva, 3–5 June 1997. World Health Organization. [https://apps.who.int/iris/bitstream/handle/10665/63854/WHO\\_NUT\\_NCD\\_98.1.%28p159-276%29.pdf?sequence=2&isAllowed=y](https://apps.who.int/iris/bitstream/handle/10665/63854/WHO_NUT_NCD_98.1.%28p159-276%29.pdf?sequence=2&isAllowed=y)

Inquiring about sleep health that addresses regularity, duration, efficiency, and satisfaction is also important. Although the mechanisms are uncertain, sleep deprivation is associated with metabolic alterations in appetite regulation, sympathetic nervous system overactivity, insulin sensitivity, and changes in circadian rhythm. Stress may also contribute to obesity, in part due to activation of the adrenal cortical axis and elevated cortisol levels and its impact on emotional health and behaviors. This historic information is best obtained by the combination of a questionnaire and an interview.

**BMI and Waist Circumference** Three key anthropometric measurements are important in evaluating the degree of obesity: weight, height, and waist circumference. The BMI, calculated as weight (kg)/height (m)<sup>2</sup> or as weight (lb)/height (in)<sup>2</sup> × 703, is used to classify weight status and risk of disease (Table 402-1). BMI is highly correlated with body fat and is related to disease risk. Lower BMI thresholds for overweight and obesity have been proposed for the Asia-Pacific region since this population appears to be at risk for glucose and lipid abnormalities at lower body weights.

Excess abdominal fat, assessed by measurement of waist circumference, is independently associated with a higher risk for metabolic syndrome, diabetes mellitus, and cardiovascular disease. Measurement of the waist circumference is a surrogate for visceral adipose tissue and should be performed in the horizontal plane above the iliac crest (Table 402-2).

**Obesity-Associated Comorbid Conditions** The evaluation of comorbid conditions should be based on presentation of symptoms,

**TABLE 402-3 Obesity-Related Organ Systems Review**

<b>Cardiovascular</b>	<b>Respiratory</b>
Hypertension	Dyspnea
Congestive heart failure	Obstructive sleep apnea
Cor pulmonale	Hypoventilation syndrome
Varicose veins	Pickwickian syndrome
Pulmonary embolism	Asthma
Coronary artery disease	
<b>Endocrine</b>	<b>Gastrointestinal</b>
Metabolic syndrome	Gastroesophageal reflux disease
Type 2 diabetes	Nonalcoholic fatty liver disease
Dyslipidemia	Cholelithiasis
Polycystic ovarian syndrome	Hernias
<b>Musculoskeletal</b>	Colon cancer
Hyperuricemia and gout	
Immobility	<b>Genitourinary</b>
Osteoarthritis (knees and hips)	Urinary stress incontinence
Low back pain	Obesity-related glomerulopathy
Carpal tunnel syndrome	Hypogonadism (male)
<b>Psychological</b>	Breast and uterine cancer
Depression/low self-esteem	Pregnancy complications
Body image disturbance	
Social stigmatization	<b>Neurologic</b>
<b>Integument</b>	Stroke
Striae distensae	Idiopathic intracranial hypertension
Stasis pigmentation of legs	Meralgia paresthetica
Lymphedema	Dementia
Cellulitis	
Intertrigo, carbuncles	
Acanthosis nigricans	
Acrochordons (skin tags)	
Hidradenitis suppurativa	

risk factors, and index of suspicion. For all patients, a fasting lipid profile (total, low-density lipoprotein, and high-density lipoprotein cholesterol and triglyceride levels), chemistry panel, and glycated hemoglobin should be performed, and blood pressure determined. Symptoms and diseases that are directly or indirectly related to obesity are listed in Table 402-3. Although individuals vary, the number and severity of organ-specific comorbid conditions usually rise with increasing levels of obesity.

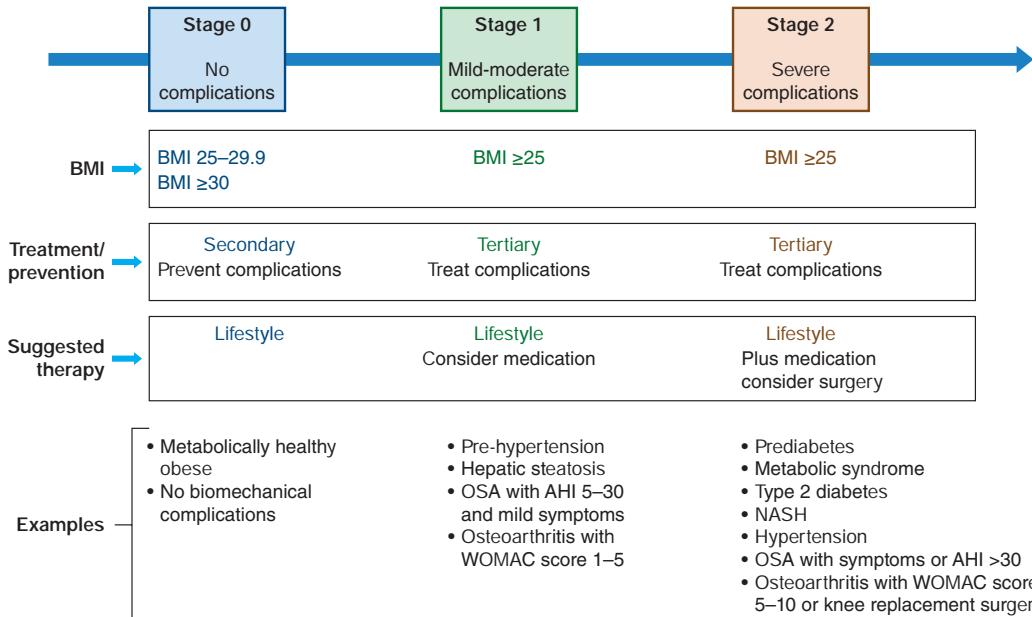
**Identifying the High-Risk Patient** Efforts are under way to develop more practical and useful assessments to identify patients who are at high risk in addition to using BMI alone. Analogous to other staging systems commonly used for congestive heart failure or chronic kidney disease, the American Association of Clinical Endocrinology (AACE) and the American College of Endocrinology (ACE) guidelines advocate a simple and clinically useful obesity disease staging system that is based on ethnic-specific BMI cutoffs in conjunction with assessment for adiposity-related complications (Fig. 402-1). Stage 0 is assigned to individuals who are overweight or obese by BMI classification but have no complications, whereas stages 1 and 2 are defined as individuals who are overweight or obese by BMI classification and have one or more mild-moderate complications (stage 1) or at least one severe complication (stage 2). A different functional staging system for obesity, called the Edmonton Obesity Staging System (EOSS), classifies individuals with obesity into five graded categories (0–4), based on their morbidity and health-risk profile along three domains—medical, functional, and mental. In this system, staging occurs independent of BMI.

**Assessing the Patient's Readiness to Change** An attempt to initiate lifestyle changes when the patient is not ready usually leads

**TABLE 402-2 Ethnic-Specific Cutpoint Values for Waist Circumference**

ETHNIC GROUP	WAIST CIRCUMFERENCE
Europeans	
Men	>94 cm (>37 in)
Women	>80 cm (>31.5 in)
South Asians and Chinese	
Men	>90 cm (>35 in)
Women	>80 cm (>31.5 in)
Japanese	
Men	>85 cm (>33.5 in)
Women	>90 cm (>35 in)
Ethnic South and Central Americans	Use South Asian recommendations until more specific data are available.
Sub-Saharan Africans	Use European data until more specific data are available.
Eastern Mediterranean and Middle Eastern (Arab) populations	Use European data until more specific data are available.

Source: KG Alberti, P Zimmet, J Shaw; IDF Epidemiology Task Force Consensus Group. The metabolic syndrome—a new worldwide definition. Lancet 366:1059, 2005.



**FIGURE 402-1** Staging the severity of obesity using the American Association of Clinical Endocrinology clinical practice guidelines. AHI, apnea-hypopnea index; BMI, body mass index; NASH, nonalcoholic steatohepatitis; OSA, obstructive sleep apnea; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index (a patient-reported outcome measure for osteoarthritis registering pain, stiffness, and function). (Data from WT Garvey et al: American Association of Clinical Endocrinologists and American College of Endocrinology comprehensive clinical practice guidelines for medical care of patients with obesity. *Endocr Pract* 22 (Suppl 3):1, 2016.)

to frustration and may hamper future weight-loss efforts. Assessment includes patient motivation and support, stressful life events, psychiatric status, time availability and constraints, and appropriateness of goals and expectations. Readiness can be viewed as the balance of two opposing forces: (1) motivation, or the patient's desire to change; and (2) resistance, or the patient's resistance to change.

A helpful method to begin a readiness assessment is to use the motivational interviewing technique of “anchoring” the patient's interest and confidence to change on a numerical scale. With this technique, the patient is asked to rate—on a scale from 0 to 10, with 0 being not so important (or confident) and 10 being very important (or confident)—his or her level of interest in and confidence about losing weight at this time. This exercise helps establish readiness to change and also serves as a basis for further dialogue.

## TREATMENT

### Obesity

#### THE GOAL OF THERAPY

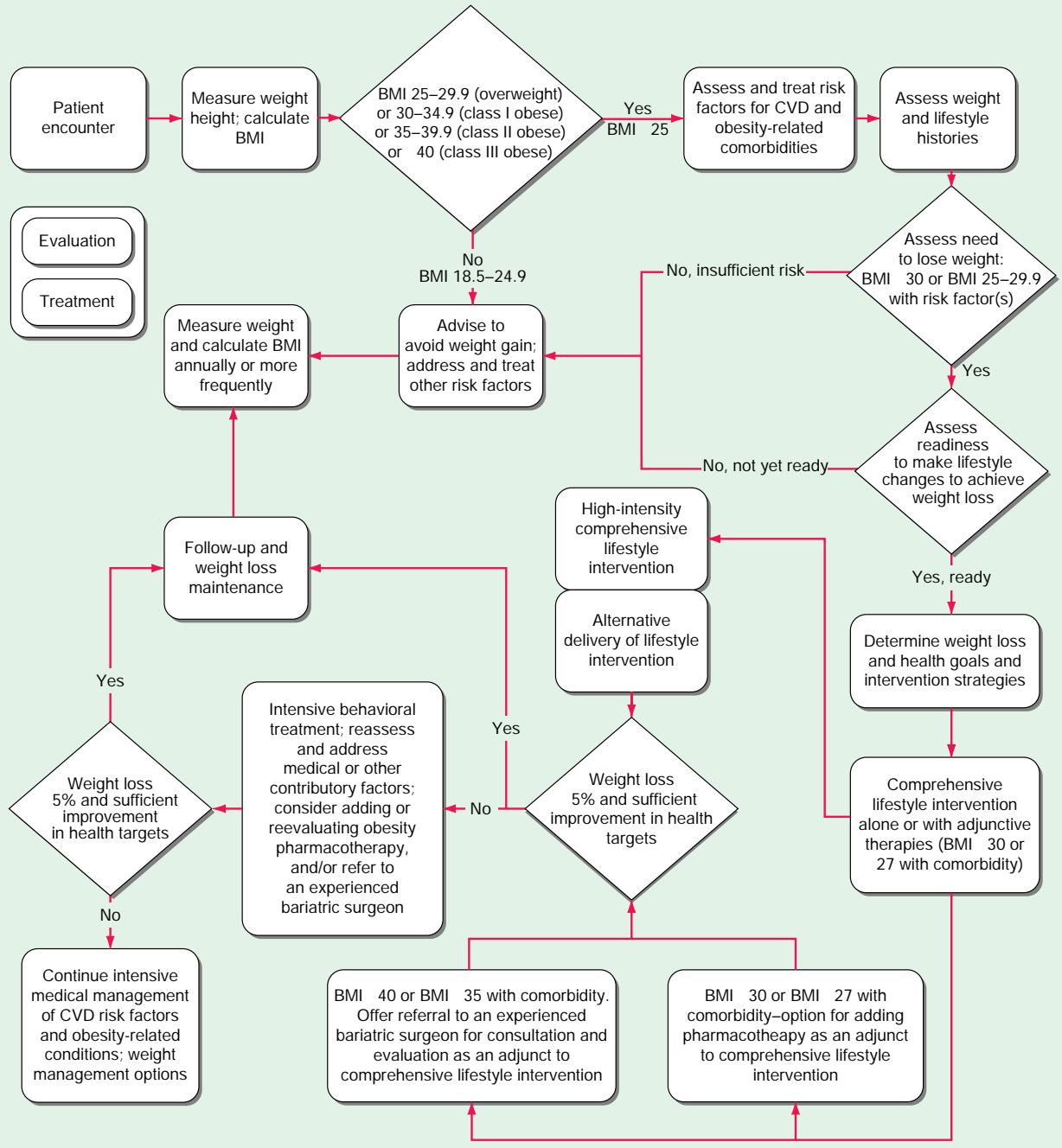
The primary goals of treatment are to improve obesity-related comorbid conditions and quality of life and reduce the risk of developing future obesity-related complications. Information obtained from the history, physical examination, and diagnostic tests is used to determine risk and develop a treatment plan (Fig. 402-2). The decision of how aggressively to treat the patient and which modalities to use is determined by the patient's risk status, expectations, and available resources. Not all patients who are deemed obese by BMI screening need to be treated, since BMI alone does not directly measure body fat, distinguish body fat distribution, or assess an individuals' health status. However, patients who present with obesity-related comorbidities and who would benefit from weight-loss intervention should be managed proactively. Therapy for obesity always begins with lifestyle management and may include pharmacotherapy or bariatric surgery, depending on BMI risk category (Table 402-4). Setting an initial weight-loss goal of 8–10% over 6 months is a realistic target.

#### LIFESTYLE MANAGEMENT

Obesity care involves attention to three essential elements of lifestyle: dietary habits, physical activity, and behavior modification. Because obesity is fundamentally a disease of energy imbalance, all patients must learn how and when energy is consumed (diet), how and when energy is expended (physical activity), and how to incorporate this information into their daily lives (behavioral therapy). Lifestyle management has been shown to result in a modest (typically 3–5 kg) weight loss when compared with no treatment or usual care.

**Diet Therapy** The primary focus of diet therapy is to reduce overall calorie consumption. Guidelines from the American Heart Association/American College of Cardiology/The Obesity Society (AHA/ACC/TOS) recommend initiating treatment with a calorie deficit of 500–750 kcal/d compared with the patient's habitual diet. Alternatively, a diet of 1200–1500 kcal/d for women and 1500–1800 kcal/d for men (adjusted for the individual's body weight) can be prescribed. This reduction is consistent with a goal of losing ~1–2 lb/week. The calorie deficit can be instituted through dietary substitutions or alternatives. Examples include choosing smaller portion sizes, eating more fruits and vegetables, consuming more whole-grain cereals, selecting leaner cuts of meat and skimmed dairy products, reducing consumption of fried foods and other foods with added fats and oils, and drinking water instead of sugar-sweetened beverages. It is important that dietary counseling remains patient centered and that the selected goals are SMART (specific, measurable, agreed upon, realistic, timely).

The macronutrient composition of the diet will vary with the patient's preference and medical condition. The 2020 U.S. Department of Agriculture Dietary Guidelines for Americans (Chap. 332), which focus on health promotion and risk reduction, can be applied to treatment of patients who are overweight or obese. The recommendations include maintaining a diet rich in whole grains, fruits, vegetables, and dietary fiber; decreasing sodium intake to <2300 mg/d; consuming fat-free or low-fat dairy products; and keeping added sugars and saturated fat intake to <10% of daily calories. Application of these guidelines to specific calorie goals can be



**FIGURE 402-2** Treatment algorithm—chronic disease management model for primary care of patients with overweight and obesity. This algorithm applies to the assessment of overweight and obesity and subsequent decisions based on that assessment. BMI indicates body mass index; CVD, cardiovascular disease; FDA, U.S. Food and Drug Administration. (Reproduced with permission from MD Jensen et al: 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: A report of the American College of Cardiology/American Heart Association task force on practice guidelines and The Obesity Society. *Circulation* 129(25 Suppl 2):S102, 2014.)

**TABLE 402-4** A Guide to Opting for Treatment for Obesity

TREATMENT	BMI CATEGORY (kg/m <sup>2</sup> )				
	25-26.9	27-29.9	30-34.9	35-39.9	40
Diet, exercise, behavioral therapy	With comorbidities	With comorbidities	+	+	+
Pharmacotherapy	—	With comorbidities	+	+	+
Surgery	—	—	—	With comorbidities	+

Source: Reproduced with permission from U.S. Department of Health and Human Services Public Health Service, National Institute of Health National Heart, Lung and Blood Institute. The Practical Guide Identification, Evaluation, and Treatment of Overweight and Obesity in Adults. NIH Publication Number 00-4084. October 2000.

found on the website [www.choosemyplate.gov](http://www.choosemyplate.gov). Since portion control is one of the most difficult strategies for patients to manage, the use of preprepared products such as meal replacements is a simple and convenient suggestion. Examples include frozen entrees, protein shakes, and bars. Use of meal replacements in the diet has been shown to result in a 7–8% weight loss.

Numerous randomized trials comparing diets of different macronutrient composition (e.g., low-carbohydrate, low-fat, Mediterranean) have shown that weight loss depends primarily on reduction of total caloric intake and adherence to the prescribed diet, not the specific proportions of carbohydrate, fat, and protein in the diet. The macronutrient composition will ultimately be determined by the patient's taste preferences, cooking style, and culture. However, the patient's underlying medical problems are also important in guiding the recommended dietary composition. The dietary prescription will vary according to the patient's metabolic profile and risk factors. A consultation with a registered dietitian for medical nutrition therapy is particularly useful in considering patient preference and treatment of comorbid diseases.

Another dietary approach to consider is based on the concept of *energy density*, which refers to the number of calories (i.e., amount of energy) a food contains per unit of weight. People tend to ingest a constant volume of food regardless of caloric or macronutrient content. Adding water or fiber to a food decreases its energy density by increasing weight without affecting caloric content. Examples of foods with low energy density include soups, fruits, vegetables, oatmeal, and lean meats. Dry foods and high-fat foods such as pretzels, cheese, egg yolks, potato chips, and red meat have a high energy density. Diets containing low-energy-dense foods have been shown to control hunger and thus to result in decreased caloric intake and weight loss.

Occasionally, very-low-calorie diets (VLCDs) are prescribed as a form of aggressive dietary therapy. The primary purpose of a VLCD is to promote a rapid and significant (13- to 23-kg) short-term weight loss over a 3- to 6-month period. The proprietary formulas designed for this purpose typically supply 800 kcal, 50–80 g of protein, and 100% of the recommended daily intake for vitamins and minerals. According to a review by the National Task Force on the Prevention and Treatment of Obesity, indications for initiating a VLCD include the involvement of well-motivated individuals who are moderately to severely obese, have failed at more conservative approaches to weight loss, and have a medical condition that would be immediately improved with rapid weight loss. These conditions include poorly controlled type 2 diabetes, hypertriglyceridemia, obstructive sleep apnea, and symptomatic peripheral edema. In the DiRECT trial of patients with type 2 diabetes and obesity, a low-energy formula diet (825–853 kcal/d) was administered for 3 months following by a structured monthly program. At 12 months, almost half of the participants achieved remission to a nondiabetic state and were not taking antidiabetic drugs. Use of formula diets should be prescribed by trained practitioners in a medical care setting where medical monitoring and high-intensity lifestyle intervention can be provided.

**Physical Activity Therapy** Although exercise alone is only moderately effective for weight loss, the combination of dietary modification and exercise is the most effective behavioral approach for the treatment of obesity. The most important role of exercise appears to be in the maintenance of the weight loss. The 2018 Physical Activity Guidelines for Americans ([www.health.gov/paguidelines](http://www.health.gov/paguidelines)) recommend that adults should engage in 150 min of moderate-intensity or 75 min a week of vigorous-intensity aerobic physical activity per week, preferably spread throughout the week. Focusing on simple ways to add physical activity into the normal daily routine through leisure activities, travel, and domestic work should be suggested. Examples include brisk walking, using the stairs, doing housework and yard work, and engaging in sports. Additionally, it is important to reduce sedentary behavior, which is associated with all-cause mortality and cardiovascular disease mortality in adults. Asking

the patient to wear a pedometer or accelerometer to monitor total accumulation of steps or kcal expended as part of the activities of daily living is a useful strategy. Step counts are highly correlated with activity level. Studies have demonstrated that lifestyle activities are as effective as structured exercise programs for improving cardiorespiratory fitness and weight loss. A high level of physical activity (>300 min of moderate-intensity activity per week) is often needed to lose weight and sustain weight loss. These exercise recommendations are daunting to most patients and need to be implemented gradually. Consultation with an exercise physiologist or personal trainer may be helpful.

**Behavioral Therapy** Cognitive behavioral therapy is used to help change and reinforce new dietary and physical activity behaviors. Strategies include self-monitoring techniques (e.g., journaling, weighing, and measuring food and activity); stress management; stimulus control (e.g., using smaller plates, not eating in front of the television or in the car); social support; problem solving; and cognitive restructuring to help patients develop more positive and realistic thoughts about themselves. When recommending any behavioral lifestyle change, the patient should be asked to identify what, when, where, and how the behavioral change will be performed. The patient should keep a record of the anticipated behavioral change so that progress can be reviewed at the next office visit. Because these techniques are time consuming to implement, their supervision is often undertaken by ancillary office staff, such as an advanced practice provider or registered dietitian.

## PHARMACOTHERAPY

Adjuvant pharmacologic treatments should be considered for patients with a BMI  $\geq 30 \text{ kg/m}^2$  or for patients with a BMI  $\geq 27 \text{ kg/m}^2$  who have concomitant obesity-related diseases and for whom dietary and physical activity therapy has not been successful. When an antiobesity medication is prescribed, patients should be actively engaged in a lifestyle program that provides the strategies and skills needed to use the drug effectively since such support increases total weight loss.

Medications for obesity fall into two major categories: those that affect appetite and those that inhibit gastrointestinal fat absorption. Since 2012, four new appetite-controlling medications were approved by the U.S. Food and Drug Administration (FDA) with an indication for chronic weight management, although one was voluntarily withdrawn in February 2020. These medications work biologically to suppress appetite, affecting hunger, satiety, and response to highly rewarding foods, thus making it easier for patients to follow their dietary intentions to restrict caloric intake. In addition, one capsule that is considered a medical device was marketed in 2020.

**Centrally Acting Medications** This class of medications affects satiety (feeling of fullness after a meal), hunger (the biologic sensation that prompts eating), and craving (intense desire for a specific food). By controlling appetite, these agents help patients reduce caloric intake without a sense of deprivation. The target site for the actions of these medications is primarily the hypothalamus and reward centers in the central nervous system (Chap. 401). The classic sympathomimetic adrenergic agents (benzphetamine, phendimetrazine, diethylpropion, mazindol, and phentermine) function by stimulating norepinephrine release or by blocking its reuptake. Among these agents, phentermine is the most commonly prescribed; there are limited long-term data on its effectiveness. A 2002 review of six randomized, placebo-controlled trials of phentermine for weight control found that patients lost 0.6–6.0 additional kg of weight over 2–24 weeks of treatment. The most common side effects of the amphetamine-derived agents are restlessness, insomnia, dry mouth, constipation, and increased blood pressure and heart rate.

PHEN/TPM is a combination drug that contains a catecholamine releaser (phentermine) and an anticonvulsant (topiramate). Topiramate is approved by the FDA as an anticonvulsant for the treatment

	PHEN/TPM		NALTREXONE SR/BUPROPION SR			LIRAGLUTIDE	
	EQUIP	CONQUER	COR-I	COR-II	COR-BMOD	SCALE	SCALE MAINTENANCE
No. of participants (ITT-LOCF)	1230	2487	1742	1496	793	3731	422
BMI (kg/m <sup>2</sup> )	35	27–45	30–45	30–45	30–45	27	27
Age (y)	18–70	18–70	18–65	18–65	18–65	18	18
Comorbid conditions (cardiovascular and metabolic)	1	2	1	1	1	1	1
Mean weight loss (%) with treatment vs placebo	10.9 vs 1.6	7.8 vs 1.2	6.1 vs 1.3	6.5 vs 1.9	9.3 vs 5.1	8.0 vs 2.6	6.2 vs 0.2
Placebo-subtracted weight loss (%)	9.3	6.6	4.8	4.6	4.2	5.4	6.0
Categorical change in 5% weight loss with treatment vs placebo	66.7 vs 17.3	62 vs 21	48 vs 16	50.5 vs 17.1	66.4 vs 42.5	63.2 vs 27.1	81.4 vs 48.9
Study completion rate, treatment vs placebo (%)	66.4 vs 52.9	69 vs 57	50	54	57.9 vs 58.4	71.9 vs 64.4	75 vs 69.5

Note: EQUIP, PHEN/TPM = 15/92 mg dose; CONQUER, PHEN/TPM = 7.5/46 mg dose.

Abbreviations: BMI, body mass index; ITT-LOCF, intention to treat, last observation carried forward; PHEN/TPM, phentermine/topiramate extended release.

of epilepsy and for the prophylaxis of migraine headaches. Weight loss was identified as an unintended side effect of topiramate during clinical trials for epilepsy. The mechanism responsible for weight loss is uncertain but is thought to be mediated through the drug's modulation of -aminobutyric acid receptors, inhibition of carbonic anhydrase, and antagonism of glutamate. PHEN/TPM has undergone two 1-year pivotal randomized, placebo-controlled, double-blind trials of efficacy and safety: EQUIP and CONQUER. In a third study, SEQUEL, 78% of CONQUER participants continued to receive their blinded treatment for an additional year. All participants received diet and exercise counseling. Participant numbers, eligibility, characteristics, and weight-loss outcomes are displayed in Table 402-5. Intention-to-treat 1-year placebo-subtracted weight loss for PHEN/TPM was 9.3% (15-mg/92-mg dose) and 6.6% (7.5-mg/46-mg dose), respectively, in the EQUIP and CONQUER trials. Clinical and statistical dose-dependent improvements were seen in selected cardiovascular and metabolic outcome measurements that were related to the weight loss. The most common adverse events experienced by the drug-randomized group were paresthesias, dry mouth, constipation, dysgeusia, and insomnia. Because of an increased risk of congenital fetal oral-cleft formation from topiramate, women of childbearing age should have a negative pregnancy test before treatment and monthly thereafter and use effective contraception consistently during medication therapy.

Lorcaserin was approved by the FDA for chronic weight management in 2012 and taken off the market in 2020. Lorcaserin was developed as a selective 5-HT<sub>2C</sub> receptor agonist with a functional selectivity ~15 times that of 5-HT<sub>2A</sub> receptors and 100 times that of 5-HT<sub>2B</sub> receptors. This selectivity is important, since the drug-induced valvulopathy documented with two other serotonergic agents that were removed from the market—fenfluramine and dexfenfluramine—was due to activation of the 5-HT<sub>2B</sub> receptors expressed on cardiac valvular interstitial cells. By activating the 5-HT<sub>2C</sub> receptor, lorcaserin is thought to decrease food intake through the pro-opiomelanocortin (POMC) system of neurons.

Lorcaserin underwent two randomized, placebo-controlled, double-blind trials for efficacy and safety. Intention-to-treat 1-year placebo-subtracted weight loss was 3.6% and 3.0%, respectively, in the two pivotal trials. Modest statistical improvements consistent with the weight loss were seen in selected cardiovascular and metabolic outcome measurements. However, a postmarketing cardiovascular outcome trial found that more patients taking lorcaserin (7.7%) were diagnosed with cancer compared to those taking a placebo (7.1%). The trial was conducted in 12,000 patients over 5 years. A range of cancer types was reported, with several different types of cancers occurring more frequently in the lorcaserin group, including pancreatic, colorectal, and lung.

Naltrexone SR/bupropion SR (NB) is a combination of an opioid antagonist and a mild reuptake inhibitor of dopamine and norepinephrine, respectively. Individually, naltrexone is approved by the FDA for the treatment of alcohol dependence and for the blockade of the effects of exogenously administered opioids, whereas bupropion is approved as an antidepressant and smoking cessation aid. As a combination drug, each component works in consort: bupropion stimulates secretion of -melanocyte-stimulating hormone (MSH) from POMC, whereas naltrexone blocks the feedback inhibitory effects of opioid receptors activated by the -endorphin released in the hypothalamus, thus allowing the inhibitory effects of MSH to reduce food intake.

The medication has undergone three randomized, placebo-controlled, double-blind trials for efficacy and safety. Participants were randomized to receive NB (8 mg/90 mg two tablets bid) or placebo in the three COR studies. Whereas participants received standardized nutritional and exercise counseling in COR-I and COR-II, a more intensive behavior modification program was provided in COR-BMOD (Table 402-5). Intention-to-treat 1-year placebo-subtracted weight loss was 4.8%, 5.1%, and 4.2%, respectively, in the COR-I, COR-II, and COR-BMOD trials. Clinical and statistical dose-dependent improvements were seen in selected cardiovascular and metabolic outcome measurements that were related to the weight loss. However, the medication led to slight increases or smaller decreases in blood pressure and pulse than placebo. The most common adverse events experienced by the drug-randomized groups were nausea, constipation, headache, vomiting, dizziness, diarrhea, insomnia, and dry mouth.

Liraglutide, the fourth new medication, is a glucagon-like peptide-1 (GLP-1) analogue with 97% homology to human GLP-1 that was previously approved for the treatment of type 2 diabetes at doses up to 1.8 mg once daily. In addition to its effect as an incretin hormone (glucose-induced insulin secretion), liraglutide inhibits both gastric emptying and glucagon secretion and stimulates GLP-1 receptors in the arcuate nucleus of the hypothalamus to reduce feeding.

Liraglutide has undergone three randomized, placebo-controlled, double-blind trials for efficacy and safety. Participants were randomized to receive liraglutide (3.0 mg SC daily) or placebo for initial weight loss—SCALE (patients without diabetes) and SCALE Diabetes (patients with diabetes)—or for weight maintenance after initial weight loss (SCALE Maintenance) (Table 402-5). All participants received diet and exercise counseling. For SCALE and SCALE Maintenance, patients were overweight or obese and had treated or untreated hypertension or dyslipidemia. Intention-to-treat 1-year placebo-subtracted weight loss was 5.4 and 6.1%, respectively, in the SCALE and SCALE Maintenance trials. Clinical and statistical dose-dependent improvements were seen in selected

cardiovascular and metabolic outcome measurements; however, there was a small increase in heart rate. The most common adverse effects include nausea, diarrhea, constipation, and vomiting. GLP-1 agonists should not be prescribed in patients with a family or personal history of medullary thyroid cancer or multiple endocrine neoplasia.

In approving the new antiobesity medications, the FDA introduced a new provision with important clinical relevance: a prescription trial period to assess effectiveness. Response to these medications should be assessed after 12 weeks of treatment for PHEN/TPM (or 16 weeks for NB and liraglutide since these medications are up titrated during the first month). Determining responsiveness at 3 or 4 months is based on the post hoc observed trial data that patients who did not lose a prespecified amount of weight early in treatment were less successful at 1 year. For PHEN/TPM, if the patient has not lost at least 3% of body weight at 3 months, the clinician can either escalate the dose and reassess progress at 6 months or discontinue treatment entirely. For NB, the medication should be discontinued if the patient has not lost at least 5% of body weight. The corresponding responsive target for liraglutide is a 4% weight loss.

**Peripherally Acting Medications** Orlistat is a synthetic hydrogenated derivative of a naturally occurring lipase inhibitor, lipostatin, that is produced by the mold *Streptomyces toxytricini*. This drug is a potent, slowly reversible inhibitor of pancreatic, gastric, and carboxylester lipases and phospholipase A<sub>2</sub>, which are required for the hydrolysis of dietary fat into fatty acids and monoacylglycerols. Orlistat acts in the lumen of the stomach and small intestine by forming a covalent bond with the active site of these lipases. Taken at a therapeutic dose of 120 mg tid, orlistat blocks the digestion and absorption of ~30% of dietary fat. After discontinuation of the drug, fecal fat content usually returns to normal within 48–72 h.

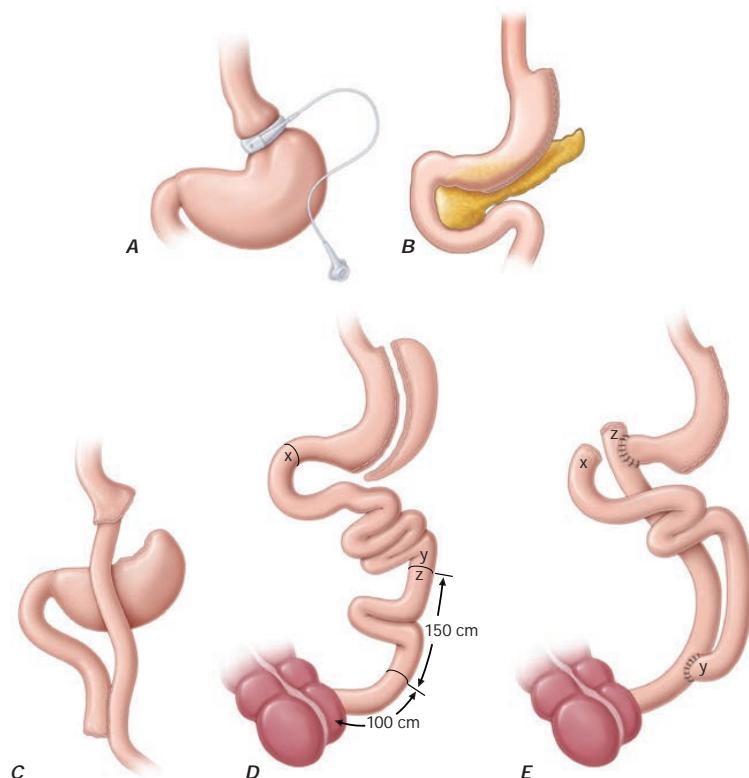
Multiple randomized, double-blind, placebo-controlled studies have shown that, after 1 year, orlistat produces a weight loss

of ~9–10%, whereas placebo recipients have a 4–6% weight loss. Because orlistat is minimally (<1%) absorbed from the gastrointestinal tract, it has no systemic side effects. The drug's tolerability is related to the malabsorption of dietary fat and the subsequent passage of fat in the feces. Adverse gastrointestinal effects, including flatus with discharge, fecal urgency, fatty/oily stool, and increased defecation, are reported in at least 10% of orlistat-treated patients. These side effects generally are experienced early, diminish as patients control their dietary fat intake, and only infrequently cause patients to withdraw from clinical trials. When taken concomitantly, psyllium mucilloid is helpful in controlling orlistat-induced gastrointestinal side effects. Because serum concentrations of the fat-soluble vitamins D and E and -carotene may be reduced by orlistat treatment, vitamin supplements are recommended to prevent potential deficiencies. Orlistat was approved for over-the-counter use in 2007.

**Oral Device** Gelesis100 is a nonsystemic, water-soluble gel that was approved by the FDA in 2019. In the stomach, the capsule releases the cellulose microgel, which absorbs water and forms a matrix with the consistency of food, occupying ~25% of the stomach. In the large intestine, it is broken down by enzymes and the cellulose is excreted. Gelesis100 and placebo were evaluated over 24 weeks in patients with BMI of 27 to 40 kg/m<sup>2</sup> and fasting plasma glucose of 90–145 mg/dL. Intention-to-treat, 24-week, placebo-subtracted weight loss was 2.1% (6.4 vs 4.4%). Gelesis100 treatment had no apparent increased safety risks. The capsules are approved for patients with a BMI of 25 kg/m<sup>2</sup>, with or without comorbidities.

## SURGERY

Bariatric surgery (Fig. 402-3) can be considered for patients with severe obesity (BMI >40 kg/m<sup>2</sup>) or for those with moderate obesity (BMI 35 kg/m<sup>2</sup>) associated with a number of comorbid conditions.



**FIGURE 402-3 Bariatric surgical procedures.** Examples of operative interventions used for surgical manipulation of the gastrointestinal tract. **A.** Laparoscopic adjustable gastric banding. **B.** Laparoscopic sleeve gastrectomy. **C.** The Roux-en-Y gastric bypass. **D.** Biliopancreatic diversion with duodenal switch. **E.** Biliopancreatic diversion.

Weight-loss surgeries have traditionally been classified into three categories on the basis of anatomic changes: restrictive, restrictive malabsorptive, and malabsorptive. More recently, however, the clinical benefits of bariatric surgery in achieving weight loss and alleviating metabolic comorbidities have been attributed largely to changes in the physiologic responses of gut hormones, bile acid metabolism, the microbiota, and adipose tissue metabolism. Metabolic effects resulting from bypassing the foregut include altered responses of ghrelin, GLP-1, peptide YY3-36, and oxyntomodulin. Additional effects on food intake and body weight control may be attributed to changes in vagal signaling. The loss of fat mass, particularly visceral fat, is associated with multiple metabolic, adipokine, and inflammatory changes that include improved insulin sensitivity and glucose disposal; reduced free fatty acid flux; increased adiponectin levels; and decreased interleukin 6, tumor necrosis factor  $\alpha$ , and high-sensitivity C-reactive protein levels.

Restrictive surgeries limit the amount of food the stomach can hold and slow the rate of gastric emptying. *Laparoscopic adjustable gastric banding* is the prototype of this category. The first banding device, the LAP-BAND, was approved for use in the United States in 2001. In contrast to previous devices, this band has a diameter that is adjustable by way of its connection to a reservoir that is implanted under the skin. Injection of saline into the reservoir and removal of saline from the reservoir tighten and loosen the band's internal diameter, respectively, thus changing the size of the gastric opening. Although the mean percentage of total body weight lost at 5 years is estimated at 20–25%, longer-term follow-up has been more disappointing, leading to near abandonment of the procedure. In the *laparoscopic sleeve gastrectomy*, the stomach is restricted by stapling and dividing it vertically, removing ~80% of the greater curvature and leaving a slim banana-shaped remnant stomach along the lesser curvature. Weight loss after this procedure is superior to that after laparoscopic adjustable gastric banding.

The three restrictive-malabsorptive bypass procedures combine the elements of gastric restriction and selective malabsorption: Roux-en-Y gastric bypass, biliopancreatic diversion, and biliopancreatic diversion with duodenal switch (Fig. 402-3). Roux-en-Y is the most commonly undertaken and most accepted bypass procedure. These procedures are routinely performed by laparoscopy.

These procedures generally produce a 30–35% average total body weight loss at 12–18 months followed by variable weight regain thereafter. Significant improvement in multiple obesity-related comorbid conditions, including type 2 diabetes, hypertension, dyslipidemia, obstructive sleep apnea, quality of life, and long-term cardiovascular events, has been reported. A meta-analysis of controlled clinical trials comparing bariatric surgery versus no surgery showed that surgery was associated with a reduced odds ratio (OR) risk of global mortality (OR = 0.55), cardiovascular death (OR = 0.58), and all-cause mortality (OR = 0.70).

Among the observed improvements in comorbidities, the prevention and treatment of type 2 diabetes resulting from bariatric surgery have garnered the most attention. Fifteen-year data from the Swedish Obese Subjects study demonstrated a marked reduction (i.e., by 78%) in the incidence of type 2 diabetes development among obese patients who underwent bariatric surgery. Multiple randomized controlled studies have shown greater weight loss and more improved glycemic control from 1 and 5 years among surgical patients than among patients receiving conventional medical therapy. A retrospective cohort study of >4000 adults with diabetes found that, overall, 68.2% of patients experienced an initial complete remission of type 2 diabetes within 5 years after surgery. However, among these patients, one-third redeveloped type 2 diabetes within 5 years. Patients with earlier-stage type 2 diabetes (i.e., those who do not need insulin, with shorter-duration disease, and with lower hemoglobin A<sub>1c</sub>) appear to have better improvement after bariatric surgery. The rapid improvement seen in diabetes after bariatric surgery is thought to be due to caloric restriction, reduced insulin resistance, and surgery-specific effects on glucose homeostasis brought about by alteration of gut hormones.

The mortality rate from bariatric surgery is generally <1% but varies with the procedure, the patient's age and comorbid conditions, and the experience of the surgical team. The most common surgical complications include stomal stenosis or marginal ulcers (occurring in 5–15% of patients) that present as prolonged nausea and vomiting after eating or inability to advance the diet to solid foods. These complications typically are treated by endoscopic balloon dilation and acid suppression therapy, respectively. For patients who undergo laparoscopic adjustable gastric banding, there are no intestinal absorptive abnormalities other than mechanical reduction in gastric size and outflow. Therefore, selective deficiencies are uncommon unless eating habits become unbalanced. In contrast, the restrictive-malabsorptive procedures carry an increased risk for micronutrient deficiencies of vitamin B<sub>12</sub>, iron, folate, calcium, and vitamin D. Patients with restrictive-malabsorptive procedures require lifelong supplementation with these micronutrients.

**Intraluminal Gastric Balloons** Three gastric balloon devices are approved for weight loss that are either placed in the stomach endoscopically (the REHAPE and ORBERA devices) or swallowed (OBALON). Efficacy of the devices at 6 months, based on a pooled weighted-mean percent weight loss, was 9.7%, and the control-subtracted percent weight loss was 5.6%. The devices are approved only for up to 6 months of use in adults with a BMI of 30–40 kg/m<sup>2</sup>. Adverse effects include nausea, vomiting, and abdominal pain.

## FURTHER READING

- Apovian CM et al: Pharmacological management of obesity: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 100:342, 2015.
- Garvey WT et al: American Association of Clinical Endocrinologists and American College of Endocrinology Comprehensive clinical practice guidelines for medical care of patients with obesity. *Endocr Pract* 22(suppl 3):1, 2016.
- Jensen MD et al: 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. *Circulation* 129(suppl 2):S102, 2014.
- Obesity Canada: Canadian Adult Obesity Clinical Guidelines (CPGs). Available at <https://obesitycanada.ca/guidelines/>. Accessed December 25, 2020.

403

## Diabetes Mellitus: Diagnosis, Classification, and Pathophysiology

Alvin C. Powers, Kevin D. Niswender,  
Carmella Evans-Molina



Diabetes mellitus (DM) refers to a group of common metabolic disorders that share the phenotype of hyperglycemia. Several distinct types of DM are caused by a complex interaction of genetics and environmental factors. Depending on the etiology of the DM, factors contributing to hyperglycemia include reduced insulin secretion, decreased glucose utilization, and increased glucose production. The metabolic dysregulation associated with DM causes secondary pathophysiologic changes in multiple organ systems that impose a tremendous burden on the individual with diabetes and on the health care system. In the United States, DM is the leading cause of end-stage renal disease

(ESRD), nontraumatic lower-extremity amputations, and adult blindness. Persons with diabetes are at increased risk for cardiovascular disease, which is the main cause of morbidity and mortality in this population.

## CLASSIFICATION

DM is classified on the basis of the pathogenic process leading to hyperglycemia (**Table 403-1**). There are two broad categories of DM, designated as either type 1 or type 2 DM. However, there is increasing recognition of other forms of diabetes in which the molecular pathogenesis is better understood and may be associated with a single gene defect. These alternative forms as well as other “atypical” forms may share features of type 1 and/or type 2 DM. Type 1 DM develops as a result of autoimmunity against the insulin-producing beta cells, resulting in insulin deficiency. Type 2 DM is a heterogeneous group of disorders characterized by variable degrees of insulin resistance, impaired insulin secretion, and increased hepatic glucose production. Defects in insulin action and/or secretion give rise to the common phenotype of hyperglycemia in type 2 DM and have important therapeutic implications now that pharmacologic agents are available to target specific metabolic derangements. Both type 1 and type 2 diabetes are preceded by a period of progressive worsening of glucose homeostasis, followed by the development of hyperglycemia that exceeds the threshold for clinical diagnosis. In terms of type 2 diabetes, this phase is referred to

**TABLE 403-1 Etiologic Classification of Diabetes Mellitus**

- I. Type 1 diabetes (immune-mediated beta cell destruction, usually leading to absolute insulin deficiency)
- II. Type 2 diabetes (may range from predominantly insulin resistance with relative insulin deficiency to a predominantly insulin secretory defect with insulin resistance)
- III. Specific types of diabetes (monogenic or MODY)
  - A. Genetic defects of beta cell development or function characterized by mutations in:
    1. Hepatocyte nuclear transcription factor (HNF) 4 $\alpha$
    2. Glucokinase
    3. HNF-1 $\alpha$
    4. Insulin promoter factor-1, HNF-1 $\beta$ , NeuroD1, and other pancreatic islet regulators/proteins such as *KLF11*, *PAX4*, *BLK*, *GATA4*, *GATA6*, *SLC2A2* (*GLUT2*), *RFX6*, *GLIS3*
    5. Insulin, leading to permanent neonatal diabetes
    6. Subunits of ATP-sensitive potassium channel, leading to permanent neonatal diabetes
    7. Mitochondrial DNA
  - B. Transient neonatal diabetes
  - C. Diseases of the exocrine pancreas—pancreatitis, pancreatectomy, neoplasia, cystic fibrosis, hemochromatosis, fibrocalculus pancreaticopathy, mutations in carboxyl ester lipase
  - D. Genetic defects in insulin action, including type A insulin resistance, Leprechaunism, Rabson-Mendenhall syndrome, lipodystrophy syndromes
  - E. Endocrinopathies—acromegaly, Cushing’s syndrome, glucagonoma, pheochromocytoma, hyperthyroidism, somafostatinoma, aldosteronoma
  - F. Drug- or chemical-induced—glucocorticoids, calcineurin and mTOR inhibitors (after organ transplantation), vorac (a rodenticide), pentamidine, nicotinic acid, diazoxide,  $\beta$ -adrenergic agonists, thiazides, hydantoin, asparaginase,  $\alpha$ -interferon, protease inhibitors, antipsychotics (atypicals and others), epinephrine
  - G. Infections—congenital rubella, cytomegalovirus, coxsackievirus
  - H. Uncommon forms of immune-mediated diabetes—"stiff-person" syndrome, anti-insulin receptor antibodies
  - I. Other genetic syndromes sometimes associated with diabetes—Wolfram syndrome, Down’s syndrome, Klinefelter’s syndrome, Turner’s syndrome, Friedreich’s ataxia, Huntington’s chorea, Laurence-Moon-Biedl syndrome, myotonic dystrophy, porphyria, Prader-Willi syndrome
- IV. Gestational diabetes mellitus (GDM)

*Abbreviation:* MODY, maturity-onset diabetes of the young or monogenic diabetes; see text.

*Source:* Data from American Diabetes Association. Standards of medical care in diabetes—2014. *Diabetes Care* 37:S14, 2014.

	Normal glucose tolerance	Hyperglycemia		
		Pre-diabetes*	Diabetes Mellitus	
		Impaired fasting glucose or impaired glucose tolerance	Not insulin required for requiring	Insulin required for control survival
FPG	<5.6 mmol/L (100 mg/dL)	5.6–6.9 mmol/L (100–125 mg/dL)	≥7.0 mmol/L (126 mg/dL) <sup>b</sup>	Symptoms of diabetes + random blood glucose concentration 11.1 mmol/L (200 mg/dL) <sup>a</sup>
HbA1C	<5.6%	5.7–6.4%	≥6.5% <sup>c</sup>	
2-h PG	<7.8 mmol/L (140 mg/dL)	7.8–11.0 mmol/L (140–199 mg/dL)	≥11.1 mmol/L (200 mg/dL) <sup>d</sup>	

**FIGURE 403-1 Spectrum of glucose homeostasis and diagnosis of diabetes mellitus (DM).** Glucose homeostasis is a spectrum from normal glucose tolerance (left portion of figure) to diabetes (right portion of figure) including type 1 DM, type 2 DM, specific types of diabetes, and gestational DM. The diagnostic criteria for diabetes are shown in the lower right portion of the figure and include the hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>), the fasting plasma glucose (FPG), and the 2-h plasma glucose (PG) after a glucose challenge. In most types of DM, the individual traverses from normal glucose tolerance to impaired glucose tolerance to overt diabetes (these should be viewed not as abrupt categories but as a spectrum). Changes in glucose tolerance may be bidirectional in some types of diabetes. For example, individuals with type 2 DM may return to the impaired glucose tolerance category with weight loss; in gestational DM, diabetes may revert to impaired glucose tolerance or normal glucose tolerance after delivery.<sup>a</sup> Random is defined as without regard to time since the last meal.<sup>b</sup> Fasting is defined as no caloric intake for at least 8 h.<sup>c</sup> Hemoglobin A<sub>1c</sub> test should be performed in a laboratory using a method approved by the National Glycohemoglobin Standardization Program and correlated to the reference assay of the Diabetes Control and Complications Trial. Point-of-care hemoglobin A<sub>1c</sub> should not be used for diagnostic purposes.<sup>d</sup> The test should be performed using a glucose load containing the equivalent of 75 g anhydrous glucose dissolved in water, not recommended for routine clinical use. Assessment of 1-h glucose may be helpful in diabetes risk prediction in individuals with cystic fibrosis or other forms of pancreatic disease. In the absence of unequivocal hyperglycemia and acute metabolic decompensation, the blood glucose criteria should be confirmed by repeat testing on a different day. These values do not apply to the diagnosis of gestational DM. Some use the term *increased risk for diabetes* or *intermediate hyperglycemia* (World Health Organization) rather than *prediabetes*. (Data from American Diabetes Association. 2. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes-2021. *Diabetes Care* 44:S15, 2021.)

as prediabetes and is more specifically classified as impaired fasting glucose (IFG) or impaired glucose tolerance (IGT) (**Fig. 403-1**). Recently, three distinct stages of type 1 DM have been defined based on the development of autoantibodies against pancreatic beta cell antigens or the development of worsening dysglycemia (discussed below).

## OTHER TYPES OF DM

Other etiologies of DM include specific genetic defects in insulin secretion or action, metabolic abnormalities that impair insulin secretion, mitochondrial abnormalities, and a host of conditions that impair glucose tolerance (Table 403-1). *Maturity-onset diabetes of the young* (MODY) and *monogenic diabetes* are subtypes of DM characterized by autosomal dominant inheritance, early onset of hyperglycemia (usually <25 years; sometimes in neonatal period), and impaired insulin secretion (discussed below). Mutations in the insulin receptor cause a group of rare disorders characterized by severe insulin resistance.

DM may also develop as a result of cystic fibrosis or chronic pancreatitis, in which the islets become damaged from a primary pathologic process originating in the pancreatic exocrine tissue. Hormones that antagonize insulin action can lead to DM. Hyperglycemia is often a feature of endocrinopathies such as acromegaly and Cushing’s disease. Viral infections have been implicated in pancreatic islet destruction but are an extremely rare cause of DM. A form of acute onset of type 1 diabetes, termed *fulminant diabetes*, has been noted in Japan and may be related to viral infection of the islets.

Glucose intolerance developing during the second or third trimester of pregnancy is classified as gestational diabetes mellitus (GDM). Insulin resistance is related to the metabolic changes of pregnancy, during which the increased insulin demands may lead to IGT or diabetes. The American Diabetes Association (ADA) recommends that diabetes diagnosed within the first trimester be classified as preexisting pregestational diabetes rather than GDM. In 2019, the International Diabetes Federation (IDF) estimated that 16% of pregnancies worldwide were affected by either GDM or preexisting DM. Most women with GDM revert to normal glucose tolerance postpartum but have a substantial risk (35–60%) of developing DM in the next 10–20 years. In addition, children born to a mother with GDM also have an increased risk of developing metabolic syndrome and type 2 DM later in life. Currently, the ADA recommends that women with a history of GDM undergo lifelong screening for the development of diabetes or prediabetes at least every 3 years.

### ATYPICAL DIABETES

It is increasingly recognized that some forms of diabetes have features of both type 1 and type 2 diabetes. These are distinct from monogenic forms (MODY) as they have not been linked to single gene defects. The development of a type 2 diabetes phenotype before puberty and a type 2 diabetes phenotype in very lean individuals are examples of atypical diabetes. An additional example is ketosis-prone diabetes, where individuals present with ketoacidosis, but do not require long-term exogenous insulin therapy. Many of these individuals are African American or Asian in heritage. Mechanisms underlying atypical forms of diabetes are being actively studied.

### EPIDEMIOLOGY AND GLOBAL CONSIDERATIONS

The worldwide prevalence of DM has risen dramatically over the past two decades, from an estimated 30 million cases in 1985 to 463 million in 2019 (Fig. 403-2). Based on current trends, the IDF projects that 642 million individuals will have diabetes by the year 2040 (see <http://www.idf.org>). Although the prevalence of both type 1 and type 2 DM is increasing worldwide, the prevalence of type 2 DM is rising much more rapidly, presumably because of dietary changes and increasing obesity, reduced activity levels as countries become more industrialized, and aging of the population. The incidence of type 1 diabetes has been increasing at a rate of 3% per year worldwide, with clear geographic differences. The cause for this increase is not well understood, but type 1 DM is increasingly being diagnosed at younger ages. In 2019, the prevalence of diabetes in individuals aged 20–79 years worldwide was 9.3%, ranging from 4.7–12.2%. The countries with the greatest number of individuals with diabetes in 2019 were China (116.4 million), India (77 million), the United States (31 million), Pakistan (19.4 million), Brazil (16.8 million), and Mexico (12.8 million). In the most recent estimate for the United States (2020), the Centers for Disease Control and Prevention (CDC) estimated that 10.5% of the population had diabetes. Diabetes affected 13% of all U.S. adults, and as many as 34% or 88 million U.S. adults had prediabetes. Approximately 21.4% of U.S. adults with diabetes in the United States were undiagnosed; globally, it is estimated that as many as 50% of individuals with diabetes may be undiagnosed. The prevalence of DM increases with age. The prevalence of DM in the United States was estimated to be 0.25% in individuals age <20 years, 4.2% in persons aged 18–44 years, and 17.5% in persons 45–64 years old. In individuals aged >65 years, the prevalence of DM was 26.8%. Similar age-related trends have been observed worldwide.

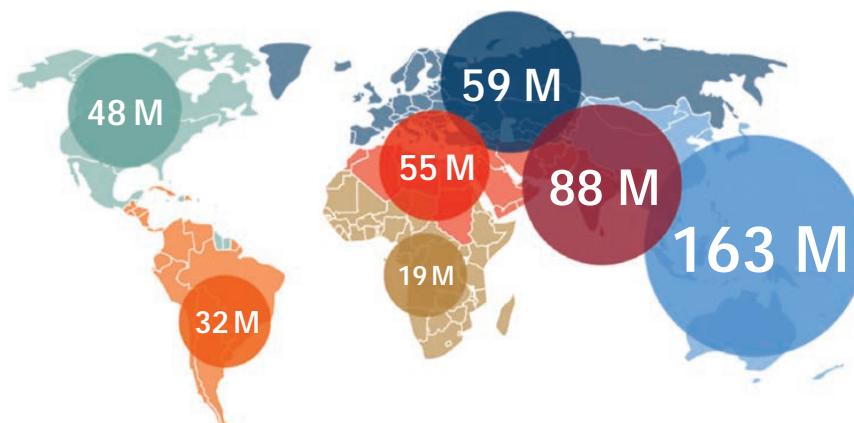
There is considerable geographic variation in the incidence of both type 1 and type 2 DM. Currently, Scandinavia, followed by Sardina and Portugal, have the highest incidence of type 1 DM; the lowest incidence is in the Pacific Rim where it is twenty- to thirtyfold lower. Northern Europe and the United States have an intermediate rate. Much of the increased risk of type 1 DM is believed to reflect the frequency of high-risk human leukocyte antigen (HLA) alleles among ethnic groups in different geographic locations. However, now populations less enriched with these classic high-risk HLA alleles are experiencing more rapid increases in type 1 DM incidence, suggesting an influence of environmental factors.

The prevalence of type 2 DM and its harbinger, IGT, is highest in certain Pacific islands and the Middle East and intermediate in countries such as India and the United States. This variability is likely due to genetic, behavioral, and environmental factors. DM prevalence also varies among different ethnic populations within a given country, with indigenous populations usually having a greater incidence of diabetes than the general population of the country. For example, the CDC estimated that the age-adjusted prevalence of DM in the United States (age >20 years; 2017–2018) was 7.5% in non-Hispanic whites, 9.2% in Asian Americans, 12.5% in Hispanics, and 11.7% in non-Hispanic blacks, but it exceeds 14% in American-Indian and Alaskan native populations. The onset of type 2 DM occurs, on average, at an earlier age in ethnic groups other than non-Hispanic whites. In Asia, the prevalence of diabetes is increasing rapidly, with an onset at a lower body mass index (BMI) and younger age, greater visceral adiposity, and reduced insulin secretory capacity.

Diabetes is a major cause of mortality. In recent years, diabetes has been listed as the seventh leading cause of death in the United States, but several studies indicate that diabetes-related deaths are likely underreported. In 2019, data from the IDF suggest that diabetes was responsible for nearly 4.2 million deaths worldwide, accounting for 11.3% of global all-cause mortality in adults aged 20–79 years. Diabetes also has important economic implications. In 2019, it was estimated that \$760 billion of health care expenditures worldwide were spent on diabetes (a range of 8–19% of total expenditures across regions). Up to 75% of individuals with diabetes live in low- or middle-income countries.

### DIAGNOSIS

Glucose tolerance is classified into three broad categories: normal glucose homeostasis, impaired glucose homeostasis, or DM (Fig. 403-1). Glucose tolerance can be assessed using the fasting plasma glucose (FPG), the response to oral glucose challenge, or the hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>). An FPG <5.6 mmol/L (100 mg/dL), a plasma glucose <7.9 mmol/L (140 mg/dL) following an oral glucose challenge, and an HbA<sub>1c</sub> <5.7% are considered to define normal glucose tolerance. The International Expert Committee with members appointed by



**FIGURE 403-2 Worldwide prevalence of diabetes mellitus.** Global estimate is 463 million individuals with diabetes in 2019. Regional estimates of the number of individuals with diabetes (20–79 years of age) are shown (2019). (Data from the *IDF Diabetes Atlas*, 9th ed. The International Diabetes Federation: 2019.)

**TABLE 403-2 Criteria for Screening for Type 2 Diabetes Mellitus in Adults**

- Consider testing in overweight or obese (BMI  $>25 \text{ kg/m}^2$ ,  $>23 \text{ kg/m}^2$  in Asian Americans, or other ethnically relevant definition who have these risk factors:
  - Family history of diabetes (i.e., parent or sibling with type 2 diabetes)
  - Race/ethnicity (e.g., African American, Latino, Native American, Asian American, Pacific Islander)
  - Hypertension (blood pressure  $>140/90 \text{ mmHg}$ )
  - HDL cholesterol level  $<35 \text{ mg/dL}$  ( $0.90 \text{ mmol/L}$ ) and/or a triglyceride level  $>250 \text{ mg/dL}$  ( $2.82 \text{ mmol/L}$ )
  - Polycystic ovary syndrome or acanthosis nigricans
  - History of cardiovascular disease
  - Physical inactivity
  - Other condition associated with insulin resistance (severe obesity, acanthosis nigricans)
- Individuals with previously identified IFG, IGT, or a hemoglobin A<sub>1c</sub> of 5.7–6.4% should be screened annually.
- Women who had GDM should be screened at least every 3 years.
- For other individuals, initiate testing at 45 years of age and repeat every 3 years.
- Individuals with HIV

**Abbreviations:** BMI, body mass index; GDM, gestational diabetes mellitus; HDL, high-density lipoprotein; IFG, impaired fasting glucose; IGT, impaired glucose tolerance; HIV, human immunodeficiency virus.

**Source:** Adapted with permission from American Diabetes Association. 2. Classification and diagnosis of diabetes: Standards of medical care in diabetes-2021. *Diabetes Care* 44:S15, 2021.

the ADA, the European Association for the Study of Diabetes, and the IDF have issued diagnostic criteria for DM (Table 403-2) based on the following premises: (1) the FPG, the response to an oral glucose challenge (oral glucose tolerance test [OGTT]), and HbA<sub>1c</sub> differ among individuals, and (2) DM is defined as the level of glycemia at which diabetes-specific complications occur rather than deviation from a population-based mean. For example, the prevalence of retinopathy in Native Americans (Pima Indian population) begins to increase at an FPG  $>6.4 \text{ mmol/L}$  (116 mg/dL) (Fig. 403-3).

Abnormal glucose homeostasis can be diagnosed by three distinct criteria (Fig. 403-1). First, *impaired fasting glucose* (IFG) is defined as a fasting plasma glucose (FPG) value of 5.6–6.9 mmol/L (100–125 mg/dL). Second, *impaired glucose tolerance* (IGT) is defined as a plasma

glucose level of 7.8–11 mmol/L (140–199 mg/dL) following an oral glucose challenge. Third, an HbA<sub>1c</sub> of 5.7–6.4% reflects dysglycemia by all mechanisms. While an HbA<sub>1c</sub> of 5.7–6.4%, IFG, and IGT do not identify the same individuals (i.e., different biologic mechanisms involved), individuals in all three groups are at greater risk of progressing to type 2 DM, have an increased risk of cardiovascular disease, and should be counseled about ways to decrease these risks (see below). Some use the terms *prediabetes*, *increased risk of diabetes*, or *intermediate hyperglycemia* (World Health Organization) and slightly different metrics for this category.

It is important to recognize that these values for the FPG, the glucose following an oral glucose challenge, and HbA<sub>1c</sub> are continuous rather than discrete variables; risk for comorbidities increases continuously rather than discretely by diagnostic category. A FPG  $7.0 \text{ mmol/L}$  (126 mg/dL), a glucose  $11.1 \text{ mmol/L}$  (200 mg/dL) 2 h after an oral glucose challenge, or an HbA<sub>1c</sub>  $6.5\%$  meets the criteria for the diagnosis of DM (Fig. 403-1). A random plasma glucose concentration  $11.1 \text{ mmol/L}$  (200 mg/dL) accompanied by classic symptoms of DM (polyuria, polydipsia, weight loss) is also sufficient for the diagnosis of DM. The current criteria for the diagnosis of DM emphasize the HbA<sub>1c</sub> and the FPG as the most reliable and convenient tests for identifying DM in asymptomatic individuals. However, some individuals may meet criteria for one test but not the other. Also, it is important to note that race and ethnicity may impact the reliability of HbA<sub>1c</sub> levels. For example, African Americans have a higher HbA<sub>1c</sub> value compared to non-Hispanic whites with a similar level of glycemia. An OGTT, although a valid means for diagnosing DM, is not often used in routine clinical care with the exception of pregnancy care and screening for gestational diabetes.

The diagnosis of DM has profound implications for an individual from both a medical and a financial standpoint. Thus, abnormalities on screening tests for diabetes should be repeated before making a definitive diagnosis of DM, unless acute metabolic derangements or a markedly elevated plasma glucose are present. These criteria also allow for the diagnosis of DM to be withdrawn in situations when the glucose intolerance reverts to normal.

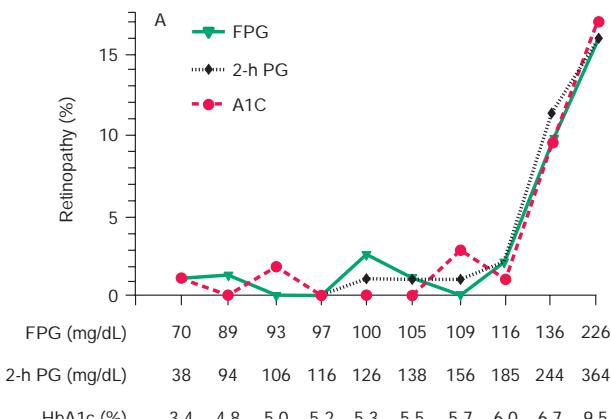
## SCREENING

Widespread use of the FPG or the HbA<sub>1c</sub> as a screening test for type 2 DM is recommended because (1) a large number of individuals who meet the current criteria for DM are asymptomatic and unaware that they have the disorder, (2) epidemiologic studies suggest that type 2 DM may be present for up to a decade before diagnosis, (3) some individuals with type 2 DM have one or more diabetes-specific complications at the time of their diagnosis, (4) treatment of type 2 DM may favorably alter the natural history of DM, and (5) diagnosis of prediabetes should spur efforts for diabetes prevention. The ADA recommends screening all individuals aged  $>45$  years every 3 years and screening individuals at an earlier age if they are overweight ( $\text{BMI} >25 \text{ kg/m}^2$  or ethnically relevant definition for overweight) and have one additional risk factor for diabetes. Although a number of immunologic markers for type 1 DM are becoming available (discussed below), their routine use outside a clinical trial is discouraged, pending the identification of clinically beneficial interventions for individuals at high risk for developing type 1 DM.

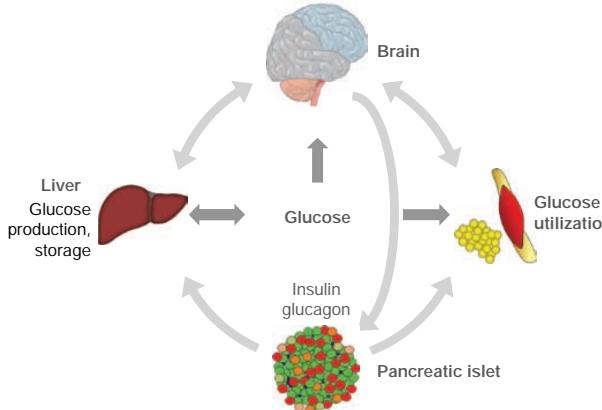
## REGULATION OF GLUCOSE HOMEOSTASIS

### OVERALL REGULATION OF GLUCOSE HOMEOSTASIS

Glucose homeostasis reflects a balance between energy intake from ingested food, hepatic glucose production (gluconeogenesis), and peripheral tissue glucose uptake and utilization. Insulin is the most important regulator of this metabolic equilibrium, but neural input, metabolic signals, and other hormones (e.g., glucagon) result in integrated control of glucose supply and utilization (Fig. 403-4). The organs that regulate glucose and lipids communicate by neural and humoral mechanisms with fat and muscle producing adipokines, myokines, and metabolites that influence liver function. In the fasting



**FIGURE 403-3 Relationship of diabetes-specific complication and glucose tolerance.** This figure shows the incidence of retinopathy in Pima Indians as a function of the fasting plasma glucose (FPG), the 2-h plasma glucose after a 75-g oral glucose challenge (2-h PG), or the hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>). Note that the incidence of retinopathy greatly increases at a FPG  $>116 \text{ mg/dL}$ , a 2-h PG of  $185 \text{ mg/dL}$ , or an HbA<sub>1c</sub>  $>6.5\%$ . (Blood glucose values are shown in mg/dL; to convert to mmol/L, divide value by 18.) (Modified with permission from Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 26:S5, 2003.)



**FIGURE 403-4** Regulation of glucose homeostasis. The organs shown contribute to glucose utilization, production, or storage. See text for a description of the communications (arrows), which can be neural or humoral. Although not shown, the GI tract and bone produce factors that influence glucose homeostasis.

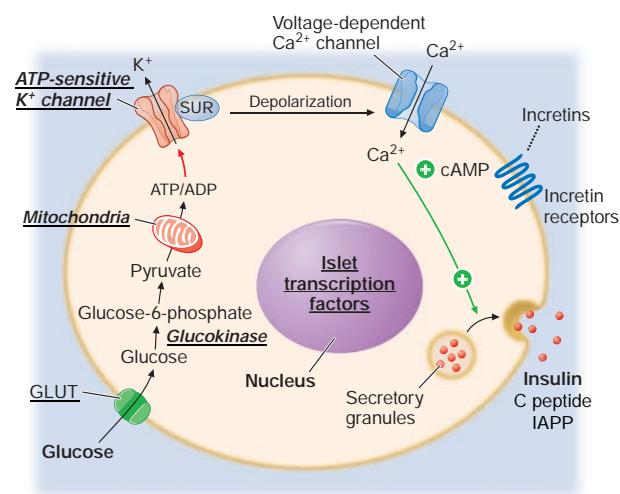
state, low insulin levels, together with modest increases in glucagon, increase glucose production by promoting hepatic gluconeogenesis and glycogen breakdown (glycogenolysis). In parallel, glucose uptake in insulin-sensitive tissues (skeletal muscle and fat) is reduced, and there is increased mobilization of gluconeogenic precursors such as amino acids and free fatty acids (lipolysis). Under normal conditions alpha cells increase glucagon secretion only when blood glucose or insulin levels are low or during exercise, but it is increased in fasting and postprandially in DM and stimulates excess glycogenolysis and gluconeogenesis by the liver and to a small degree by the renal medulla (Chap. 406). Conversely, in healthy people, the postprandial glucose load elicits a rise in insulin and fall in glucagon, leading to optimized glucose disposal. Insulin, an anabolic hormone, promotes the storage of carbohydrate and fat and protein synthesis. The major portion of postprandial glucose is utilized by skeletal muscle, an effect of insulin-stimulated glucose uptake. Other tissues, most notably the brain, utilize glucose in an insulin-independent fashion. Factors secreted by skeletal myocytes, adipocytes (e.g., leptin, resistin, adiponectin), and bone also influence glucose homeostasis.

### INSULIN BIOSYNTHESIS

Insulin, produced by the beta cells of the pancreatic islets, is initially synthesized as a single-chain 86-amino-acid precursor polypeptide, preproinsulin. Subsequent proteolytic processing removes the amino-terminal signal peptide, giving rise to proinsulin. Proinsulin is structurally related to insulin-like growth factors I and II, which bind weakly to the insulin receptor. Cleavage of an internal 31-residue fragment from proinsulin generates C-peptide with the A (21 amino acids) and B (30 amino acids) chains of insulin being connected by disulfide bonds. The mature insulin molecule and C-peptide are stored together and co-secreted from secretory granules in the beta cells. Because C-peptide is cleared more slowly than insulin, it is a useful marker of insulin secretion and allows discrimination of endogenous and exogenous sources of insulin in the evaluation of hypoglycemia (Chaps. 406 and 84). Elevated levels of serum proinsulin have been observed in both type 1 and 2 DM and are thought to be indicative of beta cell dysfunction. Pancreatic beta cells co-secrete islet amyloid polypeptide (IAPP) or amylin, a 37-amino-acid peptide, along with insulin. The role of IAPP in normal physiology is incompletely defined, but it is the major component of the amyloid fibrils found in the islets of patients with type 2 diabetes, and an analogue is sometimes used in treating type 1 and type 2 DM (Chap. 404).

### INSULIN SECRETION

Glucose is the key regulator of insulin secretion by the pancreatic beta cell, although amino acids, ketones, various nutrients, gastrointestinal peptides, and neurotransmitters also influence insulin secretion. Glucose levels  $>3.9$  mmol/L (70 mg/dL) stimulate insulin synthesis,



**FIGURE 403-5** Mechanisms of glucose-stimulated insulin secretion and abnormalities in diabetes. Glucose and other nutrients regulate insulin secretion by the pancreatic beta cell. Glucose is transported by a glucose transporter (GLUT1 and/or GLUT2 in humans, GLUT2 in rodents); subsequent glucose metabolism by the beta cell alters ion channel activity, leading to insulin secretion. The SUR receptor is the binding site for certain oral hypoglycemics (e.g., sulfonylureas, meglitinides); the other is an inwardly rectifying  $K^+$  channel protein ( $Kir6.2$ ). Inhibition of this  $K^+$  channel induces beta cell membrane depolarization, which opens voltage-dependent calcium channels (leading to an influx of calcium) and stimulates insulin secretion. Insulin secretion occurs in two phases, a rapid first-phase response, and a more prolonged second phase. Impaired first-phase insulin responses are among the earliest detectable abnormalities during the progression of both T1DM and T2DM. A number of metabolic pathways internal to the beta cell as well as external cues amplify glucose-stimulated insulin secretion. Glucagon-like peptide-1 (GLP-1) and glucose-dependent insulino tropic peptide (GIP) are incretin hormones that bind specific receptors on the beta cell to stimulate insulin secretion through cyclic AMP production, but they have this effect only when the blood glucose is above the fasting level. Incretin hormones also suppress glucagon production and secretion. Incretin analogues or pharmacologic agents that prolong the activity of endogenous GLP-1 are used to treat type 2 DM. Classically, GLP-1 release was thought to occur solely from neuroendocrine L-cells of the gastrointestinal tract following food ingestion. However, recent preclinical studies suggest that intraislet production of GLP-1 from alpha cells may play a role in the regulation of insulin secretion.

### INSULIN ACTION

Insulin is secreted into the portal venous system and acts to suppress endogenous hepatic glucose production and increase hepatic glucose uptake. A large portion (50%) of secreted insulin is cleared by the liver in this first pass, yielding a portal to peripheral insulin concentration gradient of ~2:1, with important implications for the clinical use of exogenous insulin (Ch. 404). Uncleared insulin enters the systemic circulation where it binds to receptors in peripheral target tissues such as skeletal muscle and adipose. Insulin binding to its receptor stimulates

intrinsic tyrosine kinase activity, leading to receptor autophosphorylation and the recruitment of intracellular signaling molecules, including the important insulin receptor substrates (IRS). IRS and other adaptor proteins initiate a complex cascade of phosphorylation and dephosphorylation reactions, resulting in the widespread metabolic and mitogenic effects of insulin. As an example, activation of the phosphatidylinositol-3'-kinase (PI-3-kinase) pathway stimulates translocation of a facilitative glucose transporter (e.g., GLUT4) to the cell surface, an event that is crucial for glucose uptake by skeletal muscle and fat. Activation of other insulin receptor signaling pathways induces glycogen synthesis, protein synthesis, lipogenesis, and regulation of various genes in insulin-responsive cells.

## PATHOGENESIS

### TYPE 1 DM

Type 1 DM is the result of interactions of genetic, environmental, and immunologic factors that ultimately lead to immune-mediated destruction of the pancreatic beta cells and insulin deficiency. Type 1 DM can develop at any age. Most, but not all, individuals with type 1 DM have evidence of islet-directed autoimmunity, which is detected by the presence of autoantibodies against beta cell antigens in the blood. The presence of two or more autoantibodies is now designated as stage 1 T1DM (Fig. 403-6). The temporal decline of beta cell function and mass preceding the development of type 1 DM is shown schematically in Fig. 403-6. In susceptible individuals, the autoimmune process is thought to be triggered by an infectious or environmental stimulus. In the majority of patients, autoantibodies against beta cell antigens appear after this triggering event, followed by progressive loss of insulin secretion. The rate of decline in beta cell function varies widely among individuals, with some patients progressing rapidly to clinical diabetes and others evolving to diabetes more slowly and over a period of several years. Features of diabetes do not become evident until a threshold loss of insulin secretion and beta cell mass occurs. Autopsy studies suggest the degree of loss of beta cell mass is variable at the time of disease presentation. At this point, residual, functional

beta cells exist but are insufficient in number and quality to maintain glucose tolerance. The events that trigger the transition from glucose intolerance to frank diabetes are often associated with increased insulin requirements, as might occur during infections or at puberty. After the initial clinical presentation of type 1 DM, a “honeymoon” phase may ensue during which time glycemic control is achieved with modest doses of insulin or, rarely, insulin is not needed. However, this fleeting phase of endogenous insulin production from residual beta cells disappears and the individual becomes insulin deficient. Many individuals with long-standing type 1 DM produce a small amount of insulin (as reflected by C-peptide production), and autopsy studies show that beta cells can persist in the pancreas decades after diagnosis.

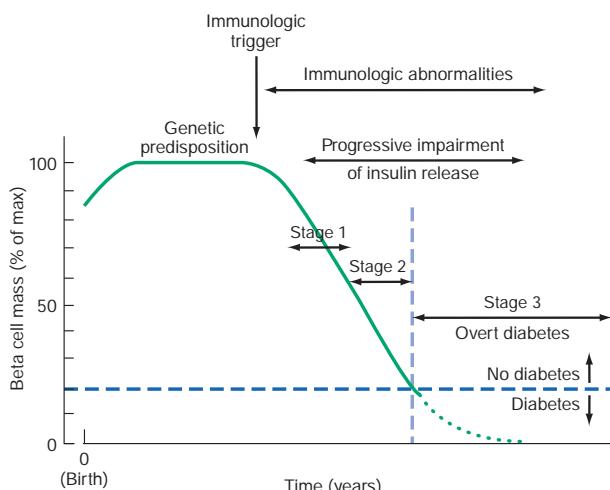
### GENETIC CONSIDERATIONS

 Susceptibility to type 1 DM involves multiple genes. The concordance of type 1 DM in identical twins ranges from 30–70%, indicating that additional modifying factors are likely involved in determining whether diabetes develops. The major susceptibility gene for type 1 DM is located in the HLA region on chromosome 6. Polymorphisms in the HLA complex account for approximately 50% of the genetic risk of developing type 1 DM. This region contains genes that encode the class II major histocompatibility complex (MHC) molecules, which present antigen to helper T cells and thus are involved in initiating the immune response (Chap. 349). The ability of class II MHC molecules to present antigen is dependent on the amino acid composition of their antigen-binding sites. Amino acid substitutions may influence the specificity of the immune response by altering the binding affinity of different antigens for class II molecules.

Many individuals with type 1 DM have the HLA DR3 and/or DR4 haplotype. Refinements in genotyping of HLA loci have shown that the haplotypes DQA1\*0301, DQB1\*0302, and DQB1\*0201 are most strongly associated with type 1 DM. These haplotypes are present in 40% of children with type 1 DM as compared to 2% of the U.S. population without type 1 DM. However, most individuals with predisposing haplotypes do not develop diabetes.

In addition to MHC class II associations, genome-wide association studies have identified more than 60 additional genetic loci that contribute susceptibility to type 1 DM (i.e., polymorphisms in the promoter region of the insulin gene, the CTLA-4 gene, interleukin 2 receptor, and PTPN22, etc.). Combined assessment of HLA and non-HLA loci using genetic risk scores has been used to improve prediction of type 1 diabetes risk. Notably, among recent cohorts of individuals with new-onset type 1 diabetes, there is a decreased representation of the highest-risk HLA alleles and increasing penetrance of disease in genotypes classically associated with lower risk, suggesting environmental factors may have an increasing role in disease pathogenesis. Genes that confer protection against the development of the disease also exist. The haplotype DQA1\*0102, DQB1\*0602 is extremely rare in individuals with type 1 DM (<1%) and appears to provide protection from type 1 DM.

Although the risk of developing type 1 DM is increased in relatives of individuals with the disease, the risk is relatively low: 1–9% if the parent has type 1 DM and 6–7% in a sibling (depending on which HLA haplotypes are shared). Hence, the majority of individuals with type 1 DM (>90%) do not have a relative with this disorder.



**FIGURE 403-6** Temporal model for development of type 1 diabetes. Individuals with a genetic predisposition are exposed to a trigger that initiates an autoimmune process, resulting in the development of islet autoantibodies and a gradual decline in beta cell function and mass. Stage 1 disease is characterized by the development of two or more islet cell autoantibodies but the maintenance of normoglycemia. Stage 2 disease is defined by continued autoimmunity and the development of dysglycemia. Stage 3 is defined by the development of hyperglycemia that exceeds the diagnostic criteria for the diagnosis of diabetes. The downward slope of the beta cell function varies among individuals and may not be continuous. A “honeymoon” phase may be seen in the first 1 or 2 years after the onset of diabetes and is associated with reduced insulin requirements. (Modified with permission from ER Kaufman: *Medical Management of Type 1 Diabetes*, 6th ed. Alexandria, VA: American Diabetes Association; 2012.)

**Pathophysiology** Pathologically, the pancreatic islets demonstrate a modest infiltration of lymphocytes (a process termed *insulitis*); however, the frequency of insulitis is heterogeneous both within and between individuals. Studies of the autoimmune process have identified the following abnormalities in the innate and adaptive arms of the immune system: (1) islet cell autoantibodies (ICAs); (2) activated lymphocytes in the islets, and peripancreatic lymph nodes; (3) T lymphocytes that proliferate when stimulated with islet proteins; and (4) release of cytokines within the insulitis. Islet cell autoantibodies (ICAs) are a composite of several different antibodies directed at pancreatic islet molecules such as GAD, insulin, IA-2/ICA-512, and ZnT-8, and serve as a marker of the autoimmune process of type 1 DM. Testing for ICAs can be useful in classifying the type of DM as type 1 as they are

present in the majority of individuals (>85%) diagnosed with new-onset type 1 DM. ICAs can also identify nondiabetic individuals at risk for developing type 1 DM, although their use for this purpose has been restricted mostly to research studies. In children with high genetic risk followed as part of several birth cohort studies, the presence of two or more ICAs was associated with a nearly 70% risk of developing type 1 DM after 10 years of follow-up and an 80% risk of developing diabetes after 15 years of follow-up. These observations led to a revision in the staging system for type 1 DM (Fig. 403-5), in which the development of multiple autoantibodies is now defined as the onset of stage 1 type 1 DM. While ICAs can be detected in the serum, and their presence is an important biomarker of type 1 diabetes risk, the antibodies do not have a direct role in beta cell death. Beta cell destruction is mediated by direct CD8+ T cell-mediated cytotoxicity. Beta cells may exacerbate this process through the development of modified proteins or “neoantigens” and through increased presentation of these antigens on their cell surface via upregulation of MHC class I molecules. In addition, beta cells may be damaged by the toxic effects of cytokines (i.e., tumor necrosis factor [TNF- $\alpha$ ], interferon, and interleukin 1 [IL-1]) as well as reactive oxygen species generated by infiltrating immune cells. Efforts to suppress the autoimmune process at the time of diagnosis of diabetes have largely been ineffective or only temporarily effective in slowing beta cell destruction. Thus, increased emphasis has now been placed on intervening earlier in the disease course (i.e., during stage 1 and 2 disease; Fig. 403-6). In support of this notion, a single 14-day course of teplizumab, an Fc receptor-nombinding anti-CD3 monoclonal antibody, delayed the onset of stage 3 T1D in high-risk individuals with multiple autoantibodies and dysglycemia (i.e., stage 2 T1D) by a median of 2.7 years.

Although other islet cell types (alpha cells [glucagon-producing], delta cells [somatostatin-producing], or PP cells [pancreatic polypeptide-producing]) are functionally and embryologically similar to beta cells, they are spared from the autoimmune destruction. However, altered patterns of hormone secretion from these other cell types in type 1 DM likely contribute to metabolic instability. Alpha cell dysfunction is reflected by fasting and post-prandial hyperglucagonemia but an impaired glucagon response to hypoglycemia.

**Environmental Factors** Numerous environmental events have been proposed to trigger the autoimmune process in genetically susceptible individuals; however, none has been conclusively linked to diabetes. Identification of an environmental trigger has been difficult because the event may precede the onset of DM by several years (Fig. 403-6). Putative environmental triggers include viruses (coxsackie, rubella, enteroviruses most prominently), bovine milk proteins, nitrosourea compounds, vitamin D deficiency, and environmental toxins. There is increasing interest in the microbiome and type 1 diabetes (Chap. 471).

## TYPE 2 DM

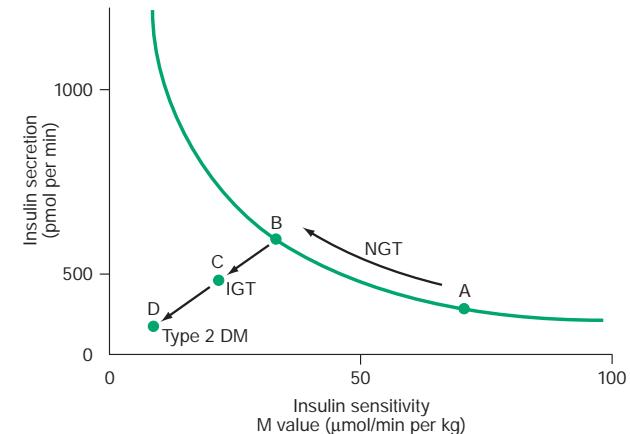
Insulin resistance and abnormal insulin secretion are central to the development of type 2 DM. Although the primary defect is controversial, most studies support the view that insulin resistance precedes an insulin secretory defect but that diabetes develops only when insulin secretion becomes inadequate. Type 2 DM likely encompasses a range of disorders with the common phenotype of hyperglycemia. Historically, our understanding of the pathophysiology and genetics is based on studies of individuals of European descent. Studies in more diverse populations have yielded unique insights into pathophysiologic differences among ethnic groups. In general, Latinos have greater insulin resistance and East Asians and South Asians have more beta cell dysfunction, but both defects are present in both populations. East and South Asians appear to develop type 2 DM at a younger age and a lower BMI. In some groups, DM that is ketosis prone (often in obese individuals) or ketosis-resistant (often lean) is sometimes seen. For example, African Americans can be more prone to nonketotic hyperosmolar presentation of diabetes exacerbations. In many forms of type 2 DM, the social determinants of health play a major role in the rates of type 2 DM.

## GENETIC CONSIDERATIONS

Type 2 DM has a strong genetic component. The concordance of type 2 DM in identical twins is between 70% and 90%. Individuals with a parent with type 2 DM have an increased risk of diabetes; if both parents have type 2 DM, the risk approaches 70%. Insulin resistance, as demonstrated by reduced glucose utilization in skeletal muscle, is present in many nondiabetic, first-degree relatives of individuals with type 2 DM. The disease is polygenic and multifactorial, because in addition to genetic susceptibility, environmental factors (such as obesity, poor nutrition, and physical inactivity) modulate the phenotype. Shared environmental and lifestyle factors also contribute to the high concordance in families. Further, the in utero environment contributes to, and either increased or reduced birth weight increases the risk of type 2 DM in adult life. Children of pregnancies complicated by gestational hyperglycemia also exhibit an increased risk of type 2 DM.

The genes that predispose to type 2 DM are incompletely identified, but genome-wide association studies have identified a large number of genes that convey a relatively small risk for type 2 DM (several hundred genes each with a relative risk of 1.06–1.5). Most prominent is a variant of the transcription factor 7-like 2 gene that has been associated with both type 2 DM and IGT in several populations. Genetic polymorphisms associated with type 2 DM have also been found in the genes encoding the peroxisome proliferator-activated receptor, inward rectifying potassium channel, zinc transporter, IRS, and calpain 10. The mechanisms by which these genetic loci increase the susceptibility to type 2 DM are not clear, but most are predicted to alter islet function or development or insulin secretion. Although the genetic susceptibility to type 2 DM is under active investigation (it is estimated that <10% of genetic risk is determined by loci identified thus far), it is currently not possible to use a combination of known genetic loci to reliably predict type 2 DM.

**Pathophysiology** Type 2 DM is characterized by impaired insulin secretion, insulin resistance, excessive hepatic glucose production, abnormal fat metabolism, and systemic low-grade inflammation. Obesity, particularly visceral or central (as evidenced by the hip-waist ratio), is very common in type 2 DM (80% of patients are obese). In the early stages of the disorder, glucose tolerance remains near-normal, despite insulin resistance, because the pancreatic beta cells compensate by increasing insulin output (Fig. 403-7). A number



**FIGURE 403-7** Metabolic changes during the development of type 2 diabetes mellitus (DM). Insulin secretion and insulin sensitivity are related, and as an individual becomes more insulin resistant (by moving from point A to point B), insulin secretion increases. A failure to compensate by increasing the insulin secretion results initially in impaired glucose tolerance (IGT; point C) and ultimately in type 2 DM (point D). NGT, normal glucose tolerance. (Data from SE Kahn: Clinical review 135: The importance of beta-cell failure in the development and progression of type 2 diabetes. *J Clin Endocrinol Metab* 86:4047, 2001 and RN Bergman, M Ader: Free fatty acids and pathogenesis of type 2 diabetes mellitus. *Trends Endocrinol Metab* 11:351, 2000.)

of pathophysiologic mechanisms contribute to type 2 DM and their relative importance varies from individual to individual. As insulin resistance and compensatory hyperinsulinemia progress, the pancreatic islets in certain individuals are unable to sustain the hyperinsulinemic state, manifesting as IGT, defined as elevations in postprandial glucose. A decline in insulin secretion and/or increased glucagon secretion causes an increase in hepatic glucose production leading to fasting hyperglycemia. Ultimately, frank beta cell failure ensues as a combination of these mechanisms leading to the manifestation of type 2 diabetes.

**Metabolic Abnormalities** Insulin resistance, the decreased ability of insulin to act effectively on target tissues (especially muscle, liver, and fat), is a prominent feature of type 2 DM and results from a combination of genetic susceptibility, obesity, and metabolic inflammation. Insulin resistance is relative, however, because supranormal levels of circulating insulin will normalize the plasma glucose. In type 2 DM, both insulin potency and efficacy are reduced leading to an overall decrease in glucose utilization under many conditions (30–60% lower than in normal individuals). Insulin resistance impairs glucose utilization by insulin-sensitive tissues (skeletal muscle) and in liver, coupled with elevated glucagon, leads to increased hepatic glucose output. Increased hepatic glucose output predominantly accounts for increased FPG levels, whereas decreased peripheral glucose utilization results in postprandial hyperglycemia.

The precise molecular mechanism leading to insulin resistance in type 2 DM has not been elucidated. Insulin receptor levels and tyrosine kinase activity in skeletal muscle are reduced, but these alterations are most likely secondary to hyperinsulinemia and are not a primary defect. Therefore, “postreceptor” defects in insulin-regulated phosphorylation/dephosphorylation appear to play the predominant role in insulin resistance. Abnormalities include the accumulation of lipid intermediates within skeletal myocytes, which may impair mitochondrial oxidative phosphorylation and reduce insulin-stimulated mitochondrial ATP production. Impaired fatty acid oxidation and lipid accumulation within skeletal myocytes also may generate reactive oxygen species such as lipid peroxides. These and other mechanisms also generate low-grade metabolic inflammation that feeds back and directly worsens insulin resistance. Of note, not all insulin signal transduction pathways are resistant to the effects of insulin (e.g., those controlling cell growth and differentiation using the mitogenic-activated protein kinase pathway). Consequently, hyperinsulinemia may increase the insulin action through these pathways, potentially accelerating diabetes-related conditions such as atherosclerosis.

The obesity accompanying type 2 DM, particularly in a central or visceral location, is thought to be part of the pathogenic process (**Chap. 401**). In addition to these white fat depots, humans have brown fat, which has much greater thermogenic capacity. Efforts are underway to increase the activity or quantity of brown fat. The increased adipocyte mass leads to increased levels of circulating free fatty acids and other fat cell products. For example, adipocytes secrete a number of biologic products (nonesterified free fatty acids, retinol-binding protein 4, leptin, TNF- $\alpha$ , resistin, IL-6, and adiponectin). Further, adipose resident macrophages are an important source of metabolic inflammation in diabetes. In addition to regulating body weight, appetite, and energy expenditure, adipokines also modulate insulin sensitivity. The increased production of free fatty acids and some adipokines may cause insulin resistance in skeletal muscle and liver. The venous drainage of the visceral adipose beds is the portal circulation and this likely contributes to hepatic dysfunction. Free fatty acids also impair glucose utilization in skeletal muscle, promote glucose production by the liver, and impair beta cell function. In contrast, the production by adipocytes of adiponectin, an insulin-sensitizing peptide, is reduced in obesity, and this may contribute to hepatic insulin resistance. Adipocyte products and adipokines also produce an inflammatory state and may explain why markers of inflammation such as IL-6 and C-reactive protein are often elevated in type 2 DM.

**IMPAIRED INSULIN SECRETION** Insulin secretion and sensitivity are interrelated (Fig. 403-7). In type 2 DM, insulin secretion initially increases in response to insulin resistance to maintain normal glucose tolerance. Initially, the insulin secretory defect is mild and selectively involves glucose-stimulated insulin secretion, including a greatly reduced first secretory phase. The response to other nonglucose secretagogues, such as arginine, is preserved, but overall beta cell function is reduced by as much as 50% at the onset of type 2 DM. Abnormalities in proinsulin processing are reflected by increased secretion of proinsulin in type 2 DM. Eventually, the insulin secretory defect is progressive.

The reason(s) for the decline in insulin secretory capacity in type 2 DM is unclear. The assumption is that a second genetic defect—superimposed upon insulin resistance—leads to defects in beta cell function, mass, and potentially cellular identity and differentiation status. Islet amyloid polypeptide or amylin, co-secreted by the beta cell, forms amyloid fibrillar deposits found in the islets of individuals with long-standing type 2 DM. Whether such islet amyloid deposits are a primary or secondary event is not known. The metabolic environment of diabetes also negatively impacts islet function. For example, chronic hyperglycemia paradoxically impairs islet function (“glucose toxicity”) and leads to a worsening of hyperglycemia. Improvement in glycemic control is often associated with improved islet function, an important clinical consideration. In addition, elevated levels of free fatty acids (“lipotoxicity”), and systemic and local elevations in pro-inflammatory cytokines from increased numbers of islet-associated macrophages, may also worsen islet function.

**INCREASED HEPATIC GLUCOSE AND LIPID PRODUCTION** In type 2 DM, insulin resistance in the liver reflects the failure of hyperinsulinemia to suppress gluconeogenesis, which results in fasting hyperglycemia and decreased glycogen storage by the liver in the postprandial state. Increased hepatic glucose production occurs early in the course of diabetes, although likely after the onset of insulin and glucagon secretory abnormalities and insulin resistance in skeletal muscle. As a result of insulin resistance in adipose tissue, lipolysis and free fatty acid flux from adipocytes are increased and efficiently cleared by liver leading to increased very-low-density lipoprotein (VLDL)-triglyceride synthesis in hepatocytes and secretion from liver. This is also responsible for the dyslipidemia found in type 2 DM (elevated triglycerides, reduced high-density lipoprotein [HDL], and increased small dense low-density lipoprotein [LDL] particles). If this lipid is retained, steatosis in the liver may lead to nonalcoholic fatty liver disease and abnormal liver function tests.

**Insulin Resistance Syndromes** The insulin resistance condition comprises a spectrum of disorders, with hyperglycemia representing one of the most readily diagnosed features. The *metabolic syndrome*, the *insulin resistance syndrome*, and *syndrome X* are terms used to describe a constellation of metabolic derangements that includes insulin resistance, hypertension, dyslipidemia (decreased HDL and elevated triglycerides), central or visceral obesity, type 2 DM or IGT/IFG, and accelerated cardiovascular disease. This syndrome is discussed in **Chap. 408**.

A number of relatively rare forms of severe insulin resistance include features of type 2 DM or IGT (Table 403-1). Mutations in the insulin receptor that interfere with binding or signal transduction are a rare cause of insulin resistance. Acanthosis nigricans and signs of hyperandrogenism (hirsutism, acne, and oligomenorrhea in women) are also common physical features. Two distinct syndromes of severe insulin resistance have been described in adults: (1) type A, which affects young women more severely and is characterized by severe hyperinsulinemia, obesity, and features of hyperandrogenism; and (2) type B, which affects middle-aged women and is characterized by severe hyperinsulinemia, features of hyperandrogenism, and autoimmune disorders. Individuals with type A insulin resistance syndrome have an undefined defect in the insulin-signaling pathway; individuals with type B insulin resistance syndrome have autoantibodies directed

**3102** at the insulin receptor. These receptor autoantibodies may block insulin binding or may stimulate the insulin receptor, leading to intermittent hypoglycemia.

Poly cystic ovary syndrome (PCOS) is a common disorder that affects premenopausal women and is characterized by chronic anovulation and hyperandrogenism (*Chap. 392*). Insulin resistance is seen in a significant subset of women with PCOS, and the disorder substantially increases the risk for type 2 DM, independent of the effects of obesity.

Lipodystrophies are a group of heterogeneous disorders characterized by selective loss of adipose tissue, leading to severe insulin resistance and hypertriglyceridemia. Lipodystrophies can be inherited or acquired and associated with variable degrees of adipose tissue loss.

**Prevention** Type 2 DM is preceded by a period of IGT or IFG, and a number of lifestyle modifications and pharmacologic agents prevent or delay the onset of DM. Individuals with prediabetes or increased risk of diabetes should be referred to a structured program to reduce body weight and increase physical activity as well as being screened for cardiovascular disease. The Diabetes Prevention Program (DPP) demonstrated that intensive changes in lifestyle (diet and exercise for 30 min/d five times/week) in individuals with IGT prevented or delayed the development of type 2 DM by 58% compared to placebo. This effect was seen in individuals regardless of age, sex, or ethnic group. In the same study, metformin prevented or delayed diabetes by 31% compared to placebo. The lifestyle intervention group lost 5–7% of their body weight during the 3 years of the study; the effects of the intervention persisted for at least 15 years. Studies in Finnish and Chinese populations noted similar efficacy of diet and exercise in preventing or delaying type 2 DM. A number of agents, including  $\alpha$ -glucosidase inhibitors, metformin, thiazolidinediones, GLP-1 receptor pathway modifiers, SGLT-2 inhibitors, and orlistat, prevent or delay type 2 DM but are not approved by the U.S. Food and Drug Administration for this purpose. Individuals with a strong family history of type 2 DM and individuals with IFG or IGT should be strongly encouraged to achieve a normal BMI and engage in regular physical activity. Pharmacologic therapy for individuals with prediabetes is currently controversial because its cost-effectiveness and safety profile are not known. The ADA suggests that metformin be considered in individuals with both IFG and IGT who are at very high risk for progression to diabetes (age <60 years, BMI >35 kg/m<sup>2</sup>, and women with a history of GDM). Individuals with IFG, IGT, or an HbA<sub>1c</sub> of 5.7–6.4% should be monitored annually to determine if diagnostic criteria for diabetes are present.

## GENETICALLY DEFINED, MONOGENIC FORMS OF DM RELATED TO REDUCED INSULIN SECRETION

Several monogenic forms of DM have been identified. Cases of maturity-onset diabetes of the young (MODY) or monogenic diabetes are caused by mutations in genes encoding islet-enriched transcription factors or glucokinase (Fig. 403-5; Table 403-1) and present with an autosomal dominant mode of transmission. MODY 1, MODY 3, and MODY 5 are caused by mutations in hepatocyte nuclear transcription factor (HNF) 4<sup>+</sup>, HNF-1<sup>+</sup>, and HNF-1<sup>+</sup>, respectively. As their names imply, these transcription factors are expressed in the liver but also in other tissues, including the pancreatic islets and kidney. These factors most likely affect islet development, the expression of genes important in glucose-stimulated insulin secretion or the maintenance of beta cell mass. For example, individuals with an HNF-1<sup>+</sup> mutation (MODY 3) have a progressive decline in glycemic control but may respond to sulfonylureas. In fact, some of these patients were initially thought to have type 1 DM but were later shown to respond to a sulfonylurea, and insulin was discontinued, a major clinical implication. Individuals with an HNF-1<sup>+</sup> mutation have progressive impairment of insulin secretion and hepatic insulin resistance, and require insulin treatment with minimal response to sulfonylureas. These individuals often have other abnormalities such as renal cysts, mild pancreatic

exocrine insufficiency, and abnormal liver function tests. Individuals with MODY 2, the result of mutations in the glucokinase gene, have mild-to-moderate, but stable hyperglycemia that does not respond to oral hypoglycemic agents, and otherwise does not require treatment. Glucokinase catalyzes the formation of glucose-6-phosphate from glucose, a reaction that is important for glucose sensing by the beta cells (Fig. 403-5) and for glucose utilization by the liver. As a result of glucokinase mutations, higher glucose levels are required to elicit insulin secretory responses, thus altering the set point for insulin secretion. MODY 4 is a rare variant caused by mutations in pancreatic and duodenal homeobox 1, a transcription factor that regulates pancreatic development and insulin gene transcription. Homozygous inactivating mutations cause pancreatic agenesis, whereas heterozygous mutations may result in DM. Studies of populations with type 2 DM suggest that mutations in MODY-associated genes are an uncommon (<5%) cause of type 2 DM.

Transient or permanent neonatal diabetes (onset <6 months of age) occurs. Permanent neonatal diabetes is a heterogeneous group of disorders caused by genetic mutations that impact beta cell function and/or pancreatic development (Fig. 403-5). Affected individuals typically require treatment with insulin and exhibit phenotypic overlap with type 1 DM. Activating mutations in the ATP-sensitive potassium channel subunits (Kir6.2 and ABCC8) impair glucose-stimulated insulin secretion. However, these individuals may respond to sulfonylureas and can be treated with these agents. Mutations in the transcription factor GATA6 are the most common cause of pancreatic agenesis. Homozygous glucokinase mutations cause a severe form of neonatal diabetes, while mutations in mitochondrial DNA are associated with diabetes and deafness. A number of mutations identified in the coding sequence of the insulin gene have been found to interfere with proinsulin folding, processing, and bioactivity and are designated as Mutant *Ins*-gene-induced Diabetes of Youth (MIDYs). Some of the neonatal diabetes syndromes are associated with a spectrum of neurologic dysfunction and a variety of extrapancreatic manifestations. Any individual who developed diabetes at 6 months of age or who has atypical features of type 1 or type 2 diabetes should be screened for forms of monogenic diabetes.

## APPROACH TO THE PATIENT

### Diabetes Mellitus

Once the diagnosis of DM is made, attention should be directed to symptoms related to diabetes (acute and chronic) and classifying the type of diabetes. DM and its complications produce a wide range of symptoms and signs; those secondary to acute hyperglycemia may occur at any stage of the disease, whereas those related to chronic hyperglycemia typically begin to appear during the second decade of hyperglycemia (*Chap. 405*). Because of long delays in clinical recognition, individuals with previously undetected type 2 DM may present with chronic complications of DM at the time of diagnosis. The history and physical examination should assess for symptoms or signs of acute hyperglycemia and screen for chronic microvascular and macrovascular complications and conditions associated with DM (*Chap. 405*).

### HISTORY

A complete medical history should be obtained with special emphasis on DM-relevant aspects such as current weight as well as any recent changes in weight, family history of DM and its complications, sleep history, risk factors for cardiovascular disease, exercise, smoking status, history of pancreatic disease, and ethanol use. Symptoms of hyperglycemia include polyuria, polydipsia, weight loss, fatigue, weakness, blurry vision, frequent superficial infections (vaginitis, fungal skin infections), and slow healing of skin lesions after minor trauma. Metabolic derangements relate mostly to hyperglycemia (osmotic diuresis) and to the catabolic state of the patient (urinary loss of glucose and calories, muscle breakdown due

to protein degradation and decreased protein synthesis). Blurred vision results from changes in the water content of the lens and resolves as hyperglycemia is controlled.

In a patient with established DM, the initial assessment should include a review of symptoms at the time of the initial diabetes diagnosis. This is an essential part of the history that can help define whether the correct type of DM has been diagnosed. Special emphasis should be placed on prior diabetes care, including types of therapies tried, the nature of any intolerance to previous therapies, prior HbA<sub>1c</sub> levels, self-monitoring blood glucose results, frequency of hypoglycemia (<3.0 mmol/L, <54 mg/dL), presence of DM-specific complications, and assessment of the patient's knowledge about diabetes, exercise, nutrition, and sleep history. Diabetes-related complications may afflict several organ systems, and an individual patient may exhibit some, all, or none of the symptoms related to the complications of DM (Chap. 405). In addition, the presence of DM-related comorbidities should be established (cardiovascular disease, hypertension, dyslipidemia). Pregnancy plans should be ascertained in women of childbearing age. The American Diabetes Association recommends that all women of childbearing age be counseled about the importance of tight glycemic control (HbA<sub>1c</sub> <6.5%) prior to conception.

#### PHYSICAL EXAMINATION

In addition to a complete physical examination, special attention should be given to DM-relevant aspects such as weight and BMI, retinal examination, orthostatic blood pressure, foot examination, peripheral pulses, and insulin injection sites. Depending on other risk factors, a blood pressure >130/80 mmHg or >140/90 mmHg is considered hypertension in individuals with diabetes. Because periodontal disease is more frequent in DM, the teeth and gums should also be examined.

An annual foot examination should (1) assess blood flow (pedal pulses), sensation (vibratory sensation [128-MHz tuning fork at the base of the great toe], the ability to sense touch with a monofilament [5.07, 10-g monofilament]), pinprick sensation, ankle reflexes, and nail care; (2) look for the presence of foot deformities such as hammer or claw toes and Charcot foot; and (3) identify sites of potential ulceration. The ADA recommends annual screening for distal symmetric polyneuropathy beginning with the initial diagnosis of diabetes and annual screening for autonomic neuropathy 5 years after diagnosis of type 1 DM and at the time of diagnosis of type 2 DM. This testing is aimed at detecting loss of protective sensation (LOPS) caused by diabetic neuropathy (Chap. 405).

#### CLASSIFICATION OF DM IN AN INDIVIDUAL PATIENT

The etiology of diabetes in an individual with new-onset disease can usually be assigned on the basis of clinical criteria. Individuals with type 1 DM are more likely to have the following characteristics: (1) lean body habitus; (2) requirement of insulin as the initial therapy; (3) propensity to develop ketoacidosis; and (4) a family or personal history of other autoimmune disorders such as autoimmune thyroid disease, adrenal insufficiency, pernicious anemia, celiac disease, and vitiligo. In contrast, individuals with type 2 DM often exhibit the following features: (1) obesity; 80% are obese, but elderly individuals may be lean; (2) may not require insulin therapy initially; and (3) may have associated conditions such as insulin resistance, hypertension, cardiovascular disease, dyslipidemia, or polycystic ovarian syndrome. In type 2 DM, insulin resistance is often associated with abdominal obesity (as opposed to hip and thigh obesity) and hypertriglyceridemia. Although most individuals diagnosed with type 2 DM are older, the age of diagnosis is declining, and there is a marked increase among overweight children and adolescents. Some individuals with phenotypic type 2 DM present with diabetic ketoacidosis but lack autoimmune markers and may be

later treated with oral glucose-lowering agents rather than insulin (this clinical picture is sometimes referred to as *ketoacidosis-prone type 2 DM*). On the other hand, some individuals (5–10%) with the phenotypic appearance of type 2 DM do not have absolute insulin deficiency but have autoimmune markers (GAD and other ICA autoantibodies) suggestive of type 1 DM (sometimes termed *latent autoimmune diabetes of the adult*). Such individuals are more likely to require insulin treatment within 5 years. Monogenic forms of diabetes should be considered in those with diabetes onset in childhood or early adulthood and especially those diagnosed within the first 6 months of life, an autosomal pattern of diabetes inheritance, diabetes without typical features of type 1 or 2 diabetes, and stable mild fasting hyperglycemia. Genetic testing should be considered in individuals suspected of having a monogenic form of diabetes as this may guide therapy selection. Despite recent advances in the understanding of the pathogenesis of diabetes, it often remains difficult to categorize some patients unequivocally. Individuals who deviate from the clinical profile of type 1 and type 2 DM, or who have other associated defects such as deafness, pancreatic exocrine disease (type 3c DM), and other endocrine disorders, should be classified accordingly (Table 403-1). A major goal is personalized or precision medicine in the diagnosis and treatment of diabetes.

#### LABORATORY ASSESSMENT

The laboratory assessment should first determine whether the patient meets the diagnostic criteria for DM (Fig. 403-1) and then assess the degree of glycemic control (Chap. 404). In addition to the standard laboratory evaluation, the patient should be screened for DM-associated conditions (e.g., albuminuria, dyslipidemia, thyroid dysfunction).

The classification of the type of DM may be facilitated by laboratory assessments. Serum C-peptide measurements may be useful but should always be interpreted with a concurrent blood glucose level. A low C-peptide in the setting of an elevated blood glucose level may confirm a patient's need for insulin. However, C-peptide levels are unable to completely distinguish type 1 from type 2 DM as many individuals with type 1 DM retain some C-peptide production. Measurement of islet cell antibodies at the time of diabetes onset may be useful if the type of DM is not clear based on the characteristics described above.

#### FURTHER READING

- Chung WK et al: Precision medicine in diabetes: A Consensus Report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetologia* 63:1671, 2020.
- Classification and Diagnosis of Diabetes: *Diabetes Care* 44:S15, 2021.
- Cole JB, Florez JC: Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol* 16:377, 2020.
- Dimeglio LA et al: Type 1 diabetes. *Lancet* 391:2449, 2018.
- Hill-Briggs F et al: Social determinants of health and diabetes: A scientific review. *Diabetes Care* 44:258, 2021.
- Insel RA et al: Staging presymptomatic type 1 diabetes: A scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care* 38:1964, 2015.
- Powers AC: Type 1 diabetes mellitus: Much progress, many opportunities. *J Clin Invest* 131:142242, 2021.
- Selph S et al: Screening for type 2 diabetes mellitus: A systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* 162:765, 2015.
- Skyler JS et al: Differentiation of diabetes by pathophysiology, natural history, and prognosis. *Diabetes* 66:241, 2017.
- Zhang H et al: Monogenic diabetes: a gateway to precision medicine in diabetes. *J Clin Invest* 131:e142244, 2021.

Alvin C. Powers, Michael J. Fowler,  
Michael R. Rickels

## OVERALL GOALS

The goals of therapy for type 1 or type 2 diabetes mellitus (DM) are to (1) eliminate symptoms related to hyperglycemia, (2) reduce or eliminate the long-term microvascular and macrovascular complications of DM ([Chap. 405](#)), and (3) allow the patient to achieve as normal a lifestyle as possible. To reach these goals, the physician should identify a target level of glycemic control for each patient, provide the patient with the educational and pharmacologic resources necessary to reach this level, and monitor/treat DM-related complications. Symptoms of diabetes usually resolve when the plasma glucose is <11.1 mmol/L (200 mg/dL), and thus most DM treatment focuses on achieving the second and third goals. This chapter first reviews the ongoing treatment of diabetes in the outpatient setting and then discusses the treatment of severe hyperglycemia, as well as the treatment of diabetes in hospitalized patients.

The care of an individual with either type 1 or type 2 DM requires a multidisciplinary team. Central to the success of this team are the patient's participation, input, and enthusiasm, all of which are essential for optimal diabetes management. Members of the health care team usually include the primary care provider and/or the endocrinologist or diabetologist, a certified diabetes educator, a nutritionist, a psychologist, and possibly a social worker. In addition, when the complications of DM arise, subspecialists (including ophthalmologists, neurologists, podiatrists, nephrologists, cardiologists, and cardiovascular surgeons) with experience in DM-related complications are essential.

## ONGOING ASPECTS OF COMPREHENSIVE DIABETES CARE

A number of names are sometimes applied to different approaches to diabetes care, such as intensive insulin therapy, intensive glycemic control, and "tight control." The current chapter, and other sources, uses the term *comprehensive diabetes care* to emphasize the fact that optimal diabetes therapy involves more than glucose management and medications and is patient-centered and individualized as advocated by the American Diabetes Association (ADA). Although glycemic control is central to optimal diabetes therapy, comprehensive diabetes care of both type 1 and type 2 DM should also detect and manage DM-specific complications ([Chap. 405](#)), and modify risk factors for DM-associated diseases. The key elements of comprehensive diabetes care are summarized in [Table 404-1](#). The morbidity and mortality of DM can be greatly reduced by timely and consistent surveillance, including the detection, prevention, and management of DM-related complications ([Table 404-1](#) and [Chap. 405](#)). Such screening procedures are indicated for all individuals with DM, but many individuals with diabetes do not receive these or comprehensive diabetes care. In addition to the physical aspects of DM, social, family, financial, cultural, and employment-related issues may impact diabetes care. The treatment goals for patients with diabetes summarized in [Table 404-2](#) should be individualized. The prevention and treatment of clinically significant hypoglycemia (<3.0 mmol/L or 54 mg/dL) is discussed in [Chap. 406](#). This chapter, while recognizing that resources available for diabetes care vary widely throughout the world, provides guidance for comprehensive diabetes care in health care settings with considerable societal resources.

**Lifestyle Management in Diabetes Care** The patient with type 1 or type 2 DM should receive education about nutrition, physical activity, psychosocial support, care of diabetes during illness, and medications to lower the plasma glucose. Patient education allows and encourages individuals with DM to assume greater responsibility for their care, leading to improved compliance.

**TABLE 404-1 Guidelines for Ongoing, Comprehensive Medical Care for Individuals with Diabetes**

- Individualized glycemic goal and therapeutic plan
- Self-monitoring at individualized frequency of blood glucose (capillary/meter) or interstitial glucose (continuous glucose monitoring)
- HbA<sub>1c</sub> testing (2–4 times/year)
- Lifestyle management in the care of diabetes, including:
  - Diabetes self-management education and support
  - Nutrition therapy
  - Physical activity
  - Psychosocial care, including evaluation for depression, anxiety
- Detection, prevention, or management of diabetes-related complications, including:
  - Diabetes-related eye examination (annual or biannual; [Chap. 405](#))
  - Diabetes-related foot examination (1–2 times/year by provider; daily by patient; [Chap. 403](#))
  - Diabetes-related neuropathy examination (annual; [Chap. 403](#))
  - Diabetes-related kidney disease testing (annual; [Chap. 405](#))
- Manage or treat diabetes-relevant conditions, including:
  - Blood pressure (assess 2–4 times/year; [Chap. 405](#))
  - Lipids (1–2 times/year; [Chap. 405](#))
  - Consider antiplatelet therapy with low-dose aspirin ([Chap. 405](#))
  - Influenza/pneumococcal/hepatitis B/coronavirus immunizations ([Chap. 6](#))

Abbreviation: HbA<sub>1c</sub>, glycosylated hemoglobin A<sub>1c</sub>.

### Diabetes Self-Management Education and Support (DSMES)

DSMES refers to ways to improve the patient's knowledge, skills, and abilities necessary for diabetes self-care and should also emphasize psychosocial issues and emotional well-being. Patient education is a continuing process with regular visits for reinforcement; it is not a process completed after one or two visits. It should receive special emphasis at the diagnosis of diabetes, annually, or at times when diabetes treatment goals are not attained, and during transitions in life or medical care. DSMES is delivered by a diabetes educator who is a health care professional (nurse, dietitian, or pharmacist) with specialized patient-education skills and who is certified in diabetes education (e.g., Association of Diabetes Care & Education Specialists). Education topics important for optimal diabetes self-care include self-monitoring of blood glucose (SMBG) and/or continuous glucose monitoring (CGM); urine or blood ketone monitoring (type 1 DM); insulin administration; guidelines for diabetes management during illnesses;

**TABLE 404-2 Treatment Goals for Adults with Diabetes<sup>a</sup>**

INDEX OF GLYCEMIC CONTROL <sup>b</sup>	GOAL (NONPREGNANT ADULTS)	GOAL (OLDER/HIGH-RISK ADULTS)
HbA <sub>1c</sub>	<7.0% (53 mmol/mol) <sup>c</sup>	<8.0% (64 mmol/mol) <sup>c</sup>
Preprandial capillary blood glucose	4.4–7.2 mmol/L (80–130 mg/dL)	5.0–7.8 mmol/L (90–140 mg/dL)
Postprandial capillary blood glucose <sup>d</sup>	<10.0 mmol/L (<180 mg/dL)	<11.1 mmol/L (200 mg/dL)
Time in range	>70%	>50%
3.9–10.0 mmol/L (70–180 mg/dL) <sup>e</sup>		
Time below 3.9 mmol/L (70 mg/dL) <sup>e</sup>	<4%	<1%
Glucose variability, % coefficient of variation <sup>e</sup>	36%	<33%

<sup>a</sup>As recommended by the American Diabetes Association; goals should be individualized for each patient (see text) with personalized goals for different patients. <sup>b</sup>HbA<sub>1c</sub> is primary goal and may also be estimated from 14 or more days of continuous glucose monitoring (CGM) data as the Glycemic Management Indicator (GMI). <sup>c</sup>Diabetes Control and Complications Trial-based assay. <sup>d</sup>1–2 h after beginning of a meal. <sup>e</sup>Derived from 14 days of CGM data.

Abbreviation: HbA<sub>1c</sub>, glycosylated hemoglobin A<sub>1c</sub>.

Source: Data from American Diabetes Association: 6. Glycemic targets: Standards of medical care in diabetes-2021. *Diabetes Care* 44(Suppl 1):S73, 2021.

prevention and management of hypoglycemia (**Chap. 406**); foot and skin care; diabetes management before, during, and after exercise; and risk factor-modifying activities. The focus is providing patient-centered, individualized education. More frequent contact between the patient and the diabetes management team (e.g., electronic, telephone, video) improves glycemic control.

**Nutrition Therapy** *Medical nutrition therapy* (MNT) is a term used by the ADA to describe the optimal coordination of caloric intake with other aspects of diabetes therapy (insulin, exercise, and weight loss). Some aspects of MNT are directed at preventing or delaying the onset of type 2 DM in high-risk individuals (obese or with prediabetes) by promoting weight reduction. Other measures of MNT are directed at improving glycemic control through limiting carbohydrate intake and avoiding simple sugars and fructose and managing diabetes-related complications (cardiovascular disease [CVD], nephropathy). Medical treatment of obesity including pharmacologic approaches that facilitate weight loss and metabolic surgery should be considered in selected patients (**Chaps. 401 and 402**).

In general, the components of optimal MNT are similar for individuals with type 1 or type 2 DM—high-quality, nutrient-dense with limits on carbohydrate intake required for glycemic control and weight management (**Table 404-3**). The data are currently inconclusive about various eating patterns (intermittent fasting, etc.). Dietary advice should be individualized, acknowledging personal preferences, culture, and religious traditions. Using the *glycemic index*, an estimate of the postprandial rise in the blood glucose when a certain amount of that food is consumed, may reduce postprandial glucose excursions and improve glycemic control.

The goal of MNT in type 1 DM is to coordinate and match the carbohydrate intake, both temporally and quantitatively, with the appropriate amount of insulin. MNT in type 1 DM is informed by SMBG and/or CGM that should be integrated to define the optimal insulin regimen. Based on the patient's estimate of the carbohydrate content of a meal, an insulin-to-carbohydrate ratio determines the bolus insulin dose for a meal or snack. MNT must be flexible enough to allow for exercise, and the insulin regimen must allow for variations in caloric

intake. An important component of MNT in type 1 DM is to minimize the weight gain often associated with intensive insulin therapy and is best achieved by placing limits on carbohydrate intake.

The goals of MNT in type 2 DM should focus on weight loss and address the greatly increased prevalence of cardiovascular risk factors (hypertension, dyslipidemia, obesity) and disease in this population. The majority of these individuals are obese, and weight loss is strongly encouraged. Very-low-carbohydrate diets that induce weight loss may result in rapid and dramatic glucose lowering in individuals with new-onset type 2 DM. MNT for type 2 DM should emphasize modest caloric reduction, increased physical activity, and weight loss (goal of at least 5–10% loss). Weight loss and exercise each independently improve insulin sensitivity.

Fasting for religious reasons, such as during Ramadan, presents a challenge for individuals with diabetes, especially those taking medications to lower the plasma glucose. Under International Diabetes Federation (IDF) guidelines on fasting during Ramadan, individuals are risk-stratified as those who can safely fast with medical evaluation and supervision and those in whom fasting is not advised. Thus, patient education and regular glucose monitoring are critical.

**Physical Activity** Exercise has multiple positive benefits including cardiovascular risk reduction, reduced blood pressure, maintenance of muscle mass, reduction in body fat, and weight loss. For individuals with type 1 or type 2 DM, exercise is also useful for lowering plasma glucose (during and following exercise) and increasing insulin sensitivity. In patients with diabetes, the ADA recommends 150 min/week (distributed over at least 3 days) of moderate aerobic physical activity with no gaps longer than 2 days. Resistance exercise, flexibility and balance training, and reduced sedentary behavior throughout the day are advised.

Despite its benefits, exercise may present challenges for some individuals with DM because they lack the normal glucoregulatory mechanisms (normally, insulin falls and glucagon rises during exercise). Skeletal muscle is a major site for metabolic fuel consumption in the resting state, and the increased muscle activity during vigorous, aerobic exercise greatly increases fuel requirements. Individuals with type 1 DM are prone to either hyperglycemia or hypoglycemia during exercise, depending on the preexercise plasma glucose, the circulating insulin level, lactate, and the level of exercise-induced catecholamines. If the insulin level is too low, the delivery of lactate to the liver and rise in catecholamines may increase the plasma glucose excessively, promote ketone body formation, and possibly lead to ketoacidosis. Conversely, if the circulating insulin level is excessive, this relative hyperinsulinemia may reduce hepatic glucose production (decreased glycogenolysis, decreased gluconeogenesis) and increase glucose entry into muscle, leading to hypoglycemia.

To avoid exercise-related hyper- or hypoglycemia, individuals with type 1 DM should (1) monitor blood glucose before, during, and after exercise; (2) delay exercise if blood glucose is >14 mmol/L (250 mg/dL) and ketones are present; (3) if the blood glucose is <5.0 mmol/L (90 mg/dL), ingest carbohydrate before exercising; (4) monitor glucose during exercise and ingest carbohydrate as needed to prevent hypoglycemia; (5) decrease insulin doses (based on previous experience) before and after exercise and inject insulin into a nonexercising area; and (6) learn individual glucose responses to different types of exercise. In individuals with type 2 DM, exercise-related hypoglycemia is less common but can occur in individuals taking either insulin or insulin secretagogues. Untreated proliferative retinopathy is a relative contraindication to vigorous exercise, because this may lead to vitreous hemorrhage or retinal detachment (**Chap. 405**).

**Psychosocial Care** Because the individual with DM faces challenges that affect many aspects of daily life, psychosocial assessment and support are a critical part of comprehensive diabetes care. The patient should view himself/herself as an essential member of the diabetes care team and not as someone who is cared for by the diabetes management team. Even with considerable effort, normoglycemia can be an elusive goal, and solutions to worsening glycemic control may not be easily identifiable. Depression, anxiety, or “diabetes distress,” defined by the ADA as “...negative psychological reactions related to

**TABLE 404-3 Nutritional Recommendations for Adults with Diabetes or Prediabetes<sup>a</sup>**

**General dietary guidelines**

- Vegetable, fruits, whole grains, legumes, low-fat dairy products and food higher in fiber and lower in glycemic content; optimal diet composition and eating patterns are not known

**Fat in diet** (optimal % of diet is not known; should be individualized)

- Mediterranean-style diet rich in monounsaturated and polyunsaturated fatty acids
- Minimal or no trans fat consumption

**Carbohydrate in diet** (optimal % of diet is not known; should be individualized)

- Monitor carbohydrate intake in regard to calories and set limits for meals to reduce postprandial glycemia
- Avoid fructose- and sucrose-containing beverages and minimize consumption of foods with added sugar that may displace healthier, more nutrient-dense food choices and elevate postprandial glycemia
- Estimate grams of carbohydrate in diet for flexible insulin dosing (type 1 DM and insulin-dependent type 2 DM)
- Consider using glycemic index to predict how consumption of a particular food may affect blood glucose

**Protein in diet** (optimal % of diet is not known; should be individualized)

**Other components**

- Reduced-calorie and nonnutritive sweeteners may be useful
- Routine supplements of vitamins, antioxidants, or trace elements not supported by evidence
- Sodium intake as advised for general population

<sup>a</sup>See text for differences for patients with type 1 or type 2 diabetes.

Source: Data from American Diabetes Association: 5. Facilitating behavior change and well-being to improve health outcomes: Standards of medical care in diabetes-2021. *Diabetes Care* 44(Suppl 1):S53, 2021.

emotional burdens...in having to manage a chronic disease like diabetes," should be recognized and may require the care of a mental health specialist. Emotional stress may provoke a change in behavior so that individuals no longer adhere to a dietary, exercise, or therapeutic regimen. Eating disorders, including binge eating disorders, bulimia, and anorexia nervosa, appear to occur more frequently in individuals with type 1 or type 2 DM.

## MONITORING THE LEVEL OF GLYCEMIC CONTROL

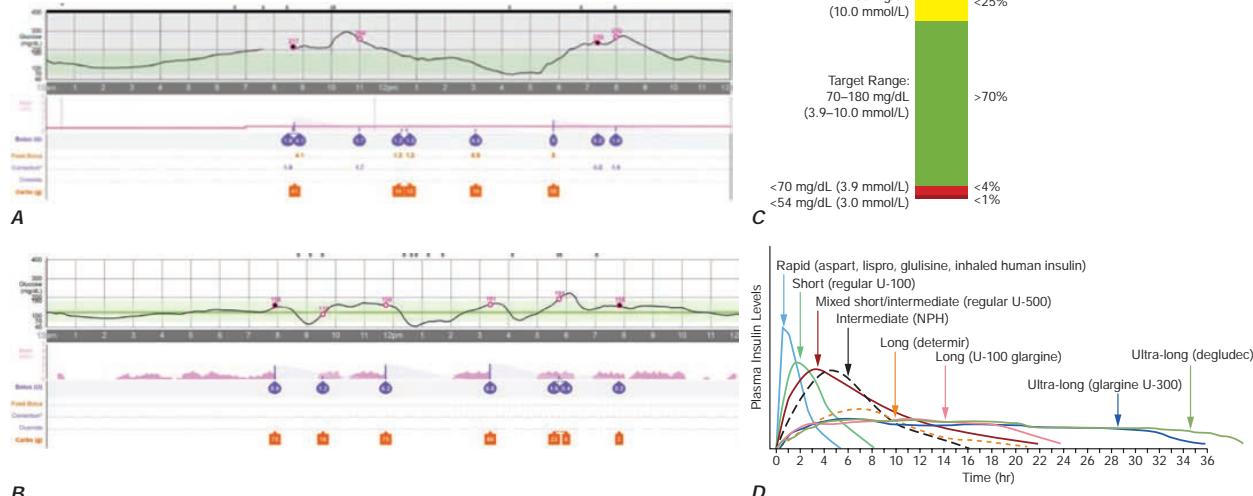
Optimal monitoring of glycemic control involves glucose measurements by the patient and an assessment of long-term control by the providers on the diabetes management team (measurement of hemoglobin A<sub>1c</sub> [HbA<sub>1c</sub>] and review of the patient's SMBG and/or CGM). These measurements are complementary: the patient's measurements provide a picture of short-term glycemic control, whereas the HbA<sub>1c</sub> reflects average glycemic control over the previous 2–3 months. Most measurements should be performed prior to a meal and supplemented with postprandial measurements to assist in reaching glucose targets (Table 404-2). By combining glucose measurements with diet and exercise history, the diabetes management team and patient can improve glycemic control. Clinical practice is changing rapidly with CGM replacing SMBG in many patients, especially those with T1 DM.

**Self-Monitoring of Blood Glucose** In SMBG, a small drop of blood (3–10 µL) and an enzymatic reaction allow rapid and accurate measurement of the capillary blood glucose by glucose monitors (calibrated to provide plasma glucose value even though blood glucose is measured). The blood is obtained from the fingertip; alternative testing sites (e.g., forearm) are less reliable. The frequency of SMBG measurements should be individualized. Individuals with type 1 DM or individuals with type 2 DM taking multiple insulin injections each day should measure their blood glucose >3 times/day (some measure >10 times/day). Most individuals with type 2 DM require less frequent monitoring, although the optimal frequency of SMBG has not been clearly defined. Individuals with type 2 DM who are taking insulin

should use SMBG more frequently than those on oral agents. Individuals with type 2 DM who are on oral medications should use SMBG as a means of assessing the efficacy of their medication and the impact of dietary choices and exercise. Because glucose levels fluctuate less in these individuals, one or two SMBG measurements per day may be sufficient.

**Continuous Glucose Monitoring** CGM technology utilizes a sensor or electrode to detect interstitial glucose, which is in equilibrium with blood glucose, but may lag behind when the blood glucose is changing. In one CGM approach, the interstitial glucose is detected and reported essentially continuously while in another approach, the sensor is in place, but the glucose is only recorded when a detector is placed over the sensor. The glucose sensors are placed subcutaneously and are replaced every 3–14 days. Some CGM requires calibration by SMBG. CGM provides unlimited glucose datapoints that can be used to define a time in a glycemic range (TIR, or time in range), the ambulatory glucose profile, the amount of time in the hypoglycemic range, and the glucose management indicator (GMI), which correlates with A1C (Fig. 404-1; Table 404-2). TIR and GMI are useful metrics but CGM also allows the patient to monitor the rate of glucose change and glucose trends that can be used to avoid predicted hyper- or hypoglycemia. CGM in type 1 DM especially in those with hypoglycemia unawareness can decrease the frequency of serious hypoglycemia (especially nocturnal hypoglycemia). The combination of an insulin-infusion device (discussed below) and a CGM can now automate insulin delivery with either predictive suspension of insulin delivery to avoid hypoglycemia or closed-loop control that automatically adjusts insulin delivery by a predictive algorithm (Fig. 404-1).

**Assessment of Long-Term Glycemic Control** Measurement of glycated hemoglobin (HbA<sub>1c</sub>) is the standard method for assessing long-term glycemic control. When plasma glucose is consistently elevated, there is an increase in nonenzymatic glycation of hemoglobin; this alteration reflects the glycemic history over the previous 2–3 months, because erythrocytes have an average life span of 120 days (glycemic



**FIGURE 404-1** Glycemic monitoring and insulin administration options for treatment of diabetes. **A.** CGM profile and delivery of rapid-acting insulin analog by continuous subcutaneous insulin infusion pump involves a basal rate (light purple line) and prandial and correction boluses (purple circles) based on estimated carbohydrate intake (orange squares) and an insulin sensitivity factor. **B.** CGM profile with sensor-communicating insulin pump that automates insulin delivery by suspending delivery for predicted hypoglycemia and increasing basal delivery for predicted hyperglycemia (light purple curves) while still requiring user input for estimated carbohydrate intake (orange squares) to provide prandial insulin boluses (purple circles). **C.** CGM profile is used to generate an estimate of time-in-range with glycemic goal shown on the left side of the bar and target % time in that glycemic range shown on the right side of the bar. **D.** Adapted with permission from JJ Neumiller: Insulin update: New and emerging insulins. American Diabetes Association, 2018.)

level in the preceding month contributes about 50% to the HbA<sub>1c</sub> value). Measurement of HbA<sub>1c</sub> at the “point of care” allows for more rapid feedback and may therefore assist in adjustment of therapy.

HbA<sub>1c</sub> should be measured in all individuals with DM during their initial evaluation and as part of their comprehensive diabetes care. As the primary predictor of long-term complications of DM, the HbA<sub>1c</sub> should mirror, to a certain extent, the short-term measurement by SMBG or CGM. Measurements of HbA<sub>1c</sub> and actual glucose levels are complementary in that recent intercurrent illnesses may impact SMBG or CGM measurements but not the HbA<sub>1c</sub>. The HbA<sub>1c</sub> may reflect postprandial or nocturnal hyperglycemia not detected by SMBG of fasting and preprandial capillary blood glucose. However, it does not detect interprandial or nocturnal hypoglycemia—these require very frequent SMBG or CGM for detection. The HbA<sub>1c</sub> is an “average” and thus does not detect glycemic variability in the way SMBG and CGM can. In standardized assays, the HbA<sub>1c</sub> approximates the following mean plasma glucose values: an HbA<sub>1c</sub> of 6% = 7.0 mmol/L (126 mg/dL), 7% = 8.6 mmol/L (154 mg/dL), 8% = 10.2 mmol/L (183 mg/dL), 9% = 11.8 mmol/L (212 mg/dL), 10% = 13.4 mmol/L (240 mg/dL), 11% = 14.9 mmol/L (269 mg/dL), and 12% = 16.5 mmol/L (298 mg/dL). However, there is interindividual variability in the HbA<sub>1c</sub> to mean glucose relationship, and in blacks the HbA<sub>1c</sub> is on average 0.4% higher than in whites for the same mean glucose. Clinical conditions leading to abnormal RBC parameters such as hemoglobinopathies, anemias, reticulocytosis, transfusions, and uremia may alter the HbA<sub>1c</sub> result. In patients achieving their glycemic goal, the ADA recommends measurement of the HbA<sub>1c</sub> at least twice per year. More frequent testing (every 3 months) is warranted when glycemic control is inadequate or when therapy has changed. Laboratory standards for the HbA<sub>1c</sub> test have been established and should be correlated to the reference assay of the Diabetes Control and Complications Trial (DCCT). The degree of glycation of other proteins, such as albumin, or measurement of 1,5-anhydroglucitol can be used as an alternative, shorter-term indicator of glycemic control when the HbA<sub>1c</sub> is inaccurate. The fructosamine assay (measuring glycated albumin) reflects the glycemic status over the prior 2 weeks.

## PHARMACOLOGIC TREATMENT OF DIABETES

Comprehensive care of type 1 and type 2 DM requires an emphasis on nutrition, exercise, and monitoring of glycemic control but also usually involves glucose-lowering medication(s). This chapter discusses classes of such medications but does not describe every glucose-lowering agent available worldwide. The initial step is to select an individualized, glycemic goal for the patient.

### ESTABLISHMENT OF TARGET LEVEL OF GLYCEMIC CONTROL

Because the complications of DM are related to glycemic control, normoglycemia or near-normoglycemia is the desired, but often elusive, goal for most patients. Normalization or near-normalization of the plasma glucose for long periods of time is extremely difficult, as demonstrated by the DCCT and United Kingdom Prospective Diabetes Study (UKPDS). Regardless of the level of hyperglycemia, improvement in glycemic control will lower the risk of diabetes-specific complications, most notably the microvascular complications (**Chap. 405**).

The target for glycemic control (as reflected by the HbA<sub>1c</sub>) must be individualized, and the goals of therapy should be developed in consultation with the patient after considering a number of medical, social, and lifestyle issues. The ADA calls this a *patient-centered approach*, and other organizations such as the IDF and American Association of Clinical Endocrinologists (AACE) also suggest an individualized glycemic goal. Important factors to consider include the patient's age and ability to understand and implement a complex treatment regimen, presence and severity of complications of diabetes, known CVD, ability to recognize hypoglycemic symptoms, presence of other medical conditions or treatments that might affect survival or the response to therapy, lifestyle and occupation (e.g., possible consequences of experiencing hypoglycemia on the job), and level of support available from family and friends.

In general, the ADA suggests that the goal is to achieve an HbA<sub>1c</sub> as close to normal as possible without significant hypoglycemia. In most individuals, the target HbA<sub>1c</sub> should be <7% (Table 404-2) with a more stringent (<6.5%) target for some patients. With modern implementation of intensive insulin therapy for type 1 DM, the level of HbA<sub>1c</sub> is no longer inversely related to the frequency and severity of hypoglycemia as seen in the DCCT; nevertheless, it may still be appropriate to set a higher HbA<sub>1c</sub> target <7.5 or 8% for patients with impaired awareness of hypoglycemia. A higher HbA<sub>1c</sub> goal may also be appropriate for the very young or old or in individuals with limited life span or comorbid conditions. For individuals using CGM, maximizing time-in-range 70–180 mg/dL, representing normoglycemia, while minimizing time-below-range <70 mg/dL, representing hypoglycemia, are shorter-term targets of therapy.

More stringent glycemic control (HbA<sub>1c</sub> <6%) is not beneficial, and may be detrimental, in patients with type 2 DM and a high risk of CVD. Large clinical trials (UKPDS, Action to Control Cardiovascular Risk in Diabetes [ACCORD], Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation [ADVANCE], Veterans Affairs Diabetes Trial [VADT]; **Chap. 405**) examined glycemic control in type 2 DM in individuals with low risk of CVD, with high risk of CVD, or with established CVD and have found that more intense glycemic control is not beneficial and, in some patient populations, may have a negative impact on some outcomes. These divergent outcomes stress the need for individualized glycemic goals based on the following general guidelines: (1) early in the course of type 2 diabetes when the CVD risk is lower, improved glycemic control likely leads to improved cardiovascular outcome, but this benefit may occur more than a decade after the period of improved glycemic control; (2) intense glycemic control in individuals with established CVD or at high risk for CVD is not advantageous, and may be deleterious, over a follow-up of 3–5 years; (3) hypoglycemia in such high-risk populations (elderly, CVD) should be avoided; and (4) improved glycemic control reduces microvascular complications of diabetes (**Chap. 405**) even if it does not improve macrovascular complications like CVD.

### TYPE 1 DIABETES MELLITUS

**General Aspects** The ADA recommendations for glycemic goals and HbA<sub>1c</sub> targets are summarized in Table 404-2. The goal is to design and implement insulin regimens that mimic physiologic insulin secretion. Because individuals with type 1 DM partially or completely lack endogenous insulin production, administration of basal insulin is essential for regulating glycogen breakdown, gluconeogenesis, lipolysis, and ketogenesis (i.e., largely fine-tuning hepatic and adipose metabolism). Likewise, insulin replacement for meals should be appropriate for the carbohydrate intake and insulin sensitivity, promoting normal glucose utilization and storage.

**Intensive Management** Intensive insulin therapy has the goal of achieving near-normal glycemia. This approach requires multiple resources, including thorough and continuing patient education, comprehensive recording of plasma glucose measurements and nutrition intake by the patient, and a variable insulin regimen that matches carbohydrate intake and insulin dose. Insulin regimens include multiple-component insulin regimens, multiple daily injections (MDIs), or continuous subcutaneous (SC) insulin infusion (CSII).

The benefits of intensive insulin therapy and improved glycemic control include a reduction in the acute metabolic and chronic microvascular complications of DM. From a psychological standpoint, the patient experiences greater control over his or her diabetes and often notes an improved sense of well-being, greater flexibility in the timing and content of meals, and the capability to alter insulin dosing with exercise. In addition, intensive insulin therapy prior to and during pregnancy reduces the risk of fetal malformations and morbidity. Intensive insulin therapy is encouraged in newly diagnosed patients with type 1 DM because it may prolong the period of C-peptide production, which may result in better glycemic control and a reduced risk of serious hypoglycemia. Although intensive management confers impressive benefits, it is also accompanied by significant personal and financial costs and therefore may not be appropriate at all times for all individuals.

TABLE 404-4 Properties of Insulin Preparations<sup>a</sup>

PREPARATION	TIME OF ACTION		
	ONSET, h	PEAK, h	EFFECTIVE DURATION, h
<b>Short-acting</b>			
Aspart <sup>b</sup>	<0.25	0.5–1.5	3–5
Glulisine	<0.25	0.5–1.5	3–5
Lispro <sup>c</sup>	<0.25	0.5–1.5	3–5
Regular <sup>d</sup>	0.5–1.0	2–3	4–8
Inhaled human insulin	<0.5	1–2	3
<b>Long-acting</b>			
Degludec	1–9	— <sup>e</sup>	42 <sup>f</sup>
Detemir	1–4	— <sup>e</sup>	12–24 <sup>f</sup>
Glargine <sup>g</sup>	2–4	— <sup>e</sup>	20–24
NPH	2–4	4–10	10–16
<b>Examples of insulin combinations<sup>h</sup></b>			
75/25–75% protamine lispro, 25% lispro	<0.25	Dual <sup>i</sup>	10–16
70/30–70% protamine aspart, 30% aspart	<0.25	Dual <sup>i</sup>	15–18
50/50–50% protamine lispro, 50% lispro	<0.25	Dual <sup>i</sup>	10–16
70/30–70% NPH, 30% regular	0.5–1	Dual <sup>i</sup>	10–16
Combination of long-acting insulin and GLP-1 receptor agonist	See text		

<sup>a</sup>Injectable insulin preparations (with exception of inhaled formulation) available in the United States; others are available in the United Kingdom and Europe. Standard formulations are U-100 (100 units of insulin per mL solution). <sup>b</sup>Formulation with niacinamide (vitamin B3) has a slightly more rapid onset and offset. <sup>c</sup>Lispro-aabc formulation has a slightly more rapid onset and offset; both formulations are also available in U-200 concentration. <sup>d</sup>Formulation also available in U-500 concentration with delayed onset and offset. <sup>e</sup>Degludec, detemir, and glargine have minimal peak activity. <sup>f</sup>Duration is dose-dependent. <sup>g</sup>Formulation also available in U-300 concentration with delayed onset and offset. <sup>h</sup>Other insulin combinations are available. <sup>i</sup>Dual: two peaks—one at 2–3 h and the second one several hours later.

**Insulin Preparations** Current insulin preparations are generated by recombinant DNA technology and consist of the amino acid sequence of human insulin or variations thereof. In the United States, most insulin is formulated as U-100 (100 units/mL); short-acting insulin formulated as U-200 (200 units/mL; lispro) and long-acting as U-300 (300 units/mL; glargine) are available in order to limit injection volumes for patients with high insulin requirements. Regular insulin formulated as U-500 (500 units/mL) is sometimes used in patients with severe insulin resistance. Human insulin has been formulated with distinctive pharmacokinetics (regular and neutral protamine Hagedorn [NPH] insulin have the native insulin amino acid sequence) or genetically modified to alter insulin absorption and hence insulin action. Insulins can be classified as short-acting or long-acting (Table 404-4; Figure 404-1D). For example, one short-acting insulin formulation, insulin lispro, is an insulin analogue in which the 28th and 29th amino acids (lysine and proline) on the insulin B chain have been reversed by recombinant DNA technology. Insulin aspart and insulin glulisine are genetically modified insulin analogues with properties similar to lispro. A biosimilar version of lispro has been approved. These insulin analogues have full biologic activity but less tendency for self-aggregation, resulting in more rapid absorption and onset of action and a shorter duration of action. These characteristics are particularly advantageous for allowing entrainment of insulin injection and action to rising plasma glucose levels following meals. The shorter duration of action also appears to be associated with a decreased number of hypoglycemic episodes, primarily because the decay of insulin action corresponds to the decline in plasma glucose after a meal. Thus, insulin aspart, lispro, or glulisine is preferred over regular insulin for prandial coverage in many patients. Insulin glargine is a long-acting biosynthetic human insulin that differs from normal insulin in that asparagine is replaced

by glycine at amino acid 21, and two arginine residues are added to the C terminus of the B chain, leading to the formation of microprecipitates at physiologic pH in subcutaneous tissue. Compared to NPH insulin, the onset of insulin glargine action is later, the duration of action is longer (~24 h), and there is a less pronounced peak. A lower incidence of hypoglycemia, especially at night, has been reported with insulin glargine when compared to NPH insulin. A biosimilar version is available. Insulin detemir has a fatty acid side chain that reversibly binds to albumin and prolongs its action by slowing absorption and catabolism, but its duration of action may only reach 12–20 h. Twice-daily injections of glargine, or especially detemir, are sometimes required to provide optimal 24-h basal insulin coverage. Because of modification and extension of the carboxy-terminus of the B chain, insulin degludec forms multihexamers in subcutaneous tissue and binds albumin, prolonging its duration of action (>42 h); it provides similar glycemic control as glargine but with less frequent nocturnal and severe hypoglycemia. Other modified insulins, such as one with a duration of action of several days, are in development.

Basal insulin requirements are provided by long-acting insulin formulations (NPH insulin, insulin glargine, insulin detemir, or insulin degludec) (Fig. 404-1D; Table 404-4). These are usually prescribed with short-acting insulin in an attempt to mimic physiologic insulin release with meals. Although mixing of NPH and short-acting insulin formulations is common practice, this mixing may alter the insulin absorption profile (especially the short-acting insulins). For example, lispro absorption is delayed by mixing with NPH. The alteration in insulin absorption when the patient mixes different insulin formulations should not prevent mixing insulins. However, the following guidelines should be followed: (1) mix the different insulin formulations in the syringe immediately before injection (inject within 2 min after mixing); (2) do not store insulin as a mixture; (3) follow the same routine in terms of insulin mixing and administration to standardize the physiologic response to injected insulin; and (4) do not mix insulin glargine, detemir, or degludec with other insulins. The miscibility of some insulins allows for the production of combination insulins that contain 70% NPH and 30% regular (70/30), or equal mixtures of NPH and regular (50/50). By including the insulin analogue mixed with protamine, several additional combinations have a short-acting and long-acting profile (Table 404-4; Fig. 404-1D). Although more convenient for the patient (only two injections/day), combination insulin formulations do not allow independent adjustment of short-acting and long-acting activity. Several insulin formulations are available as insulin “pens,” which are more convenient for some patients. Insulin delivery by inhalation to provide meal-time insulin is approved, but not widely used. Prior to its use, the forced expiratory volume in 1 second (FEV<sub>1</sub>) should be measured. Inhaled insulin can cause bronchospasm and cough and should not be used by individuals with lung disease or those who smoke. Long-acting insulin/glucagon-like peptide-1 (GLP-1) receptor agonist combinations in fixed doses (degludec + liraglutide or glargine + lixisenatide) are effective, and are associated with less weight gain.

**Insulin Regimens** There is considerable patient-to-patient variation in the peak and duration. In all regimens, long-acting insulins (NPH, glargine, detemir, or degludec) supply basal insulin, whereas regular, insulin aspart, glulisine, or lispro provide prandial insulin (Fig. 404-1D; Table 404-4). Short-acting insulin analogues should be injected just before (<10 min) and regular insulin 30–45 min prior to a meal. Sometimes short-acting insulin analogues are injected just after a meal (gastroparesis, unpredictable food intake).

A shortcoming of current insulin regimens is that injected insulin immediately enters the systemic circulation, whereas endogenous insulin is secreted into the portal venous system. Thus, exogenous insulin administration exposes the liver to subphysiologic insulin levels. No current insulin regimen reproduces the precise insulin secretory pattern of the pancreatic islet. However, the most physiologic regimens entail more frequent insulin injections, greater reliance on short-acting insulin, and more frequent SMBG and/or CGM. In general, individuals with type 1 DM require 0.3–0.7 units/kg per day of insulin divided into

multiple doses, with approximately 50% of daily insulin given as basal insulin and 50% as prandial insulin.

MDI regimens refer to the combination of basal insulin and bolus insulin (prandial short-acting insulin). The timing and dose of short-acting, preprandial insulin are altered to accommodate the SMBG or CGM results, anticipated food intake, and physical activity. Such regimens offer the patient with type 1 DM more flexibility in terms of lifestyle and the best chance for achieving near normoglycemia. Most often basal insulin with glargine, detemir, or degludec is used in conjunction with preprandial lispro, glulisine, or insulin aspart. The insulin aspart, glulisine, or lispro dose is based on individualized algorithms that integrate the preprandial glucose and the anticipated carbohydrate intake. To determine the meal component of the preprandial insulin dose, the patient uses an insulin-to-carbohydrate ratio (a common ratio for type 1 DM is 1 unit/10–15 g of carbohydrate, but this must be determined for each individual). To this insulin dose is added the supplemental or correcting insulin based on the preprandial blood glucose (one formula uses 1 unit of insulin for every 1.6–3.3 mmol/L [30–60 mg/dL] over the preprandial glucose target; this correction factor can be estimated from  $1500/[\text{total daily insulin dose}]$ ). Such calculations must be adjusted based on each individual's sensitivity to insulin. Other variations of this regimen use twice daily NPH as basal insulin but have the disadvantage that NPH has a significant peak, making hypoglycemia more common. Frequent SMBG (4 times per day) or CGM is essential for these types of insulin regimens.

CSII is a very effective insulin regimen for the patient with type 1 DM (Fig. 404-1). To the basal insulin infusion, a preprandial insulin ("bolus") is delivered by the insulin infusion device based on instructions from the patient, who uses an individualized algorithm incorporating the preprandial plasma glucose and anticipated carbohydrate intake. These sophisticated devices can accurately deliver small doses of insulin (microliters per hour) and have several advantages: (1) multiple basal infusion rates can be programmed to accommodate nocturnal versus daytime basal insulin requirement; (2) basal infusion rates can be altered during periods of exercise; (3) different waveforms of insulin infusion with meal-related bolus allow better matching of insulin depending on meal composition; and (4) programmed algorithms consider ongoing action of prior insulin administration and blood glucose values in calculating the insulin dose. These devices require instruction by a health professional with considerable experience with insulin infusion devices and frequent patient interactions with the diabetes management team. Insulin infusion devices may present unique challenges, such as infection at the infusion site, unexplained hypoglycemia because the infusion set becomes obstructed, or diabetic ketoacidosis (DKA) if the insulin infusion device becomes disconnected. Because most physicians use lispro, glulisine, or insulin aspart in CSII, the extremely short half-life of these insulins quickly leads to insulin deficiency if the delivery system is interrupted. Essential to the safe use of infusion devices is thorough patient education, frequent SMBG and/or CGM, and a backup plan for injecting long- and/or rapid-acting insulins in the event of insulin infusion device failure. CGM sensor-augmented insulin infusion devices integrate the information from the CGM to inform insulin delivery (Fig. 404-1). Currently, sensor communicating functions can interrupt basal insulin delivery during hypoglycemia (threshold suspension) or when hypoglycemia is anticipated (predictive suspension), which may be particularly useful for addressing nocturnal hypoglycemia. Hybrid closed-loop systems have recently become available that combine patient-directed preprandial boluses with automated adjustment of between-meal and basal insulin delivery based on CGM. Clinical experience with closed-loop systems is rapidly increasing and expanding. Bihormonal infusion devices that deliver both insulin and glucagon are under development.

**Other Agents That Improve Glucose Control** The role of amylin, a 37-amino-acid peptide co-secreted with insulin from pancreatic beta cells, in normal glucose homeostasis is uncertain. However, based on the rationale that patients who are insulin deficient are also amylin deficient, an analogue of amylin (pramlintide) was created and found to reduce postprandial glycemic excursions in individuals with

type 1 or type 2 DM taking insulin. Pramlintide injected just before a meal slows gastric emptying and suppresses glucagon but does not alter insulin levels. Pramlintide is approved for insulin-treated patients with type 1 or type 2 DM. Addition of pramlintide produces a modest reduction in the HbA<sub>1c</sub> and seems to dampen meal-related glucose excursions. In type 1 DM, pramlintide is started as a 15-µg SC injection before each meal and titrated up to a maximum of 30–60 µg as tolerated. In type 2 DM, pramlintide is started as a 60-µg SC injection before each meal and may be titrated up to a maximum of 120 µg. The major side effects are nausea and vomiting, and dose escalations should be slow to limit these side effects. Because pramlintide slows gastric emptying, it may influence absorption of other medications and should not be used in combination with other drugs that slow gastrointestinal (GI) motility. The short-acting insulin given before the meal should initially be reduced to avoid hypoglycemia and then titrated as the effects of the pramlintide become evident. Because pramlintide suppresses glucagon, it may worsen hypoglycemia recovery and should not be used in patients with hypoglycemia unawareness.

## TYPE 2 DIABETES MELLITUS

**General Aspects** The goals of glycemia-controlling therapy for type 2 DM are similar to those in type 1 DM. Whereas glycemic control tends to dominate the management of type 1 DM, the care of individuals with type 2 DM must also include attention to the treatment of conditions associated with type 2 DM (e.g., obesity, hypertension, dyslipidemia, CVD) and detection/management of DM-related complications (Fig. 404-2; Chap. 405). Reduction in cardiovascular risk is of paramount importance because this is the leading cause of mortality in these individuals.

Type 2 DM management should begin with MNT (discussed above). An exercise regimen to increase insulin sensitivity and promote weight loss should also be instituted. Pharmacologic approaches to the management of type 2 DM include oral glucose-lowering agents, insulin, and other agents that improve glucose control; most physicians and patients prefer oral glucose-lowering agents as the initial choice. Any therapy that improves glycemic control reduces "glucose toxicity" to beta cells and may improve endogenous insulin secretion. However, type 2 DM is a progressive disorder and ultimately requires multiple therapeutic agents and often insulin in most patients.

**Glucose-Lowering Agents** Advances in the therapy of type 2 DM have generated oral glucose-lowering agents that target different pathophysiologic processes in type 2 DM. Based on their mechanisms of action, glucose-lowering agents are subdivided into agents that increase insulin secretion, reduce glucose production, increase insulin sensitivity, enhance GLP-1 action, or promote urinary excretion of glucose (Table 404-5). Glucose-lowering agents other than insulin (with the exception of amylin analogue) are ineffective in type 1 DM and should not be used for glucose management of severely ill individuals with type 2 DM. Insulin is sometimes the initial glucose-lowering agent in type 2 DM.

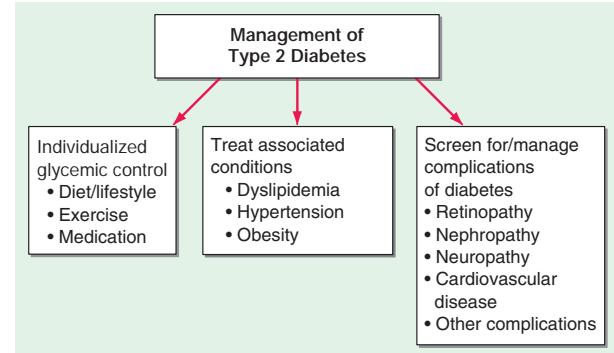


FIGURE 404-2 Essential elements in comprehensive care of type 2 diabetes.

	MECHANISM OF ACTION	EXAMPLES <sup>a</sup>	HbA <sub>1c</sub> REDUCTION (%) <sup>b</sup>	AGENT-SPECIFIC ADVANTAGES	AGENT-SPECIFIC DISADVANTAGES	CONTRAINdications
<b>Oral</b>						
Biguanides <sup>c*</sup>	↓ Hepatic glucose production, ↑ insulin sensitivity, influence gut function	Metformin	1–2	Weight neutral, do not cause hypoglycemia, inexpensive, extensive experience, ↓ CV events	Diarrhea, nausea, lactic acidosis, vitamin B12 deficiency	Renal insufficiency (see text for GFR <30 mL/min), CHF, radiographic contrast studies, hospitalized patients, acidosis
α-Glucosidase inhibitors <sup>c**</sup>	↓ GI glucose absorption	Acarbose, miglitol, voglibose	0.5–0.8	Reduce postprandial glycemia	GI flatulence, elevated liver function tests	Renal/liver insufficiency
Dipeptidyl peptidase IV inhibitors <sup>c***</sup>	Prolong endogenous GLP-1 action; ↑ Insulin, ↓ glucagon	Alogliptin, linagliptin, saxagliptin, sitagliptin, vildagliptin	0.5–0.8	Well tolerated, do not cause hypoglycemia	Angioedema/urticaria and immune-mediated dermatologic effects	Reduced dose with renal insufficiency
Insulin secretagogues: Sulfonylureas <sup>c*</sup>	↑ Insulin secretion	Glibenuride, gliclazide, glimepiride, glipizide, gliquidone, glyburide, glycyclopyramide	1–2	Short onset of action, lower postprandial glucose, inexpensive	Hypoglycemia, weight gain	Renal/liver insufficiency
Insulin secretagogues: Nonsulfonylureas <sup>c***</sup>	↑ Insulin secretion	Mitiglinide, nateglinide, repaglinide	0.5–1.0	Short onset of action, lower postprandial glucose	Hypoglycemia	Renal/liver insufficiency (except repaglinide)
Sodium-glucose cotransporter 2 inhibitors <sup>c***</sup>	↑ Renal glucose excretion	Canagliflozin, dapagliflozin, empagliflozin, ertugliflozin	0.5–1.0	Do not cause hypoglycemia, ↓ weight and BP, renal protective, ↓ CV events	Urinary and genital infections, polyuria, dehydration, exacerbate tendency to hyperkalemia and DKA; see text	Moderate renal insufficiency, insulin-deficient DM <sup>e</sup>
Thiazolidinediones <sup>c***</sup>	↓ Insulin resistance, ↑ glucose utilization	Pioglitazone, rosiglitazone	0.5–1.4	Lower insulin requirements	Peripheral edema, CHF, weight gain, fractures, macular edema	CHF, renal/liver insufficiency
<b>Parenteral/Oral</b>						
GLP-1 receptor agonists <sup>c***</sup>	↑ Insulin, ↓ glucagon, slow gastric emptying, satiety	Dulaglutide, exenatide, liraglutide, lixisenatide, semaglutide (oral formulation available)	0.5–1.0	Weight loss, do not cause hypoglycemia (unless combined with another insulin secretagogue or insulin); ↓ CV events	Injection, nausea, pancreatitis <sup>e</sup>	Renal disease, agents that also slow GI motility; medullary carcinoma of thyroid, pancreatic disease
<b>Parenteral</b>						
Amylin agonists <sup>c,d***</sup>	Slow gastric emptying, ↓ glucagon	Pramlintide	0.25–0.5	Reduce postprandial glycemia, weight loss	Injection, nausea, ↑ risk of hypoglycemia with insulin	Agents that also slow GI motility
Insulin <sup>c,d****</sup>	↑ Glucose utilization, ↓ hepatic glucose production, and other anabolic actions	See text and Table 404-4	Not limited	Known safety profile	Injection, weight gain, hypoglycemia	None
Medical nutrition therapy and physical activity <sup>c*</sup>	↓ Insulin resistance, ↑ insulin secretion	Low-calorie, carbohydrate-controlled diet, exercise	1–3	Other health benefits	Compliance difficult, long-term success low	None

<sup>a</sup>Examples are approved for use in the United States; others are available in other countries. Examples may not include all agents in the class. <sup>b</sup>HbA<sub>1c</sub> reduction (absolute) depends partly on starting HbA<sub>1c</sub>. <sup>c</sup>Used for treatment of type 2 diabetes. <sup>d</sup>Used in conjunction with insulin for treatment of type 1 diabetes. Cost of agent in the United States: "low," "moderate," "high," "variable." <sup>e</sup>Degree of risk uncertain, avoid in individuals with risk factors for pancreatitis. <sup>f</sup>Risk of euglycemic DKA in patients with insulin deficiency (e.g., type 1 diabetes).

Note: Some agents used to treat type 2 diabetes are not included in table (see text).

Abbreviations: CHF, congestive heart failure; CV, cardiovascular; GI, gastrointestinal; HbA<sub>1c</sub>, glycosylated hemoglobin A<sub>1c</sub>.

**BIGUANIDES** Metformin, representative of this class of agents, reduces hepatic glucose production and improves peripheral glucose utilization slightly (Table 404-5). Metformin activates AMP-dependent protein kinase and enters cells through organic cation transporters (polymorphisms of these may influence the response to metformin). Metformin acts in multiple tissues, but its mechanism of action remains undefined. There is evidence for reducing hepatic glucose production by antagonizing cAMP generation in hepatocytes as well as for actions in the gut. Metformin reduces fasting plasma glucose (FPG) and insulin levels, improves the lipid profile, and promotes modest weight loss. An extended-release form is available and may have fewer GI side effects

(diarrhea, anorexia, nausea, metallic taste). Because of its metformin's relatively slow onset of action and GI symptoms with higher doses, the initial dose should be low and then escalated every 1–2 weeks to a maximally tolerated dose of 2000 mg daily. Metformin is effective as monotherapy and can be used in combination with other oral agents or with insulin. Long-term use is associated with reduced micro- and macrovascular complications. The major toxicity of metformin, lactic acidosis, is very rare and can be prevented by careful patient selection. Vitamin B<sub>12</sub> levels are lower during metformin treatment and should be monitored. Metformin should not be used in patients with moderate renal insufficiency (glomerular filtration rate [GFR] <30 mL/min), any

form of acidosis, unstable congestive heart failure (CHF), liver disease, or severe hypoxemia. Metformin should be discontinued in hospitalized patients, in patients who can take nothing orally, and in those receiving radiographic contrast material. Insulin should be used until metformin can be restarted.

#### **INSULIN SECRETAGOGUES—AGENTS THAT AFFECT THE ATP-SENSITIVE K<sup>+</sup> CHANNEL**

Insulin secretagogues stimulate insulin secretion by interacting with the ATP-sensitive potassium channel on the beta cell (**Chap. 403**). These drugs are most effective in individuals with type 2 DM of relatively recent onset (<5 years) who have residual endogenous insulin production. First-generation sulfonylureas (chlorpropamide, tolazamide, tolbutamide) have a longer half-life, a greater incidence of hypoglycemia, and more frequent drug interactions, and are no longer used. Second-generation sulfonylureas have a more rapid onset of action and better coverage of the postprandial glucose rise, but the shorter half-life of some agents may require more than once-a-day dosing. Sulfonylureas reduce both fasting and postprandial glucose and should be initiated at low doses and increased at 1- to 2-week intervals based on SMBG. In general, sulfonylureas increase insulin acutely and thus should be taken shortly before a meal; with chronic therapy, though, the insulin release is more sustained. Long-term use is associated with reduced micro- and macrovascular complications. Glimepiride and glipizide can be given in a single daily dose and are preferred over glyburide, especially in the elderly. Repaglinide, nateglinide, and mitiglinide are not sulfonylureas but also interact with the ATP-sensitive potassium channel. Because of their short half-life, these glinide agents are given immediately before each meal to reduce meal-related glucose excursions.

Insulin secretagogues, especially the longer acting ones, have the potential to cause hypoglycemia, especially in elderly individuals. Hypoglycemia is usually related to delayed meals, increased physical activity, alcohol intake, or renal insufficiency. Individuals who ingest an overdose of some agents develop prolonged and serious hypoglycemia and should be monitored closely in the hospital (**Chap. 406**). Most sulfonylureas are metabolized in the liver to compounds (some of which are active, such as those of glyburide and the glinide nateglinide) that are cleared by the kidney. Thus, their use in individuals with significant hepatic or renal dysfunction is not advisable. For patients with chronic kidney disease requiring an insulin secretagogue, the shorter-acting sulfonylureas glimepiride or glipizide or the glinide repaglinide may be used with caution. Weight gain, a common side effect of sulfonylurea therapy, results from the increased insulin levels and improvement in glycemic control. Some sulfonylureas have significant drug interactions with alcohol and some medications including warfarin, aspirin, ketoconazole, -glucosidase inhibitors, and fluconazole. A related isoform of ATP-sensitive potassium channels is present in the myocardium and the brain. All of these agents except glyburide have a low affinity for this isoform. Despite concerns that this agent might affect the myocardial response to ischemia and observational studies suggesting that sulfonylureas increase cardiovascular risk, studies have not shown an increased cardiac mortality with glyburide or other agents in this class.

#### **INSULIN SECRETAGOGUES—AGENTS THAT ENHANCE GLP-1 RECEPTOR SIGNALING**

“Incretins” amplify glucose-stimulated insulin secretion (**Chap. 403**). Agents that either act as a GLP-1 receptor agonist or enhance endogenous GLP-1 activity are approved for the treatment of type 2 DM (Table 404-5). Agents in this class do not cause hypoglycemia because of the glucose-dependent nature of incretin-stimulated insulin secretion (unless there is concomitant use of an agent that can lead to hypoglycemia—sulfonylureas, etc.). GLP-1 receptor agonists increase glucose-stimulated insulin secretion, suppress glucagon, and slow gastric emptying. These agents do not promote weight gain; in fact, most patients experience modest weight loss and appetite suppression. Short-acting GLP-1 receptor agonists are exenatide twice daily, liraglutide daily, and lixisenatide daily. Long-acting GLP-1 receptor agonists include sustained-release exenatide, dulaglutide, lixisenatide, and semaglutide, all administered weekly. Short-acting GLP-1 receptor

agonists provide mostly postprandial coverage whereas the long-acting GLP-1 receptor agonists reduce both the postprandial and fasting glucose. Daily oral semaglutide is now available that depends on gastric absorption to avoid proteolytic degradation in the small intestine. All are modified to avoid enzymatic inactivation by dipeptidyl peptidase IV (DPP-IV) in the circulation.

For example, exenatide, a synthetic version of a peptide initially identified in the saliva of the Gila monster (exendin-4), is an analogue of GLP-1. Unlike native GLP-1, which has a half-life of ~2 min, differences in the exenatide amino acid sequence render it resistant to DPP-IV. Thus, exenatide has prolonged GLP-1-like action. Liraglutide, another GLP-1 receptor agonist, is almost identical to native GLP-1 except for an amino acid substitution and addition of a fatty acyl group (coupled with a -glutamic acid spacer) that promote binding to albumin and plasma proteins and prolong its half-life. Higher doses of liraglutide and semaglutide than used for glucose-lowering effects are effective for weight-loss therapy for obesity. Liraglutide treatment has also been associated with a decrease in CVD events in patients with type 2 DM and established CVD and with lower rates of diabetic kidney disease. In similar patient populations, semaglutide treatment has been associated with fewer CVD events and reduced diabetic kidney disease, but with an increased rate of retinopathy-related complications, while dulaglutide treatment has been associated with both a reduction in CVD events and in composite microvascular retinopathy and nephropathy-related complications primarily driven by prevention of renal events. Similar reductions in CVD events have not been observed with exenatide once weekly or lixisenatide. Treatment with GLP-1 receptor agonists should start at a low dose to minimize initial side effects (nausea being the limiting one). GLP-1 receptor agonists can be used as combination therapy with metformin, sulfonylureas, and thiazolidinediones. Some patients taking insulin secretagogues may require a reduction in those agents to prevent hypoglycemia. The major side effects are nausea, vomiting, and diarrhea. Some formulations carry a black box warning from the FDA because of an increased risk of thyroid C-cell tumors in rodents and are contraindicated in individuals with medullary carcinoma of the thyroid or multiple endocrine neoplasia. Because GLP-1 receptor agonists slow gastric emptying, they may influence the absorption of other drugs. Whether GLP-1 receptor agonists enhance beta cell survival or promote beta cell proliferation in humans as in rodents is not known, but these agents do not appear to alter the natural history of type 2 DM.

DPP-IV inhibitors inhibit degradation of native GLP-1 and thus enhance the incretin effect. DPP-IV, which is widely expressed on the cell surface of endothelial cells and some lymphocytes, degrades a wide range of peptides (not GLP-1 specific). DPP-IV inhibitors promote insulin secretion in the absence of hypoglycemia or weight gain and appear to have a preferential effect on postprandial blood glucose. The levels of GLP-1 action in the patient are greater with the GLP-1 receptor agonists than with DPP-IV inhibitors. DPP-IV inhibitors are used either alone or in combination with other oral agents in type 2 DM. Reduced doses should be given to patients with renal insufficiency. Allergy, including rash, hypersensitivity reactions (including anaphylaxis, angioedema, and Stevens-Johnson syndrome), and severe joint pain have been reported in association with DPP-IV inhibitors. There is evidence concerning a potentially increased risk for acute pancreatitis with GLP-1 receptor agonists and less so with DPP-IV inhibitors. For now, it is reasonable to avoid these agents in patients with pancreatic disease or with other significant risk factors for acute pancreatitis (e.g., heavy alcohol use, severely elevated serum triglycerides, hypercalcemia).

**-GLUCOSIDASE INHIBITORS** -Glucosidase inhibitors reduce postprandial hyperglycemia by delaying glucose absorption; they do not affect glucose utilization or insulin secretion (Table 404-5). Postprandial hyperglycemia, secondary to impaired hepatic and peripheral glucose disposal, contributes significantly to the hyperglycemic state in type 2 DM. These drugs, taken just before each meal, reduce glucose absorption by inhibiting the enzyme that cleaves oligosaccharides into simple sugars in the intestinal lumen. Therapy should be initiated at a

**3112** low dose with the evening meal and increased to a maximal dose over weeks to months. The major side effects (diarrhea, flatulence, abdominal distention) are related to increased delivery of oligosaccharides to the large bowel and can be reduced somewhat by gradual upward dose titration. -Glucosidase inhibitors may increase levels of sulfonylureas and increase the incidence of hypoglycemia. Simultaneous treatment with bile acid resins and antacids should be avoided. These agents should not be used in individuals with inflammatory bowel disease, gastroparesis, or a serum creatinine >177  $\mu\text{mol/L}$  (2 mg/dL). This class of agents is not as potent as other oral agents in lowering the  $\text{HbA}_{1c}$ , but is unique because it reduces the postprandial glucose rise. If hypoglycemia from other diabetes treatments occurs while taking these agents, the patient should consume glucose because the degradation and absorption of complex carbohydrates will be retarded.

**THIAZOLIDINEDIONES** Thiazolidinediones (Table 404-5) reduce insulin resistance by binding to the peroxisome proliferator-activated receptor (PPAR- $\gamma$ ) nuclear receptor (which forms a heterodimer with the retinoid X receptor). The PPAR- $\gamma$  receptor is found at highest levels in adipocytes but is expressed at lower levels in many other tissues. Agonists of this receptor regulate a large number of genes, promote adipocyte differentiation, reduce hepatic fat accumulation, and promote fatty acid storage. Thiazolidinediones promote a redistribution of fat from central to peripheral locations. Circulating insulin levels decrease with use of the thiazolidinediones, indicating a reduction in insulin resistance. Although direct comparisons are not available, the two currently available thiazolidinediones appear to have similar efficacy. The prototype of this class of drugs, troglitazone, was withdrawn from the U.S. market after reports of hepatotoxicity and an association with an idiosyncratic liver reaction that sometimes led to hepatic failure. Although rosiglitazone and pioglitazone do not appear to induce the liver abnormalities seen with troglitazone, the FDA recommends measurement of liver function tests prior to initiating therapy. Modestly increased transaminase levels related to underlying fatty liver disease should not preclude treatment as these levels may improve with thiazolidinediones due to a reduction in hepatic fat content.

Rosiglitazone raises low-density lipoprotein (LDL), high-density lipoprotein (HDL), and triglycerides slightly. Pioglitazone raises HDL to a greater degree and LDL a lesser degree but lowers triglycerides. The clinical significance of the lipid changes with these agents is not known and may be difficult to ascertain because most patients with type 2 DM are also treated with a statin.

Thiazolidinediones are associated with weight gain (2–3 kg), a small reduction in the hematocrit, and a mild increase in plasma volume. Peripheral edema and CHF are more common in individuals treated with these agents. These agents are contraindicated in patients with hepatic insufficiency or CHF (class III or IV). The FDA has issued an alert that rare patients taking these agents may experience a worsening of diabetic macular edema. An increased risk of fractures has been noted in postmenopausal women taking these agents. Thiazolidinediones have been shown to induce ovulation in premenopausal women with polycystic ovary syndrome. Women should be warned about the risk of pregnancy because the safety of thiazolidinediones in pregnancy is not established.

Concerns about increased cardiovascular risk associated with rosiglitazone led to considerable restrictions on its use and to the FDA issuing a black box warning in 2007. However, based on new information, the FDA has revised its guidelines and categorizes rosiglitazone similar to other drugs for type 2 DM. According to an FDA review, pioglitazone may be associated with an increased risk of bladder cancer. In one study, pioglitazone lowered the risk for recurrent stroke or myocardial infarction in insulin-resistant individuals without diabetes who had a prior stroke or transient ischemic attack.

**Sodium-Glucose Co-Transporter 2 (SGLT2) Inhibitors** These agents (Table 404-5) lower the blood glucose by selectively inhibiting this co-transporter, which is expressed almost exclusively in the proximal, convoluted tubule in the kidney. This inhibits glucose reabsorption, lowers the renal threshold for glucose, and leads to

increased urinary glucose excretion. Thus, the glucose-lowering effect is insulin independent and not related to changes in insulin sensitivity or secretion. The loss of urinary glucose may promote modest weight reduction. Since these agents also impair proximal reabsorption of sodium, their use is associated with a diuretic effect and 3–6 mmHg reduction in systolic blood pressure. Due to the increased urinary glucose, urinary and genital mycotic infections are more common in both men and women, and the diuretic effect can lead to reduced intravascular volume and acutely impaired kidney function. Inhibition of SGLT2 may lead to increased glucagon and consequently liver production of glucose and ketones. Euglycemic DKA may occur during illness or when ongoing glucosuria masks stress-induced requirements for insulin. These agents should not be prescribed for patients with type 1 DM or pancreatogenic forms of DM associated with insulin deficiency. Empagliflozin or canagliflozin reduces CVD events and all cause cardiovascular mortality in patients with type 2 DM and established CVD. All SGLT2 inhibitors may reduce hospitalization for CHF. Empagliflozin, canagliflozin, and dapagliflozin have all been shown to reduce progression of diabetic kidney disease but should not be initiated in patients with stage 3b CKD (eGFR <45 mL/min per 1.73  $\text{m}^2$ ) and should not be used with stage 4 CKD (eGFR <30 mL/min per 1.73  $\text{m}^2$ ). A possible increased risk of bladder cancer has been seen with dapagliflozin.

#### OTHER THERAPIES FOR TYPE 2 DM • Bile Acid-Binding Resins

Evidence indicates that bile acids, by signaling through nuclear receptors, may have a role in metabolism. Bile acid metabolism is abnormal in type 2 DM. The bile acid-binding resin colestipol has been approved for the treatment of type 2 DM (already approved for treatment of hypercholesterolemia). Because bile acid-binding resins are minimally absorbed into the systemic circulation, how bile acid-binding resins lower blood glucose is not known. The most common side effects are GI (constipation, abdominal pain, and nausea). Bile acid-binding resins can increase plasma triglycerides and should be used cautiously in patients with a tendency for hypertriglyceridemia. The role of this class of drugs in the treatment of type 2 DM is not yet defined.

**Bromocriptine** A formulation of the dopamine receptor agonist bromocriptine (Cycloset) has been approved by the FDA for the treatment of type 2 DM. However, its role in the treatment of type 2 DM is uncertain.

**INSULIN THERAPY IN TYPE 2 DM** Insulin should be considered as part of the initial therapy in type 2 DM, particularly in lean individuals or those with severe weight loss, in individuals with underlying renal or hepatic disease that precludes oral glucose-lowering agents, or in individuals who are hospitalized or acutely ill. Insulin therapy is ultimately required by a substantial number of individuals with type 2 DM because of the progressive nature of the disorder and the relative insulin deficiency that develops in patients with long-standing diabetes. Both physician and patient reluctance often delay the initiation of insulin therapy, but glucose control and patient well-being are improved by insulin therapy in patients who have not reached glycemic targets.

Because endogenous insulin secretion continues and is capable of providing some coverage of mealtime caloric intake, insulin is usually initiated in a single dose of long-acting insulin (0.1–0.4 U/kg per day), given in the evening or just before bedtime (NPH, glargin, detemir, or degludec). Because fasting hyperglycemia and increased hepatic glucose production are prominent features of type 2 DM, bedtime insulin is more effective in clinical trials than a single dose of morning insulin. Glargin given at bedtime has less nocturnal hypoglycemia than NPH insulin. Some physicians prefer a relatively low, fixed starting dose of long-acting insulin (5–15 units) or a weight-based dose (0.1 units/kg). The insulin dose may then be adjusted in 10–20% increments as dictated by SMBG results. Both morning and bedtime long-acting insulin may be used in combination with oral glucose-lowering agents. Initially, basal insulin may be sufficient, but often prandial insulin coverage with multiple insulin injections is needed as diabetes progresses (see insulin regimens used for type 1 DM). Other insulin formulations that have a combination of short-acting and long-acting

insulin (Table 404-4) are sometimes used in patients with type 2 DM because of convenience but do not allow independent adjustment of short-acting and long-acting insulin dose and often do not achieve the same degree of glycemic control as basal/bolus regimens. In selected patients with insulin-deficient type 2 DM, insulin infusion devices may be considered.

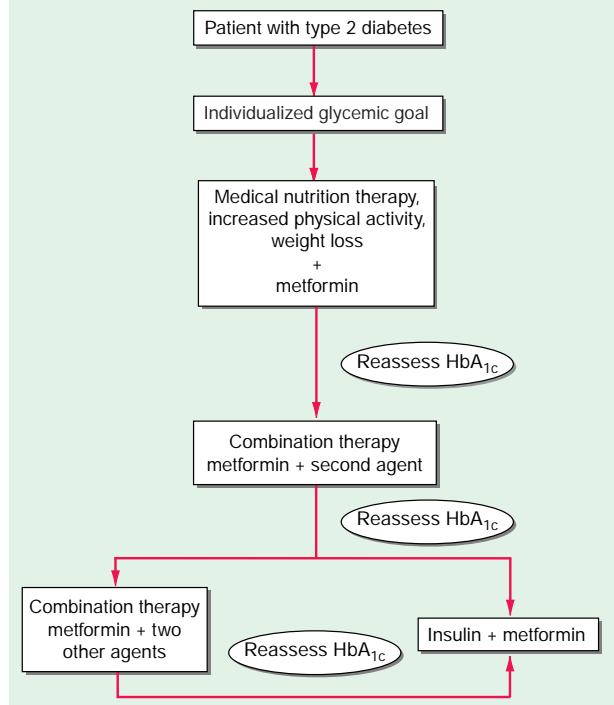
**CHOICE OF INITIAL GLUCOSE-LOWERING AGENT** The level of hyperglycemia and the patient's individualized goal (see "Establishment of Target Level of Glycemic Control") should influence the initial choice of therapy. Patients with mild hyperglycemia ( $FPG < 7.0\text{--}11.0 \text{ mmol/L}$  [ $126\text{--}199 \text{ mg/dL}$ ]) often respond well to a single, oral glucose-lowering agent, while those with moderate hyperglycemia ( $FPG 11.1\text{--}13.9 \text{ mmol/L}$  [ $200\text{--}250 \text{ mg/dL}$ ]) will usually require more than one oral agent or insulin. Patients with more severe hyperglycemia ( $FPG > 13.9 \text{ mmol/L}$  [ $250 \text{ mg/dL}$ ]) may respond partially but are unlikely to achieve normoglycemia with oral therapy. Insulin can be used as initial therapy in individuals with severe hyperglycemia ( $FPG < 13.9\text{--}16.7 \text{ mmol/L}$  [ $250\text{--}300 \text{ mg/dL}$ ]) or in those who are symptomatic from the hyperglycemia. This approach is based on the rationale that more rapid glycemic control will reduce "glucose toxicity" to the islet cells, improve endogenous insulin secretion, and possibly allow oral glucose-lowering agents to be more effective. If this occurs, the insulin may be discontinued.

Insulin secretagogues, biguanides, -glucosidase inhibitors, thiazolidinediones, GLP-1 receptor agonists, DPP-IV inhibitors, SLGT2 inhibitors, and insulin are approved for monotherapy of type 2 DM. Although each class of oral glucose-lowering agents has advantages and disadvantages (Table 404-5), certain generalizations apply: (1) insulin secretagogues, biguanides, GLP-1 receptor agonists, and thiazolidinediones improve glycemic control to a similar degree (1–2% reduction in  $\text{HbA}_{1c}$ ) and are more effective than -glucosidase inhibitors, DPP-IV inhibitors, and SLGT2 inhibitors; (2) insulin secretagogues, GLP-1 receptor agonists, DPP-IV inhibitors, -glucosidase inhibitors, and SLGT2 inhibitors begin to lower the plasma glucose immediately, whereas the glucose-lowering effects of the biguanides and thiazolidinediones are delayed by weeks; (3) not all agents are effective in all individuals with type 2 DM; (4) biguanides, -glucosidase inhibitors, GLP-1 receptor agonists, DPP-IV inhibitors, thiazolidinediones, and SLGT2 inhibitors do not directly cause hypoglycemia; (5) most individuals will eventually require treatment with more than one class of oral glucose-lowering agents or insulin, reflecting the progressive nature of type 2 DM; and (6) durability of glycemic control is slightly less for sulfonylureas compared to metformin or thiazolidinediones.

Considerable clinical experience exists with metformin and sulfonylureas because they have been available for several decades. It is assumed that the -glucosidase inhibitors, GLP-1 receptor agonists, DPP-IV inhibitors, thiazolidinediones, and SLGT2 inhibitors will reduce DM-related complications by improving glycemic control. Pioglitazone may reduce CVD events through targeting a fundamental abnormality in type 2 DM, namely insulin resistance. A reduction in CVD events and in progression of diabetic kidney disease seen with some GLP-1 agonists and SGLT2 inhibitors may also operate through glucose-independent mechanisms (Chap. 405).

Treatment algorithms by several professional societies (ADA/European Association for the Study of Diabetes [EASD], IDF, AACE) suggest metformin as initial therapy because of its efficacy, known side-effect profile, and low cost (Fig. 404-3). Initiation of pharmacologic therapy should be accompanied by an emphasis on lifestyle modification (e.g., MNT, increased physical activity, and weight loss). Metformin's advantages are that it promotes mild weight loss, lowers insulin levels, and improves the lipid profile slightly. Based on SMBG results and the  $\text{HbA}_{1c}$ , the dose of metformin should be increased until the glycemic target is achieved or maximum dose is reached.

**COMBINATION THERAPY WITH GLUCOSE-LOWERING AGENTS** A number of combinations of therapeutic agents are successful in type 2 DM: metformin + second oral agent, metformin + GLP-1 receptor agonist, metformin + insulin, or combinations of a long-acting insulin and a GLP-1 receptor agonist. Because mechanisms of action of the



**FIGURE 404-3** Glycemic management of type 2 diabetes. See text for discussion of treatment of severe hyperglycemia or symptomatic hyperglycemia. Agents that can be combined with metformin include insulin secretagogues, thiazolidinediones,  $\alpha$ -glucosidase inhibitors, DPP-IV inhibitors, GLP-1 receptor agonists, SLGT2 inhibitors, and insulin.  $\text{HbA}_{1c}$ , hemoglobin  $\text{HbA}_{1c}$ .

first and second agents should be different, the effect on glycemic control is usually additive. There are little data to support the choice of one combination over another combination. Recent results from the NIH-funded Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study (GRADE) indicated that addition of liraglutide or basal insulin to metformin leads to better glycemic control than glimepiride or sitagliptin (SLGT2 inhibitors were not studied). Based on recent demonstrations of a beneficial cardiovascular effect in certain individuals with type 2 DM and CVD, or at high risk of CVD, a GLP-1 receptor agonist or a SGLT2 inhibitor should now be considered in these populations. Medication costs vary considerably (Table 404-5), and this often factors into medication choice. Several fixed-dose combinations of oral agents are available, but evidence that they are superior to titration of a single agent to a maximum dose and then addition of a second agent is lacking. If adequate control is not achieved with the combination of two agents (based on reassessment of the  $\text{HbA}_{1c}$  every 3 months), a third oral agent, GLP-1 receptor agonist, or basal insulin should be added (Fig. 404-3). Treatment approaches vary considerably from country to country. For example, -glucosidase inhibitors are used commonly in South Asian patients (Indian), but infrequently in the United States or Europe. Whether this reflects an underlying difference in the disease or physician preference is not clear.

Treatment with insulin often becomes necessary as type 2 DM enters the phase of relative insulin deficiency and is signaled by inadequate glycemic control with one or two oral glucose-lowering agents. Insulin alone or in combination should be used in patients who fail to reach glycemic targets. For example, a single dose of long-acting insulin at bedtime is often effective in combination with metformin. As endogenous insulin production falls further, multiple injections of long-acting together with short-acting insulin are necessary to control postprandial glucose excursions. These insulin regimens are identical to the long-acting and short-acting combination regimens discussed above for type 1 DM, although usually at higher doses given insulin resistance. Weight gain and hypoglycemia are the major adverse effects

**3114** of insulin therapy. The daily insulin dose required can become quite large (1–2 units/kg per day) as endogenous insulin production falls and insulin resistance persists. Individuals who require >1 unit/kg per day of long-acting insulin should be considered for combination therapy with metformin a GLP-1 receptor agonist, or a thiazolidinedione as these can reduce insulin requirements in some individuals with type 2 DM. Insulin plus a thiazolidinedione promotes weight gain and may be associated with peripheral edema. Addition of a thiazolidinedione to a patient's insulin regimen may necessitate a reduction in the insulin dose to avoid hypoglycemia. Patients requiring large doses of insulin (>200 units/day) can be treated with a more concentrated form of insulin.

## OTHER THERAPIES FOR DIABETES

Metabolic (also referred to as bariatric) surgery for obese individuals with type 2 DM has shown considerable promise, sometimes with dramatic resolution of the diabetes or major reductions in the needed dose of glucose-lowering therapies (Chap. 402). Several large, nonrandomized clinical trials have demonstrated a much greater efficacy of metabolic surgery compared to medical management in the treatment of type 2 DM and may be considered in individuals with type 2 DM and a BMI >35 kg/m<sup>2</sup>. The ADA clinical guidelines state that metabolic surgery should be considered in individuals with type 2 DM and a body mass index >30 kg/m<sup>2</sup> if hyperglycemia is inadequately controlled despite optimal medical therapy.

Short-term intense caloric restriction (very-low-calorie diet, typically 800–1000 calories/day) can dramatically improve type 2 DM, sometimes leading to resolution of the diabetes. Such an approach is more effective in recent-onset type 2 DM and should be supervised by a provider with expertise and should be followed by a long-term, weight-maintenance program.

Whole-pancreas transplantation can normalize glucose control in type 1 DM and when performed simultaneously with or after kidney transplantation can prolong the life of the kidney transplant by offering protection against recurrent diabetic nephropathy. Pancreatic islet transplantation is available as a less invasive form of beta-cell replacement therapy for type 1 DM, but it remains investigational in the United States. Due to the risks associated with chronic immunosuppression, whole-pancreas and pancreatic islet transplantation may be considered for patients with severe metabolic instability or already requiring immunosuppression in support of a kidney or other organ transplant. Patients with chronic pancreatitis and preserved islet function who require pancreatectomy for pain relief may benefit from autologous islet transplantation as this may prevent or ameliorate postsurgical DM.

## EMERGING THERAPIES

Many individuals with long-standing type 1 DM still produce very small amounts of insulin or have insulin-positive cells within the pancreas. This suggests that beta cells may slowly regenerate but are quickly destroyed by the autoimmune process. Particularly early in the disease course, efforts to suppress the autoimmune process, for example with anti-CD3 monoclonal antibodies that target T lymphocytes, are being tested at the time of diagnosis of type 1 DM, and for prevention in autoantibody-positive individuals at stages 1 and 2 of type 1 DM (Chap. 403). Agents that target thioredoxin-interacting protein (TXNIP), especially Ca++ channel blockers, have some promise in recent-onset T1D and in rodent models of diabetes. Closed-loop insulin infusion devices that automate insulin delivery in response to changing glucose levels are progressing rapidly. New therapies under evaluation or development for type 2 DM include activators of glucokinase, inhibitors of 11-hydroxysteroid dehydrogenase-1, GPR40 agonists, dual agonists targeting the glucose-dependent insulinotropic polypeptide receptor and the GLP1-receptor, combined SLGT1 and SLGT2 inhibitors, and agents that may reduce inflammation, for example by inhibiting IL-1.

Because whole-pancreas and pancreatic islet transplantation are both limited by organ availability from deceased donors, stem cell-derived islet cells and xenogeneic sources of islets may eventually allow for a limitless supply of insulin-producing cells for transplantation.

## ADVERSE EFFECTS OF THERAPY FOR DM

As with any therapy, the benefits of efforts directed toward glycemic control must be balanced against the risks of treatment (Table 404-5). Side effects of intensive treatment include an increased frequency of serious hypoglycemia, weight gain, increased economic costs, and greater demands on the patient. In the DCCT, quality of life was very similar in the intensive and standard therapy groups. The most serious complication of therapy for DM is hypoglycemia, and its treatment with oral glucose or glucagon injection is discussed in Chap. 406. Severe, recurrent, or unexplained hypoglycemia warrants examination of treatment regimen and glycemic goal for the individual patient. Weight gain occurs with most (insulin, -glucosidase inhibitors, GLP-1 receptor agonists, DPP-IV inhibitors) therapies. The weight gain is partially due to the anabolic effects of insulin and the reduction in glucosuria.

## ACUTE DISORDERS RELATED TO SEVERE HYPERGLYCEMIA

Individuals with type 1 or type 2 DM and severe hyperglycemia (>13.9 mmol/L [250 mg/dL]) should be assessed for clinical stability, including mentation and hydration. Depending on the patient and the rapidity and duration of the severe hyperglycemia, an individual may require more intense and rapid therapy to lower the blood glucose. However, many patients with poorly controlled diabetes and hyperglycemia have few symptoms. The physician should assess if the patient is stable or if DKA or a hyperglycemic hyperosmolar state (HHS) should be considered. Ketones, an indicator of DKA, should be measured in individuals with type 1 DM when the plasma glucose is persistently >13.9 mmol/L (250 mg/dL), during a concurrent illness, or with symptoms such as nausea, vomiting, or abdominal pain. Blood measurement of -hydroxybutyrate is preferred over urine testing with nitroprusside-based assays that measure only acetoacetate and acetone.

DKA and HHS are acute, severe disorders directly related to diabetes. DKA was formerly considered a hallmark of type 1 DM, but it also occurs in individuals with type 2 DM who can sometimes subsequently be treated with oral glucose-lowering agents (frequently in individuals of Hispanic or African-American descent). HHS is primarily seen in individuals with type 2 DM. Both disorders are associated with absolute or relative insulin deficiency, volume depletion, and acid-base abnormalities. DKA and HHS exist along a continuum of hyperglycemia, with or without ketosis. The metabolic similarities and differences in DKA and HHS are highlighted in Table 404-6. Both disorders are associated with potentially serious complications if not promptly diagnosed and carefully treated.

## DIABETIC KETOACIDOSIS

**Clinical Features** The symptoms and physical signs of DKA are listed in Table 404-7 and usually develop over 24 h. DKA may be the initial symptom complex that leads to a diagnosis of type 1 DM, but more frequently, it occurs in individuals with established diabetes. Nausea and vomiting are often prominent, and their presence in an individual with diabetes warrants laboratory evaluation for DKA. Abdominal pain may be severe and can resemble acute pancreatitis or ruptured viscus. Hyperglycemia leads to glucosuria, volume depletion, and tachycardia. Hypotension can occur because of volume depletion in combination with peripheral vasodilatation. Kussmaul respirations and a fruity odor on the patient's breath (secondary to metabolic acidosis and increased acetone) are classic signs of the disorder. Lethargy and central nervous system depression may evolve into coma with severe DKA but should also prompt evaluation for other reasons for altered mental status (e.g., infection, hypoxemia). Cerebral edema, an extremely serious complication of DKA, is seen most frequently in children. Signs of infection, which may precipitate DKA, should be sought on physical examination, even in the absence of fever. Failure to augment insulin therapy during physiologic stress often compounds the problem. Tissue ischemia (heart, brain) can also be a precipitating factor. Omission of insulin because of an infusion pump delivery site

**TABLE 404-6 Laboratory Values in Diabetic Ketoacidosis (DKA), Hyperglycemic Hyperosmolar State (HHS), and Euglycemic DKA (Representative Ranges at Presentation)**

	DKA	HHS	EUGLYCEMIC DKA <sup>c</sup>
Glucose, <sup>a</sup> mmol/L (mg/dL)	13.9–33.3 (250–600)	33.3–66.6 (600–1200)	<11.1–13.9 (<200–250) <sup>c</sup>
Sodium, meq/L	125–135	135–145	~135
Potassium <sup>a,b</sup>	Normal to ↑	Normal	Normal to ↑
Magnesium <sup>a</sup>	Normal	Normal	Normal
Chloride <sup>a</sup>	Normal	Normal	Normal
Phosphate <sup>a,b</sup>	Normal	Normal	Normal
Creatinine	Slightly to moderately ↑	Moderately ↑	Slightly ↑
Osmolality (mOsm/mL)	300–320	330–380	~300
Serum/urine ketones <sup>a</sup>	++++	+/-	++++
Serum β-hydroxybutyrate, mmol/L	>2.5	<1.0	>2.5
Serum bicarbonate, <sup>a</sup> meq/L	<18	>18	<18
Arterial pH	6.8–7.3	>7.3	6.8–7.3
Arterial PCO <sub>2</sub> , <sup>a</sup> mmHg	20–30	Normal	20–30
Anion gap <sup>a</sup> (Na – [Cl + HCO <sub>3</sub> ])	↑	Normal to slightly ↑	↑

<sup>a</sup>Large changes occur during treatment of DKA. <sup>b</sup>Although plasma levels may be normal or high at presentation, total-body stores are usually depleted.

<sup>c</sup>Sometimes occurs with SGLT2 inhibitor treatment; disproportionate glucosuria is consistent with SGLT2 inhibitor effect.

occlusion or device malfunction, eating disorder, mental health disorders, or an unstable psychosocial environment may sometimes be a factor precipitating DKA. Complete omission or inadequate administration of insulin by the patient or health care team (in a hospitalized patient with type 1 DM) may precipitate DKA.

**Pathophysiology** DKA results from relative or absolute insulin deficiency combined with counterregulatory hormone excess (glucagon, catecholamines, cortisol, and growth hormone). Both insulin deficiency and glucagon excess, in particular, are necessary for DKA to develop. The decreased ratio of insulin to glucagon promotes gluconeogenesis, glycogenolysis, and ketone body formation in the liver, as well as increases in substrate delivery from fat and muscle (free fatty acids, amino acids) to the liver. Ketosis results from a marked increase in free fatty acid release from adipocytes, with a resulting shift toward ketone body synthesis in the liver. Reduced insulin levels, in combination with elevations in catecholamines and growth hormone, also increase lipolysis and the release of free fatty acids. Markers of inflammation (cytokines, C-reactive protein) are elevated in both DKA and HHS.

**Laboratory Abnormalities and Diagnosis** The timely diagnosis of DKA is crucial and allows for prompt initiation of therapy. DKA is characterized by hyperglycemia (serum glucose >13.9 mmol/L [250 mg/dL]), ketosis, and metabolic acidosis [serum bicarbonate <15–18 mmol/L with increased anion gap] along with a number of secondary metabolic derangements (Table 404-6). Occasionally, the serum glucose is only minimally elevated and may even be normal

(euglycemic DKA). This has been noted especially in individuals treated with SGLT2 inhibitors. Arterial pH usually ranges between 6.8 and 7.3, depending on the severity of the acidosis. Despite a total-body potassium deficit, the serum potassium at presentation may be mildly elevated, secondary to the acidosis and volume depletion. Total-body stores of sodium, chloride, phosphorus, and magnesium are also reduced in DKA but are not accurately reflected by their levels in the serum because of hypovolemia and hyperglycemia. Elevated blood urea nitrogen (BUN) and serum creatinine levels reflect intravascular volume depletion. Leukocytosis, hypertriglyceridemia, and hyperlipoproteinemia are commonly found as well. Hyperamylasemia may suggest a diagnosis of pancreatitis, especially when accompanied by abdominal pain. However, in DKA the amylase is usually of salivary origin and thus is not diagnostic of pancreatitis. Serum lipase should be obtained if pancreatitis is suspected.

The measured serum sodium is reduced as a consequence of the hyperglycemia (1.6-mmol/L [1.6-meq] reduction in serum sodium for each 5.6-mmol/L [100-mg/dL] rise in the serum glucose). A normal serum sodium in the setting of DKA indicates a more profound water deficit.

In DKA, the ketone body, -hydroxybutyrate, is synthesized at a threefold greater rate than acetoacetate; however, acetoacetate is preferentially detected by a commonly used ketosis detection reagent (nitroprusside). Serum ketones are present at significant levels (usually positive at serum dilution of 1:8). The nitroprusside tablet, or stick, is often used to detect urine ketones; certain medications such as captopril or penicillamine may cause false-positive reactions. Serum or plasma assays for -hydroxybutyrate are preferred because they more accurately reflect the true ketone body level.

The metabolic derangements of DKA exist along a spectrum, beginning with mild acidosis with moderate hyperglycemia evolving into more severe findings. The degree of acidosis and hyperglycemia do not necessarily correlate closely because a variety of factors determine the level of hyperglycemia (oral intake, urinary glucose loss). Ketonemia is a consistent finding in DKA and distinguishes it from simple hyperglycemia. The differential diagnosis of DKA includes starvation ketosis, alcoholic ketoacidosis (bicarbonate usually >15 meq/L), and other forms of increased anion-gap acidosis (Chap. 55).

## TREATMENT

### Diabetic Ketoacidosis

The management of DKA is outlined in Table 404-8. After initiating IV fluid replacement and insulin therapy, the agent or

**TABLE 404-7 Manifestations of Diabetic Ketoacidosis**

Symptoms	Physical Findings
Nausea/vomiting	Tachycardia
Thirst/polyuria	Dehydration/hypotension
Abdominal pain	Tachypnea/Kussmaul respirations/respiratory distress
Shortness of breath	Abdominal tenderness (may resemble acute pancreatitis or surgical abdomen)
Precipitating events	Lethargy/obtundation/cerebral edema/possibly coma
Inadequate insulin administration	
Infection (pneumonia/UTI/gastroenteritis/sepsis)	
Infarction (cerebral, coronary, mesenteric, peripheral)	
Pancreatitis	
Drugs (cocaine)	
Pregnancy	

Abbreviation: UTI, urinary tract infection.

**TABLE 404-8 Management of Diabetic Ketoacidosis**

1. Confirm diagnosis ( $\uparrow$  serum glucose,  $\uparrow$  serum  $\beta$ -hydroxybutyrate, metabolic acidosis).
2. Admit to hospital; intensive care setting may be necessary for frequent monitoring, if pH <7.00, labored respiration, or impaired level of arousal.
3. Assess:
  - Serum electrolytes ( $K^+$ ,  $Na^+$ ,  $Mg^{2+}$ ,  $Cl^-$ , bicarbonate, phosphate)
  - Acid-base status—pH,  $HCO_3^-$ ,  $PCO_2$ ,  $\beta$ -hydroxybutyrate
  - Renal function (creatinine, urine output)
4. Replace fluids: 2–3 L of 0.9% saline or lactated Ringer's over first 1–3 h (10–20 mL/kg per hour); subsequently, 0.45% saline at 250–500 mL/h; change to 5% glucose and 0.45% saline or lactated Ringer's at 150–250 mL/h when blood glucose reaches 250 mg/dL (13.9 mmol/L).
5. Administer short-acting regular insulin: IV (0.1 units/kg), then 0.1 units/kg per hour by continuous IV infusion; increase two- to threefold if no response by 2–4 h. If the initial serum potassium is <3.3 mmol/L (3.3 meq/L), do not administer insulin until the potassium is corrected. Subcutaneous insulin may be used in uncomplicated, mild-moderate DKA with close monitoring.
6. Assess patient: What precipitated the episode (noncompliance, infection, trauma, pregnancy, infarction, cocaine)? Initiate appropriate workup for precipitating event (cultures, CXR, ECG, etc.).
7. Measure blood glucose every 1–2 h; measure electrolytes (especially  $K^+$ , bicarbonate, phosphate) and anion gap every 4 h for first 24 h.
8. Monitor blood pressure, pulse, respirations, mental status, fluid intake and output every 1–4 h.
9. Replace  $K^+$ : 10 meq/h when plasma  $K^+$  <5.0–5.2 meq/L (or 20–30 meq/L of infusion fluid). ECG normal, urine flow and normal creatinine documented; administer 40–80 meq/h when plasma  $K^+$  <3.5 meq/L or if bicarbonate is given. If initial serum potassium is >5.2 mmol/L (5.2 meq/L), do not supplement  $K^+$  until the potassium is corrected.
10. See text about bicarbonate or phosphate supplementation.
11. Continue above until patient is stable, glucose goal is 8.3–11.1 mmol/L (150–200 mg/dL), and acidosis is resolved. Insulin infusion may be decreased to 0.02–0.1 units/kg per hour.
12. Administer long-acting insulin as soon as patient is eating. Allow for a 2- to 4-h overlap in insulin infusion and SC long-acting insulin injection.

**Abbreviations:** CXR, chest x-ray; ECG, electrocardiogram.

**Source:** Data from M Sperling, in *Therapy for Diabetes Mellitus and Related Disorders*, 3rd ed. Alexandria, VA: American Diabetes Association; 1998 and EA Nyenwe, AE Kitabchi: The evolution of diabetic ketoacidosis: An update of its etiology, pathogenesis and management. *Metabolism* 65:507, 2016.

event that precipitated the episode of DKA should be sought and aggressively treated. If the patient is vomiting or has altered mental status, a nasogastric tube should be inserted to prevent aspiration of gastric contents. Central to successful treatment of DKA is careful monitoring and frequent reassessment to ensure that the patient and the metabolic derangements are improving. A comprehensive flow sheet should record chronologic changes in vital signs, fluid intake and output, and laboratory values as a function of insulin administered.

After the initial bolus of normal saline or lactated Ringer's, replacement of the sodium and free water deficit is carried out over the next 24 h (fluid deficit is often 3–5 L). When hemodynamic stability and adequate urine output are achieved, IV fluids should be switched to 0.45% saline or lactated Ringer's depending on the calculated volume deficit. The change to 0.45% saline or using lactated Ringer's helps to reduce the trend toward hyperchloremia later in the course of DKA.

A bolus of IV (0.1 units/kg) short-acting regular insulin is usually administered immediately (Table 404-8), and subsequent treatment should provide continuous and adequate levels of circulating insulin. IV administration is usually preferred (0.1 units/kg of regular insulin per h) in patients with severe or complicated DKA because it ensures rapid distribution and allows adjustment of the infusion rate as the patient responds to therapy. Mild to moderate uncomplicated DKA can also be treated with SC short-acting insulin analogues. If chosen, IV regular insulin should be continued until the acidosis resolves and the patient is metabolically stable. As the acidosis and insulin resistance associated with DKA resolve,

the insulin infusion rate can be decreased (to 0.02–0.1 units/kg per h). Long-acting insulin, in combination with SC short-acting insulin, should be administered as soon as the patient resumes eating, because this facilitates transition to an outpatient insulin regimen and reduces length of hospital stay. It is crucial to continue the insulin infusion or insulin SC until adequate insulin levels are achieved by administering long-acting insulin by the SC route. Even relatively brief periods of inadequate insulin administration in this transition phase may result in DKA relapse. In euglycemic DKA associated with SGLT2 inhibitors, the pharmacologic effect may persist for 10–14 days following discontinuation of SGLT2 inhibitor therapy as evidenced by ongoing glucosuria despite normoglycemia (glucose <180 mg/dL), during which time relapse of ketoacidosis is common if nutritional intake has not advanced (e.g., in the postoperative setting).

Hyperglycemia usually improves at a rate of 4.2–5.6 mmol/L (50–100 mg/dL) per h as a result of insulin-mediated glucose disposal, reduced hepatic glucose release, and rehydration. Rehydration reduces catecholamines, increases urinary glucose loss, and expands the intravascular volume. The decline in the plasma glucose within the first 1–2 h may be more rapid and is mostly related to volume expansion. When the plasma glucose reaches 11.1–13.9 mmol/L (200–250 mg/dL), glucose should be added to the 0.45% saline infusion to maintain the plasma glucose in the 8.3–11.1 mmol/L (150–200 mg/dL) range, and the insulin infusion should be continued at a lower rate to inhibit ketogenesis. More rapid correction of the serum glucose can precipitate the development of cerebral edema. Ketoacidosis begins to resolve as insulin reduces lipolysis, increases peripheral ketone body use, suppresses hepatic ketone body formation, and promotes bicarbonate regeneration. However, the acidosis and ketosis resolve more slowly than hyperglycemia. Depending on the rise of serum chloride, the anion gap (but not bicarbonate) will normalize. A hyperchloremic acidosis (serum bicarbonate of 15–18 mmol/L [15–18 meq/L]) often follows successful treatment and gradually resolves as the kidneys regenerate bicarbonate and excrete chloride.

Potassium stores are depleted in DKA (estimated deficit 3–5 mmol/kg [3–5 meq/kg]). During treatment with insulin and fluids, various factors contribute to the development of hypokalemia. These include insulin-mediated potassium transport into cells, resolution of the acidosis (which also promotes potassium entry into cells), and urinary loss of potassium salts of organic acids. Thus, potassium repletion should commence as soon as adequate urine output and a normal serum potassium are documented. If the initial serum potassium level is elevated, then potassium repletion should be delayed until the potassium falls into the normal range. Inclusion of 20–40 meq of potassium in each liter of IV fluid is reasonable, but additional potassium supplements may also be required. To reduce the amount of chloride administered, potassium phosphate or acetate can be substituted for the chloride salt. The goal is to maintain the serum potassium at >3.5 mmol/L (3.5 meq/L).

Despite a bicarbonate deficit, bicarbonate replacement is not usually necessary. In fact, theoretical arguments suggest that bicarbonate administration and rapid reversal of acidosis may impair cardiac function, reduce tissue oxygenation, and promote hypokalemia. The results of most clinical trials do not support the routine use of bicarbonate replacement, and one study in children found that bicarbonate use was associated with an increased risk of cerebral edema. However, in the presence of severe acidosis (arterial pH <7.0), sodium bicarbonate (50 mmol [meq/L] in 200 mL of sterile water with 10 meq/L KCl per h) may be administered for the first 2 h until the pH is >7.0. Hypophosphatemia may result from increased glucose usage, but randomized clinical trials have not demonstrated that phosphate replacement is beneficial in DKA. If the serum phosphate is <0.32 mmol/L (1 mg/dL), then phosphate supplement should be considered and the serum calcium monitored. Hypomagnesemia may develop during DKA therapy and may also require supplementation.

With appropriate therapy, the mortality rate of DKA is low (<1%) and is related more to the underlying or precipitating event, such as infection or myocardial infarction. Venous thrombosis, upper GI bleeding, and acute respiratory distress syndrome occasionally complicate DKA. The major nonmetabolic complication of DKA therapy is cerebral edema, which most often develops in children as DKA is resolving. The etiology of and optimal therapy for cerebral edema are not well established, but overreplacement of free water and rapid normalization of serum glucose should be avoided.

Following treatment, the physician and patient should review the sequence of events that led to DKA to prevent future recurrences. Foremost is patient education about the symptoms of DKA, its precipitating factors, and the management of diabetes during a concurrent illness.

## HYPERGLYCEMIC HYPEROSMOLAR STATE

**Clinical Features** The prototypical patient with HHS is an elderly individual with type 2 DM, with a several-week history of polyuria, weight loss, and diminished oral intake that culminates in mental confusion, lethargy, or coma. The physical examination reflects profound dehydration and hyperosmolality and reveals hypotension, tachycardia, and altered mental status. Notably absent are symptoms of nausea, vomiting, and abdominal pain and the Kussmaul respirations characteristic of DKA. HHS is often precipitated by a serious, concurrent illness such as myocardial infarction or stroke. Sepsis, pneumonia, and other serious infections are frequent precipitants and should be sought. In addition, a debilitating condition (prior stroke or dementia) or social situation that compromises water intake usually contributes to the development of the disorder.

**Pathophysiology** Relative insulin deficiency and inadequate fluid intake are the underlying causes of HHS. Insulin deficiency increases hepatic glucose production (through glycogenolysis and gluconeogenesis) and impairs glucose utilization in skeletal muscle (see above discussion of DKA). Hyperglycemia induces an osmotic diuresis that leads to intravascular volume depletion, which is exacerbated by inadequate fluid replacement. The absence of ketosis in HHS is not understood. Presumably, the insulin deficiency is only relative and less severe than in DKA. Lower levels of counterregulatory hormones and free fatty acids have been found in HHS than in DKA in some studies. It is also possible that the liver is less capable of ketone body synthesis or that the insulin/glucagon ratio does not favor ketogenesis.

**Laboratory Abnormalities and Diagnosis** The laboratory features in HHS are summarized in Table 404-6. Most notable are the marked hyperglycemia (plasma glucose may be >55.5 mmol/L [1000 mg/dL]), hyperosmolality (>350 mOsm/L), and prerenal azotemia. The measured serum sodium may be normal or slightly low despite the marked hyperglycemia. The corrected serum sodium is usually increased (add 1.6 meq to measured sodium for each 5.6-mmol/L [100-mg/dL] rise in the serum glucose). In contrast to DKA, acidosis and ketonemia are absent or mild. A small anion-gap metabolic acidosis may be present secondary to increased lactic acid. Moderate ketonuria, if present, is secondary to starvation.

## TREATMENT

### Hyperglycemic Hyperosmolar State

Volume depletion and hyperglycemia are prominent features of both HHS and DKA. Consequently, therapy of these disorders shares several elements (Table 404-8). In both disorders, careful monitoring of the patient's fluid status, laboratory values, and insulin infusion rate is crucial. Underlying or precipitating problems should be aggressively sought and treated. In HHS, fluid losses and dehydration are usually more pronounced than in DKA due to the longer duration of the illness. The patient with HHS is usually older, more likely to have mental status changes, and more likely to have a life-threatening precipitating event with accompanying

comorbidities. Even with proper treatment, HHS has a substantially higher mortality rate than DKA (up to 15% in some clinical series).

Fluid replacement should initially stabilize the hemodynamic status of the patient (1–3 L of 0.9% normal saline over the first 2–3 h). Because the fluid deficit in HHS is accumulated over a period of days to weeks, the rapidity of reversal of the hyperosmolar state must balance the need for free water repletion with the risk that too rapid a reversal may worsen neurologic function. If the serum sodium is >150 mmol/L (150 meq/L), 0.45% saline should be used. After hemodynamic stability is achieved, the IV fluid administration is directed at reversing the free water deficit using hypotonic fluids (0.45% saline initially, then 5% dextrose in water [D/W]). The calculated free water deficit (which can be as great as 9–10 L) should be reversed over the next 1–2 days (infusion rates of 200–300 mL/h of hypotonic solution). Potassium repletion is usually necessary and should be dictated by repeated measurements of the serum potassium. In patients taking diuretics, the potassium deficit can be quite large and may be accompanied by magnesium deficiency. Hypophosphatemia may occur during therapy and can be improved by using KPO<sub>4</sub> and beginning nutrition.

As in DKA, rehydration and volume expansion lower the plasma glucose initially, but insulin is also required. A reasonable regimen for HHS begins with an IV insulin bolus of 0.1 unit/kg followed by IV insulin at a constant infusion rate of 0.1 unit/kg per hour. If the serum glucose does not fall, increase the insulin infusion rate by twofold. As in DKA, glucose should be added to IV fluid when the plasma glucose falls to 11.1–13.9 mmol/L (200–250 mg/dL), and the insulin infusion rate should be decreased to 0.02–0.1 unit/kg per h. The insulin infusion should be continued until the patient has resumed eating and can be transferred to an SC insulin regimen. The patient should be discharged from the hospital on insulin, although some patients can later switch to oral glucose-lowering agents.

## MANAGEMENT OF DIABETES IN A HOSPITALIZED PATIENT

Virtually all medical and surgical subspecialties are involved in the care of hospitalized patients with diabetes. Hyperglycemia, whether in a patient with known diabetes or in someone without known diabetes, appears to be a predictor of poor outcome in hospitalized patients. General anesthesia, surgery, infection, or concurrent illness raises the levels of counterregulatory hormones (cortisol, growth hormone, catecholamines, and glucagon) and cytokines that may lead to transient insulin resistance and hyperglycemia. These factors increase insulin requirements by increasing glucose production and impairing glucose utilization and thus may worsen glycemic control. The concurrent illness or surgical procedure may lead to variable insulin absorption and also prevent the patient with DM from eating normally and, thus, may promote hypoglycemia. Glycemic control should be assessed on admission using the HbA<sub>1c</sub>. Electrolytes, renal function, and intravascular volume status should be assessed as well. The high prevalence of CVD in individuals with DM (especially in type 2 DM) may necessitate preoperative cardiovascular evaluation ([Chap. 405](#)).

The goals of diabetes management during hospitalization are near-normoglycemia, avoidance of hypoglycemia, and transition back to the outpatient diabetes treatment regimen. Upon hospital admission, frequent glycemic monitoring should begin, as should planning for diabetes management after discharge. CGM in the hospital or ICU setting is not FDA-approved but is under study. Glycemic control appears to improve the clinical outcomes in a variety of settings, but optimal glycemic goals for the hospitalized patient are incompletely defined. In a number of cross-sectional studies of patients with diabetes, a greater degree of hyperglycemia was associated with worse cardiac, neurologic, and infectious outcomes. In some studies, patients who do not have preexisting diabetes but who develop modest blood glucose elevations during their hospitalization appear to benefit from achieving near-normoglycemia using insulin treatment. However, a large randomized clinical trial (Normoglycemia in Intensive Care Evaluation Survival Using Glucose Algorithm Regulation [NICE-SUGAR])

**3118** of individuals in the intensive care unit (ICU) (most of whom were receiving mechanical ventilation) found an increased mortality rate and a greater number of episodes of severe hypoglycemia with very strict glycemic control (target blood glucose of 4.5–6 mmol/L or 81–108 mg/dL) compared to individuals with a more moderate glycemic goal (target blood glucose of <10 mmol/L or 180 mg/dL). Currently, most data suggest that very strict blood glucose control in acutely ill patients likely worsens outcomes and increases the frequency of hypoglycemia. The ADA suggests the following glycemic goals for hospitalized patients: (1) in critically or non-critically ill patients: glucose of 7.8–10.0 mmol/L or 140–180 mg/dL; (2) in selected patients: glucose of 6.1–7.8 mmol/L or 110–140 mg/dL with avoidance of hypoglycemia; (3) the target range in the perioperative period should be 80–180 mg/dL (4.4–10.0 mmol/L).

Critical aspects for optimal diabetes care in the hospital include the following. (1) A hospital-wide system approach to treatment of hyperglycemia and prevention of hypoglycemia is needed. Inpatient diabetes management teams consisting of nurse practitioners and physicians are increasingly common. (2) Diabetes treatment plans should focus on the transition from the ICU and the transition from the inpatient to outpatient setting. (3) Adjustment of the discharge treatment regimen of patients whose diabetes was poorly controlled on admission (as reflected by the HbA<sub>1c</sub>) is important.

The physician caring for an individual with diabetes in the perioperative period, during times of infection or serious physical illness, or simply when the patient is fasting for a diagnostic procedure must monitor the plasma glucose vigilantly, adjust the diabetes treatment regimen, and provide glucose infusion as needed. Hypoglycemia is frequent in hospitalized patients, and many of these episodes are avoidable. Hospital systems should have a diabetes management protocol to avoid inpatient hypoglycemia. Measures to reduce or prevent hypoglycemia include frequent glucose monitoring, but it is also important to prevent hypoglycemia by anticipating drops in insulin requirement by factors such as decreasing renal function, decreasing glucocorticoid doses, or interruption of nutrition (parenteral or enteral or PO).

Depending on the severity of the patient's illness and the hospital setting, the physician can use either an insulin infusion or SC insulin. Insulin infusions are preferred in the ICU or in a clinically unstable setting because the half-life of the infused insulin is quite short (minutes). The absorption of SC insulin may be variable in such situations. Insulin infusions can also effectively control plasma glucose in the perioperative period and when the patient is unable to take anything by mouth, although for relatively short (<4 h) procedures most patients can remain on SC insulin. Regular insulin is used rather than insulin analogues for IV insulin infusion because it is less expensive and equally effective. The physician must consider carefully the clinical setting in which an insulin infusion will be used, including whether adequate ancillary personnel are available to monitor the blood glucose frequently and whether they can adjust the insulin infusion rate to maintain the blood glucose within the optimal range. Insulin-infusion algorithms should integrate the insulin sensitivity of the patient, frequent blood glucose monitoring, and the trend of changes in the blood glucose to determine the insulin-infusion rate. Insulin-infusion algorithms jointly developed and implemented by nursing and physician staff are advised. Because of the short half-life of IV regular insulin, it is necessary to administer long-acting insulin prior to discontinuation of the insulin infusion (2–4 h before the infusion is stopped) to avoid a period of insulin deficiency.

In patients who are not critically ill or not in the ICU, basal or "scheduled" insulin is provided by SC, long-acting insulin supplemented by prandial and/or "corrective" insulin using a short-acting insulin (insulin analogues preferred). "Sliding scale," short-acting insulin alone, where no insulin is given unless the blood glucose is elevated, is inadequate for inpatient glucose management and should not be used. The short-acting, preprandial insulin dose should include coverage for food consumption (based on anticipated carbohydrate intake) plus corrective insulin based on the patient's insulin sensitivity and the blood glucose. For example, if the patient is thin (and likely insulin-sensitive), an insulin correction factor might be 1 unit for each

2.7 mmol/L (50 mg/dL) over the glucose target. If the patient is obese and insulin-resistant, then the insulin correction factor might be 2 units for each 2.7 mmol/L (50 mg/dL) over the glucose target. It is critical to individualize the regimen and adjust the basal or "scheduled" insulin dose frequently, based on the corrective insulin required. A consistent carbohydrate-controlled diabetes meal plan for hospitalized patients provides a predictable amount of carbohydrate for a particular meal each day (but not necessarily the same amount for breakfast, lunch, and supper) and avoids concentrated sweets. Individuals with type 1 DM who are undergoing general anesthesia and surgery or who are seriously ill should receive continuous insulin, either through an IV insulin infusion, their insulin infusion device, or by SC administration of a reduced dose of long-acting insulin. Short-acting insulin alone is insufficient. Prolongation of a surgical procedure or delay in the recovery room is not uncommon and may result in periods of insulin deficiency leading to DKA. Insulin infusion is the preferred method for managing patients with type 1 DM over a prolonged (several hours) perioperative period or when serious concurrent illness is present (0.5–1.0 units/h of regular insulin). If the diagnostic or surgical procedure is brief (<4 h), a reduced dose of SC insulin may suffice (20–50% basal reduction, with short-acting bolus insulin withheld or reduced). This approach prevents interruption of insulin infusion device therapy, or for MDI, facilitates the transition back to long-acting insulin after the procedure. The blood glucose should be monitored frequently during the illness or in the perioperative period.

Individuals with type 2 DM can be managed with either an insulin infusion or SC long-acting insulin (20–50% reduction depending on clinical setting) plus preprandial, short-acting insulin. Oral glucose-lowering agents should be discontinued upon admission (or up to a week prior to planned admission for SGLT2 inhibitors) and are not useful in regulating the plasma glucose in clinical situations where the insulin requirements and glucose intake are changing rapidly. Moreover, these oral agents may be dangerous if the patient is fasting (e.g., hypoglycemia with sulfonylureas, euglycemic DKA with SGLT2 inhibitors) or at risk for declining kidney function due to, for example, radiographic contrast media or unstable CHF (lactic acidosis with metformin). Once clinically stable, oral glucose-lowering agents may be resumed in anticipation of discharge.

## SPECIAL CONSIDERATIONS IN DM

### TOTAL PARENTERAL NUTRITION (TPN)/TOTAL ENTERAL NUTRITION (TEN)

(See also Chap. 335) TPN or TEN greatly increases insulin requirements. In addition, individuals not previously known to have DM may become hyperglycemic during TPN or TEN and require insulin treatment. For TPN, IV insulin infusion is the preferred treatment for hyperglycemia, and rapid titration to the required insulin dose is done most efficiently using a separate insulin infusion. After the total insulin dose has been determined, a proportion of this insulin may be added directly to the TPN solution to cover the nutritional requirements for insulin, and adjusted based on the need for modified dosing of short-acting insulin. In TEN, hyperglycemia may be limited by using high-protein formulations, but often requires insulin treatment. Short-acting insulins should be used to cover bolus or continuous enteral feeding to minimize the risk for hypoglycemia should the TEN be interrupted or held. Patients with insulin deficiency (type 1 DM and pancreatogenic DM) should also receive long-acting insulin (0.1–0.2 units/kg per day) to cover basal insulin requirements should the TPN or TEN be interrupted or cycled.

### GLUCOCORTICOIDS

Glucocorticoids increase insulin resistance, decrease glucose utilization, increase hepatic glucose production, and impair insulin secretion. These changes lead to a worsening of glycemic control in individuals with DM and may precipitate hyperglycemia in other individuals. If new-onset hyperglycemia remains during chronic treatment with supraphysiologic doses of glucocorticoid (>5 mg of prednisone or equivalent), the DM may be called "steroid-induced diabetes." The

effects of glucocorticoids on glucose homeostasis are dose-related, usually reversible, most pronounced in the postprandial period, and dependent on the timing and type of glucocorticoid. If the FPG is near the normal range, oral diabetes agents (e.g., sulfonylureas, metformin) may be sufficient to reduce hyperglycemia. If the FPG is  $>11.1$  mmol/L (200 mg/dL), oral agents are usually not efficacious, and insulin therapy is required. Short-acting insulin may be sufficient alone or together with long-acting insulin in order to control postprandial glucose excursions.

## DIABETES MANAGEMENT IN OLDER ADULTS

Diabetes is very common in older adults, being present in ~25% of individuals over the age of 65 years. Increasingly, individuals with many years of type 1 DM are part of the patient population. As discussed above, individualized therapeutic goals and modalities in older adults should consider biologic age, other comorbidities and risk factors (hypertension, CV disease, etc.), neurocognitive and physical functional status, living arrangements, social support, and other medications. For example, the HbA<sub>1c</sub> goal for a highly functional 80-year-old should be different from that for an individual with diabetes in long-term care (skilled nursing facilities). In the former, the HbA<sub>1c</sub> goal (<7.0–7.5%) and selected therapies may be similar to younger individuals whereas in an individual with complex/poor health or cognitive impairment, an HbA<sub>1c</sub> goal of <8.0–8.5% would be reasonable. Critical to diabetes management in all older individuals is the avoidance of hypoglycemia, which can worsen underlying cognitive impairment or CV disease. For individuals using CGM, <1% of time should be spent with glucose <70 mg/dL and spending >50% of time in the target range of 70–180 mg/dL is acceptable. Thus, medications that can cause hypoglycemia (insulin secretagogues, insulin) should be used carefully. In choosing medications for diabetes, the adverse effects (Table 404-5) should be considered (especially heart failure, renal insufficiency, etc.). Hypertension and dyslipidemia should be treated in elderly individuals with diabetes because there is clear benefit of blood pressure control with the benefit for lipid-lowering medications being less clearly demonstrated.

## REPRODUCTIVE ISSUES

Reproductive capacity in either men or women with DM appears to be normal. Menstrual cycles may be associated with alterations in glycemic control in women with DM. Pregnancy is associated with marked insulin resistance; the increased insulin requirements often precipitate DM and lead to the diagnosis of gestational diabetes mellitus (GDM). Glucose, which at high levels is a teratogen to the developing fetus, readily crosses the placenta, but insulin does not. Thus, hyperglycemia from the maternal circulation may stimulate insulin secretion in the fetus. The anabolic and growth effects of insulin may result in macrosomia. GDM complicates ~7% (range 1–14%) of pregnancies. The incidence of GDM is greatly increased in certain ethnic groups, including blacks and Latinas, consistent with a similar increased risk of type 2 DM. Current recommendations advise screening for glucose intolerance between weeks 24 and 28 of pregnancy in women not known to have diabetes. Therapy for GDM is similar to that for individuals with pregnancy-associated diabetes and involves MNT and insulin, if hyperglycemia persists. Oral glucose-lowering agents are not approved for use during pregnancy, but studies using metformin or glyburide have shown efficacy and have not found toxicity. With current practices, the morbidity and mortality rates of the mother with GDM and the fetus are not different from those in the nondiabetic population. Individuals who develop GDM are at marked increased risk for developing type 2 DM in the future and should be screened periodically for DM (see screening recommendations in Chap. 403). Most individuals with GDM revert to normal glucose tolerance after delivery, but some will continue to have overt diabetes or impairment of glucose tolerance after delivery. In addition, children of women with GDM appear to be at risk for obesity and glucose intolerance and have an increased risk of diabetes beginning in the later stages of adolescence.

Pregnancy in individuals with known DM requires meticulous planning and adherence to strict treatment regimens. Intensive insulin therapy and near-normalization of the HbA<sub>1c</sub> (<6.5%) are essential

for individuals with existing DM who are planning pregnancy. Consideration should be given to insulin infusion and CGM devices that may help to improve glycemic control prior to conception since the most crucial period of glycemic control is soon after fertilization. The risk of fetal malformations is increased 4–10 times in individuals with uncontrolled DM at the time of conception, and normal blood glucose during the preconception period and throughout the periods of organ development in the fetus should be the goal, with more frequent monitoring of HbA<sub>1c</sub> every 2 months throughout gestation. Maintenance of the HbA<sub>1c</sub> <6.0–6.5% reduces the incidence and severity of fetal macrosomia and neonatal hypoglycemia related to fetal hyperinsulinism driven by elevated maternal glucose.

## LIPODYSTROPHIC DM

Lipodystrophy, or the loss of SC fat tissue, may be generalized in certain genetic conditions such as leprechaunism, or acquired as part of an autoimmune disorder. Generalized lipodystrophy is associated with leptin deficiency and severe insulin resistance and is often accompanied by acanthosis nigricans, hepatic steatosis, and severe hypertriglyceridemia. Recombinant human leptin (metreleptin) may allow for the achievement of metabolic control in generalized lipodystrophy, but is associated with the development of neutralizing antibodies and is available only through a restricted program under a Risk Evaluation and Mitigation Strategy (REMS). Partial lipodystrophy may also be caused by certain genetic or acquired (e.g., treatment of HIV infection) conditions that produce a metabolic syndrome of insulin resistance, ectopic fat accumulation (hepatic steatosis), and glucose intolerance and dyslipidemia. Treatment of early childhood cancer with total-body irradiation may affect adipose tissue development and predisposes survivors to similar metabolic syndrome of adipose tissue dysfunction with potentially severe insulin resistance, hepatic steatosis, hypertriglyceridemia, and diabetes.

**HIV-Associated Lipodystrophy** Protease inhibitors and nucleoside reverse transcriptase inhibitors used in the treatment of HIV disease (Chap. 202) have been associated with a centripetal accumulation of fat (visceral and abdominal area), accumulation of fat in the dorsocervical region, loss of extremity fat, decreased insulin sensitivity (elevations of the fasting insulin level and reduced glucose tolerance on IV glucose tolerance testing), hepatic steatosis, and dyslipidemia. Although many aspects of the physical appearance of these individuals resemble Cushing's syndrome, increased cortisol levels do not account for this appearance. The possibility remains that this is related to HIV infection or highly active antiretroviral therapy by some undefined mechanism, because the syndrome can be observed in individuals not treated with protease inhibitors. Therapy for HIV-related lipodystrophy and associated metabolic dysfunction may include metformin, especially for abdominal fat accumulation, pioglitazone, especially for lipoatrophy and hepatic steatosis. Tesamorelin, a growth hormone-releasing hormone analog, is effective for reducing excess abdominal fat but requires monitoring of the serum insulin-like growth factor-1 (IGF-1) level, and may worsen glucose tolerance or exacerbate hyperglycemia in individuals with diabetes.

## FURTHER READING

- American Diabetes Association: Lifestyle management. Diabetes Care 41:S38, 2018.
- American Diabetes Association: Comprehensive medical evaluation and assessment of comorbidities: Standards of Medical Care in Diabetes—2021. Diabetes Care 44(Suppl. 1):S40, 2021.
- American Diabetes Association: Facilitating behavior change and wellbeing to improve health outcomes: Standards of Medical Care in Diabetes—2021. Diabetes Care 44(Suppl. 1):S5, 2021.
- American Diabetes Association: Pharmacologic approaches to glycemic treatment: Standards of Medical Care in Diabetes—2021. Diabetes Care 44(Suppl. 1):S11, 2021.
- American Diabetes Association: Older adults: Standards of Medical Care in Diabetes—2021. Diabetes Care 44(Suppl. 1):S168, 2021.

3120 American Diabetes Association: Diabetes technology: Standards of Medical Care in Diabetes—2021. *Diabetes Care* 44(Suppl. 1): S85, 2021.

Evert AB et al: Nutrition therapy for adults with diabetes or prediabetes: A consensus report. *Diabetes Care* 42:731, 2019.

Hirsch IB et al: The evolution of insulin and how it informs therapy and treatment choices. *Endocr Rev* 41:733, 2020.

Ibrahim M et al: Recommendations for management of diabetes during Ramadan: update 2020, applying the principles of the ADA/EASD consensus. *BMJ Open Diabetes Res Care* 8:e001248, 2020.

Kalyani RR: Glucose-lowering drugs to reduce cardiovascular risk in type 2 diabetes. *N Engl J Med* 384:1248, 2021.

Nyenwe EA, Kitabchi AE: The evolution of diabetic ketoacidosis: An update of its etiology, pathogenesis and management. *Metabolism* 65:507, 2016.

Palmer SC et al: Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: Systematic review and network meta-analysis of randomised controlled trials. *BMJ* 372:m4573, 2021.

Powers AC: Type 1 diabetes mellitus: Much progress, many opportunities. *J Clin Invest* 131:142242, 2021.

Qaseem A et al: Oral pharmacologic treatment of type 2 diabetes mellitus: A clinical practice guideline update from the American College of Physicians. *Ann Intern Med* 166:279, 2017.

Rubino F et al: Metabolic surgery in the treatment algorithm for type 2 diabetes: A joint statement by international diabetes organizations. *Diabetes Care* 39:86, 2016.

Satin LS et al: New aspects of diabetes research and therapeutic development. *Pharmacol Rev* 73:1001, 2021.

Thabit H, Hovorka R: Coming of age: The artificial pancreas for type 1 diabetes. *Diabetologia* 59:1795, 2016.

Young-Hyman D et al: Psychosocial care for people with diabetes: A position statement of the American Diabetes Association. *Diabetes Care* 39:2126, 2016.

**TABLE 405-1 Diabetes-Related Complications**

Microvascular

Eye disease

Retinopathy (nonproliferative/proliferative)

Macular edema

Neuropathy

Sensory and motor (mono- and polyneuropathy)

Autonomic

Nephropathy (albuminuria and declining renal function)

Macrovascular

Coronary heart disease

Peripheral arterial disease

Cerebrovascular disease

Other

Gastrointestinal (gastroparesis, diarrhea)

Genitourinary (uropathy/sexual dysfunction)

Dermatologic

Infectious

Cataracts

Glaucoma

Cheiroarthropathy<sup>a</sup>

Periodontal disease

Hearing loss

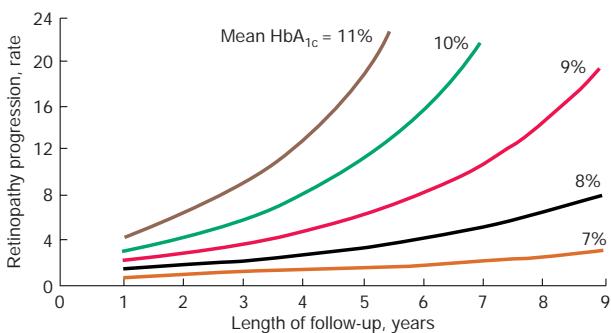
Other comorbid conditions associated with type 1 or type 2 diabetes (relationship to hyperglycemia is uncertain): depression, obstructive sleep apnea, fatty liver disease, hip fracture, osteoporosis, cognitive impairment or dementia, low testosterone in men.

<sup>a</sup>Thickened skin and reduced joint mobility.

macrovascular complications (CHD, peripheral arterial disease [PAD], cerebrovascular disease). Microvascular complications are diabetes-specific, whereas macrovascular complications have additional pathophysiological features that are shared with the general population. Non-vascular complications include infections, skin changes, hearing loss, and increased risk of dementia and impaired cognitive function.

### GLYCEMIC CONTROL AND COMPLICATIONS

The microvascular complications of both type 1 and type 2 DM result from chronic hyperglycemia (Fig. 405-1). Evidence implicating a causative role for chronic hyperglycemia in the development of macrovascular complications is less conclusive as other factors such as dyslipidemia and hypertension also play important roles in macrovascular complications. CHD events and mortality rate are two to four times greater in patients with type 2 DM, correlate with fasting and



**FIGURE 405-1 Relationship of glycemic control and diabetes duration to diabetic retinopathy.** The progression of retinopathy in individuals in the Diabetes Control and Complications Trial is graphed as a function of the length of follow-up with different curves for different hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>) values. (Modified with permission from The relationship of glycemic exposure (HbA<sub>1c</sub>) to the risk of development and progression of retinopathy in the diabetes control and complications trial. *Diabetes* 44:968, 1995.)

## 405

### Diabetes Mellitus: Complications

Alvin C. Powers, John M. Stafford,  
Michael R. Rickels

Diabetes-related complications affect many organ systems and are responsible for the majority of morbidity and mortality associated with the disease. For many years in the United States, diabetes has been the leading cause of new blindness in adults, renal failure, and nontraumatic lower extremity amputation and is a leading contributor to coronary heart disease (CHD). Diabetes-associated microvascular complications usually do not appear until the second decade of hyperglycemia. In contrast, diabetes-associated CHD risk, related in part to insulin resistance and its resultant dyslipidemia, may develop before hyperglycemia is established. Because type 2 diabetes mellitus (DM) often has a long asymptomatic period of hyperglycemia before diagnosis, many individuals with type 2 DM have both glucose-related and insulin resistance-related complications at the time of diagnosis. Fortunately, many of the diabetes-related complications can be prevented or mitigated with aggressive glycemic, lipid, and blood pressure control, as well as efforts at early detection.

Diabetes-related complications can be divided into vascular and nonvascular complications and are similar for type 1 and type 2 DM (Table 405-1). The vascular complications of DM are further subdivided into microvascular (retinopathy, neuropathy, nephropathy) and

postprandial plasma glucose levels as well the hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>), and can be reduced by intensive diabetes management as demonstrated in patients with type 1 DM.

The Diabetes Control and Complications Trial (DCCT) provided definitive proof that reduction in chronic hyperglycemia can prevent many complications of type 1 DM (Fig. 405-1). This large multicenter clinical trial randomized >1400 individuals with type 1 DM to either intensive or conventional diabetes management and prospectively evaluated the development of diabetes-related complications during a mean follow-up of 6.5 years. Individuals in the intensive diabetes management group received insulin by multiple daily injections or pump delivery along with extensive educational, psychological, and medical support, and achieved a substantially lower HbA<sub>1c</sub> (7.3%) than individuals in the conventional diabetes management group (9.1%). After the DCCT results were reported in 1993, study participants were all offered intensive therapy and continue to be followed in the Epidemiology of Diabetes Intervention and Complications (EDIC) trial, which has completed >30 years of follow-up (DCCT + EDIC). During the subsequent follow-up of >18 years, the initial separation in glycemic control disappeared with both arms maintaining a mean HbA<sub>1c</sub> of 8.0%, allowing assessment of a legacy effect of 6.5 years of near-normoglycemia on the development of long-term complications.

The DCCT phase demonstrated that improvement of glycemic control reduced nonproliferative and proliferative retinopathy (47% reduction), albuminuria (39% reduction), clinical nephropathy (54% reduction), and neuropathy (60% reduction). Improved glycemic control also slowed the progression of early diabetic complications. During the DCCT phase, weight gain (4.6 kg) and severe hypoglycemia (requiring assistance of another person to treat) were more common in the intensive therapy group. The benefits of an improvement in glycemic control occurred over the entire range of elevated HbA<sub>1c</sub> values (Fig. 405-1). The results of the DCCT predicted that individuals in the intensive diabetes management group would gain 7.7 additional years of vision, 5.8 additional years free from end-stage renal disease (ESRD), and 5.6 years free from lower extremity amputations. If all complications of DM were combined, individuals in the intensive diabetes management group would experience >15.3 more years of life without significant microvascular complications of DM, compared to individuals who received standard therapy. This translates into an additional 5.1 years of life expectancy for individuals in the intensive diabetes management group. The 30-year follow-up data in the intensively treated group show a continued reduction in retinopathy, nephropathy, and cardiovascular disease. For example, individuals in the intensive therapy group had a 42–57% reduction in cardiovascular events (non-fatal myocardial infarction [MI], stroke, or death from a cardiovascular event) at a mean follow-up of 18 years, even though their subsequent glycemic control was the same as those in the conventional diabetes management group from years 6.5 to 17. During the EDIC phase, <1% of the cohort had become blind, lost a limb to amputation, or required dialysis. Other complications of diabetes, including autonomic neuropathy, bladder and sexual dysfunction, and cardiac autonomic neuropathy, were reduced in the intensive therapy group.

The United Kingdom Prospective Diabetes Study (UKPDS) studied the course of >5000 individuals with type 2 DM for >10 years. This study used multiple treatment regimens and monitored the effect of intensive glycemic control and risk factor treatment on the development of diabetic complications. Newly diagnosed individuals with type 2 DM were randomized to (1) intensive management using various combinations of insulin, a sulfonylurea, or metformin or (2) conventional therapy using dietary modification and pharmacotherapy with the goal of symptom prevention. In addition, individuals were randomly assigned to different antihypertensive regimens. Individuals in the intensive treatment arm achieved an HbA<sub>1c</sub> of 7% compared to 7.9% in the standard treatment group. The UKPDS demonstrated that each percentage point reduction in HbA<sub>1c</sub> was associated with a 35% reduction in microvascular complications. As in the DCCT, there was a continuous relationship between glycemic control and development of complications. Improved glycemic control also reduced the cardiovascular event rate in the follow-up period of >10 years.

One of the major findings of the UKPDS was that strict blood pressure control significantly reduced both macro- and microvascular complications. In fact, the beneficial effects of blood pressure control were greater than the beneficial effects of glycemic control. Lowering blood pressure to moderate goals (144/82 mmHg) reduced the risk of DM-related death, stroke, microvascular endpoints, retinopathy, and heart failure (risk reductions between 32 and 56%). The American Diabetes Association (ADA) recommends blood pressure control <130/80 mmHg for individuals with high cardiovascular risk and <140/90 mmHg for individuals with lower cardiovascular risk.

Similar reductions in the risks of retinopathy and nephropathy were also seen in a small trial of lean Japanese individuals with type 2 DM randomized to either intensive glycemic control or standard therapy with insulin (Kumamoto study). These results demonstrate the effectiveness of improved glycemic control in individuals of different ethnicity and, presumably, a different etiology of DM (i.e., phenotypically different from those in the DCCT and UKPDS). The Action to Control Cardiovascular Risk in Diabetes (ACCORD) and Action in Diabetes and Vascular Disease: Preterax and Diamicron MR Controlled Evaluation (ADVANCE) trials also found that improved glycemic control reduced microvascular complications.

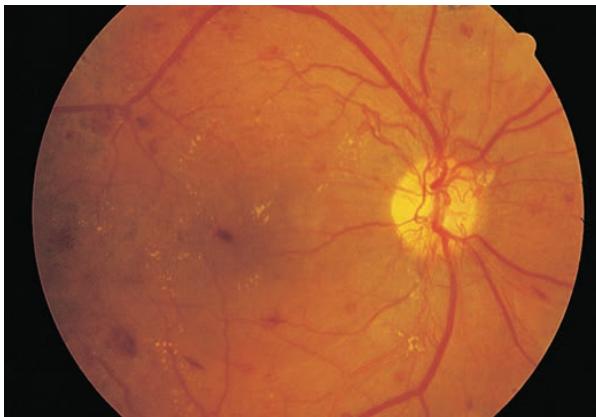
Thus, these large clinical trials in type 1 and type 2 DM indicate that chronic hyperglycemia plays a causative role in the pathogenesis of diabetic micro- and macrovascular complications. In both the DCCT and the UKPDS, cardiovascular events were reduced at follow-up of >10 years, even though the improved glycemic control was not maintained. This legacy effect for a positive impact of a period of improved glycemic control on later diabetes complications has been attributed to the benefits of *metabolic memory*. Of note, despite long-standing DM, some individuals never develop retinopathy or nephropathy, suggesting a genetic susceptibility for developing particular complications.

## MECHANISMS OF COMPLICATIONS

Chronic hyperglycemia is the important etiologic factor leading to complications of DM, but the mechanism(s) by which it leads to such diverse cellular and organ dysfunction is unknown. The complications are likely multifactorial with an emerging hypothesis that hyperglycemia leads to epigenetic changes (Chap. 466) that influence gene expression in affected cells. Chronic hyperglycemia leads to formation of advanced glycation end products (AGEs; e.g., pentosidine, glucospane, and carboxymethyllysine), which bind to specific cell surface receptor and/or the nonenzymatic glycation of intra- and extracellular proteins, leading to cross-linking of proteins, glomerular dysfunction, endothelial dysfunction, altered extracellular matrix composition, and accelerated atherosclerosis. The reduction of cellular glucose entry afforded in certain tissues such as myocardium and renal tubular epithelium through inhibition of the sodium glucose co-transporter-2 (SGLT-2) may contribute to the reduction in CHD events and renoprotective effects.

Growth factors may play an important role in some diabetes-related microvascular complications, and their production is increased by most of these proposed pathways. For example, vascular endothelial growth factor A (VEGF-A) is increased locally in diabetic proliferative retinopathy, decreases after laser photocoagulation, and is the target inhibited by intravitreous injection therapy. A possible unifying mechanism is that hyperglycemia leads to increased production of reactive oxygen species or superoxide in the mitochondria and this may activate several pathways. Although hyperglycemia serves as the initial trigger for complications of diabetes, it is still unknown whether the same pathophysiologic processes are operative in all complications or whether some pathways predominate in certain organs.

The mechanisms of diabetes-related macrovascular complications including MI and stroke are glucose-related mechanisms but also include traditional cardiovascular risk factors (dyslipidemia, hypertension) and insulin resistance. In type 2 diabetes, insulin resistance is present years prior to diagnosis and is associated with obesity and ectopic accumulation of lipids in muscle and liver. Additionally, insulin fails to appropriately suppress lipolysis from adipose tissue, which results in increased delivery of fatty acids to liver, muscle, endothelial cells, and



**FIGURE 405-2** Diabetic retinopathy results in scattered hemorrhages, yellow exudates, and neovascularization. This patient has neovascular vessels proliferating from the optic disc, requiring urgent panretinal laser photocoagulation.

cardiac tissues, leading to tissue accumulation of triglycerides, diacylglycerol, and ceramides.

### OPHTHALMOLOGIC COMPLICATIONS OF DIABETES MELLITUS

DM is the leading cause of blindness between the ages of 20 and 74 in the United States. Severe vision loss is primarily the result of progressive diabetic retinopathy, which leads to significant macular edema and new blood vessel formation. Diabetic retinopathy is classified into two stages: nonproliferative and proliferative. Nonproliferative diabetic retinopathy usually appears late in the first decade or early in the second decade of hyperglycemia and is marked by retinal vascular microaneurysms, blot hemorrhages, and cotton-wool spots (Fig. 405-2). Mild nonproliferative retinopathy may progress to more extensive disease, characterized by changes in venous vessel caliber, intraretinal microvascular abnormalities, and more numerous microaneurysms and hemorrhages. The pathophysiologic mechanisms invoked in nonproliferative retinopathy include loss of retinal pericytes, increased retinal vascular permeability, alterations in retinal blood flow, and abnormal retinal microvasculature, all of which can lead to retinal ischemia.

The appearance of neovascularization in response to retinal hypoxemia is the hallmark of proliferative diabetic retinopathy (Fig. 405-2). These newly formed vessels appear near the optic nerve and/or macula and rupture easily, leading to vitreous hemorrhage, fibrosis, and ultimately retinal detachment. Not all individuals with nonproliferative retinopathy go on to develop proliferative retinopathy, but the more severe the nonproliferative disease, the greater the chance of evolution to proliferative retinopathy within 5 years. This creates an important opportunity for early detection and treatment of diabetic retinopathy. Clinically significant macular edema can occur in the context of nonproliferative or proliferative retinopathy. Fluorescein angiography and optical coherence tomography are useful to detect macular edema, which is associated with a 25% chance of moderate visual loss over the next 3 years. Duration of DM and degree of glycemic control are the best predictors of the development of retinopathy; hypertension, nephropathy, and dyslipidemia are also risk factors. Nonproliferative retinopathy is found in many individuals who have had DM for >20 years. Although there is genetic susceptibility for retinopathy, it confers less influence than either the duration of DM or the degree of glycemic control.

## TREATMENT

### Diabetic Retinopathy

The most effective therapy for diabetic retinopathy is prevention. Intensive glycemic and blood pressure control will delay the development and slow the progression of retinopathy in individuals

with either type 1 or type 2 DM. Paradoxically, during the first 6–12 months of improved glycemic control, established diabetic retinopathy may transiently worsen. Fortunately, this progression is temporary, and in the long term, improved glycemic control is associated with less diabetic retinopathy. Individuals with known retinopathy may be candidates for prophylactic laser photocoagulation when initiating intensive therapy, and especially prior to pancreas or islet transplantation that can rapidly normalize glycemia. Women with type 1 or type 2 DM who are planning pregnancy should be screened prior to and during pregnancy. Once advanced retinopathy is present, improved glycemic control imparts less benefit, although adequate ophthalmologic care can prevent most blindness. Lowering elevated levels of triglycerides with fenofibrate may reduce the progression of retinopathy.

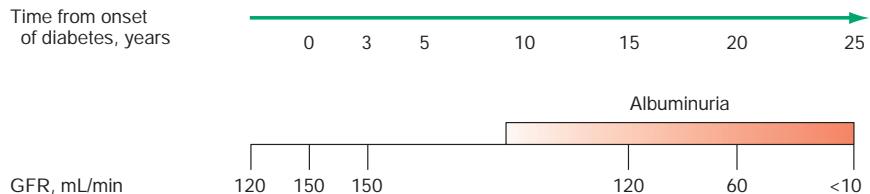
Regular, comprehensive eye examinations are essential for all individuals with DM (see Table 404-1). Most diabetic eye disease can be successfully treated if detected early. Routine, nondilated eye examinations by the primary care provider or diabetes specialist are inadequate to detect diabetic eye disease, which requires a dilated eye exam performed by an optometrist or ophthalmologist, and subsequent management by a retinal specialist. Treatment of severe nonproliferative or proliferative retinopathy or macular edema with laser photocoagulation and/or anti-VEGF therapy (intravitreous injection) usually is successful in preserving vision. Aspirin therapy (up to 650 mg/d) does not appear to influence the natural history of diabetic retinopathy, and antiplatelet agents and anticoagulation may be continued in patients receiving intravitreal injections of anti-VEGF agents. Patients with severe proliferative retinopathy with vitreous hemorrhage and/or traction involving the macula often require surgical vitrectomy.

### RENAL COMPLICATIONS OF DIABETES MELLITUS

Diabetic nephropathy is the leading cause of chronic kidney disease (CKD) and ESRD requiring renal replacement therapy. CKD in individuals with DM is associated with an increased risk of cardiovascular disease, and the prognosis of individuals with diabetes on dialysis is poor. Individuals with diabetic nephropathy commonly have diabetic retinopathy. The presence of CKD in individuals with DM and no retinopathy should prompt investigation for alternative causes of kidney disease.

Like other microvascular complications, the pathogenesis of diabetic nephropathy is related to chronic hyperglycemia. The mechanisms by which chronic hyperglycemia leads to diabetic nephropathy, although incompletely defined, involve the effects of soluble factors (growth factors, angiotensin II, endothelin, AGEs), hemodynamic alterations in the renal microcirculation (glomerular hyperfiltration or hyperperfusion, increased glomerular capillary pressure), and structural changes in the glomerulus (increased extracellular matrix, basement membrane thickening, mesangial expansion, fibrosis). Some of these effects may be mediated through angiotensin II and mineralocorticoid receptors. Smoking accelerates the decline in renal function. Because only 20–40% of patients with diabetes develop diabetic nephropathy, additional genetic or environmental susceptibility factors likely contribute. Known risk factors include a family history of diabetic nephropathy. Diabetic nephropathy and ESRD secondary to DM develop more commonly in blacks, Native Americans, and Hispanic individuals with diabetes.

The natural history of diabetic nephropathy is characterized by a sequence of events that was initially defined for individuals with type 1 DM but appears similar in type 2 DM (Fig. 405-3). Glomerular hyperperfusion and renal hypertrophy occur in the first years after the onset of DM and are associated with an increase of the estimated glomerular filtration rate (GFR). During the first 5 years of DM, thickening of the glomerular basement membrane, glomerular hypertrophy, and mesangial volume expansion occur as the GFR returns to normal. After 5–10 years of type 1 DM, many individuals begin to excrete small amounts of albumin in the urine. The ADA defines albuminuria as a persistently increased urinary albumin-to-creatinine ratio >30 mg/g



**FIGURE 405-3** Time course of development of diabetic nephropathy. The relationship of time from onset of diabetes, albuminuria, and the glomerular filtration rate (GFR) are shown. This figure is typical for type 1 diabetes; individuals with type 2 diabetes may present with a lower GFR at the time of diagnosis.

on a spot specimen. In some individuals with DM and albuminuria of short duration, the albuminuria can regress with improvement in glycemic control (Fig. 405-4) or with improvement in blood pressure control using angiotensin-aldosterone system blockade and/or SGLT-2 inhibitor therapy. Diabetic kidney disease refers to albuminuria and reduced GFR (<60 mL/min per 1.73 m<sup>2</sup>); CKD related to diabetes, which may not be accompanied by albuminuria, is also discussed in Chap. 311. Once there is marked albuminuria and a reduction in GFR, the pathologic changes are likely irreversible.

The nephropathy that develops in type 2 DM differs from that of type 1 DM in that albuminuria may be present when type 2 DM is diagnosed, reflecting its long asymptomatic period, and hypertension more often contributes to albuminuria and reduced GFR. Finally, it should be noted that albuminuria in type 2 DM may be secondary to factors unrelated to DM, such as hypertension, congestive heart failure (CHF), prostate disease, or infection.

As part of comprehensive diabetes care (Chap. 404), albuminuria should be detected at an early stage when effective therapies can be instituted. Because some individuals with type 1 or type 2 DM have a decline in GFR in the absence of albuminuria, assessment should include a spot urinary albumin-to-creatinine ratio and an estimated GFR. The urine protein measurement by routine urinalysis does not detect low levels of albumin excretion. Screening for albuminuria

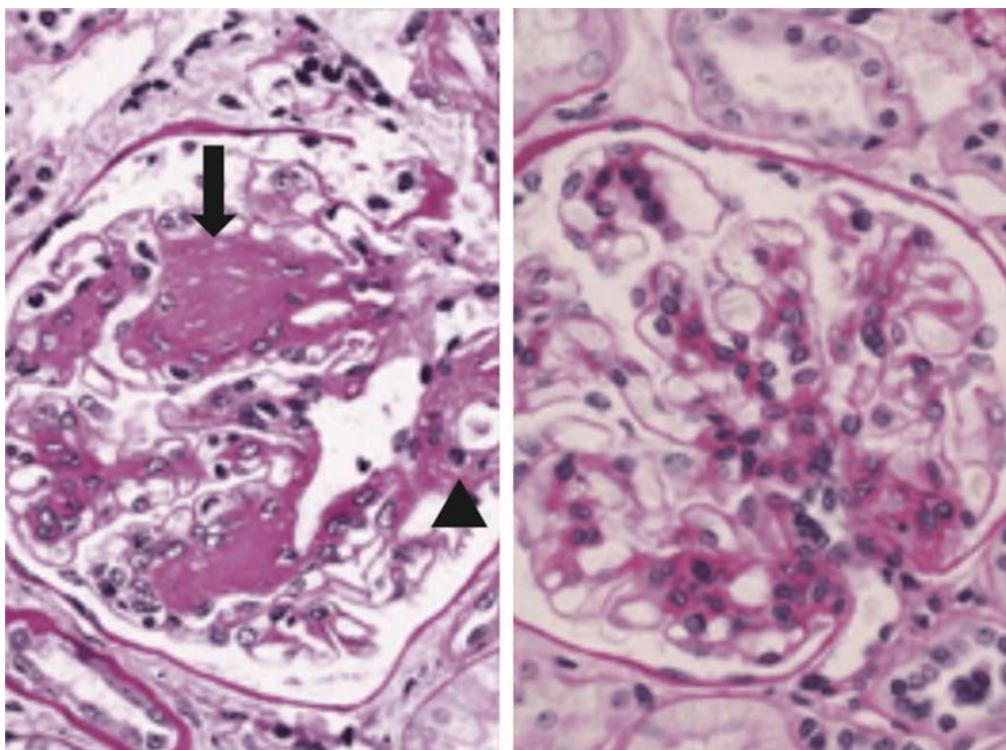
should commence 5 years after the onset of type 1 DM and at the time of diagnosis of type 2 DM.

Type IV renal tubular acidosis (hyporeninemic hypoaldosteronism) may occur in type 1 or 2 DM. These individuals develop a propensity to hyperkalemia and acidemia, which may be exacerbated by medications (especially angiotensin-converting enzyme [ACE] inhibitors, angiotensin receptor blockers [ARBs], and mineralocorticoid receptor antagonists). Patients with DM are predisposed to radiocontrast-induced nephrotoxicity. Risk factors for radiocontrast-induced nephrotoxicity are preexisting nephropathy and volume depletion. Individuals with DM undergoing radiographic procedures with contrast dye should be well hydrated before and after dye exposure, and the serum creatinine should be monitored for 24–48 h following the procedure. Metformin should be held until postintervention confirmation of preserved kidney function.

## TREATMENT

### Diabetic Nephropathy

The optimal therapy for diabetic nephropathy is prevention by control of glycemia (Chap. 404 outlines glycemic goals and approaches). Interventions effective in slowing progression of



**FIGURE 405-4** Diabetic glomerular changes in a patient with type 1 diabetes are reversed by 10 years of normoglycemia as a result of pancreas transplantation. Left panel shows diabetic glomerulosclerosis (arrow) and arteriolar hyalinosis (arrowhead) on kidney biopsy. Right panel shows a near-normal glomerulus in the same patient after 10 years of normoglycemia from pancreas transplantation. (Reproduced with permission from P Fioretto et al: Reversal of lesions of diabetic nephropathy after pancreas transplantation. *N Engl J Med* 339:69, 1998.)

albuminuria and declining kidney function include (1) improved glycemic control, (2) strict blood pressure control, (3) administration of an ACE inhibitor or ARB, and (4) in individuals with type 2 DM, administration of a SGLT-2 inhibitor. Dyslipidemia should also be treated.

Improved glycemic control reduces the rate at which albuminuria appears and progresses in type 1 and type 2 DM. However, once there is a large amount of albuminuria, it becomes more difficult for improved glycemic control to slow progression of renal disease, although 10 years of normoglycemia resulting from pancreas transplantation may lead to regression of mesangial glomerular lesions (Fig. 405-4). During the later phase of declining renal function, insulin requirements may fall as the kidney is a site of insulin degradation. As the GFR decreases with progressive nephropathy, the use and dose of glucose-lowering agents should be reevaluated (see Table 404-5). Some glucose-lowering medications (sulfonylureas and metformin) are contraindicated in advanced renal insufficiency, while others may require dose adjustment (glinides and DPP-4 inhibitors).

Many individuals with type 1 or type 2 DM develop hypertension. Numerous studies in both type 1 and type 2 DM demonstrate the effectiveness of strict blood pressure control in reducing albumin excretion and slowing the decline in renal function. Blood pressure should be maintained at <140/90 mmHg in individuals with diabetes and possibly <130/80 mmHg in individuals at increased risk for CVD and CKD progression.

Either ACE inhibitors or ARBs should be used to reduce the albuminuria and the associated decline in GFR in individuals with type 1 or type 2 DM (see "Hypertension," below). Most experts believe that the two classes of drugs are equivalent in patient with diabetes. ARBs can be used as an alternative in patients who develop ACE inhibitor-associated cough or angioedema. After initiation of therapy, some increase the dose and monitor the urinary albumin. There is no benefit of intervention prior to onset of albuminuria or using a combination of an ACE inhibitor and an ARB. If use of either ACE inhibitors or ARBs is not possible or the blood pressure is not controlled, then diuretics, calcium channel blockers (nondihydropyridine class), or beta blockers (with caution in individuals at increased risk for experiencing hypoglycemia) may be used. Mineralocorticoid receptor antagonists can help reduce blood pressure and albuminuria in refractory cases but require close monitoring of the serum potassium. SGLT-2 inhibitors can reduce albuminuria and, after an initial decline (~3 mL/min per 1.73 m<sup>2</sup>) in GFR, may slow further decline in kidney function in individuals with and without T2DM and CKD. The mechanism of action of SGLT-2 inhibitors is multifactorial and includes inducing natriuresis, reducing intraglomerular pressure through restored tubuloglomerular feedback, and potentially altering signaling pathways related to nutrient sensing (e.g., AMPK). Because of the elevated risk of euglycemic diabetic ketoacidosis with SGLT-2 inhibitors, use in individuals with type 1 DM and insulin-deficient type 2 DM is not recommended. Some glucagon-like peptide-1 (GLP-1) receptor agonists may also both improve glycemic control and reduce the progression of diabetic kidney disease in individuals with type 2 DM and established CVD (Chap. 404). The ADA suggests a protein intake of 0.8 mg/kg of body weight/day in individuals with diabetic kidney disease.

Nephrology consultation should be considered when the estimated GFR is <30 mL/min per 1.743 m<sup>2</sup> or with atypical features such as hematuria, rapidly declining renal function, or proteinuria > 3 g/day. Complications of atherosclerosis are the leading cause of death in diabetic individuals with nephropathy and hyperlipidemia should be treated aggressively. Referral for transplant evaluation should be made when the GFR approaches 20 mL/min per 1.73 m<sup>2</sup>. Preemptive (before dialysis) kidney transplantation from a living donor or simultaneous pancreas-kidney transplantation from a deceased donor both offer improved patient and kidney survival over waiting for a deceased donor kidney alone. A combined pancreas-kidney transplant offers the promise of normoglycemia and freedom from both insulin and dialysis. As compared with

nondiabetic individuals, hemodialysis in patients with DM is associated with more frequent complications, such as hypotension (due to autonomic neuropathy or loss of reflex tachycardia), more difficult vascular access, and accelerated progression of retinopathy.

## NEUROPATHY AND DIABETES MELLITUS

Diabetic neuropathy, which occurs in ~50% of individuals with long-standing type 1 and type 2 DM, manifests as a diffuse neuropathy (distal symmetrical polyneuropathy and/or autonomic neuropathy), a mononeuropathy, and/or a radiculopathy/polyradiculopathy. As with other complications of DM, the development of neuropathy correlates with the duration of diabetes and glycemic control. Additional risk factors are body mass index (BMI) (the greater the BMI, the greater the risk of neuropathy) and smoking. The presence of CVD, elevated triglycerides, and hypertension is also associated with diabetic peripheral neuropathy. Both myelinated and unmyelinated nerve fibers are lost. Because the clinical features of diabetic neuropathy are similar to those of other neuropathies, the diagnosis of diabetic neuropathy should be made only after other possible etiologies are excluded (Chap. 446).

**Distal Symmetric Polyneuropathy (DSPN)** DSPN, the most common form of diabetic neuropathy, most frequently presents with distal sensory loss and pain, but up to 50% of patients do not have symptoms of neuropathy. Symptoms may include a sensation of numbness, tingling, sharpness, or burning that begins in the feet and spreads proximally. Hyperesthesia, paresthesia, and dysesthesia also may occur. Pain typically involves the lower extremities, is usually present at rest, and worsens at night. Both an acute (lasting <12 months) and a chronic form of painful diabetic neuropathy may occur. The acute form is sometimes treatment-related, occurring in the context of improved glycemic control. As diabetic neuropathy progresses, the pain subsides and eventually disappears, but a sensory deficit persists, and motor defects may develop. Physical examination (Chap. 403) often reveals sensory loss (to 10-g monofilament and/or vibration), loss of ankle deep-tendon reflexes, abnormal position sense, and muscular atrophy or foot drop. Annual screening for DSPN should begin 5 years after diagnosis of type 1 DM and at the time of diagnosis of type 2 DM and is aimed at detecting loss of protective sensation (LOPS). LOPS and DSPN are major risk factors for foot ulceration and falls due to small and large nerve fiber dysfunction and predispose to lower extremity amputation.

**Autonomic Neuropathy** Individuals with long-standing type 1 or 2 DM may develop signs of autonomic dysfunction involving the parasympathetic (cholinergic) and sympathetic (adrenergic) systems. DM-related autonomic neuropathy can affect multiple organ systems, including the cardiovascular, gastrointestinal (GI), genitourinary, sudomotor, and metabolic systems. Cardiovascular autonomic neuropathy, reflected by decreased heart rate variability, resting tachycardia, and orthostatic hypotension, is associated with an increase in CVD. Orthostatic hypotension, a late and unusual complication of diabetes, is sometimes seen in patients with associated DPN and severe parasympathetic dysfunction. Reports of sudden death in DM have also been attributed to autonomic neuropathy affecting the cardiovascular system and predisposing to severe hypoglycemia, both of which may prolong the QTc interval. Autonomic neuropathy may reduce counterregulatory hormone release (especially epinephrine), leading to an inability to sense hypoglycemia appropriately (hypoglycemia unawareness) (Chap. 406) and subjecting the patient to the risk of severe hypoglycemia. Gastroparesis and bladder-emptying abnormalities are often caused by the autonomic neuropathy seen in DM (discussed below). Hyperhidrosis of the upper extremities and anhidrosis of the lower extremities result from sympathetic nervous system dysfunction. Anhidrosis of the feet can promote dry skin with cracking, which increases the risk of foot ulcers.

## Mononeuropathy and/or Radiculopathy/Polyradiculopathy

Mononeuropathy (dysfunction of isolated cranial or peripheral nerves) is less common than polyneuropathy in DM and presents with pain

and motor weakness in the distribution of a single nerve. Mononeuropathies can occur at entrapment sites such as carpal tunnel or be noncompressive. Involvement of the third cranial nerve is most common and is heralded by diplopia. Physical examination reveals ptosis and ophthalmoplegia with normal pupillary constriction to light. Sometimes other cranial nerves, such as IV, VI, or VII (Bell's palsy), are affected. Peripheral mononeuropathies or simultaneous involvement of more than one nerve (mononeuropathy multiplex) may also occur. Diabetic radiculopathy or polyradiculopathy is a syndrome characterized by severe disabling pain in the distribution of one or more nerve roots. It may be accompanied by motor weakness. Intercostal or truncal radiculopathy causes pain over the thorax or abdomen. Involvement of the lumbar plexus or femoral nerve may cause severe pain in the thigh or hip and may be associated with muscle weakness in the hip flexors or extensors (diabetic amyotrophy). Fortunately, diabetic polyradiculopathies are usually self-limited and resolve over 6–12 months.

## TREATMENT

### Diabetic Neuropathy

Prevention of diabetic neuropathy is critical through improved glycemic control. Treatment of diabetic neuropathy is less than satisfactory. Lifestyle modifications (exercise, diet) have some efficacy in DSPN in type 2 DM and hypertension, and hypertriglyceridemia should be treated. Efforts to improve glycemic control in long-standing diabetes may be confounded by hypoglycemia unawareness. Patients should avoid neurotoxins (including alcohol) and smoking, and consider supplementation with vitamins for possible deficiencies ( $B_{12}$ , folate; [Chap. 333](#)). Metformin may reduce intestinal absorption of vitamin  $B_{12}$  in type 2 DM, and pernicious anemia is more common in type 1 DM where it is associated with anti-parietal cell autoantibodies and may require sublingual or parenteral  $B_{12}$  replacement. Patients should be educated that loss of sensation in the foot increases the risk for ulceration and its sequelae and that prevention of such problems is paramount. Patients with symptoms or signs of neuropathy or LOPS should check their feet daily and take precautions (footwear) aimed at preventing calluses or ulcerations. If foot deformities are present, a podiatrist should be involved.

Chronic, painful diabetic neuropathy is difficult to treat with only symptomatic treatment being available; evidence of the effectiveness of improved glycemic control in painful diabetic neuropathy is lacking. Two oral agents approved by the U.S. Food and Drug Administration (FDA), duloxetine and pregabalin, or gabapentin is usually initially used for pain associated with diabetic neuropathy. Diabetic neuropathy may respond to tricyclic antidepressants, venlafaxine, carbamazepine, tramadol, or topical capsaicin products. An 8% capsaicin patch requires application by a health care provider. Tapentadol, a centrally acting opioid, is also approved by the FDA, but has only modest efficacy and poses addiction risk, making it and other opioids less desirable and not a first-line therapy. No direct comparisons of agents are available, and it is reasonable to switch agents if there is no response or if side effects develop. Referral to a pain management center may be necessary.

Therapy of orthostatic hypotension secondary to autonomic neuropathy is also difficult. Nonpharmacologic maneuvers (adequate salt intake, avoidance of dehydration and diuretics, lower extremity support hose, and physical activity) may offer some benefit. A variety of agents have limited success (midodrine and droxidopa are approved by the FDA for orthostatic hypotension of any etiology). Patients with resting tachycardia may be considered for beta blocker therapy with caution exercised if there is hypoglycemia unawareness. Patients with type 1 DM and orthostatic hypotension should be evaluated for primary adrenal insufficiency (Addison's disease) that may be associated with anti-21-hydroxylase autoantibodies as part of an autoimmune polyendocrine syndrome ([Chap. 389](#)).

## GASTROINTESTINAL/GENITOURINARY DYSFUNCTION

Long-standing type 1 and 2 DM may affect the motility and function of the GI and genitourinary systems. The most prominent GI symptoms are delayed gastric emptying (gastroparesis) and altered small- and large-bowel motility (constipation or diarrhea). Gastroparesis may present with symptoms of anorexia, nausea, vomiting, early satiety, and abdominal bloating. Microvascular complications (retinopathy and neuropathy) are usually present. Nuclear medicine scintigraphy after ingestion of a radiolabeled meal may document delayed gastric emptying but may not correlate well with the patient's symptoms. Non-invasive "breath tests" following ingestion of a radiolabeled meal are emerging as a diagnostic tool. Although parasympathetic dysfunction secondary to chronic hyperglycemia is important in the development of gastroparesis, hyperglycemia itself also impairs gastric emptying. Nocturnal diarrhea, alternating with constipation, is a feature of DM-related GI autonomic neuropathy. In type 1 DM, these symptoms should also prompt evaluation for celiac disease that is associated with anti-tissue transglutaminase autoantibodies because of its increased frequency.

Diabetic autonomic neuropathy may lead to genitourinary dysfunction including cystopathy and female sexual dysfunction (reduced sexual desire, dyspareunia, reduced vaginal lubrication). Symptoms of diabetic cystopathy begin with an inability to sense a full bladder and a failure to void completely. As bladder contractility worsens, bladder capacity and the postvoid residual increase, leading to symptoms of urinary hesitancy, decreased voiding frequency, incontinence, and recurrent urinary tract infections.

Erectile dysfunction and retrograde ejaculation are very common in DM and may be one of the earliest signs of diabetic neuropathy ([Chap. 397](#)). Erectile dysfunction, which increases in frequency with the age of the patient and the duration of diabetes, may occur in the absence of other signs of diabetic autonomic neuropathy.

## TREATMENT

### Gastrointestinal/Genitourinary Dysfunction

Current treatments for these complications of DM are inadequate and nonspecific. Improved glycemic control should be a goal but has not clearly shown benefit. Smaller, more frequent meals that are easier to digest (liquid) and low in fat and fiber may minimize symptoms of gastroparesis. Medications that slow gastric emptying (opioids, GLP-1 receptor agonists) should be avoided. Metoclopramide may be used with severe symptoms but is restricted to short-term treatment in both the United States and Europe. Symptoms of gastroesophageal reflux disease may require acid blocking therapy with a histamine-2 receptor antagonist or proton pump inhibitor. Gastric electrical stimulatory devices are available but not approved. Diabetic diarrhea in the absence of bacterial overgrowth is treated symptomatically ([Chap. 325](#)).

Diabetic cystopathy should be treated with scheduled voiding or self-catheterization. Drugs that inhibit type 5 phosphodiesterase are effective for erectile dysfunction, but their efficacy in individuals with DM is slightly lower than in the nondiabetic population ([Chap. 397](#)).

## CARDIOVASCULAR MORBIDITY AND MORTALITY

CVD is increased in individuals with type 1 or type 2 DM. The Framingham Heart Study revealed a marked increase in PAD, coronary artery disease, MI, and CHF (risk increase from one- to fivefold) in DM. In addition, the prognosis for individuals with diabetes who have coronary artery disease or MI is worse than for nondiabetics. CHD is more likely to involve multiple vessels in individuals with DM. In addition to CHD, cerebrovascular disease is increased in individuals with DM (threefold increase in stroke). Thus, after controlling for all known cardiovascular risk factors, both type 1 and type 2 DM increases the cardiovascular death rate twofold in men and fourfold in women. CHF is common in long-standing DM.

The American Heart Association considers DM as a controllable risk factor for cardiovascular disease; in some studies, type 2 DM patients without a prior MI have a similar risk for coronary artery-related events as nondiabetic individuals who have had a prior MI. Cardiovascular risk assessment in type 2 DM should encompass a more nuanced approach. Cardiovascular risk is lower and not equivalent in a younger individual with a brief duration of type 2 DM compared to an older individual with long-standing type 2 DM. In individuals without a known diagnosis of diabetes, elevated HbA<sub>1c</sub> is predictive not just of diabetes risk but also risk of CHD, stroke, and all-cause mortality. Because of the extremely high prevalence of underlying CVD in individuals with diabetes (especially in type 2 DM), evidence of atherosclerotic vascular disease (e.g., cardiac stress test) should be sought in an individual with diabetes who has symptoms, even if atypical, suggestive of cardiac ischemia or peripheral or carotid arterial disease. The screening of asymptomatic individuals with diabetes for CHD is not recommended or cost-effective. The absence of chest pain ("silent ischemia") is common in individuals with diabetes, and a thorough cardiac evaluation should be considered prior to major surgical procedures.

The increase in cardiovascular morbidity and mortality rates in diabetes appears to relate to the synergy of hyperglycemia with other cardiovascular risk factors such as dyslipidemia (elevated triglycerides, low high-density lipoprotein [HDL] cholesterol and small dense low-density lipoprotein [LDL]), hypertension, obesity, reduced physical activity, and cigarette smoking. Additional risk factors that are prevalent include CKD (albuminuria, reduced GFR), abnormal platelet function, increased markers of inflammation, and endothelial dysfunction. The results of the ACCORD trial and VADT trial, which demonstrated that tight glucose control had limited benefit on cardiovascular outcomes in individuals with established cardiovascular disease, suggesting the importance of insulin resistance and dyslipidemia.

## TREATMENT

### Cardiovascular Disease

Treatment of coronary disease in individuals with DM has substantial overlap with treatment in individuals without DM ([Chap. 273](#)). Revascularization procedures for CHD, including percutaneous coronary interventions (PCIs) and coronary artery bypass grafting (CABG), may be less efficacious in individuals with DM. Initial success rates of PCI in individuals with DM are similar to those in the nondiabetic population, but higher rates of restenosis and lower long-term patency and survival rates have been reported. CABG plus optimal medical management likely has better outcomes than PCI for individuals with diabetes.

Aggressive cardiovascular risk modification in all individuals with DM and glycemic control should be individualized, as discussed in [Chap. 404](#). In patients with known CHD and type 2 DM, an ACE inhibitor or ARB, a statin, and acetylsalicylic acid (ASA; aspirin) should be considered. Beta blockers can be used in individuals with diabetes after MI. In patients with CHF, thiazolidinediones should not be used ([Chap. 404](#)). However, metformin can be used in patients with stable CHF if the renal function is normal. Some newer glucose-lowering therapies also have cardiovascular benefit, including the GLP-1 analogues liraglutide (LEADER trial), semaglutide (SUSTAIN-6 trial), and dulaglutide (REWIND trial) and the SGLT-2 inhibitors empagliflozin (EMPA-REG trial) and canagliflozin (CANVAS trial). All SGLT-2 inhibitors have been shown to exhibit benefits on prevention of CHF exacerbations. A possible increased risk of lower limb amputation and Fournier's gangrene has been reported with SGLT-2 inhibitor therapy.

Antiplatelet therapy reduces cardiovascular events in individuals with DM who have CHD and is recommended. The ADA recommends considering the use of aspirin for primary prevention of coronary events in individuals with diabetes with an increased cardiovascular risk (>50 years old with at least one risk factor

such as hypertension, dyslipidemia, smoking, family history, or albuminuria). ASA is not recommended for primary prevention in those with a low cardiovascular risk (<50 years old with no risk factors). The aspirin dose is the same as in nondiabetic individuals.

**Cardiovascular Risk Factors • DYSLIPIDEMIA** Individuals with DM may have several forms of dyslipidemia ([Chap. 407](#)). Because of the additive cardiovascular risk of hyperglycemia and hyperlipidemia, lipid abnormalities should be assessed aggressively and treated as part of comprehensive diabetes care ([Chap. 404](#)). The most common pattern of dyslipidemia is hypertriglyceridemia and reduced HDL cholesterol levels. DM itself does not increase levels of LDL, but the small dense LDL particles found in type 2 DM are more atherogenic because they are more easily glycated and susceptible to oxidation.

Almost all treatment studies of diabetic dyslipidemia have been performed in individuals with type 2 DM because of the greater frequency of dyslipidemia in this form of diabetes. Interventional studies have shown that the beneficial effects of LDL reduction with statins are similar in the diabetic and nondiabetic populations. No prospective studies have addressed similar questions in individuals with type 1 DM. Because the frequency of CVD is low in children and young adults with diabetes, assessment of cardiovascular risk should be incorporated into the guidelines discussed below.

Based on the guidelines provided by the ADA, all individuals with diabetes should be advised about lifestyle modification, including diet, weight loss, and increased physical activity ([Chap. 404](#)). If individuals with diabetes have elevated triglyceride levels (>1.7 mmol/L [150 mg/dL]) or low HDL cholesterol (<1 mmol/L [40 mg/dL] in men and <1.3 mmol/L [50 mg/dL] in women), lifestyle modification and improved glycemic control should be further emphasized. If triglycerides remain >5.7 mmol/L (500 mg/dL), treatment with fish oil, fibrate drugs, and icosapent may reduce the risk of pancreatitis. Icosapent additionally lowers CHD risk.

In terms of pharmacologic therapy, the ADA recommends the following: (1) all patients with diabetes and atherosclerotic cardiovascular disease should receive high-intensity statin therapy; (2) in patients aged 40–75 years without cardiovascular disease, consider moderate-intensity statin therapy to target LDL cholesterol <100 mg/dL (without additional risk factors) or high-intensity statin therapy to target LDL cholesterol <70 mg/dL (with additional risk factors); and (3) in patients aged 20–39 years with additional risk factors, consider moderate-intensity statin therapy. Screening for coronary artery calcification by electron beam computed tomography (CT) scan that noninvasively detects the presence of coronary artery atherosclerosis may help guide treatment initiation or intensity in equivocal cases or ambivalent patients. Atorvastatin and rosuvastatin are generally well tolerated if started at lower doses and titrated up to meet lipid goals. Atorvastatin is the statin of choice in patients with renal disease. If statin intolerant or the LDL cholesterol goal is not met, consider the addition of ezetimibe or a PCSK9 inhibitor ([Chap. 407](#)). Icosapent results in cardiovascular risk reduction on top of statin treatment and may have a larger benefit in diabetic individuals. Statin usage is associated with a mild increase in the risk of developing type 2 DM. This risk is greatest in individuals with other risk factors for type 2 DM ([Chap. 403](#)). However, the cardiovascular benefits of statin use outweigh the mildly increased risk of diabetes. Niacin use is associated with an even greater increased risk for type 2 DM or worsening glycemic control and is not recommended because of a lack of improvement in cardiovascular outcomes.

In individuals with type 2 DM and kidney disease or type 2 DM and atherosclerotic cardiovascular disease or multiple atherosclerotic risk factors, the ADA recommends an SGLT-2 inhibitor or GLP-1 receptor agonist as a second-line agent after metformin. In individuals with type 2 DM and heart failure (reduced ejection fraction), the ADA recommends an SGLT-2 inhibitor. Individuals with atherosclerotic

cardiovascular disease and type 1 or type 2 DM should be treated with an ACE inhibitor or angiotensin receptor blocker and beta blockers and antiplatelet therapy, as in the nondiabetic population ([Chap. 273](#)).

**HYPERTENSION** Hypertension can accelerate other complications of DM, particularly CVD, nephropathy, and retinopathy. Blood pressure should be measured at every clinic visit. In targeting a goal of blood pressure of <140/90 mmHg, therapy should first emphasize lifestyle modifications such as weight loss, exercise, stress management, and sodium restriction. The blood pressure goal should be individualized. In some younger individuals or those with increased cardiovascular risk, the provider may target a blood pressure of <130/80 mmHg. Realizing that more than one agent is usually required to reach the blood pressure goal, the ADA recommends that all patients with diabetes and hypertension be treated with an ACE inhibitor or an ARB initially. Subsequently, agents that reduce cardiovascular risk (beta blockers, thiazide diuretics, and calcium channel blockers) should be incorporated into the regimen. ACE inhibitors and ARBs are likely equivalent in most patients with diabetes and renal disease but should not be combined. The addition of a potassium-sparing diuretic or mineralocorticoid receptor antagonist can help achieve blood pressure targets in refractory cases. Serum potassium and renal function should be monitored.

Because of the high prevalence of atherosclerotic disease in individuals with type 2 DM, the possibility of renovascular hypertension should be considered when the blood pressure is not readily controlled.

### LOWER EXTREMITY COMPLICATIONS

DM is the leading cause of nontraumatic lower extremity amputation in the United States. Foot ulcers and infections are also a major source of morbidity in individuals with DM. The reasons for the increased incidence of these disorders in DM involve the interaction of several pathogenic factors: neuropathy, abnormal foot biomechanics, PAD, and poor wound healing. The peripheral sensory neuropathy interferes with normal protective mechanisms and allows the patient to sustain major or repeated minor trauma to the foot, often without knowledge of the injury. Disordered proprioception causes abnormal weight bearing while walking and subsequent formation of callus or ulceration. Motor and sensory neuropathy lead to abnormal foot muscle mechanics and to structural changes in the foot (hammer toe, claw toe deformity, prominent metatarsal heads, Charcot joint). Autonomic neuropathy results in anhidrosis and altered superficial blood flow in the foot, which promote drying of the skin and fissure formation. PAD and poor wound healing impede resolution of minor breaks in the skin, allowing them to enlarge and to become infected.

Many individuals with DM develop a foot ulcer (great toe or metatarsophalangeal areas are most common), and a significant subset who develop an ulceration will ultimately undergo amputation (14–24% risk with that ulcer or subsequent ulceration). Risk factors for foot ulcers or amputation include male sex, diabetes for >10 years, peripheral neuropathy, abnormal structure of foot (bony abnormalities, callus, thickened nails), PAD, smoking, history of previous ulcer or amputation, visual impairment, poor glycemic control, and diabetic nephropathy, especially dialysis. Large calluses are often precursors to or overlie ulcerations. Aggressive treatment of LDL cholesterol with the PCSK9 inhibitor evolocumab has been shown to reduce the risk of future major adverse limb events in patients with PAD.

### TREATMENT

#### Lower Extremity Complications

The optimal therapy for foot ulcers and amputations is prevention through identification of high-risk patients, education of the patient, and institution of measures to prevent ulceration. High-risk patients should be identified during the routine, annual foot examination performed on all patients with DM (see “Ongoing Aspects of Comprehensive Diabetes Care” in [Chap. 404](#)). If the monofilament

test or one of the other tests is abnormal, the patient is diagnosed with LOPS ([Chap. 403](#)). Providers should consider screening for asymptomatic PAD in individuals >50 years of age who have diabetes and other risk factors using ankle-brachial index testing ([Chap. 281](#)). Patient education should emphasize (1) careful selection of footwear, (2) daily inspection of the feet to detect early signs of poor-fitting footwear or minor trauma, (3) daily foot hygiene to keep the skin clean and moist, (4) avoidance of self-treatment of foot abnormalities and high-risk behavior (e.g., walking barefoot), and (5) prompt consultation with a health care provider if an abnormality arises. Calluses and nail deformities should be treated by a podiatrist. Interventions directed at risk factor modification include orthotic shoes and devices, callus management, nail care, and prophylactic measures to reduce increased skin pressure from abnormal bony architecture. Attention to other risk factors for vascular disease (smoking, dyslipidemia, hypertension) and improved glycemic control are also important.

Despite preventive measures, foot ulceration and infection are common and represent a serious problem. Due to the multifactorial pathogenesis of lower extremity ulcers, management of these lesions is multidisciplinary and often demands expertise in orthopedics, vascular surgery, endocrinology, podiatry, and infectious diseases. The plantar surface of the foot is the most common site of ulceration. Ulcers may be primarily neuropathic (no accompanying infection) or may have surrounding cellulitis or osteomyelitis. Cellulitis without ulceration should be treated with antibiotics that provide broad-spectrum coverage (see below).

An infected ulcer is a clinical diagnosis, because superficial culture of any ulceration will likely find multiple bacterial species of unknown significance. The infection surrounding the foot ulcer may be due to multiple organisms, with aerobic gram-positive cocci (*staphylococci* including methicillin-resistant *Staphylococcus aureus* [MRSA], group A and B streptococci) being most common and with aerobic gram-negative bacilli and/or obligate anaerobes as co-pathogens.

Gas gangrene may develop in the absence of clostridial infection. Cultures should be obtained from the debrided ulcer base or from purulent drainage or aspiration of the wound. Wound depth should be determined by inspection and probing with a blunt-tipped sterile instrument. A wound that probes to the bone represents clinical evidence of osteomyelitis. Plain radiographs of the foot should be performed to assess the possibility of osteomyelitis in chronic ulcers that have not responded to therapy. Magnetic resonance imaging (MRI) is the most specific modality, with nuclear medicine scans and labeled white cell studies as alternatives. Surgical debridement is often necessary.

Osteomyelitis is best treated by a combination of prolonged antibiotics and debridement of infected bone when possible. The possible contribution of vascular insufficiency should be considered in all patients. Peripheral arterial bypass procedures are often effective in promoting wound healing and in decreasing the need for amputation of the ischemic limb ([Chap. 281](#)).

Interventions with demonstrated efficacy in diabetic foot ulcers or wounds include the following: (1) off-loading, (2) debridement, (3) wound dressings, (4) appropriate use of antibiotics, (5) revascularization, and (6) limited amputation. Off-loading is the complete avoidance of weight bearing on the ulcer, which removes the mechanical trauma that retards wound healing. Bed rest and a variety of orthotic devices or contact casting limit weight bearing on wounds or pressure points. Surgical debridement is important and effective, but the efficacy of other modalities for wound healing (enzymes, growth factors, cellular therapy, hyperbaric oxygen) is unclear. Dressings such as hydrocolloid dressings promote wound healing by creating a moist environment, controlling the exudate, and protecting the wound. Antiseptic agents should be avoided. Topical antibiotics are of limited value. Referral for physical therapy, orthotic evaluation, and rehabilitation should occur once the infection is controlled.

Mild or non-limb-threatening infections can be treated with oral antibiotics directed predominantly at methicillin-susceptible staphylococci and streptococci (e.g., dicloxacillin, cephalosporin, amoxicillin/clavulanate). However, in patients with a prior history of MRSA or in locations with a high prevalence of MRSA, treatment with clindamycin, doxycycline, or trimethoprim-sulfamethoxazole is preferred. Trimethoprim-sulfamethoxazole exhibits less reliable coverage of streptococci than the  $\beta$ -lactams, and individuals with diabetes may develop adverse effects including acute kidney injury and hyperkalemia. Surgical debridement of necrotic tissue, local wound care (avoidance of weight bearing over the ulcer), and close surveillance for progression of infection are crucial. More severe infections require IV antibiotics as well as offloading and local wound care. Urgent surgical debridement may be required. Optimization of glycemic control should be a goal. IV antibiotics should provide broad-spectrum coverage directed toward *S. aureus*, including MRSA, streptococci, gram-negative aerobes, and anaerobic bacteria. Initial antimicrobial regimens include vancomycin plus a  $\beta$ -lactam/ $\beta$ -lactamase inhibitor or carbapenem, or vancomycin plus a quinolone with metronidazole. In some cases, daptomycin, ceftaroline, or linezolid may be substituted for vancomycin in consultation with an Infectious Diseases expert. If the infection surrounding the ulcer is not improving with IV antibiotics, reassessment of antibiotic coverage and reconsideration of the need for surgical debridement or revascularization are indicated. With clinical improvement, oral antibiotics and local wound care can be continued on an outpatient basis with close follow-up.

## INFECTIONS

Individuals with DM have a greater frequency and severity of infection. The reasons for this include incompletely defined abnormalities in cell-mediated immunity and phagocyte function associated with hyperglycemia, as well as diminished vascularization. Hyperglycemia aids the colonization and growth of a variety of organisms (*Candida* and other fungal species). Many common infections are more frequent and severe in the diabetic population, whereas several rare infections are seen almost exclusively in the diabetic population. Examples of this latter category include rhinocerebral mucormycosis, emphysematous infections of the gallbladder and urinary tract, and "malignant" or invasive otitis externa. Invasive otitis externa is usually secondary to *Pseudomonas aeruginosa* infection in the soft tissue surrounding the external auditory canal, usually begins with pain and discharge, and may rapidly progress to osteomyelitis and meningitis. These infections should be sought, in particular, in patients presenting with severe hyperglycemia (Chap. 404).

Pneumonia, urinary tract infections, and skin and soft tissue infections are all more common in the diabetic population. In general, the organisms that cause pulmonary infections are similar to those found in the nondiabetic population; however, gram-negative organisms, *S. aureus*, and *Mycobacterium tuberculosis* are more frequent pathogens. Adults with DM should receive vaccination against pneumococcus, annually against influenza, and now also against the coronavirus SARS-CoV-2, which causes increased morbidity and mortality in obese individuals and patients with DM (Chap. 199). In addition to early antibiotic therapy for presumed bacterial infections, patients with DM should be considered for early intervention with antiviral agents (e.g., against influenza in flu or varicella-zoster virus in shingles) or with monoclonal antibodies in COVID-19. Urinary tract infections (either lower tract or pyelonephritis) are the result of common bacterial agents such as *Escherichia coli*, although several yeast species (e.g., *Candida albicans* and *Candida glabrata*) are commonly observed. Complications of urinary tract infections include emphysematous pyelonephritis and emphysematous cystitis. Bacteruria occurs frequently

in individuals with diabetic cystopathy and does not require antibiotic therapy except in specific circumstances such as pregnancy or a planned urologic procedure. Susceptibility to furunculosis, superficial candidal infections, and vulvovaginitis are increased. Poor glycemic control is a common denominator in individuals with these infections. Individuals with diabetes have an increased rate of colonization of *S. aureus* in the skinfolds and nares. Individuals with diabetes also have a greater risk of postoperative wound infections that may be mitigated by perioperative protocols for insulin administration to maintain glycemic control.

## DERMATOLOGIC MANIFESTATIONS

The most common skin manifestations of DM are xerosis and pruritus and are usually relieved by skin moisturizers. Protracted wound healing and skin ulcerations are also frequent complications. Diabetic dermopathy, sometimes termed *pigmented pretibial papules*, or "diabetic skin spots," begins as an erythematous macule or papule that evolves into an area of circular hyperpigmentation. These lesions result from minor mechanical trauma in the pretibial region and are more common in elderly men with DM. Bullous diseases, such as bullous diabeticorum (shallow ulcerations or erosions in the pretibial region), are also seen. *Necrobiosis lipoidica diabetorum* is an uncommon disorder, accompanying diabetes in predominantly young women. This usually begins in the pretibial region as an erythematous plaque or papules that gradually enlarge, darken, and develop irregular margins, with atrophic centers and central ulceration. They are often painful. Vitiligo and alopecia areata occur at increased frequency in individuals with type 1 DM. *Acanthosis nigricans* (hyperpigmented velvety plaques seen on the neck, axilla, or extensor surfaces) is sometimes a feature of severe insulin resistance and accompanying diabetes. Generalized or localized *granuloma annulare* (erythematous plaques on the extremities or trunk), *lichen planus* (violaceous papules on the cutaneous surface with or without erosions in the mouth and genitalia), and *scleredema* (areas of skin thickening on the back or neck at the site of previous superficial infections) are more common in the diabetic population. *Lipoatrophy* and *lipohypertrophy* can occur at insulin injection sites but are now unusual with the use of human insulin and avoided by rotating injection sites.

## FURTHER READING

- American Diabetes Association: Cardiovascular disease and risk management: Standards of Medical Care in Diabetes-2021. *Diabetes Care* 44(suppl 1):S125, 2021.
- American Diabetes Association: Microvascular complications and foot care: Standards of Medical Care in Diabetes-2021. *Diabetes Care* 44(suppl 1):S151, 2021.
- Drucker DJ: Diabetes, obesity, metabolism, and SARS-CoV-2 infection: The end of the beginning. *Cell Metab* 33:479, 2021.
- Hingorani A et al: The management of diabetic foot: A clinical practice guideline by the Society for Vascular Surgery in collaboration with the American Podiatric Medical Association and the Society for Vascular Medicine. *J Vasc Surg* 63:3S, 2016.
- Holman RR et al: 10-year follow-up of intensive glucose control in type 2 diabetes. *N Engl J Med* 359:1577, 2008.
- Nathan DM et al: Diabetes control and complications trial/epidemiology of diabetes interventions and complications study at 30 years: Advances and contributions. *Diabetes* 62:3976, 2013.
- Pop-Busui R et al: Diabetic neuropathy: A position statement by the American Diabetes Association. *Diabetes Care* 40:136, 2017.
- Solomon SD et al: Diabetic retinopathy: A position statement by the American Diabetes Association. *Diabetes Care* 40:412, 2017.
- Williams DM et al: Renal outcomes in type 2 diabetes: A review of cardiovascular and renal outcome trials. *Diabetes Ther* 11:369, 2020.



Hypoglycemia is most commonly caused by insulin or insulin-producing drugs used to treat diabetes mellitus or by exposure to other drugs, including alcohol. However, a number of other disorders, including critical organ failure, sepsis and inanition, hormone deficiencies, non-cell tumors, insulinoma, and prior gastric surgery, can cause hypoglycemia (**Table 406-1**). Hypoglycemia may be documented by *Whipple's triad*: (1) symptoms consistent with hypoglycemia, (2) a low plasma glucose concentration measured with a precise method, and (3) relief of symptoms after the plasma glucose level is raised. The lower limit of the fasting plasma glucose concentration is normally ~70 mg/dL (~3.9 mmol/L), but lower venous glucose levels occur normally, late after a meal, during pregnancy, and during prolonged fasting (>24 h). Severe hypoglycemia can cause serious morbidity and increase the risk for serious cardiovascular events and mortality during and after the initial hypoglycemic episode. It should be considered in any patient with episodes of confusion, an altered level of consciousness, or a seizure.

### SYSTEMIC GLUCOSE BALANCE AND GLUCOSE COUNTERREGULATION

Glucose is an obligate metabolic fuel for the brain under physiologic conditions. The brain cannot synthesize glucose or store more than a few minutes' supply as glycogen and therefore requires a continuous supply of glucose from the arterial circulation. As the arterial plasma glucose concentration falls below the physiologic range, blood-to-brain glucose transport becomes insufficient to support brain energy metabolism and function. However, multiple integrated glucose

counterregulatory mechanisms normally prevent or rapidly correct hypoglycemia.

Plasma glucose concentrations are normally maintained within a relatively narrow range—roughly 70–110 mg/dL (3.9–6.1 mmol/L) in the fasting state, with transient higher excursions after a meal—despite wide variations in exogenous glucose delivery from meals and in endogenous glucose utilization by, for example, exercising muscle. Between meals and during fasting, plasma glucose levels are maintained by endogenous glucose production, hepatic glycogenolysis, and hepatic (and renal) gluconeogenesis (**Fig. 406-1**). Although hepatic glycogen stores are usually sufficient to maintain plasma glucose levels for ~8 h, this period can be shorter if glucose demand is increased by exercise or if glycogen stores are depleted by illness or starvation.

Gluconeogenesis normally requires low insulin levels and the presence of anti-insulin (counterregulatory) hormones together with a coordinated supply of precursors from muscle and adipose tissue to the liver and kidneys. Muscle provides lactate, pyruvate, alanine, glutamine, and other amino acids. Triglycerides in adipose tissue are broken down into fatty acids and glycerol, which is a gluconeogenic precursor. Fatty acids provide an alternative oxidative fuel to tissues other than the brain (which requires glucose).

Systemic glucose balance, maintenance of the normal plasma glucose concentration, is accomplished by a network of hormones, neural signals, and substrate effects that regulate endogenous glucose production and glucose utilization by tissues other than the brain (**Chap. 403**). Among the regulatory factors, insulin plays a dominant role (**Table 406-2**; **Fig. 406-1**). As plasma glucose levels decline within the physiologic range, pancreatic  $\beta$ -cell insulin secretion decreases, thereby increasing hepatic glycogenolysis and hepatic (and renal) gluconeogenesis. Low insulin levels also reduce glucose utilization in peripheral tissues, inducing lipolysis and proteolysis and consequently releasing gluconeogenic precursors. Thus, a decrease in insulin secretion is the first defense against hypoglycemia.

As plasma glucose levels decline just below the physiologic range, glucose counterregulatory (plasma glucose-raising) hormones are released (**Table 406-2**; **Fig. 406-1**). Among these, pancreatic  $\alpha$ -cell glucagon and adrenomedullary epinephrine play a primary role. Glucagon stimulates hepatic glycogenolysis and gluconeogenesis. Adrenomedullary epinephrine also stimulates hepatic glycogenolysis and gluconeogenesis (and renal gluconeogenesis) but limits peripheral uptake of glucose and stimulates lipolysis with production of glycerol and fatty acids. Epinephrine becomes critical when glucagon is deficient. When hypoglycemia is prolonged beyond ~4 h, cortisol and growth hormone also support glucose production and restrict glucose utilization to a limited amount (both mechanisms are reduced by ~80% compared to epinephrine). Thus, cortisol and growth hormone play no role in defense against acute hypoglycemia.

As plasma glucose levels fall further, symptoms prompt behavioral defense against hypoglycemia, including the ingestion of food (**Table 406-2**; **Fig. 406-1**). The normal glycemic thresholds for these responses to decreasing plasma glucose concentrations are shown in **Table 406-2**. However, these thresholds are dynamic. They shift to higher-than-normal glucose levels in people with poorly controlled diabetes, who can experience symptoms of hypoglycemia when their glucose levels decline toward the normal range. On the other hand, thresholds shift to lower-than-normal glucose levels in people with recurrent hypoglycemia; i.e., patients with intensively treated diabetes or an insulinoma have symptoms at glucose levels lower than those that cause symptoms in healthy individuals.

**Clinical Manifestations** Neuroglycopenic manifestations of hypoglycemia are the direct result of central nervous system glucose deprivation. These features include behavioral changes, confusion, fatigue, seizure, loss of consciousness, cardiac arrhythmias, and, if hypoglycemia is severe, death. Neurogenic (or autonomic) manifestations of hypoglycemia result from the perception of physiologic changes caused by the central nervous system-mediated sympathoadrenal discharge that is triggered by hypoglycemia. They include *adrenergic* symptoms (mediated largely by norepinephrine released from

**TABLE 406-1 Causes of Hypoglycemia in Adults**

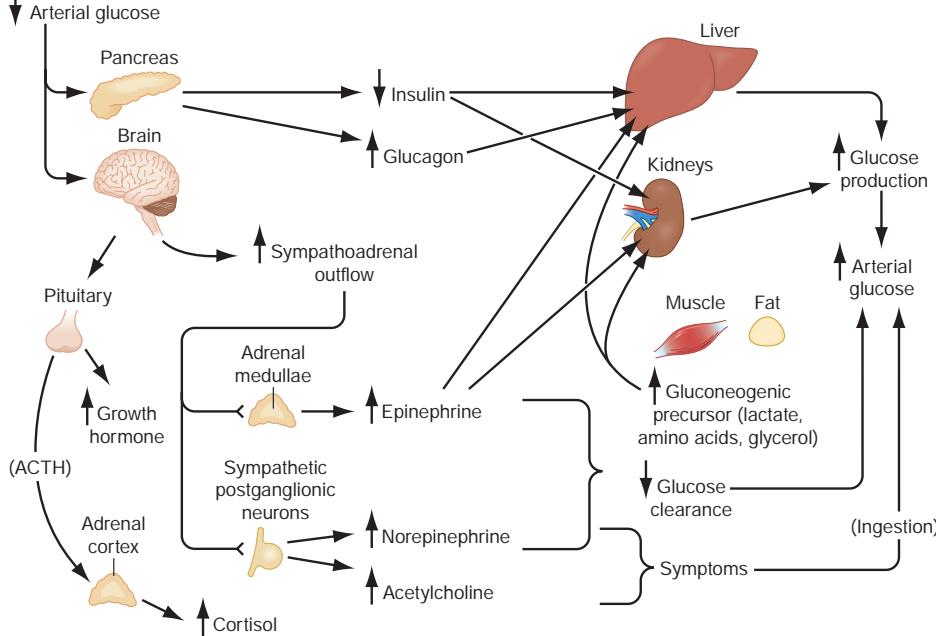
#### III or Medicated Individual

1. Drugs
  - Insulin or insulin secretagogues
  - Alcohol
  - Others
2. Critical illness
  - Hepatic, renal, or cardiac failure
  - Sepsis
  - Inanition
3. Hormone deficiency
  - Cortisol
  - Growth hormone
  - Glucagon and epinephrine (in insulin-deficient diabetes)
4. Non-islet cell tumor (e.g., mesenchymal tumors)

#### Seemingly Well Individual

5. Endogenous hyperinsulinism
  - Insulinoma
  - Functional  $\beta$ -cell disorders (nesidioblastosis)
    - Noninsulinoma pancreateogenous hypoglycemia
    - Post-gastric bypass hypoglycemia
  - Insulin autoimmune hypoglycemia
    - Antibody to insulin
    - Antibody to insulin receptor
  - Insulin secretagogues
  - Other
6. Disorders of gluconeogenesis and fatty acid oxidation
7. Exercise
8. Accidental, surreptitious, or malicious hypoglycemia

Source: Modified with permission from PE Cryer et al: Evaluation and management of adult hypoglycemic disorders: An Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 94:709, 2009.



**FIGURE 406-1** Physiology of glucose counterregulation: Mechanisms that normally prevent or rapidly correct hypoglycemia. In insulin-deficient diabetes, the key counterregulatory responses—suppression of insulin and increases in glucagon—are lost, and stimulation of sympathoadrenal outflow is attenuated. ACTH, adrenocorticotropic hormone.

sympathetic postganglionic neurons but perhaps also by epinephrine released from the adrenal medullae), such as palpitations, tremor, and anxiety, as well as *cholinergic* symptoms (mediated by acetylcholine released from sympathetic postganglionic neurons), such as sweating, hunger, and paresthesias. Clearly, these are nonspecific symptoms. Their attribution to hypoglycemia requires that the corresponding plasma glucose concentration be low and that the symptoms resolve after the glucose level is raised (as delineated by Whipple's triad).

Common signs of hypoglycemia include diaphoresis and pallor. Heart rate and systolic blood pressure are typically increased but may not be raised in an individual who has experienced repeated, recent episodes of hypoglycemia. Neuroglycopenic manifestations are often observable. Transient focal neurologic deficits occur occasionally. Permanent neurologic deficits are rare.

**Etiology and Pathophysiology** Hypoglycemia activates proinflammatory, procoagulant, and proatherothrombotic responses in type 1 diabetes mellitus (T1DM), type 2 diabetes mellitus (T2DM),

and nondiabetic individuals. These responses increase platelet aggregation, reduce fibrinolytic balance (increase plasminogen activator inhibitor-1), and increase intravascular coagulation. Hypoglycemia also reduces protective nitric oxide-mediated arterial vasodilator mechanisms in healthy, T1DM, and T2DM individuals.

## HYPOGLYCEMIA IN DIABETES

**Impact and Frequency** Hypoglycemia is the limiting factor in the glycemic management of diabetes mellitus. First, it causes recurrent morbidity in most people with T1DM and in many with advanced T2DM, and it is sometimes fatal. Second, it precludes maintenance of euglycemia over a lifetime of diabetes and, thus, full realization of the well-established microvascular benefits of glycemic control. Third, it causes a vicious cycle of recurrent hypoglycemia by producing hypoglycemia-associated autonomic failure—i.e., the clinical syndromes of defective glucose counterregulation and of hypoglycemia unawareness.

**TABLE 406-2** Physiologic Responses to Decreasing Plasma Glucose Concentrations

RESPONSE	GLYCEMIC THRESHOLD, mmol/L (mg/dL)	PHYSIOLOGIC ↓ EFFECTS	ROLE IN PREVENTION OR CORRECTION OF HYPOGLYCEMIA (GLUCOSE COUNTERREGULATION)
↓ Insulin	4.4–4.7 (80–85)	↑ R <sub>a</sub> (↓ R <sub>d</sub> ), increased lipolysis; ↑ FFA; ↑ Glycerol	Primary glucose regulatory factor/first defense against hypoglycemia
↑ Glucagon	3.6–3.9 (65–70)	↑ R <sub>a</sub>	Primary glucose counterregulatory factor/second defense against hypoglycemia
↑ Epinephrine	3.6–3.9 (65–70)	↑ R <sub>a</sub> ↓ R <sub>d</sub> , increased lipolysis; ↑ FFA and glycerol	Third defense against hypoglycemia; critical when glucagon is deficient
↑ Cortisol and growth hormone	3.6–3.9 (65–70)	↑ R <sub>a</sub> , ↓ R <sub>d</sub>	Involved in defense against prolonged hypoglycemia; not critical
Symptoms	2.8–3.1 (50–55)	Recognition of hypoglycemia	Prompt behavioral defense against hypoglycemia (food ingestion)
↓ Cognition	<2.8 (<50)	—	Compromises behavioral defense against hypoglycemia

*Note:* R<sub>a</sub>, rate of glucose appearance, glucose production by the liver and kidneys; R<sub>d</sub>, rate of glucose disappearance, glucose utilization relative to the ambient plasma glucose by insulin-sensitive tissues; R<sub>d</sub>, rate of glucose disappearance, glucose utilization by insulin-sensitive tissues such as skeletal muscle. R<sub>d</sub> by the brain is not altered by insulin, glucagon, epinephrine, cortisol, or growth hormone.

*Abbreviation:* FFA, free fatty acids.

*Source:* Reproduced with permission from PE Cryer, in S Melmed et al: *Williams Textbook of Endocrinology*, 12th ed. New York, NY: Elsevier; 2012.

Hypoglycemia is a fact of life for people with T1DM if treated with insulin, sulfonylurea, or glinides. They suffer an average of two episodes of symptomatic hypoglycemia per week and at least one episode of severe, at least temporarily disabling hypoglycemia each year. An estimated 6–10% of people with T1DM die as a result of hypoglycemia. The incidence of hypoglycemia is lower in T2DM than in T1DM. However, its prevalence in insulin-requiring T2DM is surprisingly high. Recent studies have revealed a hypoglycemia prevalence approaching 70%. In fact, as patients with T2DM outnumber those with T1DM by ten- to twentyfold, the prevalence of hypoglycemia is now greater in T2DM. Hypoglycemia can occur at any hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>) level. Although severe hypoglycemia occurs twice as frequently at lower HbA<sub>1c</sub> levels in T1DM, it still occurs at HbA<sub>1c</sub> levels >8%. In insulin-requiring T2DM, severe hypoglycemia can occur at lower HbA<sub>1c</sub> values but also importantly at values of 8–10%. Severe hypoglycemia in T2DM carries an increased risk of severe cardiovascular and cerebrovascular morbidity and mortality for up to 1 year after the event. The risk of severe hypoglycemia and a subsequent cardiovascular adverse event is, in fact, relatively increased when trying to improve glucose control in some T2DM individuals with persistently raised HbA<sub>1c</sub> values. Therefore, improvements in glycemic control in these individuals should be performed incrementally and carefully to avoid episodes of hypoglycemia. Insulin, sulfonylureas, or glinides can cause hypoglycemia in T2DM. Metformin, thiazolidinediones, -glucosidase inhibitors, glucagon-like peptide 1 (GLP-1) receptor agonists, sodium-glucose cotransporter-2 inhibitors, and dipeptidyl peptidase IV (DPP-IV) inhibitors do not cause hypoglycemia. However, they increase the risk when combined with a sulfonylurea, glinide, or insulin. Notably, the frequency of hypoglycemia approaches that in T1DM as persons with T2DM develop absolute insulin deficiency and require more complex treatment with insulin.

**Conventional Risk Factors** The conventional risk factors for hypoglycemia in diabetes are identified on the basis of relative or absolute insulin excess. This occurs when (1) insulin (or insulin secretagogue) doses are excessive, ill-timed, or of the wrong type; (2) the influx of exogenous glucose is reduced (e.g., during an overnight fast, periods of temporary fasting, or after missed meals or snacks); (3) insulin-independent glucose utilization is increased (e.g., during exercise); (4) sensitivity to insulin is increased (e.g., with improved glycemic control, in the middle of the night, after exercise, or with increased fitness or weight loss); (5) endogenous glucose production is reduced (e.g., after alcohol ingestion); and (6) insulin clearance is reduced (e.g., in renal failure). However, these conventional risk factors alone explain a minority of episodes; other factors are typically involved.

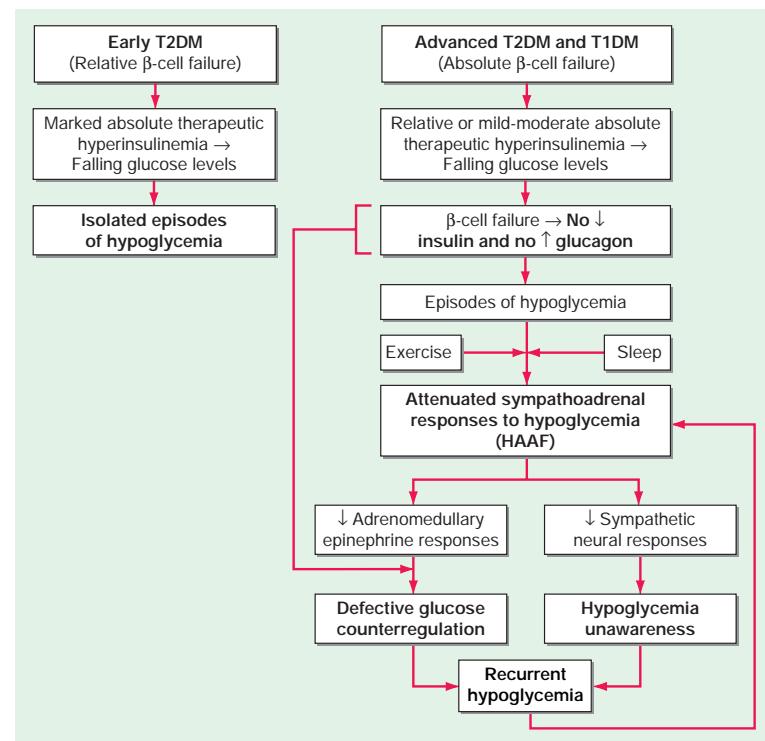
**Hypoglycemia-Associated Autonomic Failure (HAAF)** While marked insulin excess alone can cause hypoglycemia, iatrogenic hypoglycemia in diabetes (T1DM and/or T2DM) is typically the result of the interplay of relative or absolute therapeutic insulin excess and compromised physiologic and behavioral defenses against falling plasma glucose concentrations (Table 406-2; Fig. 406-2). Defective glucose counterregulation compromises physiologic defense (particularly decrements in insulin and increments in glucagon and epinephrine), and hypoglycemia unawareness compromises behavioral defense (ingestion of carbohydrate).

**DEFECTIVE GLUCOSE COUNTERREGULATION** In the setting of absolute endogenous insulin deficiency, insulin levels do not decrease as plasma glucose levels fall; thus, the first defense against hypoglycemia is lost. After a few years of disease duration in T1DM,

glucagon levels do not increase as plasma glucose levels fall; a second defense against hypoglycemia is lost. Reduced glucagon responses to hypoglycemia also occur in long-duration T2DM. However, pancreatic alpha cells that produce glucagon are present in the same number and size in T1DM as compared to age-matched nondiabetic individuals. Thus, the defect that restricts glucagon release during hypoglycemia in T1DM (and presumably in long-standing T2DM) appears to be a signaling defect, as glucagon responses to other physiologic stress in T1DM (e.g., exercise) are preserved. Finally, the increase in epinephrine levels, the third critical defense against acute hypoglycemia, is typically attenuated. The glycemic threshold for the sympathoadrenal (adrenomedullary epinephrine and sympathetic neural norepinephrine) response is shifted to lower plasma glucose concentrations. That shift is typically the result of recent antecedent iatrogenic hypoglycemia. In the setting of absent decrements in insulin and of absent increments in glucagon, the attenuated increment in epinephrine causes the clinical syndrome of defective glucose counterregulation. Affected patients are at 25-fold greater risk of severe iatrogenic hypoglycemia during intensive glycemic therapy for their diabetes than are patients with normal epinephrine responses. This functional—and potentially reversible—disorder is distinct from classic diabetic autonomic neuropathy, which also includes all of the above pathophysiologic defects, and is a structural and irreversible disorder.

**HYPOGLYCEMIA UNAWARENESS** The attenuated sympathoadrenal response (largely the reduced sympathetic neural response) to hypoglycemia causes the clinical syndrome of *hypoglycemia unawareness*—i.e., loss of the warning adrenergic and cholinergic symptoms that previously allowed the patient to recognize developing hypoglycemia and therefore to abort the episode by ingesting carbohydrates. Affected patients are at a sixfold increased risk of severe iatrogenic hypoglycemia during intensive glycemic therapy of their diabetes.

**HAAF IN DIABETES** The concept of HAAF in diabetes posits that recent antecedent iatrogenic hypoglycemia (or sleep or prior exercise)



**FIGURE 406-2** Hypoglycemia-associated autonomic failure (HAAF) in insulin-deficient diabetes. T1DM, type 1 diabetes mellitus; T2DM, type 2 diabetes mellitus. (Modified from PE Cryer: Hypoglycemia in Diabetes. Pathophysiology, Prevalence, and Prevention, 2nd ed. © American Diabetes Association, 2012.)

causes both defective glucose counterregulation (by reducing the epinephrine response to a given level of subsequent hypoglycemia in the setting of absent insulin and glucagon responses) and hypoglycemia unawareness (by reducing the sympathoadrenal response to a given level of subsequent hypoglycemia). These impaired responses, which can occur in individuals with either T1DM or T2DM, create a vicious cycle of recurrent iatrogenic hypoglycemia (Fig. 406-2). Hypoglycemia unawareness and, to some limited extent, the reduced epinephrine component of defective glucose counterregulation can be reversible by as little as 2–3 weeks of scrupulous avoidance of hypoglycemia in most affected patients.

On the basis of this pathophysiology, additional risk factors for hypoglycemia in diabetes include (1) absolute insulin deficiency, indicating that insulin levels will not decrease and glucagon levels will not increase as plasma glucose levels fall; (2) a history of severe hypoglycemia or of hypoglycemia unawareness, implying recent antecedent hypoglycemia, as well as prior exercise or sleep, indicating that the sympathoadrenal response will be attenuated; (3) impaired renal function resulting in reduced clearance of exogenous and endogenous insulin; (4) classical diabetic autonomic neuropathy; and (5) lower HbA<sub>1c</sub> or lower glycemic goals even at elevated HbA<sub>1c</sub> levels (8–10%), as they represent an increased probability of recent antecedent hypoglycemia.

**Hypoglycemia Risk Factor Reduction** Several recent, multicenter, randomized controlled trials investigating the potential benefits of tight glucose control in either inpatient or outpatient settings have reported a high prevalence of severe hypoglycemia. In the NICE-SUGAR study, attempts to control in-hospital plasma glucose values toward physiologic levels resulted in increased mortality risk. The ADVANCE and ACCORD studies and the Veterans Affairs Diabetes Trial (VADT) also found a significant incidence of severe hypoglycemia among T2DM patients. Severe hypoglycemia with accompanying serious cardiovascular morbidity and mortality also occurred in the standard (e.g., not receiving intensified treatment) control group in all of the above studies and in another large study in prediabetic and T2DM individuals (ORIGIN). Thus, as stated above, severe hypoglycemia can and does occur at HbA<sub>1c</sub> values of 8–10% in both T1DM and T2DM. Somewhat surprisingly, all three studies found little or no benefit of intensive glucose control to reduce macrovascular events in T2DM. In fact, the ACCORD study was ended early because of the increased mortality rate in the intensive glucose control arm. Whether iatrogenic hypoglycemia was the cause of the increased mortality risk is not known. In light of these findings, some new recommendations and paradigms have been formulated. Whereas there is little debate regarding the need to reduce hyperglycemia in the hospital, the glycemic maintenance goals in critical care settings have been modified to stay between 140 and 180 mg/dL. Similar glycemic targets are also recommended in non–critically ill patients by a number of expert societies, although some recommend even more strict glucose control down to 108 mg/dL. Accordingly, the benefits of insulin therapy and reduced hyperglycemia can be obtained while the prevalence of hypoglycemia is reduced.

Similarly, evidence exists that intensive glucose control can reduce the prevalence of microvascular disease in both T1DM and T2DM. These benefits need to be weighed against the increased prevalence of hypoglycemia. Certainly, the level of glucose control (i.e., the HbA<sub>1c</sub> value, symptoms of hyper- and hypoglycemia, and home glucose values) should be evaluated for each patient. Multicenter trials have demonstrated that individuals with recently diagnosed T1DM or T2DM can have better glycemic control with less hypoglycemia. In addition, there is still long-term benefit in reducing HbA<sub>1c</sub> values from higher to lower, albeit still above recommended levels. Perhaps a reasonable therapeutic goal is the lowest HbA<sub>1c</sub> level that does not cause severe hypoglycemia and that preserves awareness of hypoglycemia.

Pancreatic transplantation (both whole organ and islet cell) has been used in part as a treatment for severe hypoglycemia. Generally, rates of hypoglycemia are reduced after transplantation. This decrease appears to be due to increased physiologic insulin and glucagon responses during hypoglycemia.

The use of continuous glucose monitors (CGMs), either alone or in combination with continuous subcutaneous infusion via a wearable pump, offers promise as a method of reducing hypoglycemia while improving HbA<sub>1c</sub>. Specifically, continuous glucose monitoring coupled with temporary discontinuation of subcutaneous insulin infusion when the monitor predicts a low glucose concentration is particularly promising. Studies investigating the use of CGM during inpatient care for both insulin-requiring pediatric and adult patients with diabetes are ongoing. Furthermore, progress utilizing a portable wearable “artificial pancreas” incorporating continuous glucose sensor modulation of either insulin alone or bi-hormonal delivery of both insulin and glucagon has been established. Additionally, stem cell-derived cells also offer promise of novel therapeutic interventions to reduce hypoglycemia.

Other interventions to stimulate counterregulatory responses, such as selective serotonin reuptake inhibitors,  $\alpha$ -adrenergic receptor antagonists, opiate receptor antagonists, and fructose, remain experimental and have not been assessed in large-scale clinical trials.

Thus, intensive glycemic therapy (Chap. 404) needs to be applied along with the patient's education and empowerment, frequent self-monitoring of blood glucose, flexible insulin (and other drug) regimens (including the use of insulin analogues, both short- and longer-acting), individualized glycemic goals, and ongoing professional guidance, support, and consideration of both the conventional risk factors and those indicative of compromised glucose counterregulation. Given a history of hypoglycemia unawareness, a 2- to 3-week period of scrupulous avoidance of hypoglycemia is indicated.

## HYPOGLYCEMIA WITHOUT DIABETES

There are many causes of hypoglycemia (Table 406-1). Because hypoglycemia is common in insulin- or insulin secretagogue-treated diabetes, it is often reasonable to assume that a clinically suspicious episode is the result of hypoglycemia. On the other hand, because hypoglycemia is rare in the absence of relevant drug-treated diabetes (pregnancy and during severe episodes of morning sickness), it is reasonable to conclude that a hypoglycemic disorder is present only in patients in whom Whipple's triad can be demonstrated.

Particularly when patients are ill or medicated, the initial diagnostic focus should be on the possibility of drug involvement and then on critical illnesses, hormone deficiency, or non-islet cell tumor hypoglycemia. In the absence of any of these etiologic factors and in a seemingly well individual, the focus should shift to possible endogenous hyperinsulinism or accidental, surreptitious, or even malicious hypoglycemia.

**Drugs** Insulin and insulin secretagogues suppress glucose production and stimulate glucose utilization. Ethanol blocks gluconeogenesis but not glycogenolysis. Thus, alcohol-induced hypoglycemia typically occurs after a several-day ethanol binge during which the person eats little food, with consequent glycogen depletion. Ethanol is usually measurable in blood at the time of presentation, but its levels correlate poorly with plasma glucose concentrations. Because gluconeogenesis becomes the predominant route of glucose production during prolonged hypoglycemia, alcohol can contribute to the progression of hypoglycemia in patients with insulin-treated diabetes.

Many other drugs have been associated with hypoglycemia. These include commonly used drugs such as angiotensin-converting enzyme inhibitors and angiotensin receptor antagonists,  $\alpha$ -adrenergic receptor antagonists, quinolone antibiotics, indomethacin, quinine, and sulfonamides.

**Critical Illness** Among hospitalized patients, serious illnesses such as renal, hepatic, or cardiac failure; sepsis; and inanition are second only to drugs as causes of hypoglycemia.

Rapid and extensive hepatic destruction (e.g., toxic hepatitis) causes fasting hypoglycemia because the liver is the major site of endogenous glucose production. The mechanism of hypoglycemia in patients with cardiac failure is unknown. Hepatic congestion and hypoxia may be involved. Although the kidneys are a source of glucose production, hypoglycemia in patients with renal failure is also caused by the reduced clearance of insulin (thereby inappropriately increasing

insulin relative to the prevailing glucose levels) and the reduced mobilization of gluconeogenic precursors in renal failure.

Sepsis is a relatively common cause of hypoglycemia. Increased glucose utilization is induced by cytokine production in macrophage-rich tissues such as the liver, spleen, and lung. Hypoglycemia develops if glucose production fails to keep pace. Cytokine-induced inhibition of gluconeogenesis in the setting of nutritional glycogen depletion, in combination with hepatic and renal hypoperfusion, may also contribute to hypoglycemia.

Hypoglycemia can be seen with starvation. Due to brain conversion and utilization of alternative substrates, such as lactate, pyruvate, and ketone bodies, there is only a modest counterregulatory neuroendocrine and autonomic nervous system response. During periods of prolonged starvation (fasting) plasma glucose levels are lower in women as compared to men, perhaps because of loss of whole-body fat stores and subsequent depletion of gluconeogenic precursors (e.g., amino acids), necessitating increased glucose utilization.

**Hormone Deficiencies** Neither cortisol nor growth hormone is critical to the prevention of hypoglycemia, at least in adults. Nonetheless, hypoglycemia can occur with prolonged fasting in patients with primary adrenocortical failure (Addison's disease) or hypopituitarism. Anorexia and weight loss are typical features of chronic cortisol deficiency and likely result in glycogen depletion. Cortisol deficiency is associated with impaired gluconeogenesis and low levels of gluconeogenic precursors; these associations suggest that substrate-limited gluconeogenesis, in the setting of glycogen depletion, is the cause of hypoglycemia. Growth hormone deficiency can cause hypoglycemia in young children. In addition to extended fasting, high rates of glucose utilization (e.g., during exercise or in pregnancy) or low rates of glucose production (e.g., after alcohol ingestion) can precipitate hypoglycemia in adults with previously unrecognized hypopituitarism.

Hypoglycemia is not a feature of the epinephrine-deficient state that results from bilateral adrenalectomy when glucocorticoid replacement is adequate, nor does it occur during pharmacologic adrenergic blockade when other glucoregulatory systems are intact. Combined deficiencies of glucagon and epinephrine play a key role in the pathogenesis of iatrogenic hypoglycemia in people with insulin-deficient diabetes, as discussed earlier. Otherwise, deficiencies of these hormones are not usually considered in the differential diagnosis of a hypoglycemic disorder.

**Non- $\beta$ -Cell Tumors** Fasting hypoglycemia, often termed *non-islet cell tumor hypoglycemia*, occurs occasionally in patients with large mesenchymal or epithelial tumors (e.g., hepatomas, adrenocortical carcinomas, carcinoids). The glucose kinetic patterns resemble those of hyperinsulinism (see next), but insulin secretion is suppressed appropriately during hypoglycemia. In most instances, hypoglycemia is due to overproduction of an incompletely processed form of insulin-like growth factor II ("big IGF-II") that does not complex normally with circulating binding proteins and thus more readily gains access to target tissues. The tumors are usually apparent clinically, plasma ratios of IGF-II to IGF-I are high, and free IGF-II levels (and levels of pro-IGF-II [1–21]) are elevated. Curative surgery is seldom possible, but reduction of tumor bulk may ameliorate hypoglycemia. Therapy with a glucocorticoid, growth hormone, or both has also been reported to alleviate hypoglycemia. Hypoglycemia attributed to ectopic IGF-I production has been reported but is rare.

**Endogenous Hyperinsulinism** Hypoglycemia due to endogenous hyperinsulinism can be caused by (1) a primary  $\beta$ -cell disorder—typically a  $\beta$ -cell tumor (*insulinoma*), sometimes multiple insulinomas, or a functional  $\beta$ -cell disorder with  $\beta$ -cell hypertrophy or hyperplasia; (2) an antibody to insulin or to the insulin receptor; (3) a  $\beta$ -cell secretagogue such as a sulfonylurea; or perhaps (4) ectopic insulin secretion, among other very rare mechanisms. None of these causes are common.

The fundamental pathophysiologic feature of endogenous hyperinsulinism caused by a primary  $\beta$ -cell disorder or an insulin secretagogue is the failure of insulin secretion to fall to very low levels during

hypoglycemia. This feature is assessed by measurement of plasma insulin, C-peptide (the connecting peptide that is cleaved from proinsulin to produce insulin), proinsulin, and glucose concentrations during hypoglycemia. Insulin, C-peptide, and proinsulin levels need not be high relative to normal, euglycemic values; rather, they are inappropriately high in the setting of a low plasma glucose concentration.

Critical diagnostic findings are a plasma insulin concentration  $3 \mu\text{U/mL}$  ( $18 \text{ pmol/L}$ ), a plasma C-peptide concentration  $0.6 \text{ ng/mL}$  ( $0.2 \text{ nmol/L}$ ), and a plasma proinsulin concentration  $5.0 \text{ pmol/L}$  when the plasma glucose concentration is  $<55 \text{ mg/dL}$  ( $<3.0 \text{ mmol/L}$ ) with symptoms of hypoglycemia. A low plasma  $\beta$ -hydroxybutyrate concentration ( $2.7 \text{ mmol/L}$ ) and an increment in plasma glucose level of  $>25 \text{ mg/dL}$  ( $>1.4 \text{ mmol/L}$ ) after IV administration of glucagon ( $1.0 \text{ mg}$ ) indicate increased insulin (or IGF) actions.

The diagnostic strategy is (1) to measure plasma glucose, insulin, C-peptide, proinsulin, and  $\beta$ -hydroxybutyrate concentrations and to screen for circulating oral hypoglycemic agents during an episode of hypoglycemia and (2) to assess symptoms during the episode and seek their resolution following correction of hypoglycemia by glucose (either oral or parenteral) or by IV injection of glucagon (i.e., to document Whipple's triad). This is straightforward if the patient is hypoglycemic when seen. Since endogenous hyperinsulinemic disorders usually, but not invariably, cause fasting hypoglycemia, a diagnostic episode may develop after a relatively short outpatient fast. Serial sampling during an inpatient diagnostic fast of up to 72 h or after a mixed meal is more problematic. An alternative is to give patients a detailed list of the required measurements and ask them to present to an ambulatory care center or emergency room, with the list, during a symptomatic episode. Obviously, a normal plasma glucose concentration during a symptomatic episode indicates that the symptoms are not the result of hypoglycemia.

An *insulinoma*—an insulin-secreting pancreatic islet  $\beta$ -cell tumor—is the prototypical cause of endogenous hyperinsulinism and therefore should be sought in patients with a compatible clinical syndrome. However, insulinoma is not the only cause of endogenous hyperinsulinism. Some patients with fasting endogenous hyperinsulinemic hypoglycemia have diffuse islet involvement with  $\beta$ -cell hypertrophy and sometimes hyperplasia. This pattern is commonly referred to as *nesidioblastosis*, although  $\beta$ -cells budding from ducts are not invariably found. Other patients have a similar islet pattern but with postprandial hypoglycemia, a disorder termed *noninsulinoma pancreatogenous hypoglycemia*. Post-gastric bypass postprandial hypoglycemia, which most often follows Roux-en-Y gastric bypass, is also characterized by diffuse islet involvement and endogenous hyperinsulinism. Multiple pathophysiologic mechanisms have been suggested including exaggerated GLP-1 responses to meals resulting in hyperinsulinemia, hypoglycagonesia, and hypoglycemia. However, other mechanisms may be responsible for the relative hyperinsulinemia, such as reduced insulin clearance and reduced glucagon responses to hypoglycemia. The relevant pathogenesis has not been clearly established. However, if medical treatment with agents such as an  $\alpha$ -glucosidase inhibitor, diazoxide, or octreotide fail, partial pancreatectomy may be required. Autoimmune hypoglycemias include those caused by an antibody to insulin that binds postmeal insulin and then gradually dissociates, with consequent late postprandial hypoglycemia. Alternatively, an insulin receptor antibody can function as an agonist. The presence of an insulin secretagogue, such as a sulfonylurea or a glinide, results in a clinical and biochemical pattern similar to that of an insulinoma but can be distinguished by the presence of the circulating secretagogue. Finally, there are reports of very rare phenomena such as ectopic insulin secretion, a gain-of-function insulin receptor mutation, and exercise-induced hyperinsulinemia.

Insulinomas are uncommon, with an estimated yearly incidence of 1 in 250,000. Because >90% of insulinomas are benign, they are a treatable cause of potentially fatal hypoglycemia. The median age at presentation is 50 years in sporadic cases, but the tumor usually presents in the third decade when it is a component of multiple endocrine neoplasia type 1 (Chap. 388). More than 99% of insulinomas are within the substance of the pancreas, and the tumors are usually small ( $<2.0 \text{ cm}$  in diameter

3134 in 90% of cases). Therefore, they come to clinical attention because of hypoglycemia rather than mass effects. CT or MRI detects ~70–80% of insulinomas. These methods detect metastases in the roughly 10% of patients with a malignant insulinoma. Transabdominal ultrasound often identifies insulinomas, and endoscopic ultrasound has a sensitivity of ~90%. Somatostatin receptor scintigraphy is thought to detect insulinomas in about half of patients. Selective pancreatic arterial calcium injections, with the endpoint of a sharp increase in hepatic venous insulin levels, regionalize insulinomas with high sensitivity, but this invasive procedure is seldom necessary except to confirm endogenous hyperinsulinism in the diffuse islet disorders. Intraoperative pancreatic ultrasonography almost invariably localizes insulinomas that are not readily palpable by the surgeon. Surgical resection of a solitary insulinoma is generally curative. Diazoxide, which inhibits insulin secretion, or the somatostatin analogue octreotide can be used to treat hypoglycemia in patients with unresectable tumors; everolimus, an mTOR (mammalian target of rapamycin) inhibitor, has also been successful in combination with the above approaches.

### ACCIDENTAL, SURREPTITIOUS, OR MALICIOUS HYPOGLYCEMIA

Accidental ingestion of an insulin secretagogue (e.g., as the result of a pharmacy or other medical error) or even accidental administration of insulin can occur. Factitious hypoglycemia, caused by surreptitious or even malicious administration of insulin or an insulin secretagogue, shares many clinical and laboratory features with insulinoma. It is most common among health care workers, patients with diabetes or their relatives, and people with a history of other factitious illnesses. However, it should be considered in all patients being evaluated for hypoglycemia of obscure cause. Ingestion of an insulin secretagogue causes hypoglycemia with increased C-peptide levels, whereas exogenous insulin causes hypoglycemia with low C-peptide levels, reflecting suppression of insulin secretion.

Analytical error in the measurement of plasma glucose concentrations is rare. On the other hand, hand-held and continuous glucose monitors used to guide treatment of diabetes are not quantitative instruments, particularly at low glucose levels, and should not be used for the definitive diagnosis of hypoglycemia. Even with a quantitative method, low measured glucose concentrations can be artifactual—e.g., the result of continued glucose metabolism by the formed elements of the blood *ex vivo*, particularly in the presence of leukocytosis, erythrocytosis, or thrombocytosis or with delayed separation of the serum from the formed elements (pseudohypoglycemia).

### INBORN ERRORS OF METABOLISM CAUSING HYPOGLYCEMIA

Nondiabetic hypoglycemia also results from inborn errors of metabolism. Such hypoglycemia most commonly occurs in infancy but can also occur in adulthood. Cases in adults can be classified into those resulting in fasting hypoglycemia, postprandial hypoglycemia, and exercise-induced hypoglycemia.

**Fasting Hypoglycemia** Although rare, disorders of glycogenolysis can result in fasting hypoglycemia. These disorders include glycogen storage disease (GSD) of types 0, I, III, and IV and Fanconi-Bickel syndrome (Chap. 419). Patients with GSD types I and III characteristically have high blood lactate levels before and after meals, respectively. Both groups have hypertriglyceridemia, but ketones are high in GSD type III. Defects in fatty acid oxidation also result in fasting hypoglycemia. These defects can include (1) defects in the carnitine cycle; (2) fatty-acid  $\beta$ -oxidation disorders; (3) electron transfer disturbances; and (4) ketogenesis disorders. Finally, defects in gluconeogenesis (fructose-1,6-biphosphatase) have been reported to result in recurrent hypoglycemia and lactic acidosis.

**Postprandial Hypoglycemia** Inborn errors of metabolism resulting in postprandial hypoglycemia are also rare. These errors include (1) glucokinase, SUR1, and Kir6.2 potassium channel mutations; (2) congenital disorders of glycosylation; and (3) inherited fructose intolerance.

**Exercise-Induced Hypoglycemia** Exercise-induced hypoglycemia, by definition, follows exercise. It results in hyperinsulinemia caused by increased activity of monocarboxylate transporter 1 in cells.

## APPROACH TO THE PATIENT

### Hypoglycemia

In addition to the recognition and documentation of hypoglycemia as well as its treatment (often on an urgent basis), diagnosis of the hypoglycemic mechanism is critical for the selection of therapy that prevents, or at least minimizes, recurrent hypoglycemia.

#### RECOGNITION AND DOCUMENTATION

Hypoglycemia is suspected in patients with typical symptoms; in the presence of confusion, an altered level of consciousness, or a seizure; or in a clinical setting in which hypoglycemia is known to occur. Blood should be drawn, whenever possible, before the administration of glucose to allow documentation of a low plasma glucose concentration. Convincing documentation of hypoglycemia requires the fulfillment of Whipple's triad. Thus, the ideal time to measure the plasma glucose level is during a symptomatic episode. A normal glucose level excludes hypoglycemia as the cause of the symptoms. A low glucose level confirms that hypoglycemia is the cause of the symptoms, provided the latter resolve after the glucose level is raised. When the cause of the hypoglycemic episode is obscure, additional measurements—made while the glucose level is low and before treatment—should include plasma insulin, C-peptide, proinsulin, and  $\beta$ -hydroxybutyrate levels; also critical are screening for circulating oral hypoglycemic agents and assessment of symptoms before and after the plasma glucose concentration is raised.

When the history suggests prior hypoglycemia and no potential mechanism is apparent, the diagnostic strategy is to evaluate the patient as just described and assess for Whipple's triad during and after an episode of hypoglycemia. On the other hand, while it cannot be ignored, a distinctly low plasma glucose concentration measured in a patient without corresponding symptoms raises the possibility of an artifact (pseudohypoglycemia).

#### DIAGNOSIS OF THE HYPOGLYCEMIC MECHANISM

In a patient with documented hypoglycemia, a plausible hypoglycemic mechanism can often be deduced from the history, physical examination, and available laboratory data (Table 406-1). Drugs, particularly alcohol or agents used to treat diabetes, should be the first consideration—even in the absence of known use of a relevant drug—given the possibility of surreptitious, accidental, or malicious drug administration. Other considerations include evidence of a relevant critical illness, hormone deficiencies (less commonly), and a non- $\beta$ -cell tumor that can be pursued diagnostically (rarely). Absent one of these mechanisms in an otherwise seemingly well individual, the care provider should consider endogenous hyperinsulinism and proceed with measurements and assessment of symptoms during spontaneous hypoglycemia or under conditions that might elicit hypoglycemia.

#### URGENT TREATMENT

If the patient is able and willing, oral treatment with glucose tablets or glucose-containing fluids, candy, or food is appropriate. A reasonable initial dose is 15–20 g of glucose. If the patient is unable or unwilling (because of neuroglycopenia) to take carbohydrates orally, parenteral therapy is necessary. IV administration of glucose (25 g) should be followed by a glucose infusion guided by serial plasma glucose measurements. If IV therapy is not practical, SC or IM glucagon (1.0 mg in adults) can be used, particularly in patients with T1DM. Because it acts by stimulating glycogenolysis, glucagon is ineffective in glycogen-depleted individuals (e.g., those with alcohol-induced hypoglycemia). Glucagon also stimulates insulin secretion and is therefore less useful in T2DM. The

somatostatin analogue octreotide can be used to suppress insulin secretion in sulfonylurea-induced hypoglycemia. These treatments raise plasma glucose concentrations only transiently, and patients should therefore be urged to eat as soon as is practical to replete glycogen stores.

#### PREVENTION OF RECURRENT HYPOGLYCEMIA

Prevention of recurrent hypoglycemia requires an understanding of the hypoglycemic mechanism. Offending drugs can be discontinued or their doses reduced. Hypoglycemia caused by a sulfonylurea can persist for hours or even days. Underlying critical illnesses can often be treated. Cortisol and growth hormone can be replaced if levels are deficient. Surgical, radiotherapeutic, or chemotherapeutic reduction of a non-islet cell tumor can alleviate hypoglycemia even if the tumor cannot be cured; glucocorticoid or growth hormone administration also may reduce hypoglycemic episodes in such patients. Surgical resection of an insulinoma is curative; medical therapy with diazoxide or octreotide can be used if resection is not possible and in patients with a nontumor -cell disorder. Partial pancreatectomy may be necessary in the latter patients. The treatment of autoimmune hypoglycemia (e.g., with glucocorticoid or immunosuppressive drugs) is problematic, but these disorders are sometimes self-limited. Failing these treatments, frequent feedings and avoidance of fasting may be required. Administration of uncooked cornstarch at bedtime or even an overnight intragastric infusion of glucose may be necessary for some patients.

#### FURTHER READING

- Cryer PE: *Hypoglycemia in Diabetes*, 3rd ed. Alexandria, VA, American Diabetes Association, 2016.
- Cryer PE: Hypoglycemia, in *Williams Textbook of Endocrinology*, 13th ed, S Melmed et al (eds). Philadelphia, Saunders, 2016, pp. 1582–1607.
- Lee AK et al: The association of severe hypoglycemia with incident cardiovascular events and mortality in adults with type 2 diabetes. *Diabetes Care* 41:104, 2018.
- Russell SJ et al: Outpatient glycemic control with a bionic pancreas in type 1 diabetes. *N Engl J Med* 371:313, 2014.
- Salehi M et al: Hypoglycemia after gastric bypass surgery: Current concepts and controversies. *J Clin Endocrinol Metab* 103:2815, 2018.

pathophysiology of disorders of lipoprotein metabolism and clinical approaches to their diagnosis and management. 3135

#### LIPOPROTEIN STRUCTURE AND METABOLISM

Lipoproteins contain an “oil droplet” core of hydrophobic lipids (TGs and cholesteryl esters) surrounded by a shell of hydrophilic lipids (phospholipids, unesterified cholesterol) and proteins (called apolipoproteins) that interact with body fluids (Fig. 407-1). The plasma lipoproteins are divided into major classes based on their relative density: chylomicrons, very-low-density lipoproteins (VLDLs), intermediate-density lipoproteins (IDLs), low-density lipoproteins (LDLs), and high-density lipoproteins (HDLs). Each lipoprotein class comprises a family of particles that vary in density, size, and protein composition. Because lipid is less dense than water, the density of a lipoprotein particle is primarily determined by the amount of lipid per particle. Chylomicrons are the most lipid-rich and therefore least dense lipoprotein particles, whereas HDLs have the least lipid and are therefore the most dense. Lipoprotein particles vary widely in size, with the largest particles being the most lipid-rich (chylomicrons) and the smallest particles being the most dense (HDL).

The proteins associated with lipoproteins, called *apolipoproteins* (Table 407-1), are required for the assembly, structure, function, and metabolism of lipoproteins. Apolipoproteins provide a structural basis for lipoproteins, activate enzymes important in lipoprotein metabolism, and act as ligands for cell surface receptors. ApoB is the major structural protein of chylomicrons, VLDLs, IDLs, and LDLs (collectively known as apoB-containing lipoproteins). One molecule of apoB, either apoB-48 (chylomicrons) or apoB-100 (VLDL, IDL, or LDL), is present on each lipoprotein particle. The human liver synthesizes the full-length apoB-100 (one of the largest proteins in humans), whereas the intestine makes the shorter apoB-48, which is derived from transcription of the same *APOB* gene after posttranscriptional mRNA editing. HDLs lack apoB and have different apolipoproteins that define this lipoprotein class, most importantly apoA-I, which is synthesized in both the liver and intestine and is found on virtually all HDL particles. ApoA-II is the second most abundant HDL apolipoprotein and is on approximately two-thirds of the HDL particles. ApoC-II, apoC-III, and apoA-V regulate the metabolism of TG-rich lipoproteins. ApoE plays a critical role in the metabolism and clearance of TG-rich particles. Most apolipoproteins, other than apoB, exchange actively among lipoprotein particles in the blood. Apolipoprotein(a) [apo(a)] is a distinctive apolipoprotein that results in the formation of a lipoprotein known as lipoprotein(a) [Lp(a)] and is discussed more below.

#### TRANSPORT OF INTESTINALLY DERIVED DIETARY LIPIDS BY CHYLOMICRONS

The critical role of chylomicrons is the efficient transport of absorbed dietary lipids from the intestine to tissues that require fatty acids for energy or storage and then return of cholesterol to the liver (Fig. 407-2). Dietary lipids are hydrolyzed by lipases within the intestinal lumen and emulsified with bile acids to form micelles. Dietary cholesterol, fatty acids, and fat-soluble vitamins are absorbed in the proximal small intestine. Cholesterol and retinol are esterified (by the addition of a fatty acid) in the enterocyte to form cholesteryl esters and retinyl esters, respectively. Longer-chain fatty acids (>12 carbons) are incorporated into TGs and packaged with apoB-48, phospholipids, cholesteryl esters, retinyl esters, and -tocopherol (vitamin E) in a process that requires the action of the microsomal TG transfer protein (MTP) to form chylomicrons. Nascent chylomicrons are secreted into the intestinal lymph and delivered via the thoracic duct directly to the systemic circulation, where they are extensively processed by peripheral tissues before reaching the liver. The particles encounter lipoprotein lipase (LPL), which is anchored to the endothelial surfaces of capillaries in adipose tissue and heart and skeletal muscle (Fig. 407-2). ApoC-II and apoA-V are apolipoproteins that are transferred to circulating chylomicrons from HDL in the postprandial state; apoC-II acts as a required cofactor for LPL activation, and apoA-V serves as a facilitator of LPL activity. The TGs in chylomicrons

**407**

## Disorders of Lipoprotein Metabolism

Daniel J. Rader

Lipoproteins are complexes of lipids and proteins that are essential for transport of cholesterol, triglycerides (TGs), and fat-soluble vitamins in the blood. Lipoproteins play essential roles in the absorption of dietary cholesterol, long-chain fatty acids, and fat-soluble vitamins; the transport of TGs, cholesterol, and fat-soluble vitamins from the liver to peripheral tissues; and the transport of cholesterol from peripheral tissues back to the liver and intestine for excretion. Disorders of lipoprotein metabolism can be primary (caused by genetic conditions) or secondary (to other medical conditions or environmental exposures) and involve either a substantial increase or decrease in specific circulating lipids or lipoproteins. Lipoprotein disorders can have a number of clinical consequences, most notably premature atherosclerotic cardiovascular disease (ASCVD), and are therefore important to appropriately diagnose and treat. This chapter reviews the etiology and

**TABLE 407-1 Major Apolipoproteins**

APOLIPROTEIN	PRIMARY SOURCE	LIPOPROTEIN ASSOCIATION	FUNCTION
ApoA-I	Intestine, liver	HDL, chylomicrons	Core structural protein for HDL, promotes cellular lipid efflux via ABCA1, activates LCAT
ApoA-II	Liver	HDL, chylomicrons	Structural protein for HDL
ApoA-V	Liver	VLDL, chylomicrons	Promotes LPL-mediated triglyceride lipolysis
Apo(a)	Liver	Lp(a)	Structural protein for Lp(a)
ApoB-48	Intestine	Chylomicrons, chylomicron remnants	Core structural protein for chylomicrons
ApoB-100	Liver	VLDL, IDL, LDL, Lp(a)	Core structural protein for VLDL, LDL, IDL, Lp(a); ligand for binding to LDL receptor
ApoC-II	Liver	Chylomicrons, VLDL, HDL	Cofactor for LPL
ApoC-III	Liver, intestine	Chylomicrons, VLDL, HDL	Inhibits LPL activity and lipoprotein binding to receptors
ApoE	Liver	Chylomicron remnants, IDL, HDL	Ligand for binding to LDL receptor and other receptors

Abbreviations: HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LCAT, lecithin-cholesterol acyltransferase; LDL, low-density lipoprotein; Lp(a), lipoprotein(a); LPL, lipoprotein lipase; VLDL, very-low-density lipoprotein.

are hydrolyzed by LPL, and free fatty acids are released and taken up by adjacent myocytes or adipocytes and are either oxidized to generate energy or reesterified and stored as TG. Some of the released free fatty acids bind albumin before entering cells and are transported to other tissues, especially the liver. The chylomicron particle progressively shrinks in size as the hydrophobic TG core is hydrolyzed and the excess hydrophilic lipids (cholesterol and phospholipids) and apolipoproteins on the particle surface are transferred to HDL, ultimately creating chylomicron remnants.

Chylomicron remnants contain apoB-48, which lacks the region in apoB-100 that binds to the LDL receptor. Nevertheless, they are rapidly removed from the circulation by the liver through a process that critically requires apoE as a ligand for receptors in the liver. Few, if any, chylomicrons or chylomicron remnants are generally present in the blood after a 12-h fast, except in patients with certain disorders of lipoprotein metabolism.

### TRANSPORT OF HEPATICALLY DERIVED LIPIDS BY VLDL AND LDL

Another key role of lipoproteins is the transport of hepatic lipids from the liver to the periphery (Fig. 407-2) to provide an energy source during fasting. During the fasting state, lipolysis of adipose TGs generates fatty acids that are transported to the liver, and the liver is also capable of synthesizing fatty acids through de novo lipogenesis. These fatty acids are esterified by the liver into TGs, which are packaged into VLDL particles along with apoB-100, phospholipids, cholesterol esters, and vitamin E in a process that also requires MTP. VLDL thus resemble chylomicrons in that they are “triglyceride-rich lipoproteins,” but they contain apoB-100 rather than apoB-48, are smaller and less buoyant,

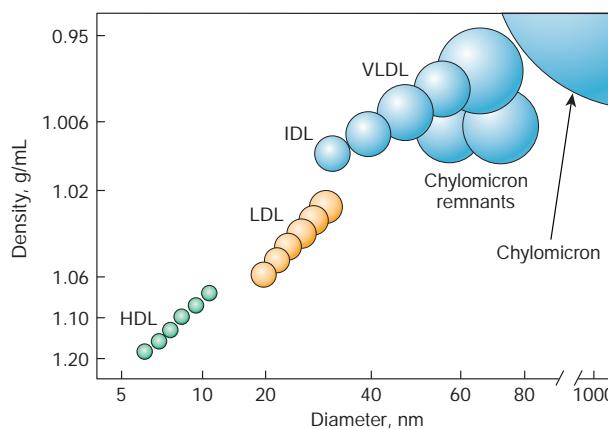
and have a higher ratio of cholesterol to TG (~1 mg of cholesterol for every 5 mg of TG, whereas in chylomicrons, this ratio is closer to ~1:8). After secretion by the liver into the plasma, the circulating TGs in VLDL are hydrolyzed by LPL. After the relatively TG-depleted VLDL remnants dissociate from LPL, they are referred to as IDLs, which contain roughly similar amounts of cholesterol and TG by mass. The liver removes ~40–60% of IDL by receptor-mediated endocytosis via binding to apoE, which is acquired through transfer of this protein from HDL. The remainder of IDL is further remodeled by hepatic lipase (HL) to form LDL. During this process, phospholipids and TGs in the particle are hydrolyzed, and most of the remaining apolipoproteins except apoB-100 are transferred to other lipoproteins. LDL is primarily a by-product of fatty acid energy transport by VLDL with little true physiologic role; one exception is that LDL may be partially responsible for delivery of vitamin E to the retina and brain. LDL is ultimately removed from the circulation by receptor-mediated endocytosis (primarily via the LDL receptor) in the liver, with a region of apoB-100 serving as the specific ligand for binding to the LDL receptor. It should be noted that apoB-48 does not contain the LDL receptor-binding ligand region, and therefore, clearance of apoB-48-containing chylomicron remnants is dependent on apoE-mediated clearance as noted above. Some LDL particles are lipolytically processed to small dense LDL particles that are believed to be especially atherogenic.

Lp(a) is a lipoprotein similar to LDL in lipid and protein composition, but it contains an additional distinctive protein called apo(a). Apo(a) is synthesized in the liver and attached to apoB-100 by a disulfide linkage. The major site of clearance of Lp(a) is the liver, but the uptake pathway is not known. Lp(a) is now established as causal factor for ASCVD, and an elevated level of Lp(a) serves as an independent risk factor and merits more aggressive therapy to reduce LDL cholesterol levels (see below).

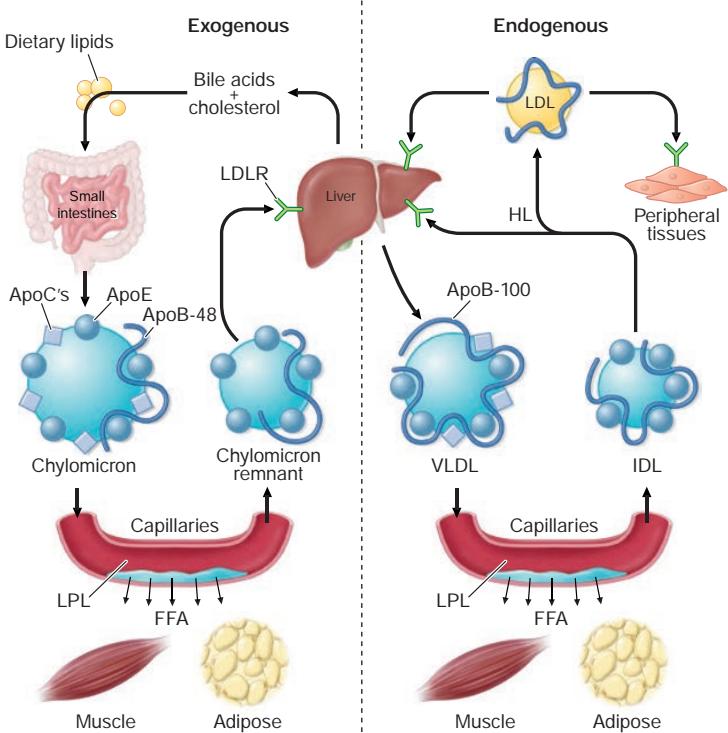
### HDL METABOLISM AND REVERSE CHOLESTEROL TRANSPORT

All nucleated cells synthesize cholesterol, but only hepatocytes and enterocytes can effectively excrete cholesterol from the body, into either the bile or the gut lumen, respectively. In the liver, cholesterol is secreted into the bile, either directly or after conversion to bile acids. Cholesterol in peripheral cells is transported from the plasma membranes of peripheral cells to the liver and intestine by a process termed *reverse cholesterol transport* that is facilitated by HDL (Fig. 407-3).

Nascent HDL particles are synthesized by the intestine and the liver. Newly secreted apoA-I rapidly acquires phospholipids and unesterified cholesterol from its site of synthesis (intestine or liver) via cellular efflux promoted by the membrane protein ATP-binding cassette protein A1 (ABCA1). This process results in the formation of discoidal HDL particles, which then recruit additional unesterified cholesterol from cells or circulating lipoproteins. Within the HDL particle, the cholesterol is esterified to cholesterol ester (CE) through the addition of a fatty acid by lecithin-cholesterol acyltransferase (LCAT), a plasma enzyme associated with HDL; the hydrophobic CE forms the core



**FIGURE 407-1** The density and size distribution of the major classes of lipoprotein particles. Lipoproteins are classified by density and size, which are inversely related. HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; VLDL, very-low-density lipoprotein.



**FIGURE 407-2** The exogenous and endogenous lipoprotein metabolic pathways. The exogenous pathway transports dietary lipids to the periphery and the liver. The endogenous pathway transports hepatic lipids to the periphery. FFA, free fatty acid; HL, hepatic lipase; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; LDLR, low-density lipoprotein receptor; LPL, lipoprotein lipase; VLDL, very-low-density lipoprotein.

of the mature HDL particle. As HDL acquires more CE, it becomes spherical, and additional apolipoproteins and lipids are transferred to the particles from the surfaces of chylomicrons and VLDLs during lipolysis.

HDL cholesterol in the blood is transported to hepatocytes by two major pathways. HDL CE can be “selectively” taken up by hepatocytes

via the scavenger receptor class B1 (SR-B1), a cell surface HDL receptor that mediates the selective transfer of CE from HDL with subsequent dissociation and “recycling” of the HDL particle. In addition, HDL CE can be transferred to apoB-containing lipoproteins in exchange for TG by the cholesteryl ester transfer protein (CETP). The CE esters are then removed from the circulation by LDL receptor-mediated endocytosis. HDL-derived CE taken up by the hepatocyte through these pathways is hydrolyzed, and much of the cholesterol is ultimately excreted directly into the bile or converted to bile acids with excretion to bile, providing a biliary route into the intestinal lumen. There is also evidence that, under certain conditions, HDL cholesterol can be transported directly into the intestinal lumen without requiring a transhepatobiliary route, a process known as *transintestinal cholesterol excretion*.

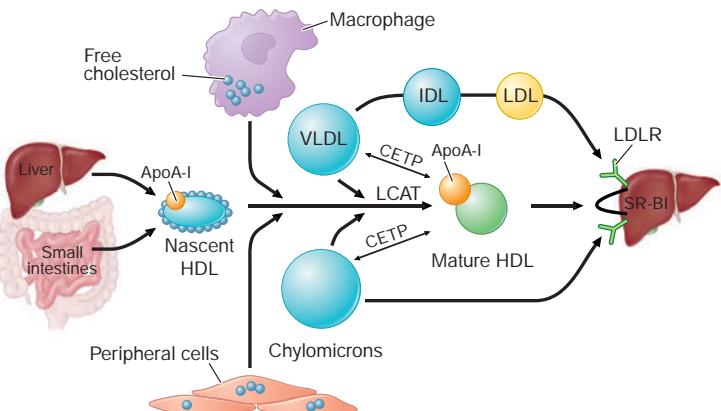
HDL particles undergo extensive remodeling within the plasma compartment by a variety of lipid transfer proteins and lipases. The phospholipid transfer protein (PLTP) transfers phospholipids from other lipoproteins to HDL or among different classes of HDL particles and is a regulator of HDL metabolism. After CETP- and PLTP-mediated lipid exchange, the TG-enriched HDL becomes a much better substrate for HL, which hydrolyzes the TGs and phospholipids to generate smaller HDL particles. A related enzyme called endothelial lipase (EL) hydrolyzes HDL phospholipids, generating smaller HDL particles that are catabolized faster. Remodeling of HDL influences the metabolism, function, and plasma concentrations of HDL.

## SCREENING

Dyslipidemia is an important causal factor in ASCVD, and treatment has been proven to substantially reduce cardiovascular risk. Therefore, all adults (and many children) should be actively screened for plasma lipids. A lipid panel should be measured, preferably after an overnight fast. In most clinical laboratories, the total cholesterol and TGs in the plasma are measured enzymatically, and then after precipitation of apoB-containing lipoproteins, the cholesterol in the supernatant is measured to determine the HDL cholesterol (HDL-C). The LDL cholesterol (LDL-C) is then estimated using the following equation (the Friedewald formula):

$$\text{LDL-C} = \text{total cholesterol} - (\text{TG}/5) - \text{HDL-C}$$

(The VLDL cholesterol content is estimated by dividing the plasma TG by 5, reflecting the ratio of TG to cholesterol in VLDL particles.) This formula is reasonably accurate if test results are obtained on fasting plasma and if the TG level does not exceed ~200 mg/dL; by convention, it cannot be used if the TG level is >400 mg/dL. LDL-C can be directly measured by a number of methods. The non-HDL-C can be easily calculated by subtracting the HDL-C from the total cholesterol. It has the advantage of incorporating the cholesterol contained within VLDL and IDL, which in most cases is also atherogenic and associated with increased ASCVD risk. There is increasing evidence that measurement of plasma apoB levels may provide a better assessment of cardiovascular risk than the LDL-C level, and even the non-HDL-C level, and is recommended by some experts. While this has not yet become standard clinical practice, the data supporting the use of apoB as a risk marker and guide to therapeutic intervention are quite strong. There is also increasing interest in Lp(a), an independent ASCVD risk factor that is highly heritable and may be helpful



**FIGURE 407-3** High-density lipoprotein (HDL) metabolism and reverse cholesterol transport. The HDL pathway transports excess cholesterol from the periphery back to the liver for excretion in the bile. The liver and the intestine produce nascent HDLs. Free cholesterol is acquired from macrophages and other peripheral cells and esterified by lecithin-cholesterol acyltransferase (LCAT), forming mature HDLs. HDL cholesterol can be selectively taken up by the liver via SR-B1 (scavenger receptor class B1). Alternatively, HDL cholesteryl ester can be transferred by cholesteryl ester transfer protein (CETP) from HDLs to very-low-density lipoproteins (VLDLs) and chylomicrons, which can then be taken up by the liver. IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; LDLR, low-density lipoprotein receptor.

3138 in risk stratification. In patients with evidence of dyslipidemia, further evaluation and treatment are based on evidence of preexisting ASCVD and clinical assessment of cardiovascular risk using risk calculators such as the American Heart Association (AHA)/American College of Cardiology (ACC) risk calculator as well as, in some cases, based on additional approaches to risk assessment such as apoB and Lp(a) (see “Approach to the Patient” for more detailed discussion).

## DISORDERS ASSOCIATED WITH ELEVATED APOB-CONTAINING LIPOPROTEINS

Disorders of lipoprotein metabolism that cause elevated levels of apoB-containing lipoproteins are among the most common and clinically important of the dyslipoproteinemias. They are generally characterized by increased plasma levels of total cholesterol, accompanied by increased TGs, LDL-C, or both. Many patients with hyperlipidemia have some combination of genetic predisposition (often polygenic) and medical or environmental contribution (medical condition, diet, lifestyle, or drug). Many, but not all, patients with hyperlipidemia are at increased risk for ASCVD, which is the primary reason for making the diagnosis, as intervention can substantially reduce this risk. In addition, patients with severe hypertriglyceridemia may be at risk for acute pancreatitis and require intervention to reduce this risk.

Although hundreds of proteins influence lipoprotein metabolism, and genetic variants in most of the genes that encode them interact with each other and the environment to produce dyslipidemia, there are a limited number of discrete “nodes” or pathways that regulate lipoprotein metabolism and are dysfunctional in specific dyslipidemias. These include (1) lipolysis of TG-rich lipoproteins by LPL; (2) receptor-mediated uptake of apoB-containing lipoproteins by the liver; (3) cellular cholesterol metabolism in the hepatocyte and the enterocyte; (4) assembly and secretion of VLDLs by the liver; and (5) neutral lipid transfer and phospholipid hydrolysis in the plasma. Primary genetic disorders of lipoprotein metabolism caused by single-gene mutations (**Table 407-2**) have taught us a great deal about the physiologic roles of specific proteins in these pathways in humans and are clinically important to diagnose and treat.

### SEVERE HYPERTRIGLYCERIDEMIA

Severe hypertriglyceridemia (HTG) is defined by fasting TG levels >500 mg/dL and is usually accompanied by moderately elevated total cholesterol levels and reduced levels of HDL-C, usually without important elevation in LDL-C or apoB. It is medically important because it is associated with risk of acute pancreatitis and, in some cases, is also associated with increased risk of ASCVD. Severe HTG is usually caused by impaired lipolysis of TGs in TG-rich lipoproteins (TRLs) by the enzyme LPL. LPL is synthesized by adipocytes, skeletal myocytes, and cardiomyocytes, and its posttranslational maturation and folding require the action of lipase maturation factor 1 (LMF1). After secretion, it is transported from the subendothelial to the vascular endothelial surfaces by GPIHBP1, which docks it to the endothelial surface. ApoC-II is a required cofactor for LPL, and apoA-V promotes LPL activity, and both are transported to the bound LPL on the TRLs. Single-gene Mendelian disorders that reduce LPL activity have been described (**Table 407-3**) as reviewed below; the majority of patients with severe HTG have a polygenic predisposition to secondary factors like obesity or insulin resistance.

**Primary (Genetic) Causes of Severe Hypertriglyceridemia • FAMILIAL CHYLOMICRONEMIA SYNDROME (FCS)** LPL is required for the hydrolysis of TGs in chylomicrons and VLDLs. Genetic deficiency or inactivity of LPL results in impaired lipolysis and profound elevations in plasma TGs, mostly in chylomicrons. While chylomicronemia predominates, in fact, these patients often have elevated plasma levels of VLDL as well. The fasting plasma is turbid, and if left undisturbed for several hours, the chylomicrons float to the top and form a creamy supernatant layer. Fasting TG levels are >500 mg/dL and usually >1000 mg/dL. Because chylomicrons contain cholesterol, fasting total cholesterol levels are also elevated. There are five

genes in which mutations can result in FCS (Table 407-2). FCS has an estimated frequency of ~1 in 200,000–300,000, although its true prevalence is unknown. The most common molecular cause of FCS involves mutations in the *LPL* gene. *LPL deficiency* has autosomal recessive inheritance (loss-of-function mutations in both alleles). Heterozygotes with *LPL* mutations often have moderate elevations in plasma TG levels and increased risk for coronary heart disease (CHD). FCS can also be caused by mutations in genes that affect LPL processing or activity. For example, apoC-II is a required cofactor for LPL. *APOC2 deficiency* due to loss-of-function mutations in both *APOC2* alleles results in functional lack of LPL activity and severe hyperchylomicronemia that is indistinguishable from LPL deficiency. It is also recessive in inheritance pattern and much rarer than LPL deficiency. Another apolipoprotein, apoA-V, facilitates the association of TRLs with LPL and promotes hydrolysis of the TGs. Individuals harboring loss-of-function mutations in both *APOA5* alleles causing *APOA5 deficiency* develop a form of FCS. GPIHBP1 is required for transport and tethering of LPL to the endothelial luminal surface. Homozygosity for mutations in *GPIHBP1* that interfere with its synthesis or folding cause FCS. Autoantibodies to GPIHBP1 have also been reported to cause severe hyperchylomicronemia. Finally, *LMF1* is required for appropriate processing and folding of LPL, and biallelic loss-of-function mutations can cause FCS.

FCS can present in childhood or adulthood with severe abdominal pain due to acute pancreatitis. In this setting, the diagnosis should be suspected if a fasting TG level is >500 mg/dL. Eruptive xanthomas, which are small, yellowish-white papules, may appear in clusters on the back, buttocks, and extensor surfaces of the arms and legs. On funduscopic examination, the retinal blood vessels may be opalescent (lipemia retinalis). Hepatosplenomegaly is sometimes noted as a result of uptake of circulating chylomicrons by reticuloendothelial cells in the liver and spleen. Premature ASCVD is not generally a feature of FCS.

The diagnosis of FCS is a clinical diagnosis based on persistence and severity of HTG, with a history of acute pancreatitis or eruptive xanthomas increasing the suspicion. While LPL activity can be measured in “postheparin plasma” obtained after an IV heparin injection to release the endothelial-bound LPL, this assay is not widely available. Genetic testing of a panel of candidate FCS genes can be used to confirm the diagnosis but is not required for making the clinical diagnosis.

Because of the risk of pancreatitis, it is important to consider the diagnosis and institute therapeutic interventions in FCS. The goal is to prevent pancreatitis by reducing fasting TG levels to <500 mg/dL. Consultation with a registered dietitian familiar with this disorder is essential. Dietary fat intake should be markedly restricted (to as little as 15 g/d), often with fat-soluble vitamin supplementation. Strict adherence to dietary fat restriction can be successful at controlling the chylomicronemia; fish oils or fibrates (such as fenofibrate) may be tried but are unlikely to be effective. A new therapeutic approach involving the silencing of *APOC3* with an antisense oligonucleotide is approved in Europe for patients with FCS. In patients with *APOC2* deficiency, apoC-II can be provided exogenously by infusing fresh-frozen plasma to resolve the chylomicronemia in the setting of severe acute pancreatitis. Management of patients with FCS is particularly challenging during pregnancy when VLDL production is increased.

**FAMILIAL PARTIAL LIPODYSTROPHY (FPLD)** FPLD is a genetic condition in which the generation of adipose tissue in certain fat depots is impaired and in others is excessive. FPLD is an underrecognized monogenic cause of severe HTG, which is likely due to both increased lipid synthesis and VLDL production, as well as reduced LPL-mediated clearance of TRLs. FPLD is typically a dominantly inherited disorder caused by mutations in several different genes, including lamin A/C (*LMNA*), PPAR (*PPARG*), perilipin (*PLIN1*), *AKT2*, and *ADRA2A* (Table 407-2). FPLD is characterized by loss of subcutaneous fat in the extremities and buttocks, often accompanied by increased visceral fat. Because of the reduced or absent subcutaneous fat in the arms and legs, patients are often described as having a “muscular” appearance. In addition to severe HTG, FPLD patients usually have insulin resistance, often quite severe, accompanied by type 2 diabetes and hepatosteatosis.

**TABLE 407-2 Primary Dyslipoproteinemias Caused by Known Single-Gene Mutations**

GENETIC DISORDER	GENES MUTATED	LIPOPROTEINS AFFECTED	CLINICAL FINDINGS	GENETIC TRANSMISSION	ESTIMATED PREVALENCE
<b>Severe Hypertriglyceridemia</b>					
Familial chylomicronemia syndrome (FCS)	Biallelic LoF mutations in: <i>LPL</i> , <i>APOC2</i> , <i>APOA5</i> , <i>GPIHBP1</i> , <i>LMF1</i>	Elevated: Chylomicrons, VLDL Reduced: HDL	Pancreatitis, eruptive xanthomas, hepatosplenomegaly	AR	~1/200,000–300,000
Familial partial lipodystrophy (FPLD)	Heterozygous LoF mutations in: <i>LMNA</i> , <i>PPARG</i> , <i>PLIN1</i> , <i>AKT2</i> , <i>ADRA2A</i>	Elevated: Chylomicrons, VLDL, LDL Reduced: HDL	Insulin resistance, fatty liver disease, pancreatitis, central obesity, lack of subcutaneous adipose in extremities	AD	<1/1,000,000
<b>Hypercholesterolemia</b>					
Familial hypercholesterolemia (FH)	Heterozygous LoF mutations in <i>LDLR</i>	Elevated: LDL	Tendon xanthomas, premature atherosclerotic cardiovascular disease (ASCVD)	AD	~1/250
Familial defective apoB-100 (FDB)	Heterozygous LoF receptor binding region mutations in <i>APOB</i>	Elevated: LDL	Tendon xanthomas, premature ASCVD	AD	~1/1500
Autosomal dominant hypercholesterolemia (ADH), type 3	Heterozygous GoF mutations in <i>PCSK9</i>	Elevated: LDL	Tendon xanthomas, premature ASCVD	AD	<1/1,000,000
Autosomal recessive hypercholesterolemia (ARH)	Biallelic LoF mutations in <i>LDLRAP1</i>	Elevated: LDL	Tendon xanthomas, premature ASCVD	AR	<1/1,000,000
Sitosterolemia	Biallelic LoF mutations in <i>ABCG5</i> , <i>ABCG8</i>	Elevated: LDL	Tendon xanthomas, premature ASCVD	AR	<1/1,000,000
Lysosomal acid lipase deficiency	Biallelic LoF mutations in <i>LIPA</i>	Elevated: LDL Reduced: HDL	Fatty liver disease, micronodular cirrhosis	AR	<1/1,000,000
<b>Mixed Dyslipidemia</b>					
Familial dysbetalipoproteinemia (FDBL)	Biallelic carriers of the <i>APOE2</i> variant	Elevated: Chylomicron remnants, IDL	Palmar and tuberoeruptive xanthomas, premature ASCVD	AR	~1/10,000
Hepatic lipase deficiency	Biallelic LoF mutations in <i>LIPC</i>	Elevated: Chylomicron remnants, IDL, HDL	Premature ASCVD	AR	<1/1,000,000
<b>Hypolipidemic Syndromes</b>					
Abetalipoproteinemia	Biallelic LoF mutations in <i>MTTP</i>	Absent: LDL Reduced: TG, HDL	Spinocerebellar degeneration, retinal degeneration	AR	<1/1,000,000
Familial hypobetalipoproteinemia	Heterozygous truncating mutations in <i>APOB</i>	Reduced: LDL	Fatty liver, reduced risk of ASCVD	AD	<1/1,000,000
Familial PCSK9 deficiency	Heterozygous LoF mutations in <i>PCSK9</i>	Reduced: LDL	Reduced risk of ASCVD	AD	~1/1,000
Familial combined hypolipidemia	Heterozygous LoF mutations in <i>ANGPTL3</i>	Reduced: TG, LDL, HDL	Reduced risk of ASCVD	AD	<1/1,000,000
<b>Primary Low HDL Cholesterol Syndromes</b>					
ApoA-I deletions/mutations	Heterozygous structural mutations in <i>APOA1</i>	Reduced: HDL	Variable depending on mutation: premature ASCVD, systemic amyloidosis	AD	<1/1,000,000
Tangier disease	Biallelic LoF mutations in <i>ABCA1</i>	Nearly absent: HDL Reduced: LDL Elevated: TG	Peripheral neuropathy, hepatosplenomegaly	AR	<1/1,000,000
Familial LCAT deficiency (FLD); fish eye disease (FED)	Biallelic LoF mutations in <i>LCAT</i>	Markedly reduced: HDL	Corneal opacities (both FLD and FED), progressive chronic kidney disease (FLD only)	AR	<1/1,000,000

*Abbreviations:* AD, autosomal dominant; apo, apolipoprotein; AR, autosomal recessive; ARH, autosomal recessive hypercholesterolemia; CHD, coronary heart disease; GoF, gain of function; HDL, high-density lipoprotein; IDL, intermediate-density lipoprotein; LCAT, lecithin-cholesterol acyltransferase; LDL, low-density lipoprotein; LoF, loss of function; LPL, lipoprotein lipase; PVD, peripheral vascular disease; TG, triglyceride; VLDL, very-low density lipoprotein.

Pancreatitis secondary to HTG can be a complication; in addition, ASCVD risk is increased in FPLD patients. The diagnosis of FPLD is a clinical diagnosis based on the constellation of metabolic findings accompanied by the distinctive distribution of adipose tissue. Genetic testing of a panel of candidate FPLD genes can be used to confirm the diagnosis but is not required for making the clinical diagnosis. Because FPLD is a dominant disorder, the finding of a causal mutation should lead to family-based screening.

The dyslipidemia of FPLD can be difficult to manage clinically. Patients should be treated aggressively not only to reduce TG levels but also with statins and, if necessary, additional LDL-lowering therapies to reduce atherogenic lipoproteins. The insulin-resistant diabetes often requires aggressive management as well. Some patients have progression of fatty liver disease to nonalcoholic steatohepatitis and fibrosis. A different group of very rare patients have congenital generalized lipodystrophy, a recessive disorder caused by mutations in the *AGPAT2*

TABLE 407-3 Secondary Causes of Altered Lipid and Lipoprotein Levels

TG ELEVATED	LDL-C		HDL-C		LP(a) ELEVATED
	ELEVATED	REDUCED	ELEVATED	REDUCED	
High-carbohydrate diet	Hypothyroidism	Vegan diet	High-fat diet	Hypertriglyceridemia	Chronic kidney disease
Alcohol	Cholestasis	Malabsorption	Alcohol	Vegan diet	Nephrotic syndrome
Obesity	Nephrotic syndrome	Malnutrition	Exercise	Malabsorption	Inflammation
Insulin resistance	Cushing's syndrome	Severe liver disease	Drugs: estrogen	Malnutrition	Menopause
Type 2 diabetes	Acute intermittent porphyria	Gaucher's disease		Sedentary lifestyle	Orchiectomy
Lipodystrophy	Drugs: corticosteroids, cyclosporin, sirolimus, carbamazepine	Chronic infectious disease		Smoking	Hypothyroidism
Chronic kidney disease		Hyperthyroidism		Obesity	Acromegaly
Nephrotic syndrome				Gaucher's disease	Drugs: growth hormone, isotretinoin
Viral hepatitis				LAL deficiency	
Sepsis				Drugs: anabolic steroids, testosterone, beta blockers	
Cushing's syndrome					
Acromegaly					
Glycogen storage disease					
Pregnancy					
Drugs: estrogen, glucocorticoids, isotretinoin, bexarotene, other retinoids, beta blockers, bile acid binding resins					

Abbreviations: HDL-C, high-density lipoprotein cholesterol; LAL, lysosomal acid lipase; LDL-C, low-density lipoprotein cholesterol; Lp(a), lipoprotein(a); TG, triglyceride.

and *BSCL2* genes. These patients have nearly complete absence of subcutaneous fat, accompanied by profound leptin deficiency, insulin resistance, severe HTG, and accumulation of TGs in multiple tissues including the liver. Patients with generalized lipodystrophy can be effectively treated with recombinant leptin administration, which often manages the multiple metabolic issues in these patients.

**Multifactorial Severe Hypertriglyceridemia** Most patients with severe HTG do not have a single-gene mutation but instead have a multifactorial etiology that includes genetics and environment. The prevalence of this phenotype is ~1 in 1000. HTG often runs in families, and the term *familial HTG* has been employed; however, except for the genes in which mutations cause FCS or FPLD, reviewed above, no other classic Mendelian causes of HTG have been identified to date. Instead, extensive human genetic studies have clearly established a polygenic basis to this phenotype that consists of two categories: (1) rare heterozygous variants in the five genes discussed earlier that cause FCS in the homozygous state, and (2) a high burden of common variants that have small individual effects at raising TGs. Patients who inherit some combination of rare and common TG-raising alleles often have environmental factors that exacerbate their HTG. These “secondary” factors are reviewed in detail below, but the quantitatively most important factors promoting HTG include obesity, type 2 diabetes, insulin resistance, and alcohol use. Multifactorial HTG is characterized by elevated fasting TGs but average to below average LDL-C levels and low HDL-C levels; apoB levels are not generally elevated. This condition is not generally associated with a significantly increased risk of ASCVD. However, if the HTG is exacerbated by environmental factors, medical conditions, or drugs, the TGs can rise to a level at which acute pancreatitis is a risk. Indeed, management of patients with this condition is mostly focused on reduction of TGs to prevent pancreatitis. It is important to consider and rule out secondary causes of the HTG. Patients who are at high risk for ASCVD due to other risk factors should be treated with statin therapy. In patients who are otherwise not at high risk for ASCVD, lipid-lowering drug therapy can frequently be avoided with appropriate dietary and lifestyle changes. Patients with plasma TG levels >500 mg/dL after a trial of diet and exercise should be considered for drug therapy with a fibrate or fish oil to reduce TGs in order to prevent pancreatitis. These patients should also be carefully evaluated for ASCVD risk and may be candidates for statin therapy to further reduce cholesterol and cardiovascular risk.

#### HYPERCHOLESTEROLEMIA (ELEVATED LDL-C)

Elevated LDL-C is common and is medically important because it is associated with risk of premature ASCVD. Elevated LDL-C is often

caused by impaired uptake of LDL by the liver. As discussed above, the LDL receptor is the major receptor responsible for uptake of LDL, and most causes of elevated LDL-C converge on reduced expression or activity of the LDL receptor in the liver. One major environmental factor that reduces LDL receptor activity is a diet high in saturated and trans fats. Other medical conditions that reduce LDL receptor activity include hypothyroidism and estrogen deficiency. Single-gene Mendelian disorders involving several genes that influence LDL clearance should be considered in patients with LDL-C levels >190 mg/dL (Table 407-2). However, the majority of patients with elevated LDL-C have a polygenic predisposition exacerbated by secondary factors like a diet high in saturated and trans fats.

**Primary (Genetic) Causes of Elevated LDL-C • FAMILIAL HYPERCHOLESTEROLEMIA (FH)** FH is an autosomal dominant disorder characterized by elevated plasma levels of LDL-C usually with relatively normal TG levels. FH is caused by mutations that lead to reduced function of the LDL receptor, with the most common being mutations in the *LDLR* gene itself. The reduction in LDL receptor activity in the liver results in a reduced rate of clearance of LDL from the circulation. The plasma level of LDL increases to a level such that the rate of LDL production equals the rate of LDL clearance by residual LDL receptor as well as non-LDL receptor mechanisms. Individuals with two mutated *LDLR* alleles (homozygotes or compound heterozygotes) have much higher LDL-C levels than those with one mutant allele, causing a condition known as *homozygous FH*.

Although mutations in *LDLR* are the most common cause of FH (and originally the term *FH* was used specifically for patients with *LDLR* mutations), mutations in at least two other genes, *APOB* and *PCSK9*, can also cause FH. ApoB-100 is the critical structural protein in LDL and contains a domain that serves as the ligand for binding to the LDL receptor. Mutations in the LDL receptor-binding domain of apoB-100 reduce the affinity of apoB/LDL binding to the LDL receptor, such that LDL is removed from the circulation at a reduced rate. This condition has also been termed *familial defective apoB* (FDB). Of note, truncating mutations in *APOB* cause low LDL-C levels (see below). The proprotein convertase subtilisin/kexin type 9 (PCSK9) is a secreted protein that binds to the LDL receptor and targets it for lysosomal degradation. Normally, after LDL binds to the LDL receptor, it is internalized along with the receptor, and in the low pH of the endosome, the LDL receptor dissociates from the LDL and recycles to the cell surface. When circulating PCSK9 binds the receptor, the complex is internalized and the receptor is directed to the lysosome, rather than to the cell surface, reducing the number of active LDL receptors. *Gain-of-function*

mutations in *PCSK9* that enhance the activity of PCSK9 cause a form of FH, also known as ADH type 3. Of note, loss-of-function mutations in *PCSK9* reduce LDL-C levels (see below).

The population frequency of heterozygous FH was originally estimated to be 1 in 500 individuals, but recent data suggest it may be as high as ~1 in 250 individuals, making it one of the most common single-gene disorders in humans. FH has a much higher prevalence in certain founder populations, such as South African Afrikaners, Christian Lebanese, French Canadians, and Lancaster County Amish. Heterozygous FH is characterized by elevated plasma levels of LDL-C (usually >190 mg/dL) and relatively normal levels of TGs. Patients with FH have hypercholesterolemia from birth, and FH diagnosis is often based on detection of hypercholesterolemia on routine lipid screening; this serves as the basis for the recommendation to screen children between the ages of 9 and 11. A family history of hypercholesterolemia or premature ASCVD should prompt targeted screening. Inheritance of FH is dominant, meaning that the condition is inherited from one parent, and ~50% of the patient's siblings and children can be expected to have FH. For this reason, family-based "cascade screening" can be very effective in identifying additional persons with FH. Physical findings in some, but not all, patients with FH may include corneal arcus and/or tendon xanthomas, particularly involving the dorsum of the hands and the Achilles tendons. Untreated heterozygous FH is associated with a markedly increased risk of cardiovascular disease; untreated men with heterozygous FH have an ~50% chance of having a myocardial infarction before age 60 years, and women with heterozygous FH are at substantially increased risk as well. The age of onset of cardiovascular disease is highly variable and depends on the specific molecular defect, the level of LDL-C, and coexisting cardiovascular risk factors.

The diagnosis of FH is generally a clinical diagnosis based on hypercholesterolemia with LDL-C >190 mg/dL in the absence of a secondary etiology and ideally with a family history of hypercholesterolemia and/or premature ASCVD. Secondary causes of significant hypercholesterolemia such as hypothyroidism, nephrotic syndrome, and obstructive liver disease should be excluded. Sequencing of an FH gene panel (*LDLR*, *APOB*, *PCSK9*) to confirm the diagnosis is widely available and worthy of consideration; persons with molecularly confirmed FH are at higher risk of ASCVD and therefore may benefit from more aggressive treatment, and the finding of a specific causal variant has implications for family-based cascade screening.

FH patients should be actively treated to lower plasma levels of LDL-C, preferably starting in childhood. Initiation of a diet low in saturated and trans fats is recommended, but heterozygous FH patients almost always require pharmacologic therapy for effective control of their LDL-C levels. Statins are the initial drug class of choice, and usually "high-intensity" statin therapy is needed. Many FH patients cannot achieve adequate control of their LDL-C levels even with high-intensity statin therapy, and a cholesterol absorption inhibitor (ezetimibe), a PCSK9 inhibitor, an ACL inhibitor (bempedoic acid), and a bile acid sequestrant are other classes of drugs that can be added to statins (Table 407-4). Some patients with severe heterozygous FH cannot be adequately managed using existing therapies and are candidates for LDL apheresis, a physical method of purging the blood of LDL in which the LDL particles are selectively removed from the circulation. Other novel approaches for these patients are under development.

Homozygous FH (HoFH) is caused by loss-of-function mutations in both alleles of the LDL receptor or double heterozygosity for mutations in two FH genes. Patients with HoFH have been classified into those with virtually no detectable LDL receptor activity (*receptor negative*) and patients with markedly reduced but detectable LDL receptor activity (*receptor defective*). Untreated LDL-C levels in patients with HoFH range from ~400 to >1000 mg/dL, with receptor-defective patients at the lower end and receptor-negative patients at the higher end of the range. TGs are usually relatively normal. Some patients with HoFH, particularly receptor-negative patients, present in childhood with cutaneous planar xanthomas on the hands, wrists, elbows, knees, heels, or buttocks. The devastating consequence of HoFH is accelerated ASCVD, which often presents in childhood or early adulthood. Atherosclerosis often develops first in the aortic root, where it can cause

aortic valvular or supravalvular stenosis, and typically extends into the coronary ostia, which become stenotic. Symptoms can be atypical, and sudden death is not uncommon. Untreated, receptor-negative patients with HoFH rarely survive beyond the second decade; patients with receptor-defective LDL receptor defects have a better prognosis but almost invariably develop clinically apparent atherosclerotic vascular disease by age 30 and often much sooner.

HoFH should be suspected in a child or young adult with LDL >400 mg/dL without secondary cause. Cutaneous xanthomas, evidence of ASCVD, and/or hypercholesterolemia in both parents all are supportive of the diagnosis. While the diagnosis is usually made on clinical grounds, genetic testing should be performed to identify specific causal variants. Patients with HoFH must be treated aggressively to delay the onset and progression of CVD. Although receptor-negative patients have no response to statins and PCSK9 inhibitors, receptor-defective patients can have modest responses to these medicines, and they should be tried in patients with HoFH. Two drugs that reduce the hepatic production of VLDL and thus LDL, a small-molecule inhibitor of the microsomal TG transfer protein (MTP) and an antisense oligonucleotide to apoB, and an antibody that inhibits ANGPTL3 are approved for the treatment of patients with HoFH and should be considered in patients who have insufficient response to statins and PCSK9 inhibitors. LDL apheresis should be considered in HoFH patients who have persistently elevated LDL-C levels despite drug therapy. Liver transplantation is effective in decreasing plasma LDL-C levels in this disorder and is sometimes used as a last resort. Liver-directed gene therapy is under development for HoFH, as are other new therapeutic approaches intended to address the remaining unmet medical need.

FH is an autosomal dominant disorder. There are a few rare conditions that cause an FH-like phenotype in an autosomal recessive manner and should be considered in patients with severe hypercholesterolemia who do not report a family history of hypercholesterolemia or premature CHD.

**AUTOSOMAL RECESSIVE HYPERCHOLESTEROLEMIA (ARH)** ARH is a very rare autosomal recessive disorder that was originally reported in individuals of Sardinian descent. The disease is caused by mutations in the gene *LDLRAP1* encoding the protein LDLR adaptor protein (also called the ARH protein), which is required for LDL receptor-mediated endocytosis in the liver. *LDLRAP1* binds to the cytoplasmic domain of the LDL receptor and links the receptor to the endocytic machinery. In the absence of *LDLRAP1*, LDL binds to the extracellular domain of the LDL receptor, but the lipoprotein-receptor complex fails to be internalized. ARH, like HoFH, is characterized by hypercholesterolemia, tendon xanthomas, and premature coronary artery disease (CAD). The levels of plasma LDL-C tend to be intermediate between the levels present in FH homozygotes and FH heterozygotes, and CAD is not usually symptomatic until the third decade. LDL receptor function in cultured fibroblasts is normal or only modestly reduced in ARH, whereas LDL receptor function in the liver is negligible. Unlike FH homozygotes, the hyperlipidemia responds to treatment with statins, but these patients often require additional therapy to lower plasma LDL-C to acceptable levels.

**SITOSTEROLEMIA** Sitosterolemia is a rare autosomal recessive disease that is caused by biallelic loss-of-function mutations in either of two members of the ATP-binding cassette (ABC) half transporter family, *ABCG5* and *ABCG8*. These genes are expressed in both enterocytes and hepatocytes. The proteins heterodimerize to form a functional complex that transports plant sterols such as sitosterol and campesterol, and animal sterols, predominantly cholesterol, across the apical biliary membrane of hepatocytes into the bile and across the apical luminal membrane of enterocytes into the gut lumen, thus reducing their (re)absorption and promoting their excretion. In normal individuals, <5% of dietary plant sterols are absorbed by the proximal small intestine. The small amounts of plant sterols that enter the circulation are preferentially excreted into the bile, and thus, levels of plant sterols are kept very low in tissues. In sitosterolemia, the intestinal absorption of sterols is increased and biliary and fecal excretion of the sterols is reduced, resulting in increased plasma and tissue levels of both plant sterols and cholesterol. The increase in hepatic sterol

DRUG	MAJOR INDICATIONS	STARTING DOSE	MAXIMAL DOSE	MECHANISM	ADVERSE EFFECTS
<b>LDL-Lowering Drugs</b>					
HMG-CoA reductase inhibitors (statins)	Elevated LDL-C; increased CV risk			↓ Inhibition of cholesterol synthesis → ↑ Hepatic LDL receptors	Myalgias and myopathy, ↑ transaminases, ↑ diabetes risk
Lovastatin		20–40 mg daily	80 mg daily		
Pravastatin		40–80 mg daily	80 mg daily		
Simvastatin		20–40 mg daily	80 mg daily		
Fluvastatin		20–40 mg daily	80 mg daily		
Atorvastatin		20–40 mg daily	80 mg daily		
Rosuvastatin		5–20 mg daily	40 mg daily		
Pitavastatin		1–2 mg daily	4 mg daily		
Cholesterol absorption inhibitor	Elevated LDL-C			↓ Cholesterol absorption → ↑ LDL receptors	Elevated transaminases
Ezetimibe		10 mg daily	10 mg daily		
Bile acid sequestrants	Elevated LDL-C			↑ Bile acid excretion → ↑ LDL receptors	Bloating, constipation, elevated triglycerides
Cholestyramine		4 g daily	32 g daily		
Colestipol		5 g daily	40 g daily		
Colesevelam		3750 mg daily	4375 mg daily		
PCSK9 inhibitors	Elevated LDL-C			↓ PCSK9 activity due to Ab inhibition → ↑ LDL receptors	Injection site reactions
Evolocumab (Ab)		140 mg SC every 2 weeks	420 mg SC every 1 month (HoFH)		
Alirocumab (Ab)		75 mg SC every 2 weeks	150 mg SC every 2 weeks		
Inclisiran (siRNA)		300 mg SC every 6 months	300 mg SC every 6 months	↓ PCSK9 synthesis due to siRNA silencing → ↑ LDL receptors	Injection site reactions
ATP citrate lyase inhibitor	Elevated LDL-C	180 mg daily	180 mg daily	↓ Inhibition of cholesterol synthesis → ↑ LDL receptors	↑ uric acid and gout Tendon rupture
Bempedoic acid					
MTP inhibitor	HoFH	5 mg daily	60 mg daily	MTP inhibition → ↓ VLDL assembly and secretion	Nausea, diarrhea, increased hepatic fat
Lomitapide					
ApoB inhibitor (ASO)	HoFH	200 mg SC weekly	200 mg SC weekly	↓ ApoB synthesis due to ASO silencing → ↓ ApoB/VLDL secretion	Injection site reactions, flu-like symptoms, increased hepatic fat
Mipomersen					
ANGPTL3 inhibitor (Ab)	HoFH	15 mg/kg IV q 4 weeks	15 mg/kg IV q 4 weeks	↓ ANGPTL3 activity due to Ab inhibition → ↑ LPL activity, ↑ LDL catabolism	Reduced HDL-C levels
Evinacumab					
<b>TG-Lowering Drugs</b>					
Fibric acid derivatives (fibrates)	Elevated TG	600 mg bid	600 mg bid	↑ LPL, ↓ VLDL synthesis	Dyspepsia, myalgia, gallstones, elevated transaminases
Gemfibrozil		40–160 mg daily depending on product	40–160 mg daily depending on product		
Fenofibrate					
Omega-3 fatty acids	Elevated TG	4 g daily	4 g daily	↑ TG catabolism	Dyspepsia, fishy odor to breath
Acid ethyl esters					
Icosapent ethyl		4 g daily	4 g daily		

Abbreviations: Ab, antibody; GI, gastrointestinal; HDL-C, high-density lipoprotein cholesterol; HoFH, homozygous familial hypercholesterolemia; LDL, low-density lipoprotein; LDL-C, LDL cholesterol; LPL, lipoprotein lipase; TG, triglyceride; VLDL, very-low-density lipoprotein.

levels results in transcriptional suppression of the expression of the LDL receptor, resulting in reduced uptake of LDL and substantially increased LDL-C levels. In addition to the clinical picture of severe hypercholesterolemia, often accompanied by tendon xanthomas and premature ASCVD, these patients also have anisocytosis and poikilocytosis of erythrocytes and megathrombocytes due to the incorporation of plant sterols into cell membranes. Episodes of hemolysis and splenomegaly are a distinctive clinical feature of this disease compared to other genetic forms of hypercholesterolemia and can be a clue to the diagnosis. Sitosterolemia should be suspected in a patient with severe hypercholesterolemia without a family history of such or who fails to respond to statin therapy. Sitosterolemia can be diagnosed by a laboratory finding of a substantial increase in plasma sitosterol and/or other plant sterols and should be confirmed by gene sequencing of *ABCG5* and *ABCG8*. It is important to make the diagnosis, because diet, bile acid sequestrants, and cholesterol-absorption inhibitors are

the most effective agents to reduce LDL-C and plasma plant sterol levels in these patients. Of note, heterozygosity for mutations in *ABCG5* or *ABCG8* is now recognized to cause a moderate form of hypercholesterolemia.

**LYSOSOMAL ACID LIPASE DEFICIENCY (LALD)** LALD, also known as *cholesteryl ester storage disease*, is an autosomal recessive disorder caused by loss-of-function variants in both alleles of the gene *LIPA* encoding the enzyme lysosomal acid lipase (LAL). LAL is responsible for hydrolyzing neutral lipids, particularly TGs and CEs, after delivery to the lysosome by cell surface receptors such as the LDL receptor. It is particularly important in the liver, which clears large amounts of lipoproteins from the circulation. LALD is characterized by elevated LDL-C, usually in association with low HDL-C and with variably elevated TG levels, together with progressive fatty liver ultimately leading to hepatic fibrosis. Genetic deficiency of LAL results in accumulation

of neutral lipid in the hepatocytes, leading to hepatosplenomegaly, microvesicular steatosis, and ultimately fibrosis and end-stage liver disease. The most severe form of this disorder, Wolman's disease, presents in infancy and is rapidly fatal. The etiology of the elevated LDL-C levels is primarily due to impaired LDL receptor-mediated clearance of LDL. LALD should be suspected in nonobese patients with elevated LDL-C, low HDL-C, and evidence of fatty liver in the absence of overt insulin resistance. The diagnosis can be made with a dried blood spot assay of LAL activity and confirmed by DNA genotyping for the most common mutation, followed if necessary by sequencing of the gene to find the second mutation. Liver biopsy is required to assess the degree of inflammation and fibrosis. LALD is underdiagnosed; it is critically important to suspect it and make the diagnosis because enzyme replacement therapy with sebelipase alfa is now available and is highly effective in treating this condition.

The above conditions primarily cause elevations in LDL due to impaired catabolism of LDL from the blood. There are a few forms of primary dyslipidemia that impair the catabolism of "remnant" TRLs (after their processing by LPL) and therefore cause elevations in both cholesterol and TGs due to remnant accumulation.

**Multifactorial Hypercholesterolemia** Most patients with elevated LDL-C do not have a single-gene disorder, as described above, but instead have a multifactorial etiology that includes genetics and environment. Genetic variation contributes substantially to elevated LDL-C levels in the general population. It has been estimated that at least 50% of variation in LDL-C is genetically determined. Many patients with elevated LDL-C have *polygenic hypercholesterolemia* due to multiple common genetic variants exerting modest LDL-raising effects. Individuals at the tail of the highest burden of polygenic risk score for LDL-C often have LDL-C levels that are similar to those with FH. In patients who are genetically predisposed to higher LDL-C levels, diet plays a key exacerbating role; indeed, increased saturated and trans fats in the diet shift the entire distribution of LDL-C levels in the population to the right. As described in more detail below, patients with elevated LDL-C should be carefully assessed for their risk of ASCVD and managed with lifestyle modification and LDL-lowering medications as needed to reduce LDL-C and risk of ASCVD.

### MIXED HYPERLIPIDEMIA (ELEVATED TG AND LDL-C)

Mixed hyperlipidemia can be defined as fasting TGs >150 mg/dL and evidence of elevated cholesterol-containing lipoproteins (such as LDL-C >130 mg/dL or non-HDL-C >160 mg/dL). It is one of the most common types of lipid disorders seen in clinical practice, due both to genetic predisposition and influence of medical conditions and environmental factors (see below). It is generally associated with elevated risk of ASCVD, and therefore, patients with mixed hyperlipidemia should be carefully evaluated and managed to reduce this risk.

**Primary (Genetic) Causes of Mixed Hyperlipidemia • FAMILIAL DYSBETALIPOPROTEINEMIA (FDBL)** FDBL (also known as *type III hyperlipoproteinemia*) is a recessive disorder characterized by a mixed hyperlipidemia due to the accumulation of remnant lipoprotein particles (chylomicron remnants and VLDL remnants, or IDL). ApoE is present in multiple copies on chylomicron remnants and IDL and mediates their removal via hepatic lipoprotein receptors (Fig. 407-2). The *APOE* gene is polymorphic in sequence, resulting in the expression of three common isoforms: apoE3, which is the most common (~78% global allele frequency [AF]), apoE4 (~14% global AF), and apoE2 (~8% global AF). The apoE4 allele, which has an arginine instead of a cysteine at position 112, is widely known for being the major genetic risk factor for Alzheimer's disease. It is associated with slightly higher LDL-C levels and increased ASCVD risk but is not associated with FDBL. The apoE2 allele, which has a cysteine at position 158 instead of an arginine, is the cause of FDBL when present on both alleles. ApoE2 has a lower affinity for the LDL receptor; therefore, chylomicron remnants and IDL containing apoE2 are removed from plasma at a slower rate, leading to their accumulation in blood.

Approximately 0.5% of the general population are apoE2/E2 homozygotes, but only a small minority of these individuals actually develop hyperlipidemia characteristic of FDBL (which has a prevalence of ~1 in 10,000). Thus, an additional, sometimes identifiable, factor precipitates the development of overt dysbetalipoproteinemia in apoE2/E2 homozygotes. The most common precipitating factors are a high-fat diet, sedentary lifestyle, obesity, alcohol use, menopause, diabetes mellitus, hypothyroidism, renal disease, HIV infection, or certain drugs. Certain dominant-negative mutations in apoE can cause a dominant form of FDBL where the hyperlipidemia is fully manifest in the heterozygous state, but these mutations are very rare.

Patients with FDBL usually present in adulthood with hyperlipidemia, xanthomas, or premature coronary or peripheral vascular disease. In FDBL, in contrast to other disorders of elevated TGs, the plasma levels of cholesterol and TG are often elevated to a similar degree, and the level of HDL-C is usually normal. Two distinctive types of xanthomas, tuberoeruptive and palmar, are seen in FDBL patients. Tuberoeruptive xanthomas begin as clusters of small papules on the elbows, knees, or buttocks and can grow to the size of small grapes. Palmar xanthomas (alternatively called *xanthomata striata palmaris*) are orange-yellow discolorations of the creases in the palms and wrists. Both of these xanthoma types are virtually pathognomonic for FDBL. Subjects with FDBL have premature ASCVD and tend to have more peripheral vascular disease than is typically seen in FH.

The definitive diagnosis of FDBL can be made either by the documentation of very high levels of remnant lipoproteins or by identification of the apoE2/E2 genotype. A variety of methods are used to identify remnant lipoproteins in the plasma, including "-quantification" by ultracentrifugation (ratio of directly measured VLDL cholesterol to total plasma TG >0.30), lipoprotein electrophoresis (broad band), or nuclear magnetic resonance lipoprotein profiling. The Friedewald formula for calculation of LDL-C is not valid in FDBL because the VLDL particles are depleted in TG and enriched in cholesterol. The plasma levels of LDL-C are actually low in this disorder due to defective metabolism of VLDL to LDL. DNA-based apoE genotyping can be performed to confirm homozygosity for apoE2, which is diagnostic for FDBL. However, absence of the apoE2/E2 genotype does not strictly rule out the diagnosis of FDBL, because other mutations in apoE can (rarely) cause this condition.

Because FDBL is associated with increased risk of premature ASCVD, it should be treated aggressively. Other metabolic conditions that can exacerbate the hyperlipidemia (see above) should be managed. Patients with FDBL are typically diet-responsive and can respond favorably to low-cholesterol, low-fat diets and weight reduction. Alcohol intake should be curtailed. Pharmacologic therapy is often required, and statins are the first line in management. In the event of statin intolerance or insufficient control of hyperlipidemia, cholesterol absorption inhibitors, PCSK9 inhibitors, and fibrates are also effective in the treatment of FDBL.

**HEPATIC LIPASE DEFICIENCY** Hepatic lipase (HL; gene name *LIPC*) is a member of the same gene family as LPL and hydrolyzes TGs and phospholipids in remnant lipoproteins and HDL. Hydrolysis of lipids in remnant particles by HL contributes to their hepatic uptake via an apoE-mediated process. HL deficiency is a very rare autosomal recessive disorder caused by biallelic loss-of-function mutations in *LIPC*. It is characterized by elevated plasma levels of cholesterol and TGs (mixed hyperlipidemia) due to the accumulation of lipoprotein remnants, accompanied by elevated plasma level of HDL-C. The diagnosis is confirmed by confirmation of pathogenic mutations in both alleles of *LIPC*. Due to the small number of patients with HL deficiency, the association of this genetic defect with ASCVD is not entirely clear, although anecdotally, patients with HL deficiency who have premature CVD have been described. As with FDBL, statin therapy is recommended to reduce remnant lipoproteins and cardiovascular risk.

**FAMILIAL COMBINED HYPERLIPIDEMIA (FCHL)** FCHL is one of the most common familial lipid disorders; it is estimated to occur in ~1 in 100–200 individuals. FCHL is characterized by elevations in plasma levels of TGs (VLDL) and LDL-C (including especially a small dense

**3144** form of LDL) and reduced plasma levels of HDL-C. This disorder is an important contributor to premature CHD; ~20% of patients who develop CHD under age 60 have FCHL. FCHL can manifest in childhood but is usually not fully expressed until adulthood. The disease clusters in families, and affected family members typically have one of three possible phenotypes: (1) elevated plasma levels of LDL-C, (2) elevated plasma levels of TGs due to elevation in VLDL, or (3) elevated plasma levels of both LDL-C and TG. The lipoprotein profile can switch among these three phenotypes in the same individual over time and may depend on factors such as diet, exercise, weight, and insulin sensitivity. Patients with FCHL have substantially elevated plasma levels of apoB, often disproportionately high relative to the plasma LDL-C concentration, indicating the presence of small dense LDL particles, which are characteristic of this syndrome.

Individuals with this phenotype generally share the same metabolic defect, namely overproduction of VLDL and apoB by the liver. The molecular etiology of this condition remains poorly understood, and no single gene has been identified in which mutations convincingly cause this disorder in a simple Mendelian fashion. It is likely that defects in a combination of genes can cause the condition, suggesting that a more appropriate term for the disorder might be *polygenic combined hyperlipidemia*.

The presence of a mixed dyslipidemia (plasma TG levels between 150 and 500 mg/dL and total cholesterol levels between 200 and 400 mg/dL, usually with HDL-C levels <40 mg/dL in men and <50 mg/dL in women) and a family history of mixed dyslipidemia and/or premature CHD suggests the diagnosis. Measurement of plasma apoB levels can help support the diagnosis if they are substantially elevated, particularly relative to the LDL-C level. Individuals with this disorder should be treated aggressively due to significantly increased risk of premature CHD, often disproportionate to the LDL-C level. Decreased dietary intake of simple carbohydrates, increased aerobic exercise, and weight loss can all have beneficial effects on the lipid profile. Patients with type 2 diabetes should be aggressively treated to maintain good glucose control. Virtually all patients with FCHL merit lipid-lowering drug therapy to reduce apoB-containing lipoprotein levels and lower the risk of ASCVD. High-intensity statins are first line, but many patients with FCHL require combination therapy that includes ezetimibe, a PCSK9 inhibitor, and/or bempedoic acid.

### SECONDARY CONTRIBUTORS TO ELEVATED LEVELS OF APOB-CONTAINING LIPOPROTEINS

There are many “secondary” factors that contribute to dyslipidemia (Table 407-3), often acting in concert with polygenic predisposition as reviewed above. Some primarily affect TGs, some primarily affect LDL-C, and some influence both, with a great deal of variability. Here the major secondary contributors are reviewed.

**Secondary Factors That Primarily Elevate TG Levels • HIGH-CARBOHYDRATE DIET** Dietary carbohydrates are utilized as a substrate for fatty acid synthesis in the liver. Some of the newly synthesized fatty acids are esterified, forming TGs, and secreted in VLDL. Thus, excessive intake of calories as carbohydrates, which is frequent in Western societies, leads to increased hepatic VLDL-TG secretion and elevated TG levels. Reduction in carbohydrate consumption can have a substantial effect in reducing TG levels, although replacing carbohydrates with saturated fat can elevate LDL-C levels.

**OBESEITY, INSULIN RESISTANCE, AND TYPE 2 DIABETES (See also Chaps. 401–403)** Obesity, insulin resistance, and type 2 diabetes mellitus are the most frequent contributors to dyslipidemia, primarily by influencing TGs. The increase in adipocyte mass and accompanying decreased insulin sensitivity associated with obesity have multiple effects on lipid metabolism, with one of the major effects being excessive hepatic VLDL production. More free fatty acids are delivered from the expanded and insulin-resistant adipose tissue to the liver, where they are reesterified in hepatocytes to form TGs, which are packaged into VLDLs for secretion into the circulation. In addition, the increased insulin levels promote increased fatty acid synthesis in the liver. In insulin-resistant

patients who progress to type 2 diabetes mellitus, dyslipidemia remains common, even when the patient is under relatively good glycemic control. In addition to increased VLDL production, insulin resistance can also result in decreased LPL activity, resulting in reduced catabolism of chylomicrons and VLDLs and more severe HTG. This may be due in part to the effects of tissue insulin resistance leading to reduced transcription of LPL in skeletal muscle and adipose, as well as to increased production of the LPL inhibitor apoC-III by the liver. This reduction in LPL activity often exacerbates the effects of increased VLDL production and contributes to the dyslipidemia seen in these patients. The dyslipidemia in this setting is almost invariable associated with low HDL-C levels as well. A cluster of metabolic risk factors are often found together, including obesity, insulin resistance, hypertension, high TGs, and low HDL-C (the so-called “metabolic syndrome,” [Chap. 408](#)).

**ALCOHOL CONSUMPTION** Excessive alcohol consumption inhibits hepatic oxidation of free fatty acids, thus promoting hepatic TG synthesis and VLDL secretion and leading to increased plasma TG levels. Regular alcohol use also raises plasma levels of HDL-C and should be considered in patients with the relatively unusual combination of elevated TGs and normal or elevated HDL-C. A careful history of alcohol use should be taken in patients with elevated TGs. Reduction in alcohol consumption can often have a substantial effect in reducing TG levels.

**CHRONIC KIDNEY DISEASE (See also Chap. 311)** Chronic kidney disease (CKD) is often associated with mild HTG (150–400 mg/dL) due to the accumulation of VLDLs and remnant lipoproteins in the circulation. TG lipolysis and remnant clearance are both reduced in patients with renal failure. Because the risk of ASCVD is increased in CKD, patients should usually be treated with lipid-lowering agents, particularly statins.

**ESTROGEN AND OTHER DRUGS** Many drugs have an impact on lipid metabolism and can result in significant alterations in the lipoprotein profile (Table 407-3). Estrogens often elevate TG levels, and TG levels can also increase during pregnancy. In women with HTG, plasma TG levels should be monitored when birth control pills or postmenopausal estrogen therapy is initiated and during pregnancy. Use of low-dose preparations of estrogen or the estrogen patch can minimize the effect of exogenous estrogen on lipids. Isotretinoin therapy for acne can cause substantial elevations in TGs, and TG levels should be checked at baseline and after initiation of therapy. Bexarotene therapy for cutaneous T-cell lymphoma often causes substantial increases in TGs, and patients should be monitored accordingly.

**Secondary Factors That Elevate LDL-C Levels • DIET HIGH IN SATURATED AND TRANS FATS** Dietary saturated and trans fats act to downregulate LDL receptor expression in the liver, leading to elevation in LDL-C levels and increased ASCVD risk. A careful dietary history should be taken in individuals with elevated LDL-C with a focus on sources of saturated and trans fats. Reduction in consumption of saturated and trans fats can sometimes have a substantial effect in reducing LDL-C levels and is a cornerstone of the initial nonpharmacologic management of hypercholesterolemia.

**HYPOTHYROIDISM (See also Chap. 382)** Hypothyroidism is the most important secondary factor causing elevated LDL-C levels. It causes elevated plasma LDL-C levels due to downregulation of the hepatic LDL receptor, which is normally increased by the action of thyroid hormone. Because hypothyroidism is often subtle and therefore easily overlooked, all patients presenting with elevated plasma levels of LDL-C, especially if there has been an unexplained increase in LDL-C, should be screened for hypothyroidism by measuring thyroid-stimulating hormone (TSH). Thyroid replacement therapy usually reduces LDL-C levels; if not, the patient probably has a primary lipoprotein disorder and may require lipid-lowering drug therapy with a statin.

**LIVER DISORDERS (See also Chap. 336)** Cholestasis is almost invariably associated with hypercholesterolemia due to elevated LDL-C levels and sometimes particles called Lp-X. A major pathway by which cholesterol is excreted from the body is via secretion into bile, either directly or after conversion to bile acids, and cholestasis blocks this

critical excretory pathway. The increase in hepatocellular cholesterol results in downregulation of the LDL receptor, leading to increased plasma LDL-C levels. In severe cholestasis, excess free cholesterol, coupled with phospholipids, is shed into the plasma as a constituent of a lamellar particle called Lp-X. These unusual particles, which are not lipoproteins, lack apoB, and have an aqueous and not neutral lipid core, are rich in free cholesterol, and can deposit in the skin, producing xanthomas sometimes seen in patients with cholestasis. Some liver disorders can affect plasma lipid levels in other ways. Viral hepatitis can increase TGs, and liver failure can result in reduction in plasma cholesterol and TGs.

**NEPHROTIC SYNDROME (See also Chap. 311)** Nephrotic syndrome is a classic cause of excessive VLDL production leading to elevation in both TGs and LDL-C. The molecular mechanism of VLDL overproduction remains poorly understood but has been attributed to the effects of hypoalbuminemia leading to increased hepatic protein synthesis. Effective treatment of the underlying renal disease may normalize the lipid profile, but many patients with chronic nephrotic syndrome require lipid-lowering drug therapy with statins and sometimes additional drugs.

**CUSHING'S SYNDROME (See also Chap. 386)** Endogenous glucocorticoid excess in Cushing's syndrome is associated with increased VLDL synthesis and secretion leading to dyslipidemia characterized by HTG and elevated LDL-C. Treatment of the underlying cause is often sufficient to manage the dyslipidemia, but sometimes lipid-lowering drug therapy is needed.

**IMMUNOSUPPRESSIVE THERAPY AND CORTICOSTEROIDS** Several of the immunosuppressants used after solid organ transplantation, including cyclosporin and sirolimus, can cause substantial elevation in LDL-C and TG levels. These patients can present a difficult clinical management problem. Chronic corticosteroid use, whether after transplant or in other inflammatory conditions, can also result in elevations in LDL-C and TG levels, sometimes producing a substantial mixed dyslipidemia. When the immunosuppressant or steroid must be continued, which is often the case, drug therapy with statins may be indicated in certain patients, with careful attention to the potential for untoward muscle-related side effects.

### DISORDERS ASSOCIATED WITH REDUCED APOB-CONTAINING LIPOPROTEINS

Plasma concentrations of LDL-C <60 mg/dL are unusual. Although in some cases, LDL-C levels in this range may be reflective of malnutrition or serious chronic illness, LDL-C <60 mg/dL in an otherwise healthy individual suggests an inherited condition. The major inherited causes of low LDL-C are reviewed here and listed in Table 407-2.

**Abetalipoproteinemia** The synthesis and secretion of apoB-containing lipoproteins in the enterocytes of the proximal small bowel and in the hepatocytes of the liver involve a complex series of events that coordinate the coupling of various lipids with apoB-48 and apoB-100, respectively. Abetalipoproteinemia is a rare autosomal recessive disease caused by loss-of-function mutations in the gene encoding microsomal TG transfer protein (MTP; gene name *MTPA*), a protein that transfers lipids to nascent chylomicrons and VLDLs in the intestine and liver, respectively. Plasma levels of cholesterol and TG are extremely low in this disorder, and chylomicrons, VLDLs, LDLs, and apoB are undetectable in plasma. The parents of patients with abetalipoproteinemia (obligate heterozygotes) have normal plasma lipid and apoB levels. Abetalipoproteinemia usually presents in early childhood with diarrhea and failure to thrive due to fat malabsorption. The initial neurologic manifestations are loss of deep tendon reflexes, followed by decreased distal lower extremity vibratory and proprioceptive sense, dysmetria, ataxia, and the development of a spastic gait, often by the third or fourth decade. Patients with abetalipoproteinemia also develop a progressive pigmented retinopathy presenting with decreased night and color vision, followed by reductions in daytime visual acuity and ultimately progressing to near-blindness. The presence of spinocerebellar degeneration and pigmented retinopathy in this disease has

resulted in some patients with abetalipoproteinemia being misdiagnosed as having Friedreich's ataxia.

Most of the clinical manifestations of abetalipoproteinemia result from defects in the absorption and transport of fat-soluble vitamins. Vitamin E and retinyl esters are normally transported from enterocytes to the liver by chylomicrons, and vitamin E is dependent on VLDL for transport out of the liver and into the circulation. As a consequence of the inability of these patients to secrete apoB-containing particles, patients with abetalipoproteinemia are markedly deficient in vitamin E and are also mildly to moderately deficient in vitamins A and K. Patients with abetalipoproteinemia should be referred to specialized centers for confirmation of the diagnosis and appropriate therapy. Treatment consists of a low-fat, high-caloric, vitamin-enriched diet accompanied by large supplemental doses of vitamin E. It is imperative that treatment be initiated as soon as possible to prevent development of neurologic sequelae, which can progress even with high-dose vitamin E therapy. New therapies for this serious, albeit rare, disease are needed. The discovery that genetic loss of MTP causes absent LDL-C led to the development of an MTP inhibitor to treat homozygous FH (see below).

**Familial Hypobetalipoproteinemia (FHBL)** FHBL generally refers to a condition of low total cholesterol, LDL-C, and apoB due to mutations in the *APOB* gene. Most of the mutations causing FHBL result in a truncated apoB protein, resulting in impaired assembly and secretion of chylomicrons from enterocytes and VLDL from the liver. Any secreted VLDL particles containing a truncated apoB protein are cleared from the circulation at an accelerated rate, which also contributes to the low levels of LDL-C and apoB. Individuals heterozygous for these mutations usually have LDL-C levels <60–80 mg/dL and also tend to have low levels of plasma TG. Many FHBL patients have elevated levels of hepatic fat (due to reduced VLDL export) and sometimes have increased levels of liver transaminases, although it appears that these patients infrequently develop associated hepatic inflammation and fibrosis.

Truncating mutations in both apoB alleles cause homozygous FHBL, an extremely rare disorder resembling abetalipoproteinemia with nearly undetectable LDL-C and apoB. The neurologic defects in homozygous hypobetalipoproteinemia are similar to those seen in abetalipoproteinemia, but tend to be less severe. Homozygous hypobetalipoproteinemia can be distinguished from abetalipoproteinemia by examining the inheritance pattern of the plasma LDL-C level. The levels of LDL-C and apoB are normal in the parents of patients with abetalipoproteinemia, a classic recessive condition, and low in those of patients with homozygous hypobetalipoproteinemia, a co-dominant condition. The discovery that truncating mutations in apoB reduce LDL-C led to the development of an antisense oligonucleotide to treat HOFH (see below).

**Familial PCSK9 Deficiency** Another inherited cause of low LDL-C results from loss-of-function mutations in *PCSK9*. PCSK9 is a secreted protein that binds to the extracellular domain of the LDL receptor in the liver and promotes the degradation of the receptor. Heterozygosity for nonsense mutations in *PCSK9* that interfere with the synthesis of the protein are associated with increased hepatic LDL receptor activity and reduced plasma levels of LDL-C. Such mutations are more frequent in individuals of African descent. Individuals who are heterozygous for a loss-of-function mutation in *PCSK9* have an ~30–40% reduction in plasma levels of LDL-C and have a substantial protection from CHD relative to those without a *PCSK9* mutation, presumably due to having lower plasma cholesterol levels since birth. Homozygotes for these nonsense mutations have been reported and have extremely low LDL-C levels (<20 mg/dL) but appear otherwise healthy. A sequence variation of somewhat higher frequency (R46L) is found predominantly in individuals of European descent. This mutation impairs, but does not completely destroy, PCSK9 function. As a consequence, the plasma levels of LDL-C in individuals carrying this mutation are more modestly reduced (~15–20%); individuals with these mutations have a 45% reduction in CHD risk. The discovery of this condition led to the development of therapies that antagonize or silence PCSK9, thus reducing LDL-C levels and risk of CHD (see below).

**3146 Familial Combined Hypolipidemia** Nonsense mutations in both alleles of the gene angiopoietin-like 3 (*ANGPTL3*) lead to low plasma levels of all three major lipid fractions—TG, LDL-C, and HDL-C—a phenotype termed *familial combined hypolipidemia*. *ANGPTL3* is a protein synthesized by the liver and secreted into the bloodstream. It inhibits LPL, thus delaying clearance of TRLs from the blood and increasing TRL blood concentrations. Deficiency of *ANGPTL3*, therefore, raises LPL activity and lowers blood TG; it also lowers LDL-C and raises HDL-C levels apparently related to the effects of *ANGPTL3* on endothelial lipase. *ANGPTL3* deficiency is associated with a reduced risk for CHD. The discovery of this condition led to the development of therapies that antagonize or silence *ANGPTL3* to reduce LDL-C and TG levels (see below).

## DISORDERS ASSOCIATED WITH REDUCED HIGH-DENSITY LIPOPROTEINS

Low levels of HDL-C, generally defined as <50 mg/dL in women and <40 mg/dL in men, are very common in clinical practice. Low HDL-C is an important independent predictor of increased cardiovascular risk and has been used regularly in standardized risk calculators. As an independent risk factor, it has clinical value in the assessment of cardiovascular risk, and a patient with low HDL-C should generally be considered at higher risk of ASCVD. However, it is now considered doubtful that low HDL-C is directly causal for the development of ASCVD. Thus, while HDL-C remains an important biomarker for assessing cardiovascular risk, it is not considered a particularly attractive target for therapeutic intervention to raise HDL-C levels in order to reduce cardiovascular risk. HDL-targeted therapies that, remain in clinical development include a CETP inhibitor and an intravenous infusion of a lipidated apoA-I particle.

HDL metabolism is strongly influenced by TG metabolism, insulin resistance, and inflammation, among other environmental and medical factors. Thus, the HDL-C measurement integrates a number of cardiovascular risk factors, potentially explaining its strong inverse association with ASCVD. The majority of patients with low HDL-C have some combination of genetic predisposition and secondary factors. Variants in hundreds of genes have been shown to influence HDL-C levels. Even more important quantitatively, obesity and insulin resistance have strong suppressive effects on HDL-C, and low HDL-C in these conditions is widely observed. Furthermore, the vast majority of patients with elevated TGs have reduced levels of HDL-C due to the substantial interplay between the metabolism of TRLs and HDL (see above). Most patients with low HDL-C who have been studied in detail have accelerated catabolism of HDL and its associated apoA-I protein as the physiologic basis for the low HDL-C. Single-gene Mendelian disorders that reduce HDL-C activity have been described (Table 407-2) but are rare; the vast majority of patients with low HDL-C have a polygenic predisposition with secondary factors like obesity, insulin resistance, or HTG.

### PRIMARY (GENETIC) CAUSES OF LOW HDL-C

Mutations in three key genes encoding proteins that play critical roles in HDL synthesis and catabolism result in hypoalphalipoproteinemia (primary low levels of HDL-C). Unlike the genetic forms of hypercholesterolemia, which are invariably associated with premature coronary atherosclerosis, genetic forms of hypoalphalipoproteinemia are usually not associated with clearly increased risk of ASCVD. Nevertheless, in the clinical setting of an HDL-C level <20 mg/dL without accompanying severe HTG, these rare conditions should be considered.

**Gene Deletions and Missense Mutations in *APOA1*** Complete genetic deficiency of apoA-I due to a complete deletion of the *APOA1* gene results in the virtual absence of circulating HDL, proving the critical role of apoA-I in HDL biogenesis. The *APOA1* gene is part of a gene cluster on chromosome 11 that includes *APOA5*, *APOC3*, and *APOA4*. Some patients with no apoA-I have large genomic deletions that include other genes in the cluster. The rare patient lacking apoA-I can have cholesterol deposits in the cornea and in the skin, and in contrast to the other genetic disorders of low HDL-C, premature CHD has been reported. Heterozygotes for apoA-I deletions have reduced HDL-C levels but no obvious clinical sequelae.

More common, but still rare, are heterozygous missense mutations in the *APOA1* gene associated with low plasma levels of HDL-C. The first example reported, and still the best known, is an Arg173Cys substitution in apoA-I (so-called apoA-I<sub>Milano</sub>), found in multiple residents of a town in northern Italy. Heterozygotes for this mutation have very low plasma levels of HDL-C (<25 mg/dL) due to impaired LCAT activation and accelerated clearance of the HDL particles containing the abnormal apoA-I. Despite having very low plasma levels of HDL-C, these individuals do not appear to have an increased risk of premature CHD (neither are they protected against CHD as was initially believed). Multiple other rare *APOA1* missense mutations causing low HDL-C have been reported. A few of these mutations in *APOA1* (as well as some mutations in *APOA2*) promote the formation of amyloid fibrils, causing systemic amyloidosis.

**Tangier Disease (*ABCA1* Deficiency)** Tangier disease is a rare autosomal co-dominant form of extremely low plasma HDL-C levels that is caused by mutations in the *ABCA1* gene encoding ABCA1, a cellular transporter that facilitates efflux of unesterified cholesterol and phospholipids from cells to apoA-I as an acceptor (Fig. 407-3). Through transporting cellular lipids, ABCA1 in the hepatocytes and intestinal enterocytes promotes the extracellular lipidation of the apoA-I secreted from the basolateral membranes of these tissues. In the genetic absence of ABCA1, the nascent, poorly lipidated apoA-I is rapidly cleared from the circulation. Thus, patients with Tangier disease (both *ABCA1* alleles mutated) have extremely low circulating plasma levels of HDL-C (<5 mg/dL) and apoA-I (<5 mg/dL). Cholesterol accumulates in the reticuloendothelial system of these patients, resulting in hepatosplenomegaly and pathognomonic enlarged, grayish yellow or orange tonsils. An intermittent peripheral neuropathy (mononeuritis multiplex) or a sphingomyelia-like neurologic disorder can also be seen in this disorder. Tangier disease may be associated with some increased risk of ASCVD, although the association is not as robust as might have been anticipated, given the extremely low levels of HDL-C in these patients. Patients with Tangier disease also have low plasma levels of LDL-C, which may attenuate the atherosclerotic risk. Heterozygotes for *ABCA1* mutations have moderately reduced plasma HDL-C levels (~15–40 mg/dL), and the effect on risk of ASCVD remains uncertain.

**Familial LCAT Deficiency** This rare autosomal recessive disorder is caused by mutations in LCAT, an enzyme synthesized in the liver and secreted into the plasma, where it circulates associated with lipoproteins (Fig. 407-3). As reviewed above, the enzyme is activated by apoA-I and mediates the esterification of cholesterol to form CEs. Consequently, in familial LCAT deficiency, the proportion of free cholesterol in circulating lipoproteins is greatly increased (from ~25% to >70% of total plasma cholesterol). Deficiency in this enzyme interferes with the maturation of HDL particles and results in rapid catabolism of circulating apoA-I.

Two genetic forms of familial LCAT deficiency have been described in humans: complete deficiency (also called *classic LCAT deficiency*) and partial deficiency (also called *fish eye disease*). Progressive corneal opacification due to the deposition of free cholesterol in the cornea, very low plasma levels of HDL-C (usually <10 mg/dL), and variable HTG are characteristic of both disorders. In partial LCAT deficiency, there are no other known clinical sequelae. In contrast, patients with complete LCAT deficiency have hemolytic anemia and progressive renal insufficiency that eventually leads to end-stage renal disease. Remarkably, despite the extremely low plasma levels of HDL-C and apoA-I, premature ASCVD is not a consistent feature of either LCAT deficiency or fish eye disease. The diagnosis can be confirmed in a specialized laboratory by assaying plasma LCAT activity or by sequencing the *LCAT* gene.

**Primary Hypoalphalipoproteinemia** Primary hypoalphalipoproteinemia is defined as a plasma HDL-C level below the tenth percentile in the setting of relatively normal cholesterol and TG levels, no apparent secondary causes of low plasma HDL-C, and no clinical signs of LCAT deficiency or Tangier disease. This syndrome is often referred

to as *isolated low HDL*. A family history of low HDL-C suggests an inherited condition and may trigger an evaluation of one of the Mendelian causes of hypoalphalipoproteinemia. However, most patients with isolated low HDL do not have an identifiable single-gene disorder and likely have a polygenic etiology, possibly exacerbated by a secondary factor. The physiologic defect appears to be accelerated catabolism of HDL and its apolipoproteins. Several kindreds with primary hypoalphalipoproteinemia and an increased incidence of premature CHD have been described, although it is not clear if the low HDL-C level is the cause of the accelerated atherosclerosis in these families.

### SECONDARY FACTORS THAT REDUCE HDL-C LEVELS

**Hypertriglyceridemia** Low HDL-C is very commonly found in association with elevated TG levels. The lipolysis of TRLs generates lipids that transfer to HDL, and therefore, any impairment of lipolysis (the most common cause of elevated TGs) leads to reduced HDL biosynthesis. In settings of elevated TGs, where the HDL-C is not reduced, alternative explanations (e.g., FDBL, alcohol, estrogens) should be considered. Conversely, an isolated low HDL-C in the presence of normal TGs should prompt consideration of a primary genetic etiology (as above) or specific secondary factors (see below).

**Very-Low-Fat Diet** Dietary fat is positively associated with HDL-C levels. Individuals who eat very-low-fat vegan diets or who have anorexia or severe fat malabsorption often have low levels of HDL-C that are secondary to low dietary fat. In this setting, LDL-C levels are also usually low as well. There is no known harm to low HDL-C levels in this setting and no indication for liberalizing the diet solely for the purpose of raising the HDL-C.

**Sedentary Lifestyle and Obesity** Physical activity is known to have a (generally modest) effect in raising HDL-C levels, and conversely, a sedentary lifestyle is often associated with low HDL-C levels. Concordant with that observation, obesity is frequently associated with low HDL-C levels even when overt insulin resistance or HTG is not present. Increased physical activity and weight loss usually have some effect in raising HDL-C, which is not the primary reason for recommending these interventions but can have a motivating influence on the patient.

**ANABOLIC STEROIDS AND TESTOSTERONE** Anabolic steroids have a well-established effect on lowering HDL-C levels, sometimes quite dramatically. Testosterone supplementation can also reduce HDL-C levels, although not to the degree caused by anabolic steroids. In a young male patient who presents with unexplained very low HDL-C, a careful history of medication and supplement use should be taken.

## APPROACH TO THE PATIENT

### Lipoprotein Disorders

The major goals in the diagnosis and clinical management of lipoprotein disorders are (1) prevention of acute pancreatitis in patients with severe HTG and (2) prevention of CVD and related cardiovascular events. Given the high prevalence of dyslipidemia and the proven clinical benefits of early diagnosis and initiation of therapy, it is essential that physicians screen lipids systematically, rule out secondary causes of dyslipidemia, suspect inherited disorders of lipoprotein metabolism where appropriate, actively promote family-based cascade screening, carefully assess risk for ASCVD and consider additional risk stratification approaches, and be knowledgeable about the wide range of existing therapeutic options for dyslipidemia. The field of clinical lipidology has matured and is moving toward a more systematic clinical application of genomic medicine. Diagnostic DNA sequencing or genotyping in patients with suspected FCS, FPLD, FH, and FDBL has the potential to enhance molecular diagnosis, facilitate appropriate therapeutic interventions, and promote family-based cascade screening.

### DIAGNOSIS

A critical first step in managing a lipoprotein disorder is to attempt to determine the class or classes of lipoproteins that are increased or decreased in the patient. Once the dyslipidemia is accurately classified, efforts should be directed to identify or rule out any possible secondary causes (Table 407-3). A careful social, medical, and family history should be obtained. In patients with elevated TG levels ( $>150$  mg/dL), a fasting glucose and/or hemoglobin A<sub>1c</sub> should be obtained to rule out diabetes. In patients with elevated LDL-C levels ( $>160$  mg/dL), a TSH should be obtained to rule out hypothyroidism and consideration should be given to the possibility of liver or kidney disease. Once secondary causes have been ruled out, attempts should be made to diagnose a primary lipid disorder, because the underlying genetic defect can provide important prognostic information regarding the risk of pancreatitis in severe HTG and the risk of ASCVD in other dyslipidemias, as well as impact on the choice of drug therapy and the screening of other family members. Obtaining the correct diagnosis often requires a detailed family history, lipid analyses in family members, and sometimes specialized or genetic testing.

**Severe Hypertriglyceridemia** If the fasting plasma TG level is  $>500$  mg/dL, the patient has severe HTG and may be at risk for pancreatitis. If the TG levels are persistently severely elevated, especially if they are  $>1000$  mg/dL, and the total cholesterol-to-TG ratio is  $>8$ , FCS should be considered, and genetic testing of an FCS gene panel may be indicated (Table 407-2). If central obesity, insulin resistance, and/or fatty liver disease are also present, consideration should be given to the possibility of FPLD, and an FPLD gene panel may be indicated (Table 407-2). However, most individuals with severe HTG do not have a single-gene disorder but have increased polygenic risk for high TGs often exacerbated by secondary factors (diet, alcohol, obesity, insulin resistance, medications). Such patients are still at risk for acute pancreatitis and should be treated to reduce their TG levels and thus their risk of pancreatitis (see below).

**Hypercholesterolemia** If the LDL-C levels are  $>190$  mg/dL, the patient has severe hypercholesterolemia and is at risk for premature ASCVD. In absence of secondary causes, FH should be considered, particularly if there is a family history of hypercholesterolemia and/or premature CHD, and genetic testing of an FH gene panel may be indicated (Table 407-2). While FH is a clinical diagnosis, a finding of a causal mutation may appropriately result in earlier and more aggressive therapy to lower LDL-C and should also promote family-based cascade screening as per the Centers for Disease Control and Prevention guidelines labeling FH as a Tier 1 condition. Recessive forms of severe hypercholesterolemia are rare, but if a patient with severe hypercholesterolemia has parents with normal cholesterol levels, ARH, sitosterolemia, and LALD should be considered, and genetic testing may be indicated (Table 407-2). Patients without an identified genetic variant or who have more moderate hypercholesterolemia are likely to have polygenic hypercholesterolemia but should still be considered at risk and eligible for treatment (see below).

**Mixed Hyperlipidemia** Elevations in fasting plasma levels of both TGs ( $>150$  mg/dL) and LDL-C ( $>130$  mg/dL), often accompanied by reduced levels of HDL-C ( $<40$  mg/dL in men and  $<50$  mg/dL in women), are common, and such patients are often diagnosed as having "mixed hyperlipidemia." Most such patients are at increased risk of ASCVD and merit consideration of lifestyle and/or pharmacologic interventions. Secondary factors, particularly obesity, insulin resistance, and type 2 diabetes, are common in such patients, who often also have increased polygenic risk for dyslipidemia. The presence of palmar or tuberous xanthomas or an unusual lipid profile of total cholesterol and TG levels in the same range with an HDL-C that is not reduced should prompt consideration of FDBL, or type III hyperlipidemia, and can be diagnosed by a nuclear magnetic

resonance (NMR) lipoprotein profile or genetic testing for the *APOE2* genotype. FDBL patients should be managed aggressively due to substantially increased risk of ASCVD. More commonly, patients with mixed hyperlipidemia, particularly those with family histories of dyslipidemia or premature ASCVD, have familial combined hyperlipidemia (FCHL). ApoB should be measured in such patients, and the finding of substantially elevated apoB levels can help identify patients with FCHL, who are at especially increased risk of ASCVD and require more aggressive treatment.

## TREATMENT

### Severe Hypertriglyceridemia

There is a well-established observational relationship between severe HTG, particularly chylomicronemia, and acute pancreatitis; however, there has never been a clinical trial designed or powered to definitively prove that intervention to reduce TGs reduces the risk of pancreatitis. Nevertheless, it is generally considered appropriate medical practice to intervene in patients with TGs >500 mg/dL in order to reduce the risk of pancreatitis. It remains uncertain whether chylomicronemia increases risk for ASCVD per se. Importantly, moderate HTG (TG 150–500 mg/dL) is associated with increased ASCVD risk; management of these patients is focused on reducing risk of ASCVD and on reducing LDL-C, non-HDL-C, and apoB.

#### LIFESTYLE AND MODIFIABLE FACTORS

In patients with severe HTG, lifestyle modification can be associated with a significant reduction in plasma TG level. Patients who drink alcohol should be encouraged to decrease or preferably eliminate their intake. Patients with severe HTG often benefit from a formal dietary consultation with a dietitian intimately familiar with counseling patients on the dietary management of high TGs. Dietary fat intake should be restricted to reduce the formation of chylomicrons in the intestine. The excessive intake of simple carbohydrates should be discouraged because insulin drives TG production in the liver. Aerobic exercise and even increase in regular physical activity can have a positive effect in reducing TG levels and should be strongly encouraged. For patients who are overweight, weight loss can help to reduce TG levels. In extreme cases, bariatric surgery has been shown to not only produce effective weight loss but also substantially reduce plasma TG levels. Many patients with diabetes have HTG, and better control of diabetes can result in lowering of TGs. Finally, certain medications can exacerbate HTG (Table 407-3).

#### PHARMACOLOGIC THERAPY

Despite lifestyle interventions, many patients with severe HTG require pharmacologic therapy (Table 407-4). Patients who persist in having fasting TG >500 mg/dL despite active lifestyle management are candidates for pharmacologic therapy. The two major classes of drugs used for management of these patients are fibrates and omega-3 fatty acids (fish oils). In addition, statins can reduce plasma TG levels and also reduce ASCVD risk and should be used in patients with severe HTG who are at increased risk of ASCVD.

**Fibrates** Fibrac acid derivatives, or fibrates, are agonists of PPAR $\alpha$ , a nuclear receptor involved in the regulation of lipid metabolism. Fibrates stimulate LPL activity (enhancing TG hydrolysis), reduce apoC-III synthesis (enhancing lipoprotein remnant clearance), promote  $\omega$ -oxidation of fatty acids, and may reduce VLDL TG production. Fibrates reduce TG levels by ~30% in individuals with severe HTG and are often used as first-line therapy. They sometimes modestly raise LDL-C levels. Fibrates are generally well tolerated but can cause myopathy, especially when combined with statins, can raise creatinine, and are associated with an increase in gallstones. Fibrates can potentiate the effect of warfarin and certain oral hypoglycemic agents.

**Omega-3 Fatty Acids (Fish Oils)** Omega-3 fatty acids, or omega-3 polyunsaturated fatty acids (n-3 PUFAs), commonly known as fish oils, are present in high concentration in fish and in flaxseed. The n-3 PUFAs used for the treatment of HTG are eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA). n-3 PUFAs have been concentrated into capsules and in doses of 3–4 g/d are effective at lowering fasting TG levels by ~30%. Fish oils can cause an increase in plasma LDL-C levels in some patients. Icosapent ethyl is an EPA-only product that has been shown to reduce cardiovascular events in patients with HTG. In general, fish oils are well tolerated, with the major side effect being dyspepsia. They appear to be safe, at least at doses up to 3–4 g, but can be associated with a prolongation in the bleeding time. Fish oils can be first-line therapy for the treatment of severe HTG or can be used in combination with fibrates.

***APOC3 Silencing*** ApoC-III inhibits LPL and TRL uptake, and genetic variants in the *APOC3* gene reduce TG levels and risk of ASCVD. Volanesorsen is an antisense oligonucleotide targeted to the *APOC3* mRNA in the liver; it significantly reduces plasma apoC-III and TG levels and is approved in Europe for patients with FCS. It has been associated with severe thrombocytopenia. Additional therapeutic approaches to *APOC3* and other targets for TG lowering are in development.

### Hypercholesterolemia (Elevated LDL-C with or without Elevated TG)

There are abundant and compelling data that intervention to reduce LDL-C substantially reduces the risk of ASCVD, including myocardial infarction and stroke, as well as total mortality. Thus, it is imperative that patients with hypercholesterolemia be carefully assessed for cardiovascular risk and need for intervention. It is also worth emphasizing that patients with or at high risk for ASCVD who have plasma LDL-C levels in the “normal” or average range also benefit from intervention to reduce LDL-C levels.

#### LIFESTYLE AND MODIFIABLE FACTORS

In patients with elevated LDL-C, lifestyle modifications can be effective but are often less effective than in HTG. Patients should receive dietary counseling to reduce the content of saturated fats and trans fats in the diet. Obese patients should make an effort to lose weight. Regular aerobic exercise has relatively little impact on reducing plasma LDL-C levels, although it has cardiovascular benefits independent of LDL-C lowering. Patients with hypothyroidism should be optimally controlled. Finally, certain medications can elevate LDL-C levels (Table 407-3).

#### PHARMACOLOGIC THERAPY

The decision to use LDL-lowering drug therapy (Table 407-4)—with a statin being first-line therapy—depends on the presence of ASCVD or, if absent, the level of LDL-C as well as the level of cardiovascular risk. In patients with established ASCVD, drug therapy to reduce LDL-C is well supported by clinical trial data to reduce LDL-C as long as it remains >70 mg/dL, using combination drug therapy if necessary. In the absence of ASCVD, patients with FH must be treated to reduce the very high lifetime risk of ASCVD, and treatment should be initiated as early as possible, ideally during childhood. Otherwise, the decision to initiate LDL-lowering drug therapy is generally determined by the level of cardiovascular risk. For patients >40 years old without clinical CVD, the AHA/ACC risk calculator can be used to determine the 10-year absolute risk for CVD, and current guidelines suggest that a 10-year risk >7.5% merits consideration of statin therapy regardless of plasma LDL-C level. For younger patients, the assessment of lifetime risk of CVD may help inform the decision to start a statin, as well as a careful assessment of family history of ASCVD. In patients for whom the decision to start a statin is uncertain due to borderline ASCVD risk and/or borderline LDL-C levels, additional risk stratification might be considered. Blood tests that predict ASCVD risk beyond

traditional risk factors include apoB, Lp(a), and high-sensitivity C-reactive protein (hs-CRP). In patients who are of a sufficient age (men >40 years and women >50 years), a coronary artery calcium (CAC) score has been shown to provide independent information about risk of CAD. Elevated levels of one or more of these biomarkers or an elevated CAC score might be used to justify initiation of statin therapy in primary prevention for patients who are in a borderline zone with regard to treatment. Finally, given the strong polygenic contribution to ASCVD, there is increasing interest in the concept that a polygenic risk score for CAD might eventually be of clinical utility in lifetime risk assessment and decision-making regarding statin therapy in primary prevention.

**HMG-CoA Reductase Inhibitors (Statins)** Statins inhibit HMG-CoA reductase, a key enzyme in cholesterol biosynthesis. By inhibiting cholesterol synthesis in the liver, statins lead to a counter-regulatory increase in the expression of the LDL receptor and thus accelerated clearance of circulating LDL, resulting in a dose-dependent reduction in plasma levels of LDL-C. The magnitude of LDL-C lowering associated with statin treatment (~30–55%) varies by statin and among individuals, but once a patient is on a statin, the doubling of the statin dose produces a ~6% further reduction in the level of plasma LDL-C. An extensive body of randomized clinical trials has clearly established that statin therapy significantly reduces major cardiovascular events (and in some cases total mortality) in both primary and secondary prevention settings. The seven statins currently available differ in their LDL-C-reducing potency (Table 407-4). Current recommendations are to use high-intensity statin therapy in patients with ASCVD or deemed at high risk of ASCVD. Statins also reduce plasma TGs in a dose-dependent fashion, which is roughly proportional to their LDL-C-lowering effects.

Statins, taken in tablet form once a day, are remarkably safe and well tolerated. The most important side effect associated with statin therapy is muscle pain, or myalgia, which occurs in 3–5% of patients, some of whom are unable to tolerate any statin. Severe myopathy (associated with an increase in plasma creatine kinase [CK]) and even rhabdomyolysis can occur rarely with statin treatment. The risk of statin-associated myalgia or myopathy is increased by the presence of older age, frailty, renal insufficiency, and coadministration of drugs that interfere with the metabolism of statins, such as erythromycin and related antibiotics, antifungal agents, immunosuppressive drugs, and fibric acid derivatives (particularly gemfibrozil). In the event of muscle symptoms, a plasma CK level may be obtained to differentiate myopathy from myalgia. Serum CK levels need not be monitored on a routine basis in patients taking statins because an elevated CK in the absence of symptoms does not predict the development of myopathy and does not necessarily suggest the need for discontinuing the drug. Statins can result in elevation in liver transaminases (alanine aminotransferase [ALT] and aspartate aminotransferase [AST]), but it is usually mild and transient and generally does not require discontinuation. Finally, meta-analyses of large randomized controlled clinical trials with statins indicate a slight excess in incident type 2 diabetes, an observation as yet not fully understood. However, the cardiovascular benefits associated with statin therapy far outweigh the slight increase in incident diabetes. Based on their safety and extensively documented benefit with regard to cardiovascular outcomes, statins are the clear drug class of choice for LDL-C reduction and are by far the most widely used class of lipid-lowering drugs.

**Cholesterol Absorption Inhibitor** Cholesterol within the lumen of the small intestine is derived from the diet (about one-third) and the bile (about two-thirds) and is actively absorbed by the enterocyte through a process that involves the protein NPC1L1. Ezetimibe (Table 407-4) is a cholesterol absorption inhibitor that binds directly to and inhibits NPC1L1 and blocks the intestinal absorption of cholesterol. Ezetimibe (10 mg taken once daily) inhibits cholesterol absorption by almost 60%, resulting in a reduction in delivery of dietary sterols in the liver and a compensatory increase

in hepatic LDL receptor expression. The mean reduction in plasma LDL-C on ezetimibe (10 mg) is 18%, and the effect is additive when used in combination with a statin. Effects on TG and HDL-C levels are negligible. Ezetimibe added to a statin has been shown to significantly reduce major cardiovascular events compared with statin alone. It is generally considered the second-line option for adding to a statin in order to achieve further LDL-C reduction. Ezetimibe is very safe and well-tolerated. When used in combination with a statin, monitoring of liver transaminases is recommended. The only roles for ezetimibe in monotherapy are in patients who do not tolerate statins and in some patients with sitosterolemia.

**PCSK9 Inhibitors** Circulating PCSK9 targets the LDL receptor for lysosomal degradation, thus reducing its recycling and abundance at the surface of the hepatocyte. Genetic loss of function of PCSK9 results in low levels of LDL-C and protection from CAD. Antibodies to PCSK9 (Table 407-4) sequester it and functionally increase the number of LDL receptors available to remove LDL from the blood. They are highly effective in lowering LDL-C, with an ~60% reduction in LDL-C. They also reduce plasma levels of Lp(a) modestly. Both PCSK9 antibodies have been shown to significantly reduce cardiovascular events when added to a statin in patients with existing CAD. These antibodies are administered subcutaneously every 2 weeks. They are generally well tolerated, with the major side effect being injection site reactions. They are generally indicated as second-line (added to statin) or third-line (added to statin plus ezetimibe) therapy in patients with FH or ASCVD in whom LDL-C is not reduced to acceptable levels with a statin (with or without ezetimibe) alone. An alternative approach to silencing PCSK9, inclisiran, is a therapeutic siRNA molecule that targets the PCSK9 mRNA in the liver. In contrast to the antibodies, it is administered subcutaneously every 6 months. It is effective in reducing LDL-C by ~60% and appears to be well tolerated and safe; a cardiovascular outcomes trial is ongoing.

**ATP Citrate Lyase Inhibitor** Bempedoic acid is a first-in-class competitive inhibitor of ATP citrate lyase (ACL), which acts on mitochondrial-derived citrate to generate production of acetyl-CoA, which is subsequently used for cholesterol synthesis. Thus, it reduces cholesterol synthesis through a different mechanism than statins, ultimately upregulating the hepatic LDL receptor. In phase 3 trials, bempedoic acid 180 mg daily reduced LDL-C by ~18% when added to a statin and by ~23% as monotherapy. A cardiovascular outcomes trial is ongoing. Bempedoic acid is a prodrug that requires activation by very-long-chain acyl-CoA synthetase-1 (ASCVL1), which is not expressed in skeletal muscle, potentially explaining why it has less association with myalgias than statins; indeed, it has been shown to be relatively well tolerated in patients with statin intolerance. It is available in a fixed-dose combination with ezetimibe, which reduced LDL-C by ~36%, for patients who are statin intolerant. It can be used in combination with statins but should not be used with simvastatin in a dose >20 mg. Bempedoic acid is associated with increased uric acid levels and incidence of gout; it was also associated with increased incidence of tendon rupture in phase 3 trials. Unlike statins, it was not associated with increased incidence of diabetes.

**Bile Acid Sequestrants (Resins)** Bile acid sequestrants (BAS) bind bile acids in the intestine and promote their excretion rather than reabsorption in the ileum. To maintain the bile acid pool size, the liver diverts cholesterol to bile acid synthesis. The decreased hepatic intracellular cholesterol content results in upregulation of the LDL receptor and enhanced LDL clearance from the plasma. BAS, including cholestyramine, colestipol, and colesevalelam (Table 407-4), primarily reduce plasma LDL-C levels but can cause an increase in plasma TGs. Therefore, patients with HTG generally should not be treated with bile acid-binding resins. Cholestyramine and colestipol are insoluble resins that must be suspended in liquids. Colesevelam is available as tablets but generally requires up to six to seven tablets per day for effective LDL-C lowering. BAS are effective

in combination with statins and in combination with ezetimibe. Side effects of resins are limited to the gastrointestinal tract and include bloating and constipation. Because BAS are not systemically absorbed, they are very safe and are the cholesterol-lowering drug of choice in children and in women who are pregnant, lactating, or actively trying to conceive. However, they are otherwise fourth- or fifth-line drugs for LDL-C reduction in other settings.

**Specialized Drugs for HoFH** Three “orphan” drugs are approved specifically for the management of HoFH, a rare condition caused by biallelic mutations in the major genes causing FH in which patients respond poorly to traditional LDL-lowering medications. Lomitapide is a small-molecule inhibitor of MTP that reduces LDL-C by ~50%, and mipomersen is an antisense oligonucleotide against apoB that reduces LDL-C by ~25%. Both of these drugs reduce hepatic VLDL production and thus LDL-C levels; however, due to their mechanism of action, each drug causes an increase in hepatic fat, the long-term consequences of which are unknown. In addition, lomitapide is associated with gastrointestinal-related side effects, and mipomersen is associated with skin reactions and flulike symptoms. Finally, an antibody inhibitor of ANGPTL3, evinacumab, was approved in 2021 for the treatment of HoFH. In a phase 3 trial, an intravenous infusion every 4 weeks reduced LDL-C levels in patients with HoFH by ~50% and was well tolerated. One of these three drugs should be considered in HoFH patients after a trial of a high-intensity statin, and possibly a PCSK9 inhibitor, is shown to be insufficient to reduce LDL-C levels.

**LDL Apheresis** Patients with severe hypercholesterolemia who cannot reduce their LDL-C to acceptable levels despite optimally tolerated combination drug therapy are candidates for LDL apheresis. In this process, the patient's plasma is passed over a column that selectively removes the LDL, and the LDL-depleted plasma is returned to the patient. LDL apheresis is indicated for patients on maximally tolerated combination drug therapy (including a PCSK9 inhibitor) who have CHD and a plasma LDL-C level >200 mg/dL or no CHD and a plasma LDL-C level >300 mg/dL; LDL apheresis could be considered in high-risk patients who have an LDL-C >160 mg/dL on maximal therapy.

## FURTHER READING

- Baass A et al: Familial chylomicronemia syndrome: An under-recognized cause of severe hypertriglyceridaemia. *J Intern Med* 287:340, 2020.
- Brown EE et al: Genetic testing in dyslipidemia: A scientific statement from the National Lipid Association. *J Clin Lipidol* 14:398, 2020.
- Feingold KR: Approach to the patient with dyslipidemia. In *Endotext*. Available at <https://www.ncbi.nlm.nih.gov/books/NBK326736/>. Last updated May 11, 2020.
- Hussain I et al: Lipodystrophies, dyslipidaemias and atherosclerotic cardiovascular disease. *Pathology* 51:202, 2019.
- Li F, Zhang H: Lysosomal acid lipase in lipid metabolism and beyond. *Arterioscler Thromb Vasc Biol* 39:850, 2019.
- Luijink IK et al: 20-Year follow-up of statins in children with familial hypercholesterolemia. *N Engl J Med* 381:1547, 2019.
- Schmidt AF et al: PCSK9 monoclonal antibodies for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev* 10:CD011748, 2020.
- Sniderman AD et al: Apolipoprotein B particles and cardiovascular disease: A narrative review. *JAMA Cardiol* 4:1287, 2019.
- Sturm AC et al: Clinical genetic testing for familial hypercholesterolemia: JACC scientific expert panel. *J Am Coll Cardiol* 72:662, 2019.
- Trinder M et al: Association of monogenic vs polygenic hypercholesterolemia with risk of atherosclerotic cardiovascular disease. *JAMA Cardiol* 5:390, 2020.



The metabolic syndrome (syndrome X, insulin resistance syndrome) consists of a constellation of metabolic abnormalities that confer increased risk of cardiovascular disease (CVD) and diabetes mellitus. Evolution of the criteria for the metabolic syndrome since the original definition by the World Health Organization in 1998 reflects growing clinical evidence and analysis by a variety of consensus conferences and professional organizations. The major features of the metabolic syndrome include central obesity, hypertriglyceridemia, low levels of high-density lipoprotein (HDL) cholesterol, hyperglycemia, and hypertension (Table 408-1).

## GLOBAL HEALTH/EPIDEMILOGY

The most challenging feature of the metabolic syndrome to define is waist circumference. Intraabdominal circumference (visceral adipose tissue) is most strongly related to insulin resistance and risk of diabetes and CVD, and for any given waist circumference, the distribution of adipose tissue between subcutaneous (SC) and visceral depots varies substantially. Thus, within and between populations, there is a lesser versus greater risk at the same waist circumference. These differences in populations reflect the range of waist circumferences considered to confer risk in different geographic locations (Table 408-1).

The prevalence of the metabolic syndrome varies around the world, in part reflecting the age and ethnicity of the populations studied and the diagnostic criteria applied. In general, the prevalence of the metabolic syndrome increases with age. The prevalence of metabolic syndrome in the U.S. adult population meeting the criteria of the National Cholesterol Education Program (NCEP) and Adult Treatment Panel III (ATP III) is ~35%. Greater global industrialization is associated with rising rates of obesity and expected increase in the prevalence of the metabolic syndrome, especially as the population ages. Moreover, the rising prevalence and severity of obesity among children reflect features of the metabolic syndrome in a younger population, now estimated to be 12 and 30% among obese and overweight children, respectively.

Using National Health and Nutrition Examination Survey (NHANES) data from 2012–2016, the prevalence of metabolic syndrome in the United States was 34.7% and not different between men (35.1%) and women (34.3%). The highest prevalence was in “other” race/ethnicity (39.0%) followed by Hispanic (36.3%) and non-Hispanic white (36.0%). Although increased from 32.5% in 2010–2012, the increase was not significant; however, the prevalence did increase significantly among those aged 20–39 years (from 16.2 to 21.3%) and in women (from 31.7 to 36.6%), Asian participants (from 19.9 to 26.2%), and Hispanic participants (from 32.9 to 40.4%). As in previous data, prevalence of metabolic syndrome increased with increasing age among all subgroups, i.e., from 19.5% among those aged 20–39 years to 48.6% among those aged 60 years.

The frequency distribution of the five components of the syndrome for the U.S. population (NHANES III) is summarized in Fig. 408-1. Increases in waist circumference predominate among women, whereas increases in fasting plasma triglyceride levels (i.e., >150 mg/dL), reductions in HDL cholesterol levels, and hyperglycemia are more likely in men.

## RISK FACTORS

**Overweight/Obesity** The metabolic syndrome was first described in the early twentieth century; however, the worldwide overweight/obesity epidemic has recently been the force driving its increasing recognition. Central adiposity is a key feature of the syndrome, and the syndrome's prevalence reflects the strong relationship between waist circumference and increasing adiposity. However, despite the importance of obesity, patients who are of normal weight may also be

**TABLE 408-1 NCEP:ATPIII<sup>a</sup> 2001 and Harmonizing Definition Criteria for the Metabolic Syndrome**

NCEP:ATPIII 2001	HARMONIZING DEFINITION <sup>b</sup>												
<p>Three or more of the following:</p> <ul style="list-style-type: none"> <li>Central obesity: waist circumference &gt;102 cm (males), &gt;88 cm (females)</li> <li>Hypertriglyceridemia: triglyceride level 150 mg/dL or specific medication</li> <li>Low HDL cholesterol: &lt;40 mg/dL and &lt;50 mg/dL for men and women, respectively, or specific medication</li> <li>Hypertension: blood pressure 130 mmHg systolic or 85 mmHg diastolic or specific medication</li> <li>Fasting plasma glucose level 100 mg/dL or specific medication or previously diagnosed type 2 diabetes</li> </ul>	<p>Three of the following:</p> <p>Waist circumference (cm)</p> <table border="1"> <thead> <tr> <th>Men</th> <th>Women</th> <th>Ethnicity</th> </tr> </thead> <tbody> <tr> <td>94</td> <td>80</td> <td>Europid, sub-Saharan African, Eastern and Middle Eastern</td> </tr> <tr> <td>90</td> <td>80</td> <td>South Asian, Chinese, and ethnic South and Central American</td> </tr> <tr> <td>85</td> <td>90</td> <td>Japanese</td> </tr> </tbody> </table> <ul style="list-style-type: none"> <li>Fasting triglyceride level &gt;150 mg/dL or specific medication</li> <li>HDL cholesterol level &lt;40 mg/dL and &lt;50 mg/dL for men and women, respectively, or specific medication</li> <li>Blood pressure &gt;130 mm systolic or &gt;85 mm diastolic or previous diagnosis or specific medication</li> <li>Fasting plasma glucose level 100 mg/dL (alternative indication: drug treatment of elevated glucose levels)</li> </ul>	Men	Women	Ethnicity	94	80	Europid, sub-Saharan African, Eastern and Middle Eastern	90	80	South Asian, Chinese, and ethnic South and Central American	85	90	Japanese
Men	Women	Ethnicity											
94	80	Europid, sub-Saharan African, Eastern and Middle Eastern											
90	80	South Asian, Chinese, and ethnic South and Central American											
85	90	Japanese											

<sup>a</sup>National Cholesterol Education Program and Adult Treatment Panel III. <sup>b</sup>In this analysis, the following thresholds for waist circumference were used: white men, 94 cm; African-American men, 94 cm; Mexican-American men, 90 cm; white women, 80 cm; African-American women, 80 cm; Mexican-American women, 80 cm. For participants whose designation was "other race—including multiracial," thresholds that were once based on Europid cutoffs (94 cm for men and 80 cm for women) and on South Asian cutoffs (90 cm for men and 80 cm for women) were used. For participants who were considered "other Hispanic," the International Diabetes Federation thresholds for ethnic South and Central Americans were used. <sup>c</sup>High-density lipoprotein.

insulin-resistant and may have the metabolic syndrome. This phenotype is particularly evident for populations in India, Southeast Asia, and Central America.

**Sedentary Lifestyle** Physical inactivity and less cardiorespiratory fitness are predictors of CVD events and the related risk of death. Many components of the metabolic syndrome are associated with a sedentary lifestyle, including increased adipose tissue (predominantly central), reduced HDL cholesterol, and increased triglycerides, blood pressure, and glucose in genetically susceptible persons. Compared with individuals who watch television or videos or use the computer <1 h daily, those who do so for >4 h daily have a twofold increased risk of the metabolic syndrome.

**Genetics** No single gene explains the complex phenotype called the metabolic syndrome. However, using genome-wide association and candidate gene approaches, several genetic variants are associated with the metabolic syndrome. Although many of the loci have unknown function, many others relate to body weight and composition, insulin resistance, and unfavorable disturbances in lipid and lipoprotein metabolism.

**Aging** The metabolic syndrome affects nearly 50% of the U.S. population aged >60, and at >60 years of age, women are more often affected. The age dependency of the syndrome's prevalence is seen in most populations around the world.

**Diabetes Mellitus** Diabetes mellitus can be included in both the NCEP and the harmonizing definitions of the metabolic syndrome, but the greatest value of the metabolic syndrome, and especially fasting glucose, is predicting type 2 diabetes. The great majority (~75%) of patients with type 2 diabetes or impaired glucose tolerance have the metabolic syndrome. The presence of the metabolic syndrome in these populations relates to a higher prevalence of CVD than in patients who have type 2 diabetes or impaired glucose tolerance but do not have the syndrome.

**Cardiovascular Disease** Individuals with the metabolic syndrome are twice as likely to die of CVD as those who do not, and their risk of an acute myocardial infarction or stroke is threefold higher. The approximate prevalence of the metabolic syndrome among patients with coronary heart disease (CHD) is 60%, with a prevalence of ~35% among patients with premature coronary artery disease (age 45) and a particularly high prevalence among women. With appropriate cardiac rehabilitation and changes in lifestyle (e.g., nutrition, physical activity, weight reduction, and—in some cases—pharmacologic therapy), the prevalence of the syndrome can be reduced.

**Lipodystrophy** Lipodystrophic disorders in general are associated with the metabolic syndrome. Moreover, it is quite common for such patients to present with the metabolic syndrome. Both genetic

lipodystrophy (e.g., Berardinelli-Seip congenital lipodystrophy, Dunnigan familial partial lipodystrophy) and acquired lipodystrophy (e.g., HIV-related lipodystrophy and in HIV patients receiving certain antiretroviral therapies) may give rise to severe insulin resistance and many of the components of the metabolic syndrome.

## ETIOLOGY

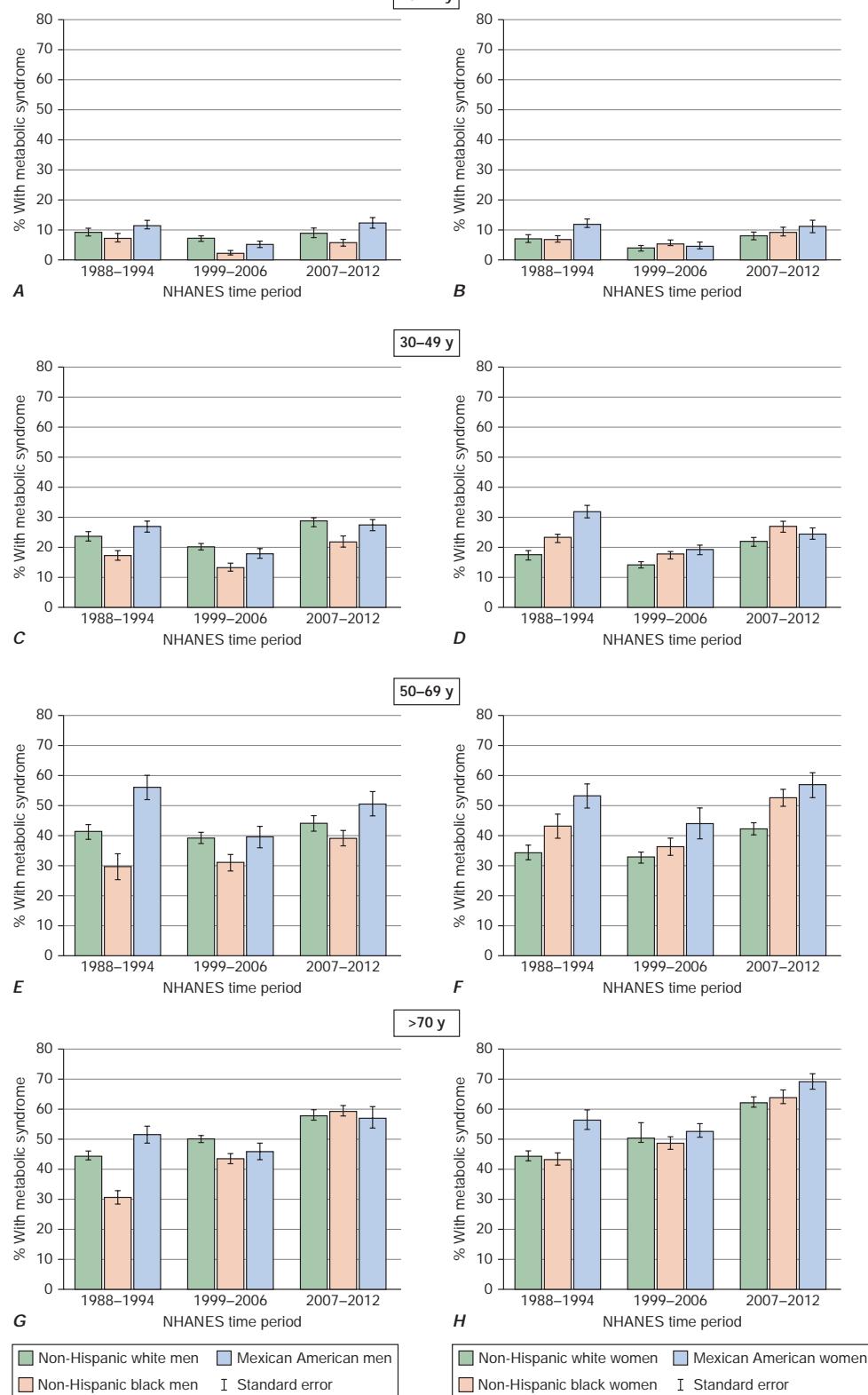
**Insulin Resistance** The most accepted and unifying hypothesis to describe the pathophysiology of the metabolic syndrome is insulin resistance, caused systemically by an incompletely understood defect in insulin action (**Chap. 403**). The onset of insulin resistance is heralded by postprandial hyperinsulinemia, which is followed by fasting hyperinsulinemia and ultimately by hyperglycemia.

An early major contributor to the development of insulin resistance is an overabundance of circulating fatty acids (**Fig. 408-2**). Plasma albumin-bound free fatty acids are derived predominantly from adipose-tissue triglyceride stores released by intracellular lipolytic enzymes. The lipolysis of triglyceride-rich lipoproteins in tissues by lipoprotein lipase also produces free fatty acids. Insulin mediates both anti-lipolysis and the stimulation of lipoprotein lipase in adipose tissue. Of note, the inhibition of lipolysis in adipose tissue is the most sensitive pathway of insulin action. Thus, when insulin resistance develops, increased lipolysis produces more fatty acids, which further decrease the anti-lipolytic effect of insulin. Excessive fatty acids enhance substrate availability and create insulin resistance by modifying downstream signaling. Fatty acids impair insulin-mediated glucose uptake and accumulate as triglycerides in both skeletal and cardiac muscle, whereas increased fatty acid flux increases glucose production and triglyceride production and accumulation in the liver.

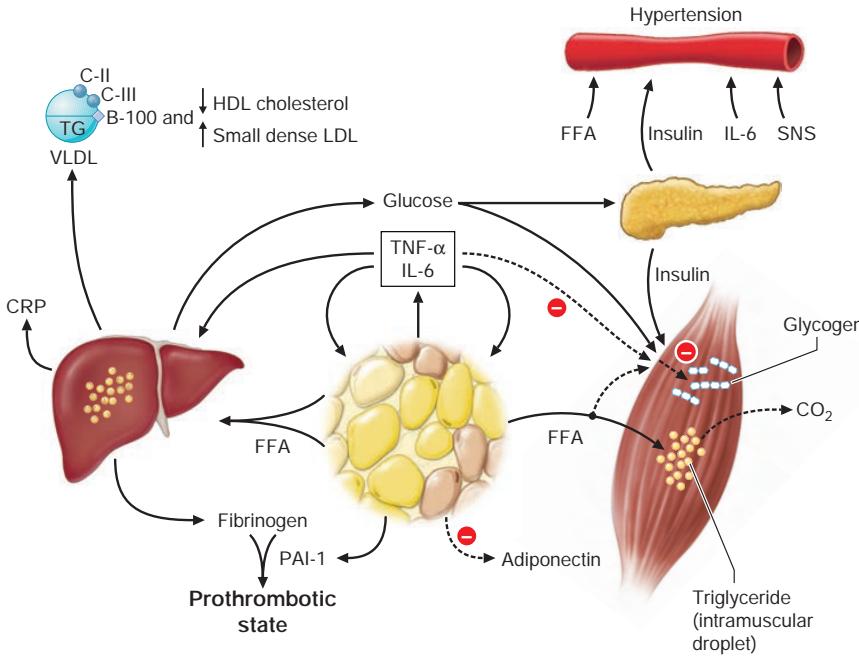
Leptin resistance also may be a pathophysiologic mechanism to explain the metabolic syndrome. Physiologically, leptin reduces appetite, promotes energy expenditure, and enhances insulin sensitivity when insulin resistance is associated with leptin deficiency. In addition, leptin may regulate cardiac and vascular function through a nitric oxide-dependent mechanism. However, when obesity develops, hyperleptinemia ensues, with evidence of leptin resistance in the brain and other tissues resulting in inflammation, insulin resistance, hyperlipidemia, and a plethora of cardiovascular disorders, such as hypertension, atherosclerosis, CHD, and heart failure.

The oxidative stress hypothesis provides a unifying theory for aging and the predisposition to the metabolic syndrome. In studies of insulin-resistant individuals with obesity or type 2 diabetes, the offspring of patients with type 2 diabetes, and the elderly, a defect in mitochondrial oxidative phosphorylation leads to the accumulation of triglycerides and related lipid molecules in muscle, liver, and perhaps other tissues, i.e., -cells.

Recently, the gut microbiome has emerged as an important contributor to the development of obesity and related metabolic disorders,



**FIGURE 408-1** Prevalence of metabolic syndrome among US adults over time by race/ethnicity-sex and age group, National Health and Nutrition Examination Survey (NHANES), 1988-2012. Metabolic syndrome was defined by using the criteria agreed to jointly by the International Diabetes Federation; the US National Heart, Lung, and Blood Institute in the United States; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity SE, standard error. (Reproduced with permission from JX Moore, N Chaudhary, T Akinyemiju. Metabolic syndrome prevalence by race/ethnicity and sex in the United States, National Health and Nutrition Examination Survey, 1988-2012. *Prev Chronic Dis* 14:E24, 2017.)



**FIGURE 408-2 Pathophysiology of the metabolic syndrome.** Free fatty acids (FFAs) are released in abundance from an expanded adipose tissue mass. In the liver, FFAs result in increased production of glucose and triglycerides and secretion of very-low-density lipoproteins (VLDLs). Associated lipid/lipoprotein abnormalities include reductions in high-density lipoprotein (HDL) cholesterol and an increased low-density lipoprotein (LDL) particle number. FFAs also reduce insulin sensitivity in muscle by inhibiting insulin-mediated glucose uptake. Associated defects include a reduction in glucose partitioning to glycogen and increased lipid accumulation in triglyceride (TG). The increase in circulating glucose, and to some extent FFAs, increases pancreatic insulin secretion, resulting in hyperinsulinemia. Hyperinsulinemia may result in enhanced sodium reabsorption and increased sympathetic nervous system (SNS) activity and contribute to hypertension, as might higher levels of circulating FFAs. The proinflammatory state is superimposed and contributory to the insulin resistance produced by excessive FFAs. The enhanced secretion of interleukin 6 (IL-6) and tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ) produced by adipocytes and monocyte-derived macrophages results in more insulin resistance and lipolysis of adipose tissue triglyceride stores to circulating FFAs. IL-6 and other cytokines also enhance hepatic glucose production, VLDL production by the liver, hypertension, and insulin resistance in muscle. Insulin resistance also contributes to increased triglyceride accumulation in the liver (nonalcoholic fatty liver disease). Cytokines and FFAs also increase hepatic production of fibrinogen and adipocyte production of plasminogen activator inhibitor 1 (PAI-1), resulting in a pro-thrombotic state. Higher levels of circulating cytokines stimulate hepatic production of C-reactive protein (CRP). Reduced production of the anti-inflammatory and insulin-sensitizing cytokine adiponectin is also associated with the metabolic syndrome. (Reproduced with permission from RH Eckel et al: The metabolic syndrome. Lancet 365:1415, 2005.)

including inflammation and components of the metabolic syndrome. Although the mechanisms remain uncertain, interaction among genetic predisposition, diet, bile acid metabolism, and the intestinal flora is important.

**Increased Waist Circumference** Waist circumference is an important component of the most recent and frequently applied diagnostic criteria for the metabolic syndrome. However, measuring waist circumference does not reliably distinguish increases in SC abdominal adipose tissue from that in visceral fat; this distinction requires CT or MRI. With increases in visceral adipose tissue, adipose tissue-derived free fatty acids reach the liver more readily. In contrast, increases in abdominal SC fat release lipolysis products into the systemic circulation and therefore have fewer direct effects on hepatic metabolism. Relative increases in visceral versus SC adipose tissue with increasing waist circumference in Asians and Asian Indians may explain the greater prevalence of the syndrome in those populations than in African-American men, in whom SC fat predominates. It is also possible that visceral fat is a marker for—but not the source of—excess postprandial free fatty acids in obesity.

**Dyslipidemia** (See also Chap. 407) In general, free fatty acid flux to the liver results in increased production of apolipoprotein (apo) B-containing, triglyceride-rich, very-low-density lipoproteins (VLDLs). The effect of insulin on this process is complex, but *hypertriglyceridemia* is an excellent marker of the insulin-resistant condition. Not only is hypertriglyceridemia a feature of the metabolic syndrome, but patients with the metabolic syndrome have elevated levels of apoC-III carried on VLDLs and other lipoproteins. This increase in apoC-III is inhibitory to lipoprotein lipase, reduces triglyceride-rich lipoprotein

remnant removal further contributing to hypertriglyceridemia, and confers more risk for atherosclerotic cardiovascular disease (ASCVD).

The other major lipoprotein disturbance in the metabolic syndrome is a *reduction in HDL cholesterol*. This reduction is a consequence of changes in HDL composition and metabolism. In the presence of hypertriglyceridemia, a decrease in the cholesterol content of HDL is a consequence of reduced cholesteryl ester content of the lipoprotein core in combination with cholesteryl ester transfer protein-mediated alterations in triglycerides that make the HDL particle small and dense. This change in lipoprotein composition also results in increased clearance of HDL from the circulation. These changes in HDL have a relationship to insulin resistance that is probably indirect, occurring in concert with the changes in triglyceride-rich lipoprotein metabolism.

In addition to HDLs, low-density lipoproteins (LDLs) have alterations in composition in the metabolic syndrome. With fasting serum triglycerides at  $>2.0$  mM ( $\sim 180$  mg/dL), there is usually a predominance of small dense LDLs, which are thought to be more atherogenic, although their association with hypertriglyceridemia and low HDLs make their independent contribution to ASCVD difficult to assess. Individuals with hypertriglyceridemia often have increases in cholesterol content of both VLDL1 and VLDL2 subfractions and in LDL particle number. Both lipoprotein changes may contribute to atherosclerosis in patients with the metabolic syndrome.

**Glucose Intolerance** (See also Chap. 403) Defects in insulin action in the metabolic syndrome lead to impaired suppression of glucose production by the liver (and kidney) and reduced glucose uptake and metabolism in insulin-sensitive tissues—i.e., muscle and adipose tissue. There is a strong relationship between impaired fasting

glucose or impaired glucose tolerance and insulin resistance in studies of humans, nonhuman primates, and rodents. To compensate for defects in insulin action, insulin secretion and/or clearance increases or decreases, respectively, so that euglycemia remains. Ultimately, this compensatory mechanism fails because of defects in insulin secretion, resulting in progression from impaired fasting glucose and/or impaired glucose tolerance to type 2 diabetes mellitus.

**Hypertension** The relationship between insulin resistance and hypertension is well established. Paradoxically, under normal physiologic conditions, insulin is a vasodilator with secondary effects on sodium reabsorption in the kidney. However, in the setting of insulin resistance, the vasodilatory effect of insulin is lost but the renal effect on sodium reabsorption is preserved. Sodium reabsorption is increased in Caucasians with the metabolic syndrome but not in Africans or Asians. Insulin also increases the activity of the sympathetic nervous system, an effect that is preserved in the setting of insulin resistance. Insulin resistance is also associated with pathway-specific impairment in phosphatidylinositol-3-kinase signaling. In the endothelium, this impairment may cause an imbalance between the production of nitric oxide and the secretion of endothelin 1, with a consequent decrease in blood flow. In addition, increases in angiotensinogen gene expression in adipose tissue of obese subjects results in increases in circulating angiotensin II and vasoconstriction. Although these mechanisms are provocative, the inadequate evaluation of insulin action by measurement of fasting insulin levels or by homeostasis model assessment shows that insulin resistance contributes only partially to the increased prevalence of hypertension in the metabolic syndrome.

Another possible mechanism underlying hypertension in the metabolic syndrome is the vasoactive role of perivascular adipose tissue. Reactive oxygen species released by NADPH oxidase impair endothelial function and result in local vasoconstriction. Other paracrine effects such as leptin or other proinflammatory cytokines released from adipose tissue, such as tumor necrosis factor (TNF- $\alpha$ ) may also be important.

Hyperuricemia is another consequence of insulin resistance in the metabolic syndrome. There is growing evidence not only that uric acid is associated with hypertension but also that reduction of uric acid normalizes blood pressure in hyperuricemic adolescents with hypertension. The mechanism appears to be in part related to an adverse effect of uric acid on nitric oxide synthase in the macula densa of the kidney and stimulation of the renin-angiotensin-aldosterone system.

**Proinflammatory Cytokines** The increases in proinflammatory cytokines—including interleukins 1, 6, and 18; resistin; TNF- $\alpha$ ; and the systemic biomarker C-reactive protein—reflect overproduction by the expanded adipose tissue mass (Fig. 408-2). Adipose tissue-derived macrophages may be the primary source of proinflammatory cytokines locally and in the systemic circulation. It remains unclear, however, how much of the insulin resistance is caused by the paracrine effects of these cytokines and how much by the endocrine effects.

**Adiponectin** Adiponectin is an anti-inflammatory cytokine produced exclusively by adipocytes. Adiponectin enhances insulin sensitivity and inhibits many steps in the inflammatory process. In the liver, adiponectin inhibits the expression of gluconeogenic enzymes and the rate of glucose production. In muscle, adiponectin increases glucose transport and enhances fatty acid oxidation, partially through the activation of AMP kinase. Reductions in adiponectin levels are common in the metabolic syndrome. The relative contributions of adiponectin deficiency and overabundance of the proinflammatory cytokines are unclear.

## CLINICAL FEATURES

**Symptoms and Signs** The metabolic syndrome typically is not associated with symptoms. On physical examination, waist circumference and blood pressure are often elevated. The presence of either or both signs should prompt the clinician to search for other biochemical abnormalities that may be associated with the metabolic syndrome. Much less frequently, lipoatrophy or acanthosis nigricans is present

on examination. Because these physical findings characteristically are associated with severe insulin resistance, other components of the metabolic syndrome are much more common.

**Associated Diseases • CARDIOVASCULAR DISEASE** The relative risk for new-onset CVD in patients with the metabolic syndrome who do not have diabetes averages 1.5- to 3-fold. However, in INTERHEART, a study of 26,903 subjects from 52 countries, the risk for acute myocardial infarction in subjects with the metabolic syndrome (World Health Organization or International Diabetes Federation definition) is comparable to that conferred by some, but not all, of the component risk factors. Diabetes mellitus (odds ratio [OR], 2.72) and hypertension (OR, 2.60) are stronger than other risk factors. Although congestive heart failure and the metabolic syndrome can occur together, typically this consequence is secondary to metabolic syndrome-related ASCVD or hypertension. Metabolic syndrome is also associated with increases in the risk for stroke, peripheral vascular disease, and Alzheimer's disease. However, as for myocardial infarction, the risk beyond the additive role of the components of the metabolic syndrome remains debatable. In the Reasons for Geographic and Racial Differences in Stroke (REGARDS) cohort, an observational study of black and white adults 45 years old across the United States, there were 9741 participants, and 41% had the metabolic syndrome. After adjustment for multiple confounders, metabolic syndrome was associated with increases in high-sensitivity C-reactive protein (hsCRP), and this relationship was associated with a 1.34 relative risk for all-cause mortality, but <50% of deaths were from CVD. The population-attributable risk was 9.5% for the metabolic syndrome alone and 14.7% for both metabolic syndrome and increased hsCRP. The relationship of metabolic syndrome and hsCRP to mortality was greater for whites than blacks.

**TYPE 2 DIABETES** Overall, the risk for type 2 diabetes among patients with the metabolic syndrome is increased three- to fivefold. In the Framingham Offspring Study's 8-year follow-up of middle-aged participants, the population-attributable risk of the metabolic syndrome for developing type 2 diabetes was 62% among men and 47% among women, yet increases in fasting plasma glucose explained most, if not all, of this increased risk.

**Other Associated Conditions** In addition to the features specifically used to define the metabolic syndrome, other metabolic alterations are secondary to or accompany insulin resistance. Those alterations include increases in apoB and apoC-III, uric acid, prothrombotic factors (fibrinogen, plasminogen activator inhibitor 1), serum viscosity, asymmetric dimethylarginine, homocysteine, white blood cell count, proinflammatory cytokines, C-reactive protein, urine albumin/creatinine ratio, nonalcoholic fatty liver disease (NAFLD) and/or nonalcoholic steatohepatitis (NASH), polycystic ovary syndrome, and obstructive sleep apnea.

**NONALCOHOLIC FATTY LIVER DISEASE** NAFLD has become the most common liver disease, in part a consequence of the insulin resistance of the metabolic syndrome. The mechanism relates to increases in free fatty acid flux and reductions in intrahepatic fatty acid oxidation with resultant increases in triglyceride biosynthesis and hepatocellular accumulation, with variable inflammation and oxidative stress. The more serious NASH, a consequence of NAFLD in some patients and a precursor of cirrhosis and end-stage liver disease, includes a more substantial proinflammatory contribution. NAFLD affects ~25% of the global population and up to 45% of patients with the metabolic syndrome; over half of these patients have NASH. As the prevalence of overweight/obesity and the metabolic syndrome increases, NASH may become one of the more common causes of end-stage liver disease and hepatocellular carcinoma.

**HYPURICEMIA** (See also Chap. 417) Hyperuricemia reflects defects in insulin action on the renal tubular reabsorption of uric acid and may contribute to hypertension through its effect on the endothelium. An increase in asymmetric dimethylarginine, an endogenous inhibitor of nitric oxide synthase, also relates to endothelial dysfunction. In addition, increases in the urine albumin/creatinine ratio may relate to altered endothelial pathophysiology in the insulin-resistant state.

**POLYCYSTIC OVARY SYNDROME (See also Chap. 392)** Polycystic ovary syndrome is highly associated with insulin resistance (50–80%) and the metabolic syndrome, with a prevalence of the syndrome between 12 and 60% based on phenotypes D through A.

**OBSTRUCTIVE SLEEP APNEA (See also Chap. 31)** Obstructive sleep apnea is commonly associated with obesity, hypertension, increased circulating cytokines, impaired glucose tolerance, and insulin resistance. In fact, obstructive sleep apnea may predict metabolic syndrome, even in the absence of excess adiposity. Moreover, when biomarkers of insulin resistance are compared between patients with obstructive sleep apnea and weight-matched controls, insulin resistance is found to be more severe in those with apnea. Continuous positive airway pressure treatment improves insulin sensitivity in patients with obstructive sleep apnea.

## DIAGNOSIS

The diagnosis of the metabolic syndrome relies on fulfillment of the criteria listed in Table 408-1, as assessed using tools at the bedside and in the laboratory. The medical history should include evaluation of symptoms for obstructive sleep apnea in all patients and polycystic ovary syndrome in premenopausal women. Family history will help determine risk for CVD and diabetes mellitus. Blood pressure and waist circumference measurements provide information necessary for the diagnosis.

**Laboratory Tests** Measurement of fasting lipids and glucose is needed in determining whether the metabolic syndrome is present. The measurement of additional biomarkers associated with insulin resistance can be individualized. Such tests might include those for apoB, hsCRP, fibrinogen, uric acid, urinary albumin/creatinine ratio, and liver function. A sleep study should be performed if symptoms of obstructive sleep apnea are present. If polycystic ovary syndrome is suspected based on clinical features and anovulation, testosterone, luteinizing hormone, and follicle-stimulating hormone should be measured. NAFLD can be further assessed by the NAFLD fibrosis score (FIB4) or elastography.

## TREATMENT

### The Metabolic Syndrome

#### LIFESTYLE (SEE ALSO CHAP. 402)

Obesity, particularly abdominal, is the driving force behind the metabolic syndrome. Thus, weight reduction is the primary approach to the disorder. With at least a 5% and more so with 10% weight reduction, improvement in insulin sensitivity results in favorable modifications in many components of the metabolic syndrome. In general, recommendations for weight loss include a combination of caloric restriction, increased physical activity, and behavior modification. Caloric restriction is the most important component, whereas increases in physical activity are important for maintenance of weight loss. Some but not all evidence suggests that the addition of exercise to caloric restriction may promote greater weight loss from the visceral depot. The tendency for weight regain after successful weight reduction underscores the need for long-lasting behavioral changes.

**Diet** Before prescribing a weight-loss diet, it is important to emphasize that it has taken the patient a long time to develop an expanded fat mass; thus, the correction need not occur quickly. Given that  $\sim 3500 \text{ kcal} = 1 \text{ lb of fat}$ , an  $\sim 500\text{-kcal}$  restriction daily equates to weight reduction of 1 lb per week. Diets restricted in carbohydrate typically provide a more rapid initial weight loss. However, after 1 year, the amount of weight reduction is minimally more reduced or no different from that with caloric restriction alone. Thus, adherence to the diet is more important than the chosen diet. Moreover, there is concern about low-carbohydrate diets enriched in saturated fat, particularly for patients at risk for ASCVD. Therefore, a high-quality dietary pattern—i.e., a diet enriched in fruits, vegetables, whole grains, lean poultry, and fish—should be encouraged to maximize overall health benefit.

**Physical Activity** Before prescribing a physical activity program to patients with the metabolic syndrome, it is important to ensure that the increased activity does not incur risk. Some high-risk patients should undergo formal cardiovascular evaluation before initiating an exercise program. For an inactive participant, gradual increases in physical activity should be encouraged to enhance adherence and avoid injury. Although increases in physical activity can lead to modest weight reduction, 60–90 min of moderate- to high-intensity daily activity is required to achieve this goal. Even if an overweight or obese adult is unable to undertake this level of activity, a health benefit will follow from at least 30 min of moderate-intensity activity daily. The caloric value of 30 min of a variety of activities can be found at <https://www.health.harvard.edu/diet-and-weight-loss/calories-burned-in-30-minutes-of-leisure-and-routine-activities>. Of note, a variety of routine activities, such as gardening, walking, and housecleaning, require moderate caloric expenditure. Thus, physical activity should not be defined solely in terms of formal exercise such as jogging, swimming, or tennis.

**Behavior Modification** Behavioral treatment typically includes recommendations for dietary restriction and more physical activity that predicts sufficient weight loss that benefits metabolic health. The subsequent challenge is the duration of the program because weight regain so often follows successful weight reduction. Improved long-term outcomes often follow a variety of methods, such as a personal or group counselor, the Internet, social media, and telephone follow-up to maintain contact between providers and patients.

**Obesity (See also Chap. 402)** In some patients with the metabolic syndrome, treatment options need to extend beyond lifestyle intervention. Weight-loss drugs come in two major classes: appetite suppressants and absorption inhibitors. Appetite suppressants approved by the U.S. Food and Drug Administration (FDA) include phentermine (for short-term use [3 months] only) as well as the more recent additions phentermine/topiramate, naltrexone/bupropion, high-dose (3.0 mg) liraglutide (rather than 1.8 mg, the maximum for treatment of type 2 diabetes), semaglutide (2.4 mg) which are approved without restrictions on the duration of therapy. In clinical trials, the phentermine/topiramate extended-release combination resulted in ~8% weight loss relative to placebo in 50% of patients. Side effects include palpitations, headache, paresthesias, constipation, and insomnia. Naltrexone/bupropion extended release reduces body weight by 10% in ~20% of patients; however, the drug combination is contraindicated in patients with seizure disorders or any condition that predisposes to seizures. Naltrexone/bupropion also increases pulse and blood pressure and should not be given to patients with uncontrolled hypertension. High-dose liraglutide, a glucagon-like peptide 1 (GLP-1) receptor agonist, results in ~6% weight loss relative to placebo with ~33% of patients with >10% weight loss. Common side effects are limited to the upper gastrointestinal tract, including nausea and, less frequently, emesis. Semaglutide (2.4 mg weekly), recently FDA approved has been shown to produce an average weight loss of 14.9% over 68 weeks.

Orlistat inhibits fat absorption by ~30% and is moderately effective compared with placebo (~4% more weight loss). Moreover, orlistat reduced the incidence of type 2 diabetes, an effect that was especially evident among patients with impaired glucose tolerance at baseline. This drug is often difficult to take because of oily leakage per rectum. In general, for all weight-loss drugs, greater weight reduction leads to greater improvement in metabolic syndrome components, including the conversion from prediabetes to type 2 diabetes.

Metabolic or bariatric surgery is an important option for patients with the metabolic syndrome who have a body mass index of  $>40 \text{ kg/m}^2$  or  $>35 \text{ kg/m}^2$  with comorbidities. An evolving application for metabolic surgery includes patients with a body mass index as low as  $30 \text{ kg/m}^2$  and type 2 diabetes. Gastric bypass or vertical sleeve gastrectomy results in dramatic weight reduction

and improvement in most features of the metabolic syndrome. A survival benefit with gastric bypass has also been realized.

#### **LDL CHOLESTEROL (SEE ALSO CHAP. 407)**

The rationale for the development of criteria for the metabolic syndrome by NCEP was to go beyond LDL cholesterol in identifying and reducing the risk of ASCVD. The working assumption by the panel was that LDL cholesterol goals had already been achieved and that increasing evidence supports a linear reduction in ASCVD events because of progressive lowering of LDL cholesterol with statins with subsequent benefit using additional LDL cholesterol-lowering agents. The 2019 American College of Cardiology (ACC)/American Health Association (AHA) Cholesterol Guidelines have no specific recommendations for patients with the metabolic syndrome; however, they recommend that patients aged 20–75 years with LDL cholesterol levels  $>190$  mg/dL should use a high-intensity statin (e.g., atorvastatin 40–80 mg or rosuvastatin 20–40 mg daily) and those with type 2 diabetes aged 40–75 years should use a moderate-intensity statin and, if or when risk estimate is high, a high-intensity statin. For patients with the metabolic syndrome but without diabetes, the 10-year ASCVD risk estimator should be employed, and patients with a risk 7.5% and 20% or patients aged 20–59 with elevated lifetime risk should have a discussion with their provider about initiating statin therapy for primary prevention of ASCVD. A coronary calcium score may help in making this decision.

Diets restricted in saturated fats (<6% of calories) and trans fats (as few as possible) should be applied aggressively. Although evidence is controversial, dietary cholesterol can also be restricted. If LDL cholesterol remains elevated, pharmacologic intervention is needed. Based on substantial evidence, treatment with statins, which lower LDL cholesterol by 15–60%, is the first-choice medication intervention. Of note, for each doubling of the statin dose, LDL cholesterol is further lowered by only ~6%. Hepatotoxicity (more than a threefold increase in hepatic aminotransferases) is rare, but myopathy occurs in ~10–20% of patients. The cholesterol absorption inhibitor ezetimibe is well tolerated and should be the second-choice medication intervention. Ezetimibe typically reduces LDL cholesterol by 15–20%. Bempedoic acid alone or in combination with ezetimibe is another option, with up to a 35% lowering of LDL cholesterol with the combination. Bempedoic acid can increase plasma uric acid. Proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors are potent LDL cholesterol-lowering drugs (~45–60%) but are not needed for most patients with the metabolic syndrome. Of course, if these patients also have familial hypercholesterolemia or insufficient LDL cholesterol lowering on statins with or without ezetimibe, a PCSK9 inhibitor should be considered. The bile acid sequestrants colestipol, and coleselvam may be more effective than ezetimibe alone, but because they can increase triglyceride levels, they must be used with caution in patients with the metabolic syndrome when fasting triglycerides are  $>300$  mg/dL. Side effects include gastrointestinal symptoms (palatability, bloating, belching, constipation, anal irritation). Nicotinic acid has similar LDL cholesterol-lowering capabilities (<20%). Fibrates are best employed to lower LDL cholesterol when triglycerides are not elevated. Fenofibrate may be more effective than gemfibrozil in this setting.

#### **TRIGLYCERIDES (SEE ALSO CHAP. 407)**

The 2019 ACC/AHA Cholesterol Guidelines stated that fasting triglycerides  $>500$  mg/dL should be treated to prevent more serious hypertriglyceridemia and pancreatitis. Although a fasting triglyceride value of  $>150$  mg/dL is a component of the metabolic syndrome, post hoc analyses of multiple fibrate trials have suggested reduction in the primary ASCVD outcome in patients (with or without concomitant statin therapy) with fasting triglycerides  $>200$  mg/dL, often in the setting of reduced levels of HDL cholesterol. It remains uncertain whether triglycerides cause ASCVD or if levels are just associated with increased ASCVD risk.

A fibrate (gemfibrozil or fenofibrate) is one drug class of choice to lower fasting triglyceride levels, which are typically reduced by 30–45%. Concomitant administration with drugs metabolized by the 3A4 cytochrome P450 system (including some statins) increases the risk of myopathy. In these cases, fenofibrate may be preferable to gemfibrozil. In the Veterans Affairs HDL Intervention Trial, gemfibrozil was administered to men with known CHD and levels of HDL cholesterol  $<40$  mg/dL. A coronary disease event and mortality rate benefit was experienced predominantly among men with hyperinsulinemia and/or diabetes, many of whom were identified retrospectively as having the metabolic syndrome. Of note, the degree of triglyceride lowering in this trial or other fibrate trials did not predict benefit.

Other drugs that lower triglyceride levels include statins, nicotinic acid, and prescription omega-3 fatty acids. For this purpose, an intermediate or high dose of the “more potent” statins (atorvastatin, rosuvastatin) is needed. The effect of nicotinic acid on fasting triglycerides is dose related and ~20–35%, an effect that is less pronounced than that of fibrates. In patients with the metabolic syndrome and diabetes, nicotinic acid may increase fasting glucose levels, and clinical trials with nicotinic acid plus a statin have failed to reduce ASCVD events. Prescriptions of omega-3 fatty acid preparations that include high doses of eicosapentaenoic acid (EPA) with or without docosahexaenoic acid (DHA) (~1.5–4.5 g/d) lower fasting triglyceride levels by ~25–40%. The two omega-3 randomized controlled trials associated with ASCVD risk reduction, JELIS and REDUCE-IT, used EPA only, whereas STRENGTH, which was terminated prematurely because of futility, used EPA plus DHA. Here, no drug interactions with fibrates or statins occur, and the main side effect of their use is eructation with a fishy taste. Freezing the nutraceutical can partially block this unpleasant side effect. Importantly, lowering triglycerides with any of the pharmaceuticals has not proven to be an independent predictor of CVD outcomes.

#### **HDL CHOLESTEROL (SEE ALSO CHAP. 407)**

Very few lipid-modifying compounds increase HDL cholesterol levels. Statins, fibrates, and bile acid sequestrants have modest effects (5–10%), whereas ezetimibe and omega-3 fatty acids have no effect. Nicotinic acid is the only currently available drug with predictable HDL cholesterol-raising properties. The response is dose related, and nicotinic acid can increase HDL cholesterol by up to 30% above baseline. After several trials of nicotinic acid versus placebo in statin-treated patients, there is no evidence that raising HDL cholesterol with nicotinic acid beneficially affects ASCVD events in patients with or without the metabolic syndrome.

#### **BLOOD PRESSURE (SEE ALSO CHAP. 277)**

The direct relationship between blood pressure and all-cause mortality rate has been well established in studies comparing patients with hypertension ( $>140/90$  mmHg), patients with prehypertension ( $>120/80$  mmHg but  $<140/90$  mmHg), and individuals with normal blood pressure ( $<120/80$  mmHg). In patients who have the metabolic syndrome without diabetes, the best choice for the initial antihypertensive medication is an angiotensin-converting enzyme (ACE) inhibitor or an angiotensin II receptor blocker, as these two classes of drugs are effective and well tolerated. Additional agents include a diuretic, calcium channel blocker, beta blocker, and mineralocorticoid inhibitor, such as the recently FDA approved mineralocorticoid receptor antagonist finerenone. In all patients with hypertension, a sodium-restricted dietary pattern enriched in fruits and vegetables, whole grains, and low-fat dairy products should be advocated. Home monitoring of blood pressure may assist in maintaining good blood pressure control.

#### **IMPAIRED FASTING GLUCOSE (SEE ALSO CHAP. 403)**

In patients with the metabolic syndrome and type 2 diabetes, aggressive glycemic control may favorably modify fasting levels of triglycerides and/or HDL cholesterol. In patients with impaired fasting glucose who do not have diabetes, a lifestyle intervention that includes weight reduction, dietary saturated fat restriction, and increased physical activity has been shown to reduce the incidence

of type 2 diabetes. Metformin also reduces the incidence of diabetes, although the effect is less pronounced than that of lifestyle intervention.

#### INSULIN RESISTANCE (SEE ALSO CHAP. 404)

Several drug classes (biguanides, thiazolidinediones [TZDs]) increase insulin sensitivity. Because insulin resistance is the primary pathophysiologic mechanism for the metabolic syndrome, representative drugs in these classes reduce its prevalence. Both metformin and TZDs enhance insulin action in the liver and suppress endogenous glucose production. TZDs, but not metformin, also improve insulin-mediated glucose uptake in muscle and adipose tissue. In a meta-analysis of nine trials involving 12,026 participants, the TZD pioglitazone versus placebo was associated with reduction in ASCVD events in patients with insulin resistance (metabolic syndrome), prediabetes and type 2 diabetes. However, adverse effects including weight gain, bone fracture, and congestive heart failure with or without edema were seen. Benefit of TZDs has been seen in patients with NAFLD, and with metformin in women with polycystic ovary syndrome, and both drug classes have been shown to reduce markers of inflammation.

#### FURTHER READING

- Alberti KG et al: Harmonizing the metabolic syndrome: A joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation* 120:1640, 2009.
- Brown AE, Walker M: Genetics of insulin resistance and the metabolic syndrome. *Curr Cardiol Rep* 18:75, 2016.
- Eckel RH et al: The metabolic syndrome. *Lancet* 365:1415, 2005.
- Genser L et al: Obesity, type 2 diabetes, and the metabolic syndrome: Pathophysiologic relationships and guidelines for surgical intervention. *Surg Clin North Am* 96:681, 2016.
- Lechner K et al: High-risk atherosclerosis and metabolic phenotype: The roles of ectopic adiposity, atherogenic dyslipidemia, and inflammation. *Metab Syndr Relat Disord* 18:176, 2020.
- Lim S, Eckel RH: Pharmacological treatment and therapeutic perspective of metabolic syndrome. *Rev Endocr Metab Disord* 15:329, 2014.
- Neel and JJ et al: Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: A position statement. *Lancet Diabetes Endocrinol* 7:715, 2019.

calcium, magnesium, phosphorus, sodium, and other ions necessary for homeostatic functions. Bone also hosts and regulates hematopoiesis by providing niches for hematopoietic cell proliferation and differentiation. The skeleton is highly vascular and receives ~10% of the cardiac output. Remodeling of bone is accomplished by two distinct cell types: osteoblasts produce bone matrix, and osteoclasts resorb the matrix. The activities of these cells are coordinated by osteocytes, long-lived regulatory cells embedded within bone matrix.

The extracellular components of bone consist of a solid mineral phase in close association with an organic matrix, of which 90–95% is type I collagen (Chap. 413). The noncollagenous portion of the organic matrix is heterogeneous and contains serum proteins such as albumin as well as many locally produced proteins, whose functions are incompletely understood. Those proteins include cell attachment/signaling proteins such as thrombospondin, osteopontin, and fibronectin; calcium-binding proteins such as matrix gla protein and osteocalcin; and proteoglycans such as biglycan and decorin. Some of the proteins organize collagen fibrils; others influence mineralization and binding of the mineral phase to the matrix.

The mineral phase is made up of calcium and phosphate and is best characterized as a poorly crystalline hydroxyapatite. The mineral phase of bone is deposited initially in intimate relation to the collagen fibrils and is found in specific locations in the “holes” between the collagen fibrils. This architectural arrangement of mineral and matrix results in a two-phase material well suited to withstand mechanical stresses. The organization of collagen influences the amount and type of mineral phase formed in bone. Although the primary structures of type I collagen in skin and bone tissues are similar, there are differences in posttranslational modifications and distribution of intermolecular cross-links. The holes in the packing structure of the collagen are larger in mineralized collagen of bone and dentin than in unmineralized collagens such as those in tendon. Single amino acid substitutions in the helical portion of either the 1 (*COL1A1*) or 2 (*COL1A2*) chains of type I collagen disrupt the organization of bone in the disease, osteogenesis imperfecta. The severe skeletal fragility associated with this group of disorders highlights the importance of the fibrillar matrix in the structure of bone (Chap. 413).

Osteoblasts synthesize and secrete the organic matrix and regulate its mineralization. They are derived from cells of mesenchymal origin (Fig. 409-1A). Active osteoblasts are found on the surface of newly forming bone. As an osteoblast secretes matrix, which then is mineralized, the cell may become an osteocyte, still connected with its nutrient supply through a series of canaliculi. Osteocytes account for the vast majority of the cells in bone. They are thought to be the mechanosensors in bone that communicate signals to surface osteoblasts and osteoclasts and their progenitors through the canalicular network and thereby serve as master regulators of bone formation and resorption. Remarkably, osteocytes also secrete fibroblast growth factor 23 (FGF23), a major hormonal regulator of phosphate metabolism (see below). Mineralization of the matrix, both in trabecular bone and in osteones of compact cortical bone (*Haversian systems*), begins soon after the matrix is secreted (primary mineralization) but is not completed for several weeks or even longer (secondary mineralization). Although this mineralization takes advantage of the high concentrations of calcium and phosphate, already near saturation in serum, mineralization is a carefully regulated process that is dependent on the activity of osteoblast-derived alkaline phosphatase, which probably works by hydrolyzing inhibitors of mineralization, such as pyrophosphate.

Genetic studies in humans and mice have identified several key genes that control osteoblast development. *Runx2* is a transcription factor expressed specifically in chondrocyte (cartilage cells) and osteoblast progenitors as well as in hypertrophic chondrocytes and mature osteoblasts. *Runx2* regulates the expression of several important osteoblast proteins, including osterix (another transcription factor needed for osteoblast maturation), osteopontin, bone sialoprotein, type I collagen, osteocalcin, and receptor-activator of nuclear factor (NF)-B (RANK) ligand. *Runx2* expression is regulated in part by bone morphogenic proteins (BMPs). *Runx2*-deficient mice are devoid of

## Section 4 Disorders of Bone and Mineral Metabolism

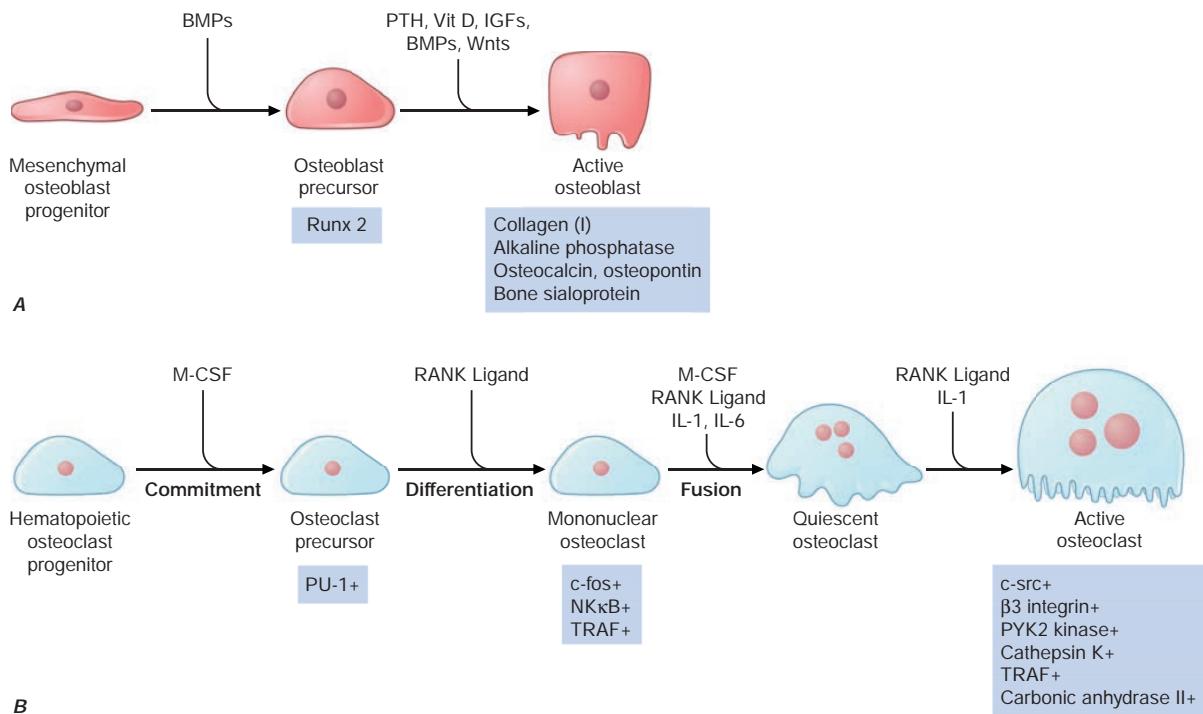
### 409

### Bone and Mineral Metabolism in Health and Disease

F. Richard Bringhurst, Henry M. Kronenberg, Eva S. Liu

#### BONE STRUCTURE AND METABOLISM

Bone is a dynamic tissue that is remodeled constantly throughout life. The arrangement of compact and cancellous bone provides strength and density suitable for both mobility and protection. Compact or cortical bone forms the roughly cylindrical shell of long bones; cancellous or trabecular bone forms the plate-like meshwork that internally supports the cortical shell. In addition, bone provides a reservoir for



**FIGURE 409-1** Pathways regulating development of (A) osteoblasts and (B) osteoclasts. Hormones, cytokines, and growth factors that control cell proliferation and differentiation are shown above the arrows. Transcription factors and other markers specific for various stages of development are depicted below the arrows. BMPs, bone morphogenic proteins; IGFs, insulin-like growth factors; IL-1, interleukin 1; IL-6, interleukin 6; M-CSF, macrophage colony-stimulating factor; NFKB, nuclear factor- $\kappa$ B; PTH, parathyroid hormone; PU-1, a monocyte- and B lymphocyte-specific ets family transcription factor; RANK ligand, receptor activator of NFKB ligand; Runx2, Runt-related transcription factor 2; TRAF, tumor necrosis factor receptor-associated factor; Vit D, vitamin D; wnts, wingless-type mouse mammary tumor virus integration site. (Modified with permission from T Suda et al: Modulation of osteoclast differentiation and function by the new members of the tumor necrosis factor receptor and ligand families. *Endocr Rev* 20:345, 1999.)

osteoblasts, whereas mice with a deletion of only one allele (*Runx2*<sup>+/−</sup>) exhibit a delay in formation of the clavicles and some cranial bones. The latter abnormalities are similar to those in the human disorder *cleidocranial dysplasia*, which is also caused by heterozygous inactivating mutations in *Runx2*.

The paracrine signaling molecule, Indian hedgehog (Ihh), also plays a critical role in osteoblast development, as evidenced by Ihh-deficient mice that lack osteoblasts in the type of bone formed on a cartilage mold (endochondral ossification). Signals originating from members of the wnt (an amalgam of “wingless,” the *Drosophila* developmental gene and “int-1,” an analogous mammalian gene activated by integration of a mouse tumor viral genome nearby) family of paracrine factors are also important for osteoblast proliferation and differentiation. Osteocytes regulate osteoblasts partly by secreting a potent inhibitor of wnt signaling called sclerostin. Numerous other growth-regulatory factors affect osteoblast function, including the three closely related transforming growth factor  $\beta$ s, fibroblast growth factors (FGFs) 2 and 18, platelet-derived growth factor, and insulin-like growth factors (IGFs) I and II. Hormones such as parathyroid hormone (PTH) and 1,25-dihydroxyvitamin D [1,25(OH)<sub>2</sub>D] activate receptors expressed by osteoblasts to assure mineral homeostasis and influence a variety of bone cell functions. Osteoclasts that resorb bone (see below) also regulate osteoblasts by releasing growth factors from bone matrix and by synthesizing proteins that can directly regulate osteoblastogenesis.

Resorption of bone is carried out mainly by *osteoclasts*, multinucleated cells that are formed by fusion of cells derived from the common precursor of macrophages and osteoclasts. Thus, these cells derive from the hematopoietic lineage, quite different from the mesenchymal lineage cells that become osteoblasts. Multiple factors that regulate osteoclast development have been identified (Fig. 409-1B). Factors produced by osteocytes, osteoblasts, and marrow stromal cells allow cells of the osteoblast lineage to control osteoclast development and

activity. Macrophage colony-stimulating factor (M-CSF) plays a critical role during several steps in the pathway and ultimately leads to fusion of osteoclast progenitor cells to form multinucleated, active osteoclasts. RANK ligand, a member of the tumor necrosis factor (TNF) family, is expressed on the surface of osteocytes, osteoblasts, and stromal fibroblasts. In a process involving cell-cell interactions, RANK ligand binds to the RANK receptor on osteoclast progenitors, stimulating osteoclast differentiation and activation. Alternatively, a soluble decoy receptor, referred to as osteoprotegerin, can bind RANK ligand and inhibit osteoclast differentiation. Several growth factors and cytokines (including interleukins 1, 6, and 11; TNF; and interferon  $\gamma$ ) modulate osteoclast differentiation and function. Most hormones that influence osteoclast function do not target these cells directly but instead target cells of the osteoblast lineage to increase production of M-CSF and RANK. Both PTH and 1,25(OH)<sub>2</sub>D increase osteoclast number and activity by this indirect mechanism. Calcitonin, in contrast, binds to its receptor on the basal surface of osteoclasts and directly inhibits osteoclast function. Estradiol has multiple cellular targets in bone, including osteoclasts, immune cells, and osteoblasts; actions on all these cells serve to decrease osteoclast number and decreased bone resorption.

Osteoclast-mediated resorption of bone takes place in scalloped spaces (*Howship's lacunae*) where the osteoclasts are attached through a specific  $\alpha$  v  $\beta$  3 integrin to components of the bone matrix such as osteopontin. The osteoclast forms a tight seal to the underlying matrix and secretes protons, chloride, and proteinases into a confined space that has been likened to an extracellular lysosome. The active osteoclast surface forms a ruffled border that contains a specialized proton pump ATPase that secretes acid that solubilizes the mineral phase. Carbonic anhydrase (type II isoenzyme) within the osteoclast generates the needed protons. The bone matrix is resorbed in the acid environment adjacent to the ruffled border by proteases, such as cathepsin K, that act at low pH.

In the embryo and the growing child, bone develops mostly by replacing previously calcified cartilage (endochondral bone formation) with subsequent remodeling or, in a few bones, is formed without a cartilage matrix (intramembranous bone formation). During endochondral bone formation, chondrocytes proliferate, secrete and mineralize a matrix, enlarge (hypertrophy), and then die, enlarging bone and providing the matrix and factors that stimulate endochondral bone formation. This program is regulated by both local factors such as IGF-I and -II, Ihh, parathyroid hormone-related peptide (PTHrP), BMPs, and FGFs and by systemic hormones such as growth hormone, glucocorticoids, and estrogen. Some hypertrophic chondrocytes do not die but, instead, through still poorly understood steps, can become osteoblasts and bone stromal cells.

New bone, whether formed in infants or in adults during repair, has a relatively high ratio of cells to matrix and is characterized by coarse fiber bundles of collagen that are interlaced and randomly dispersed (woven bone). In adults, the more mature bone is organized with fiber bundles regularly arranged in parallel or concentric sheets (lamellar bone). In long bones, deposition of lamellar bone in a concentric arrangement around blood vessels forms the Haversian systems. Growth in length of bones is dependent on proliferation of cartilage cells and the endochondral sequence at the growth plate. Growth in width and thickness is accomplished by formation of bone at the periosteal surface and by resorption at the endosteal surface, with the rate of formation exceeding that of resorption. In adults, after the growth plates of cartilage close through the actions of estrogen, growth in length and endochondral bone formation cease. Even in adults, however, remodeling of bone (within Haversian systems as well as along the surfaces of trabecular bone) continues throughout life. In adults, ~4% of the surface of trabecular bone (such as iliac crest) is involved in active resorption, whereas 10–15% of trabecular surfaces are covered with osteoid, unmineralized new bone formed by osteoblasts. Radioisotope studies indicate that as much as 18% of the total skeletal calcium is deposited and removed each year. Thus, bone is an active metabolizing tissue that requires an intact blood supply. The cycle of bone resorption and formation is a highly orchestrated process, directed by osteocytes and carried out by the basic multicellular unit, which is composed of a group of osteoclasts and osteoblasts (Fig. 409-2).

The response of bone to fractures, infection, and interruption of blood supply and to expanding lesions is relatively limited. Dead bone must be resorbed, and new bone must be formed, a process carried out in association with growth of new blood vessels into the involved area. In injuries that disrupt the organization of the tissue, such as a fracture, in which apposition of fragments is poor or when motion exists at the fracture site, progenitor stromal cells recapitulate the endochondral bone formation of early development and form cartilage

that is replaced by bone and, variably, fibrous tissue. When there is good apposition with fixation and little motion at the fracture site, repair occurs predominantly by formation of new bone without other mediating tissue.

Remodeling of bone occurs along lines of force generated by mechanical stress. The signals from these mechanical stresses are sensed by osteocytes, which transmit signals to osteoclasts and osteoblasts or their precursors. One such signal made by osteocytes is sclerostin, an inhibitor of wnt signaling. Mechanical forces suppress sclerostin production and thus increase bone formation by osteoblasts. Expanding lesions in bone such as tumors induce resorption at the surface in contact with the tumor by producing ligands such as PTHrP that stimulate osteoclast differentiation and function. Thus, bone plasticity reflects the interaction of cells with each other and with the environment.

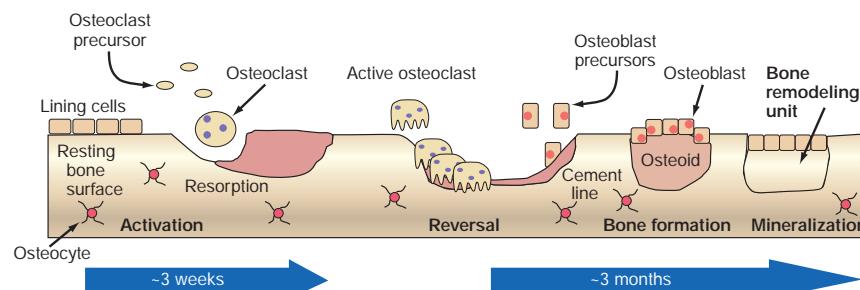
Measurement of the products of osteoblast and osteoclast activity can assist in the diagnosis and management of bone diseases. Osteoblast activity can be assessed by measuring serum bone-specific alkaline phosphatase. Similarly, osteocalcin, a protein secreted from osteoblasts, is made virtually only by osteoblasts. Measurement of an amino-terminal fragment of procollagen I is also an effective index of bone formation. Osteoclast activity can be assessed by measurement of products of collagen degradation. Collagen molecules are covalently linked to each other in the extracellular matrix through the formation of hydroxypyridinium cross-links (Chap. 413). After digestion by osteoclasts, these cross-linked peptides can be measured both in urine and in blood.

## CALCIUM METABOLISM

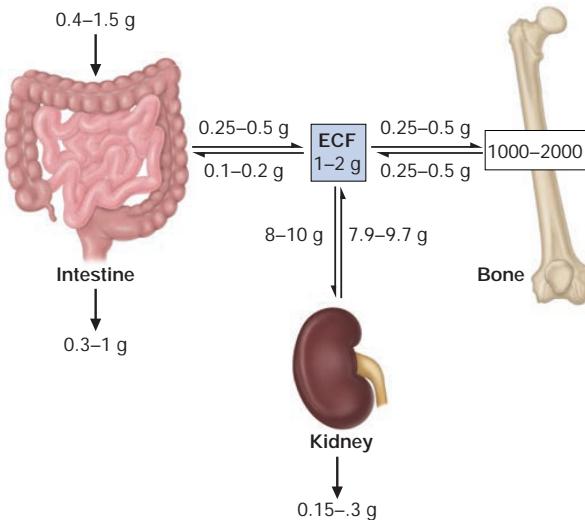
Over 99% of the 1–2 kg of calcium present normally in the adult human body resides in the skeleton, where it provides mechanical stability and serves as a reservoir sometimes needed to maintain extracellular fluid (ECF) calcium concentration (Fig. 409-3). Skeletal calcium accretion first becomes significant during the third trimester of fetal life, accelerates throughout childhood and adolescence, reaches a peak in early adulthood, and gradually declines thereafter at rates that rarely exceed 1–2% per year. These slow changes in total skeletal calcium content contrast with relatively high daily rates of closely matched fluxes of calcium into and out of bone (~250–500 mg each), a process mediated by coupled osteoblastic and osteoclastic activity. Another 0.5–1% of skeletal calcium is freely exchangeable (e.g., in chemical equilibrium) with that in the ECF.

The concentration of ionized calcium in the ECF must be maintained within a narrow range because of the critical role calcium plays in a wide array of cellular functions, especially those involved in neuromuscular activity, secretion, and signal transduction. Intracellular cytosolic free calcium levels are ~100 nmol/L and are 10,000-fold lower than ionized calcium concentration in the blood and ECF (1.1–1.3 mmol/L).

Cytosolic calcium does not play the structural role played by extracellular calcium; instead, it serves a signaling function. The steep chemical gradient of calcium from outside to inside the cell promotes rapid calcium influx through various membrane calcium channels that can be activated by hormones, metabolites, or neurotransmitters, swiftly changing cellular function. In blood, total calcium concentration is normally 2.2–2.6 mM (8.5–10.5 mg/dL), of which ~50% is ionized. The remainder is bound ionically to negatively charged proteins (predominantly albumin and immunoglobulins) or loosely complexed with phosphate, citrate, sulfate, or other anions. Alterations in serum protein concentrations directly affect the total blood calcium concentration even if the ionized



**FIGURE 409-2 Schematic representation of bone remodeling.** The cycle of bone remodeling is carried out by the basic multicellular unit (BMU), which consists of a group of osteoclasts and osteoblasts. In cortical bone, the BMU moves through the tissue, whereas in cancellous bone, they move across the trabecular surface. The process of bone remodeling is initiated by the recruitment of osteoclast precursors, perhaps to sites of microdamage. These precursors fuse to form multinucleated, active osteoclasts that mediate bone resorption. Osteoclasts adhere to bone and subsequently remove it by acidification and proteolytic digestion. As the BMU advances, osteoclasts leave the resorption site and osteoblasts, derived from marrow precursors and previously inactive bone lining cells, move in to cover the excavated area and begin the process of new bone formation by secreting osteoid, which eventually is mineralized into new bone. After osteoid mineralization, osteoblasts flatten and form a layer of lining cells over new bone, become osteocytes, or die.



**FIGURE 409-3 Calcium homeostasis.** Schematic illustration of calcium content of extracellular fluid (ECF) and bone as well as of diet and feces; magnitude of calcium flux per day as calculated by various methods is shown at sites of transport in intestine, kidney, and bone. Ranges of values shown are approximate and were chosen to illustrate certain points discussed in the text. In conditions of calcium balance, rates of calcium release from and uptake into bone are equal.

calcium concentration remains normal. An algorithm to correct for protein changes adjusts the total serum calcium (in mg/dL) upward by 0.8 times the deficit in serum albumin (g/dL) or by 0.5 times the deficit in serum immunoglobulin (in g/dL). Such corrections provide only rough approximations of actual free calcium concentrations, however, and may be misleading, particularly during acute illness. Acidosis also alters ionized calcium by reducing its association with proteins. The best practice is to measure blood ionized calcium directly by a method that employs calcium-selective electrodes in acute settings during which calcium abnormalities might occur.

Control of the ionized calcium concentration in the ECF ordinarily is accomplished by adjusting the rates of calcium movement across intestinal and renal epithelia and into and out of bone. These adjustments are mediated mainly via changes in blood levels of the hormones PTH and  $1,25(\text{OH})_2\text{D}$ . Acting via binding to calcium-sensing receptors (CaSRs) on the surface of parathyroid cells, blood ionized calcium suppresses PTH secretion by reducing levels of PTH mRNA and promoting the cleavage of PTH to inactive peptides. Also, ionized calcium indirectly affects PTH secretion by lowering  $1,25(\text{OH})_2\text{D}$  production. This active vitamin D metabolite inhibits PTH production by an incompletely understood mechanism of negative feedback (Chap. 410).

Normal dietary calcium intake in the United States varies widely, ranging from 10 to 37 mmol/d (400–1500 mg/d). A National Academy of Medicine (formerly, Institute of Medicine) analysis recommends a daily allowance of 25–30 mmol (1000–1200 mg) for most adults. Intestinal absorption of ingested calcium involves both active (transcellular) and passive (paracellular) mechanisms. Passive calcium absorption is nonsaturable and approximates 5% of daily calcium intake, whereas active absorption involves apical calcium entry via specific ion channels (TRPV5 in the kidney and TRPV6 in the intestine), whose expression is controlled principally by  $1,25(\text{OH})_2\text{D}$ . This active transport mechanism normally accounts for absorption of 20–70% of dietary calcium. Active calcium transport occurs mainly in the proximal small bowel (duodenum and proximal jejunum), although some active calcium absorption occurs in most segments of the small intestine. Optimal rates of calcium absorption require gastric acid. This is especially true for weakly dissociable calcium supplements such as calcium carbonate. In fact, large boluses of calcium carbonate are poorly absorbed because of their neutralizing effect on gastric acid. In achlorhydric subjects and for those taking drugs that inhibit gastric

acid secretion, supplements should be taken with meals to optimize their absorption. Use of calcium citrate may be preferable in these circumstances. Calcium absorption may also be blunted in disease states such as pancreatic or biliary insufficiency, in which ingested calcium remains bound to unabsorbed fatty acids or other food constituents. At high levels of calcium intake, synthesis of  $1,25(\text{OH})_2\text{D}$  is reduced; this decreases the rate of active intestinal calcium absorption. The opposite occurs with dietary calcium restriction. Some calcium, ~2.5–5 mmol/d (100–200 mg/d), is excreted as an obligate component of intestinal secretions and is not regulated by calcitropic hormones.

The feedback-controlled hormonal regulation of intestinal absorptive efficiency results in a relatively constant daily net calcium absorption of ~5–10 mmol/d (200–400 mg/d) despite large changes in daily dietary calcium intake. This daily load of absorbed calcium is excreted by the kidneys in a manner that is also tightly regulated by the concentration of ionized calcium in the blood. Approximately 8–10 g/d of calcium is filtered by the glomeruli, of which only 2–3% appears in the urine. Most filtered calcium (65%) is reabsorbed in the proximal tubules via a passive, paracellular route that is coupled to concomitant NaCl reabsorption and not specifically regulated. The cortical thick ascending limb of Henle's loop (cTAL) reabsorbs roughly another 20% of filtered calcium, also via a paracellular mechanism. Calcium reabsorption in the cTAL requires a tight-junctional protein called paracellin-1 and is inhibited by increased blood concentrations of calcium or magnesium, acting via the CaSR, which is highly expressed on basolateral membranes in this nephron segment. Operation of the renal CaSR provides a mechanism, independent of those engaged directly by PTH or  $1,25(\text{OH})_2\text{D}$ , by which serum ionized calcium can control renal calcium reabsorption. Finally, ~10% of filtered calcium is reabsorbed in the distal convoluted tubules (DCTs) by a transcellular mechanism. Calcium enters the luminal surface of the cell through specific apical calcium channels (TRPV5), whose number is regulated. It then moves across the cell in association with a specific calcium-binding protein (calbindin-D28k) that buffers cytosolic calcium concentrations from the large mass of transported calcium.  $\text{Ca}^{2+}$ -ATPases and  $\text{Na}^+/\text{Ca}^{2+}$  exchangers actively extrude calcium across the basolateral surface and thereby maintain the transcellular calcium gradient. All these processes are stimulated directly or indirectly by PTH. The DCT is also the site of action of thiazide diuretics, which lower urinary calcium excretion by inducing sodium depletion and thereby augmenting proximal calcium reabsorption. Conversely, dietary sodium loads, or increased distal sodium delivery caused by loop diuretics or saline infusion, induce calciuresis.

The homeostatic mechanisms that normally maintain a constant serum ionized calcium concentration may fail at extremes of calcium intake or when the hormonal systems or organs involved are compromised. Thus, even with maximal activity of the vitamin D-dependent intestinal active transport system, sustained calcium intakes <5 mmol/d (<200 mg/d) cannot provide enough net calcium absorption to replace obligate losses via the intestine, the kidney, sweat, and other secretions. In this case, increased blood levels of PTH and  $1,25(\text{OH})_2\text{D}$  activate osteoclastic bone resorption to obtain needed calcium from bone, which leads to progressive bone loss and negative calcium balance. Increased PTH and  $1,25(\text{OH})_2\text{D}$  also enhance renal calcium reabsorption, and  $1,25(\text{OH})_2\text{D}$  enhances calcium absorption in the gut. At very high calcium intakes (>100 mmol/d [>4 g/d]), passive intestinal absorption continues to deliver calcium into the ECF despite maximally downregulated intestinal active transport and renal tubular calcium reabsorption. This can cause severe hypercalcioria, nephrocalcinosis, progressive renal failure, and hypercalcemia (e.g., "milk-alkali syndrome"). Deficiency or excess of PTH or vitamin D, intestinal disease, and renal failure represent other commonly encountered challenges to normal calcium homeostasis (Chap. 410).

## PHOSPHORUS METABOLISM

Although 85% of the ~600 g of body phosphorus is present in bone mineral, phosphorus is also a major intracellular constituent both as the free anion(s) and as a component of numerous organophosphate compounds, including structural proteins, enzymes, transcription

factors, carbohydrate and lipid intermediates, high-energy stores (ATP [adenosine triphosphate], creatine phosphate), and nucleic acids. Unlike calcium, phosphorus exists intracellularly at concentrations close to those present in ECF (e.g., 1–2 mmol/L). In cells and in the ECF, phosphorus exists in several forms, predominantly as  $H_2PO_4^-$  or  $NaHPO_4^-$ , with perhaps 10% as  $HPO_4^{2-}$ . This mixture of anions will be referred to here as “phosphate.” In serum, ~12% of phosphorus is bound to proteins. Concentrations of phosphates in blood and ECF generally are expressed in terms of elemental phosphorus, with the normal range in adults being 0.75–1.45 mmol/L (2.5–4.5 mg/dL). Because the volume of the intracellular fluid compartment is twice that of the ECF, measurements of ECF phosphate may not accurately reflect phosphate availability within cells that follows even modest shifts of phosphate from one compartment to the other.

Phosphate is widely available in foods and is absorbed efficiently (65%) by the small intestine even in the absence of vitamin D. However, phosphate absorptive efficiency may be enhanced (to 85–90%) via active transport mechanisms that are stimulated by  $1,25(OH)_2D$ . These mechanisms involve activation of  $Na^+/PO_4^{2-}$  co-transporters, such as Npt2b, that move phosphate into intestinal cells against an unfavorable electrochemical gradient. Daily net intestinal phosphate absorption varies widely with the composition of the diet but is generally in the range of 500–1000 mg/d. Phosphate absorption can be inhibited by large doses of calcium salts or by sevelamer hydrochloride (Renagel), strategies commonly used to control levels of serum phosphate in renal failure. Aluminum hydroxide antacids also reduce phosphate absorption but are used less commonly because of the potential for aluminum toxicity. Low serum phosphate stimulates renal proximal tubular synthesis of  $1,25(OH)_2D$ , perhaps by suppressing blood levels of FGF23 (see below).

Serum phosphate levels vary by as much as 50% on a normal day. This reflects the effect of food intake but also an underlying circadian rhythm that produces a nadir between 7 and 10 a.m. Carbohydrate administration, especially as IV dextrose solutions in fasting subjects, can decrease serum phosphate by >0.7 mmol/L (2 mg/dL) during treatment of ketoacidosis or during metabolic or respiratory alkalosis. Because of this wide variation in serum phosphate, it is best to perform measurements in the basal, fasting state.

Control of serum phosphate is determined mainly by the rate of renal tubular reabsorption of the filtered load, which is ~4–6 g/d. Because intestinal phosphate absorption is highly efficient, urinary excretion is not constant but varies directly with dietary intake. The fractional excretion of phosphate (ratio of phosphate to creatinine clearance) is generally in the range of 10–15%. The proximal tubule is the principal site at which renal phosphate reabsorption is regulated. This is accomplished by changes in the levels of apical expression and activity of specific  $Na^+/PO_4^{2-}$  co-transporters (NaPi-2a and NaPi-2c) in the proximal tubule. Levels of these transporters at the apical surface of these cells are reduced rapidly by PTH, a major hormonal regulator of renal phosphate excretion. FGF23 can impair phosphate reabsorption dramatically by a similar mechanism. Activating *FGF23* mutations cause the rare disorder autosomal dominant hypophosphatemic rickets (ADHR). In contrast to PTH, FGF23 also leads to reduced synthesis of  $1,25(OH)_2D$ , which may worsen the resulting hypophosphatemia by lowering intestinal phosphate absorption. Renal reabsorption of phosphate is responsive to changes in dietary intake such that experimental dietary phosphate restriction leads to a dramatic lowering of urinary phosphate within hours, preceding any decline in serum phosphate (e.g., filtered load). This physiologic renal adaptation to changes in dietary phosphate availability occurs independently of PTH and may be mediated in part by changes in levels of serum FGF23. Findings in *FGF23*-knockout mice suggest that FGF23 normally acts to lower blood phosphate and  $1,25(OH)_2D$  levels. In turn, elevation of blood phosphate increases blood levels of FGF23.

Renal phosphate reabsorption is impaired by hypocalcemia, hypomagnesemia, and severe hypophosphatemia. Phosphate clearance is enhanced by ECF volume expansion and impaired by dehydration. Phosphate retention is an important pathophysiologic feature of renal insufficiency (Chap. 311).

## HYPOPHOSPHATEMIA

**Causes** Hypophosphatemia can occur by one or more of three primary mechanisms: (1) inadequate intestinal phosphate absorption, (2) excessive renal phosphate excretion, and (3) rapid redistribution of phosphate from the ECF into bone or soft tissue (Table 409-1). Because phosphate is so abundant in foods, inadequate intestinal absorption is almost never observed now that aluminum hydroxide antacids, which bind phosphate in the gut, are no longer widely used. Fasting or starvation, however, may result in depletion of body phosphate and predispose to subsequent hypophosphatemia during refeeding, especially if this is accomplished with IV glucose alone.

Chronic hypophosphatemia usually signifies the presence of a persistent renal tubular phosphate-wasting disorder. Excessive activation of PTH/PTHRP receptors in the proximal tubule as a result of primary or secondary hyperparathyroidism or because of the PTHRP-mediated hypercalcemia syndrome in malignancy (Chap. 410) is among the more common causes of renal hypophosphatemia. Familial hypocalciuric hypercalcemia and Jansen's chondrodystrophy are rare examples of genetic disorders in this category (Chap. 410).

Several genetic and acquired diseases cause PTH/PTHRP receptor-independent tubular phosphate wasting with associated rickets and osteomalacia. All these diseases manifest severe hypophosphatemia; renal phosphate wasting, sometimes accompanied by aminoaciduria; inappropriately low blood levels of  $1,25(OH)_2D$ ; low-normal serum levels of calcium; and evidence of impaired cartilage or bone mineralization. Analysis of these diseases led to the discovery of the hormone FGF23, which is an important physiologic regulator of phosphate metabolism. FGF23 decreases phosphate reabsorption in the proximal tubule and also suppresses the 1'-hydroxylase responsible for synthesis of  $1,25(OH)_2D$ . FGF23 is synthesized by cells of the osteoblast lineage, primarily osteocytes. High-phosphate diets increase FGF23 levels, and low-phosphate diets decrease them. ADHR was the first disease linked to abnormalities in FGF23. ADHR results from activating mutations in the gene that encodes FGF23. These mutations alter a cleavage site that ordinarily allows for inactivation of intact FGF23. Several other genetic disorders feature elevated FGF23 and hypophosphatemia. The most common of these is X-linked hypophosphatemic rickets (XLH), which results from inactivating mutations in an endopeptidase termed *PHEX* (phosphate-regulating gene with homologies to endopeptidases on the X chromosome) that is expressed most abundantly on the surface of osteocytes and mature osteoblasts. Patients with XLH usually have high FGF23 levels, and ablation of the *FGF23* gene reverses the hypophosphatemia found in the mouse version of XLH. How inactivation of *PHEX* leads to increased levels of FGF23 has not been determined. Two rare autosomal recessive hypophosphatemic syndromes associated with elevated FGF23 are due to inactivating mutations of dentin matrix protein 1 (DMP1) and ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1), respectively, both of which normally are highly expressed in bone and presumably regulate FGF23 production. An unusual hypophosphatemic disorder, tumor-induced osteomalacia (TIO), is an acquired disorder in which tumors, usually of mesenchymal origin and generally histologically benign, secrete FGF23 that induces renal phosphate wasting. The hypophosphatemic syndrome resolves completely within hours to days after successful resection of the responsible tumor. Such tumors typically express large amounts of FGF23 mRNA, and patients with TIO usually exhibit elevations of FGF23 in their blood.

Dent's disease is an X-linked recessive disorder caused by inactivating mutations in *CLCN5*, a chloride transporter expressed in endosomes of the proximal tubule; features include hypercalciuria, hypophosphatemia, and recurrent kidney stones. Renal phosphate wasting is common among poorly controlled diabetic patients and alcoholics, who therefore are at risk for iatrogenic hypophosphatemia when treated with insulin or IV glucose, respectively. Diuretics and certain other drugs and toxins can cause defective renal tubular phosphate reabsorption (Table 409-1).

In hospitalized patients, hypophosphatemia is often attributable to massive redistribution of phosphate from the ECF into cells. Insulin

**TABLE 409-1 Causes of Hypophosphatemia**

- I. Reduced renal tubular phosphate reabsorption
  - A. PTH/PTHRP-dependent
    - 1. Primary hyperparathyroidism
    - 2. Secondary hyperparathyroidism
      - a. Vitamin D deficiency/resistance
      - b. Calcium starvation/malabsorption
      - c. Bartter's syndrome
      - d. Autosomal recessive renal hypercalcemia with hypomagnesemia
    - 3. PTHRP-dependent hypercalcemia of malignancy
    - 4. Familial hypocalciuric hypercalcemia
  - B. PTH/PTHRP-independent
    - 1. Excess FGF23 or other "phosphatonin"
      - a. X-linked hypophosphatemic rickets (XLH)
      - b. Autosomal recessive hypophosphatemia (ARHP)
      - c. Autosomal recessive hypophosphatemic rickets (ARHR) (DMP1, ENPP1 deficiency)
      - d. Tumor-induced osteomalacia syndrome (TIO)
      - e. McCune-Albright syndrome (fibrous dysplasia)
      - f. Epidermal nevus syndrome
    - 2. Intrinsic renal disease
      - a. Fanconi's syndrome(s)
      - b. Cystinosis
      - c. Wilson's disease
      - d. NaPi-2a or NaPi-2c mutations
    - 3. Other systemic disorders
      - a. Poorly controlled diabetes mellitus
      - b. Alcoholism
      - c. Hyperaldosteronism
      - d. Hypomagnesemia
      - e. Amyloidosis
      - f. Hemolytic-uremic syndrome
      - g. Renal transplantation or partial liver resection
      - h. Rewarming or induced hyperthermia
    - 4. Drugs or toxins
      - a. Ethanol
      - b. Acetazolamide, other diuretics
      - c. High-dose estrogens or glucocorticoids
      - d. Heavy metals (lead, cadmium, saccharated ferric oxide)
      - e. Toluene, *N*-methyl formamide
      - f. Cisplatin, ifosfamide, fosfarnet, rapamycin
  - II. Impaired intestinal phosphate absorption
    - A. Aluminum-containing antacids
    - B. Sevelamer
  - III. Shifts of extracellular phosphate into cells
    - A. Intravenous glucose
    - B. Insulin therapy for prolonged hyperglycemia or diabetic ketoacidosis
    - C. Catecholamines (epinephrine, dopamine, albuterol)
    - D. Acute respiratory alkalosis
    - E. Gram-negative sepsis, toxic shock syndrome
    - F. Recovery from starvation or acidosis
    - G. Rapid cellular proliferation
      - 1. Leukemic blast crisis
      - 2. Intensive erythropoietin, other growth factor therapy
  - IV. Accelerated net bone formation
    - A. After parathyroidectomy
    - B. Treatment of vitamin D deficiency, Paget's disease
    - C. Osteoblastic metastases

*Abbreviations:* PTH, parathyroid hormone; PTHRP, parathyroid hormone-related peptide.

therapy for diabetic ketoacidosis is a paradigm for this phenomenon, in which the severity of the hypophosphatemia is related to the extent of antecedent depletion of phosphate and other electrolytes (**Chap. 404**). The hypophosphatemia is usually greatest at a point many hours after initiation of insulin therapy and is difficult to predict from baseline measurements of serum phosphate at the time of presentation, when prerenal azotemia can obscure significant phosphate depletion. Other factors that may contribute to such acute redistributive hypophosphatemia include antecedent starvation or malnutrition, administration of IV glucose without other nutrients, elevated blood catecholamines (endogenous or exogenous), respiratory alkalosis, and recovery from metabolic acidosis.

Hypophosphatemia also can occur transiently (over weeks to months) during the phase of accelerated net bone formation that follows parathyroidectomy for severe primary hyperparathyroidism or during treatment of vitamin D deficiency or lytic Paget's disease. This is usually most prominent in patients who preoperatively have evidence of high bone turnover (e.g., high serum levels of alkaline phosphatase). Osteoblastic metastases can also lead to this syndrome.

**Clinical and Laboratory Findings** The clinical manifestations of severe hypophosphatemia reflect a generalized defect in cellular energy metabolism because of ATP depletion, a shift from oxidative phosphorylation toward glycolysis, and associated tissue or organ dysfunction. Acute, severe hypophosphatemia occurs mainly or exclusively in hospitalized patients with underlying serious medical or surgical illness and preexisting phosphate depletion due to excessive urinary losses, severe malabsorption, or malnutrition. Chronic hypophosphatemia tends to be less severe, with a clinical presentation dominated by musculoskeletal complaints such as bone pain, osteomalacia, pseudofractures, and proximal muscle weakness or, in children, rickets and short stature.

Neuromuscular manifestations of severe hypophosphatemia are variable but may include muscle weakness, lethargy, confusion, disorientation, hallucinations, dysarthria, dysphagia, oculomotor palsies, anisocoria, nystagmus, ataxia, cerebellar tremor, ballismus, hyporeflexia, impaired sphincter control, distal sensory deficits, paresthesia, hyperesthesia, generalized or Guillain-Barré-like ascending paralysis, seizures, coma, and even death. Serious sequelae such as paralysis, confusion, and seizures are likely only at phosphate concentrations  $<0.25$  mmol/L ( $<0.8$  mg/dL). Rhabdomyolysis may develop during rapidly progressive hypophosphatemia. The diagnosis of hypophosphatemia-induced rhabdomyolysis may be overlooked, as up to 30% of patients with acute hypophosphatemia ( $<0.7$  mM) have creatine phosphokinase elevations that peak 1–2 days after the nadir in serum phosphate, when the release of phosphate from injured myocytes may have led to a near normalization of circulating levels of phosphate.

Respiratory failure and cardiac dysfunction, which are reversible with phosphate treatment, may occur at serum phosphate levels of 0.5–0.8 mmol/L (1.5–2.5 mg/dL). Renal tubular defects, including tubular acidosis, glycosuria, and impaired reabsorption of sodium and calcium, may occur. Hematologic abnormalities correlate with reductions in intracellular ATP and 2,3-diphosphoglycerate and may include erythrocyte microspherocytosis and hemolysis; impaired oxyhemoglobin dissociation; defective leukocyte chemotaxis, phagocytosis, and bacterial killing; and platelet dysfunction with spontaneous gastrointestinal hemorrhage.

## TREATMENT

### Hypophosphatemia

Severe hypophosphatemia ( $<0.75$  mmol/L [ $<2$  mg/dL]), particularly in the setting of underlying phosphate depletion, constitutes a dangerous electrolyte abnormality that should be corrected promptly. Unfortunately, the cumulative deficit in body phosphate cannot be predicted directly from knowledge of the circulating level of phosphate, and therapy must be approached empirically. The threshold for IV phosphate therapy and consequently the dose of phosphate to be administered should reflect consideration of renal function, the likely severity and duration of the underlying phosphate depletion,

**TABLE 409-2** Intravenous Therapy for Hypophosphatemia

CONSIDER			
SERUM PHOSPHORUS, MM (MG/DL)	RATE OF INFUSION, MMOL/H	DURATION, H	TOTAL ADMINISTERED, MMOL
<0.8 (<2.5)	2	6	12
<0.5 (<1.5)	4	6	24
<0.3 (<1)	8	6	48

**Note:** Rates shown are calculated for a 70-kg person; levels of serum calcium and phosphorus must be measured every 6–12 h during therapy; infusions can be repeated to achieve stable serum phosphorus levels >0.8 mmol/L (>2.5 mg/dL); most formulations available in the United States provide 3 mmol/mL of sodium or potassium phosphate.

and the presence and severity of symptoms consistent with those of hypophosphatemia. In adults, phosphate may be safely administered IV as neutral mixtures of sodium or potassium phosphate salts at initial doses of 0.2–0.8 mmol/kg of elemental phosphorus over 6 h (e.g., 10–50 mmol over 6 h), with doses >20 mmol/6 h reserved for those who have serum levels <0.5 mmol/L (1.5 mg/dL) and normal renal function. A suggested approach is presented in **Table 409-2**. Serum levels of phosphate and calcium must be monitored closely (every 6–12 h) throughout treatment. It is necessary to avoid a serum calcium-phosphorus product >50 mg<sup>2</sup>/dL<sup>2</sup> to reduce the risk of heterotopic calcification. Hypocalcemia, if present, should be corrected before administering IV phosphate. Less severe hypophosphatemia, in the range of 0.5–0.8 mmol/L (1.5–2.5 mg/dL), usually can be treated with oral phosphate in divided doses of 750–2000 mg/d as elemental phosphorus; higher doses can cause bloating and diarrhea.

Management of chronic hypophosphatemia requires knowledge of the cause(s) of the disorder. Hypophosphatemia related to the secondary hyperparathyroidism of vitamin D deficiency usually responds to treatment with vitamin D and calcium alone. XLH, ADHR, TIO, and related renal tubular disorders usually are managed with divided oral doses of phosphate, often with calcium and 1,25(OH)<sub>2</sub>D supplements to bypass the block in renal 1,25(OH)<sub>2</sub>D synthesis and prevent secondary hyperparathyroidism caused by suppression of ECF calcium levels. Care must be taken to be sure that oral calcium and phosphate are not administered at the same time, to avoid precipitation before absorption. Thiazide diuretics may be used to prevent nephrocalcinosis in patients who are managed this way. Complete normalization of hypophosphatemia is generally not possible in these conditions. Burosumab, a human monoclonal antibody that inhibits FGF23, has been approved for the treatment of XLH. It corrects hypophosphatemia, improves bone pain, and heals fractures in both adults and children.

Optimal therapy for TIO is extirpation of the responsible tumor, which may be localized by radiographic skeletal survey or bone scan (many are located in bone) or by radionuclide scanning using sestamibi or labeled octreotide. Successful treatment of TIO-induced hypophosphatemia with octreotide has been reported in a small number of patients. Burosumab treatment, originally used for XLH, also shows promise as a treatment for TIO.

## HYPERPHOSPHATEMIA

**Causes** When the filtered load of phosphate and glomerular filtration rate (GFR) are normal, control of serum phosphate levels is achieved by adjusting the rate at which phosphate is reabsorbed by

**TABLE 409-3** Causes of Hyperphosphatemia

I. Impaired renal phosphate excretion
A. Renal insufficiency
B. Hypoparathyroidism
1. Developmental
2. Autoimmune
3. After neck surgery or radiation
4. Activating mutations of the calcium-sensing receptor
C. Parathyroid suppression
1. Parathyroid-independent hypercalcemia
a. Vitamin D or vitamin A intoxication
b. Sarcoidosis, other granulomatous diseases
c. Immobilization, osteolytic metastases
d. Milk-alkali syndrome
2. Severe hypermagnesemia or hypomagnesemia
D. Pseudohypoparathyroidism
E. Acromegaly
F. Tumoral calcinosis
G. Heparin therapy
II. Massive extracellular fluid phosphate loads
A. Rapid administration of exogenous phosphate (intravenous, oral, rectal)
B. Extensive cellular injury or necrosis
1. Crush injuries
2. Rhabdomyolysis
3. Hyperthermia
4. Fulminant hepatitis
5. Cytotoxic therapy
6. Severe hemolytic anemia
C. Transcellular phosphate shifts
1. Metabolic acidosis
2. Respiratory acidosis

the proximal tubular NaPi-2 co-transporters. The principal hormonal regulators of NaPi-2 activity are PTH and FGF23. Hyperphosphatemia, defined in adults as a fasting serum phosphate concentration >1.8 mmol/L (5.5 mg/dL), usually results from impaired glomerular filtration, hypoparathyroidism, excessive delivery of phosphate into the ECF (from bone, gut, or parenteral phosphate therapy), or a combination of these factors (**Table 409-3**). The upper limit of normal serum phosphate concentrations is higher in children and neonates (2.4 mmol/L [7 mg/dL]). It is useful to distinguish hyperphosphatemia caused by impaired renal phosphate excretion from that which results from excessive delivery of phosphate into the ECF (Table 409-3).

In chronic renal insufficiency, reduced GFR leads to phosphate retention. Hyperphosphatemia in turn further impairs renal synthesis of 1,25(OH)<sub>2</sub>D, increases FGF23 levels, and stimulates PTH secretion and parathyroid gland hypertrophy both directly and indirectly (by lowering blood ionized calcium levels). Thus, hyperphosphatemia is a major cause of the secondary hyperparathyroidism of renal failure and must be addressed early in the course of the disease (**Chaps. 311 and 410**).

Hypoparathyroidism leads to hyperphosphatemia via increased expression of NaPi-2 co-transporters in the proximal tubule. Hypoparathyroidism, or parathyroid suppression, has multiple potential causes, including autoimmune disease; developmental, surgical, or radiation-induced absence of functional parathyroid tissue; vitamin D intoxication or other causes of PTH-independent hypercalcemia; cellular PTH resistance (pseudohypoparathyroidism or hypomagnesemia); infiltrative disorders such as Wilson's disease and hemochromatosis; and impaired PTH secretion caused by hypermagnesemia, severe hypomagnesemia, or activating mutations in the CaSR. Hypocalcemia may also contribute directly to impaired phosphate clearance, as calcium infusion can induce phosphaturia in hypoparathyroid subjects. Increased tubular phosphate reabsorption also occurs in acromegaly, during heparin administration, and in tumoral calcinosis. Tumoral

calcinoses is caused by a rare group of genetic disorders in which FGF23 is processed in a way that leads to low levels of active FGF23 in the bloodstream. This may result from mutations in the FGF23 sequence or via inactivating mutations in the *GALNT3* gene, which encodes a galactosaminyl transferase that normally adds sugar residues to FGF23 that slow its proteolysis. A similar syndrome results from FGF23 resistance due to inactivating mutations of the FGF23 co-receptor Klotho. These abnormalities cause elevated serum 1,25(OH)<sub>2</sub>D, parathyroid suppression, increased intestinal calcium absorption, and focal hyperostosis with large, lobulated periarticular heterotopic ossifications (especially at shoulders or hips) and are accompanied by hyperphosphatemia. In some forms of tumoral calcinosis, serum phosphorus levels are normal.

When large amounts of phosphate are delivered rapidly into the ECF, hyperphosphatemia can occur despite normal renal function. Examples include overzealous IV phosphate therapy, oral or rectal administration of large amounts of phosphate-containing laxatives or enemas (especially in children), extensive soft tissue injury or necrosis (crush injuries, rhabdomyolysis, hyperthermia, fulminant hepatitis, cytotoxic chemotherapy), extensive hemolytic anemia, and transcellular phosphate shifts induced by severe metabolic or respiratory acidosis.

**Clinical Findings** The clinical consequences of acute, severe hyperphosphatemia are due mainly to the formation of widespread calcium phosphate precipitates and resulting hypocalcemia. Thus, tetany, seizures, accelerated nephrocalcinosis (with renal failure, hyperkalemia, hyperuricemia, and metabolic acidosis), and pulmonary or cardiac calcifications (including development of acute heart block) may occur. The severity of these complications relates to the elevation of serum phosphate levels, which can reach concentrations as high as 7 mmol/L (20 mg/dL) in instances of massive soft tissue injury or tumor lysis syndrome.

## TREATMENT

### Hyperphosphatemia

Therapeutic options for management of severe hyperphosphatemia are limited. Volume expansion may enhance renal phosphate clearance. Aluminum hydroxide antacids or sevelamer may be helpful in chelating and limiting absorption of offending phosphate salts present in the intestine. Hemodialysis is the most effective therapeutic strategy and should be considered early in the course of severe hyperphosphatemia, especially in the setting of renal failure and symptomatic hypocalcemia.

## MAGNESIUM METABOLISM

Magnesium is the major intracellular divalent cation. Normal concentrations of extracellular magnesium and calcium are crucial for normal neuromuscular activity. Intracellular magnesium forms a key complex with ATP and is an important cofactor for a wide range of enzymes, transporters, and nucleic acids required for normal cellular function, replication, and energy metabolism. The concentration of magnesium in serum is closely regulated within the range of 0.7–1 mmol/L (1.5–2 meq/L; 1.7–2.4 mg/dL), of which 30% is protein-bound and another 15% is loosely complexed to phosphate and other anions. One-half of the 25 g (1000 mmol) of total body magnesium is located in bone, only one-half of which is insoluble in the mineral phase. Almost all extraskeletal magnesium is present within cells, where the total concentration is 5 mM, 95% of which is bound to proteins and other macromolecules. Because only 1% of body magnesium resides in the ECF, measurements of serum magnesium levels may not accurately reflect the level of total body magnesium stores.

Dietary magnesium content normally ranges from 6 to 15 mmol/d (140–360 mg/d), of which 30–40% is absorbed, mainly in the jejunum and ileum. Intestinal magnesium absorptive efficiency is stimulated by 1,25(OH)<sub>2</sub>D and can reach 70% during magnesium deprivation. Urinary magnesium excretion normally matches net intestinal absorption and is ~4 mmol/d (100 mg/d). Regulation of serum magnesium

concentrations is achieved mainly by control of renal magnesium reabsorption. Only 20% of filtered magnesium is reabsorbed in the proximal tubule, whereas 60% is reclaimed in the cTAL and another 5–10% in the DCT. Magnesium reabsorption in the cTAL occurs via a paracellular route that requires both a lumen-positive potential, created by NaCl reabsorption, and tight-junction proteins encoded by members of the Claudin gene family. Magnesium reabsorption in the cTAL is increased by PTH but inhibited by hypercalcemia or hypermagnesemia, both of which activate the CaSR in this nephron segment.

## HYPOMAGNESEMIA

**Causes** Hypomagnesemia usually signifies substantial depletion of body magnesium stores (0.5–1 mmol/kg). Hypomagnesemia can result from intestinal malabsorption; protracted vomiting, diarrhea, or intestinal drainage; defective renal tubular magnesium reabsorption; or rapid shifts of magnesium from the ECF into cells, bone, or third spaces (Table 409-4). Dietary magnesium deficiency is unlikely except possibly in the setting of alcoholism. A rare genetic disorder that causes selective intestinal magnesium malabsorption has been described (primary infantile hypomagnesemia). Another rare inherited disorder (hypomagnesemia with secondary hypocalcemia) is caused by mutations in the gene encoding TRPM6, a protein that, along with TRPM7, forms a channel important for both intestinal and distal-tubular renal transcellular magnesium transport. Malabsorptive states, often compounded by vitamin D deficiency, can critically limit magnesium absorption and produce hypomagnesemia despite the compensatory effects of secondary hyperparathyroidism and of hypocalcemia and hypomagnesemia to enhance cTAL magnesium reabsorption. Diarrhea or surgical drainage fluid may contain 5 mmol/L of magnesium. Proton pump inhibitors (omeprazole and others) may produce hypomagnesemia by an unknown mechanism that does not involve renal wasting of magnesium.

Several genetic magnesium-wasting syndromes have been described, including inactivating mutations of genes encoding the DCT NaCl co-transporter (Gitelman's syndrome), proteins required for cTAL Na-K-2Cl transport (Bartter's syndrome), claudin 16 or claudin 19 (autosomal recessive renal hypomagnesemia with hypercalcemia), a DCT Na<sup>+</sup>,K<sup>+</sup>-ATPase -subunit (autosomal dominant renal hypomagnesemia with hypocalciuria), DCT K<sup>+</sup> channels (Kv1.1, Kir4.1), and a mitochondrial gene encoding a tRNA. Activating mutations of the CaSR can cause hypomagnesemia as well as hypocalcemia. ECF expansion, hypercalcemia, and severe phosphate depletion may impair magnesium reabsorption, as can various forms of renal injury, including those caused by drugs such as cisplatin, cyclosporine, aminoglycosides, and pentamidine as well as the epidermal growth factor (EGF) receptor inhibitory antibody cetuximab (EGF action is required for normal DCT apical expression of TRPM6) (Table 409-4). A rising blood concentration of ethanol directly impairs tubular magnesium reabsorption, and persistent glycosuria with osmotic diuresis leads to magnesium wasting and probably contributes to the high frequency of hypomagnesemia in poorly controlled diabetic patients. Magnesium depletion is aggravated by metabolic acidosis, which causes intracellular losses as well.

Hypomagnesemia due to rapid shifts of magnesium from ECF into the intracellular compartment can occur during recovery from diabetic ketoacidosis, starvation, or respiratory acidosis. Less acute shifts may be seen during rapid bone formation after parathyroidectomy, with treatment of vitamin D deficiency, or with osteoblastic metastases. Large amounts of magnesium may be lost with acute pancreatitis, extensive burns, or protracted and severe sweating and during pregnancy and lactation.

**Clinical and Laboratory Findings** Hypomagnesemia may cause generalized alterations in neuromuscular function, including tetany, tremor, seizures, muscle weakness, ataxia, nystagmus, vertigo, apathy, depression, irritability, delirium, and psychosis. Patients are usually asymptomatic when serum magnesium concentrations are >0.5 mmol/L (1 meq/L; 1.2 mg/dL), although the severity of symptoms may not correlate well with serum magnesium levels. Cardiac

**TABLE 409-4 Causes of Hypomagnesemia**

I.	Impaired intestinal absorption
A.	Hypomagnesemia with secondary hypocalcemia ( <i>TRPM6</i> mutations)
B.	Malabsorption syndromes
C.	Vitamin D deficiency
D.	Proton pump inhibitors
II.	Increased intestinal losses
A.	Protracted vomiting/diarrhea
B.	Intestinal drainage, fistulas
III.	Impaired renal tubular reabsorption
A.	Genetic magnesium-wasting syndromes
1.	Gitelman's syndrome
2.	Bartter's syndrome
3.	Claudin 16 or 19 mutations
4.	Potassium channel mutations ( <i>Kv1.1, Kir4.1</i> )
5.	Na <sup>+</sup> -K <sup>+</sup> -ATPase $\gamma$ -subunit mutations ( <i>FXYD2</i> )
B.	Acquired renal disease
1.	Tubulointerstitial disease
2.	Postobstruction, ATN (diuretic phase)
3.	Renal transplantation
C.	Drugs and toxins
1.	Ethanol
2.	Diuretics (loop, thiazide, osmotic)
3.	Cisplatin
4.	Pentamidine, fosfarnet
5.	Cyclosporine
6.	Aminoglycosides, amphotericin B
7.	Cetuximab
D.	Other
1.	Extracellular fluid volume expansion
2.	Hyperaldosteronism
3.	SIADH
4.	Diabetes mellitus
5.	Hypercalcemia
6.	Phosphate depletion
7.	Metabolic acidosis
8.	Hyperthyroidism
IV.	Rapid shifts from extracellular fluid
A.	Intracellular redistribution
1.	Recovery from diabetic ketoacidosis
2.	Refeeding syndrome
3.	Correction of respiratory acidosis
4.	Catecholamines
B.	Accelerated bone formation
1.	Post-parathyroideectomy
2.	Treatment of vitamin D deficiency
3.	Osteoblastic metastases
C.	Other
1.	Pancreatitis, burns, excessive sweating
2.	Pregnancy (third trimester) and lactation

**Abbreviations:** ATN, acute tubular necrosis; SIADH, syndrome of inappropriate antidiuretic hormone.

arrhythmias may occur, including sinus tachycardia, other supraventricular tachycardias, and ventricular arrhythmias. Electrocardiographic abnormalities may include prolonged PR or QT intervals, T-wave flattening or inversion, and ST straightening. Sensitivity to digitalis toxicity may be enhanced.

Other electrolyte abnormalities often seen with hypomagnesemia, including hypocalcemia (with hypocalciuria) and hypokalemia, may not be easily corrected unless magnesium is administered as well. The hypocalcemia may be a result of concurrent vitamin D deficiency, although hypomagnesemia can cause impaired synthesis of 1,25(OH)<sub>2</sub>D, cellular resistance to PTH, and, at very low serum

magnesium (<0.4 mmol/L [ $<0.8 \text{ meq/L}$ ;  $<1 \text{ mg/dL}$ ]), a defect in PTH secretion; these abnormalities are reversible with therapy.

## TREATMENT

### Hypomagnesemia

Mild, asymptomatic hypomagnesemia may be treated with oral magnesium salts (MgCl<sub>2</sub>, MgO, Mg[OH]<sub>2</sub>) in divided doses totaling 20–30 mmol/d (40–60 meq/d). Diarrhea may occur with larger doses. More severe hypomagnesemia should be treated parenterally, preferably with IV MgCl<sub>2</sub>, which can be administered safely as a continuous infusion of 50 mmol/d (100 meq Mg<sup>2+</sup>/d) if renal function is normal. If GFR is reduced, the infusion rate should be lowered by 50–75%. Use of IM MgSO<sub>4</sub> is discouraged; the injections are painful and provide relatively little magnesium (2 mL of 50% MgSO<sub>4</sub> supplies only 4 mmol). MgSO<sub>4</sub> may be given IV instead of MgCl<sub>2</sub>, although the sulfate anions may bind calcium in serum and urine and aggravate hypocalcemia. Serum magnesium should be monitored at intervals of 12–24 h during therapy, which may continue for several days because of impaired renal conservation of magnesium (only 50–70% of the daily IV magnesium dose is retained) and delayed repletion of intracellular deficits, which may be as high as 1–1.5 mmol/kg (2–3 meq/kg).

It is important to consider the need for calcium, potassium, and phosphate supplementation in patients with hypomagnesemia. Vitamin D deficiency frequently coexists and should be treated with oral or parenteral vitamin D or 25(OH)D (but not with 1,25(OH)<sub>2</sub>D, which may impair tubular magnesium reabsorption, possibly via PTH suppression). In severely hypomagnesemic patients with concomitant hypocalcemia and hypophosphatemia, administration of IV magnesium alone may worsen hypophosphatemia, provoking neuromuscular symptoms or rhabdomyolysis, due to rapid stimulation of previously suppressed PTH secretion. This is avoided by administering both calcium and magnesium.

## HYPERMAGNESEMIA

**Causes** Hypermagnesemia is rarely seen in the absence of renal insufficiency as normal kidneys can excrete large amounts (250 mmol/d) of magnesium. Mild hypermagnesemia due to excessive reabsorption in the cTAL occurs with CaSR mutations in familial hypocalciuric hypercalcemia and has been described in some patients with adrenal insufficiency, hypothyroidism, or hypothermia. Massive exogenous magnesium exposures, usually via the gastrointestinal tract, can overwhelm renal excretory capacity and cause life-threatening hypermagnesemia (Table 409-5). A notable example of this is prolonged retention of even normal amounts of magnesium-containing cathartics in patients with intestinal ileus, obstruction, or perforation. Extensive soft tissue injury or necrosis also can deliver large amounts of magnesium into the ECF in patients who have suffered trauma, shock,

**TABLE 409-5 Causes of Hypermagnesemia**

I.	Excessive magnesium intake
A.	Cathartics, urologic irritants
B.	Parenteral magnesium administration
II.	Rapid mobilization from soft tissues
A.	Trauma, shock, sepsis
B.	Cardiac arrest
C.	Burns
III.	Impaired magnesium excretion
A.	Renal failure
B.	Familial hypocalciuric hypercalcemia
IV.	Other
A.	Adrenal insufficiency
B.	Hypothyroidism
C.	Hypothermia

**3166** sepsis, cardiac arrest, or severe burns. Further, infusion of magnesium in pregnant women with eclampsia can lead to hypocalcemia.

**Clinical and Laboratory Findings** The most prominent clinical manifestations of hypermagnesemia are vasodilation and neuromuscular blockade, which may appear at serum magnesium concentrations  $>2$  mmol/L ( $>4$  meq/L;  $>4.8$  mg/dL). Hypotension that is refractory to vasopressors or volume expansion may be an early sign. Nausea, lethargy, and weakness may progress to respiratory failure, paralysis, and coma, with hypoactive tendon reflexes, at serum magnesium levels  $>4$  mmol/L. Other findings may include gastrointestinal hypomotility or ileus; facial flushing; pupillary dilation; paradoxical bradycardia; prolongation of PR, QRS, and QT intervals; heart block; and, at serum magnesium levels approaching 10 mmol/L, asystole.

Hypermagnesemia, acting via the CaSR, causes hypocalcemia and hypercalcuria due to both parathyroid suppression and impaired cTAL calcium reabsorption.

## TREATMENT

### Hypermagnesemia

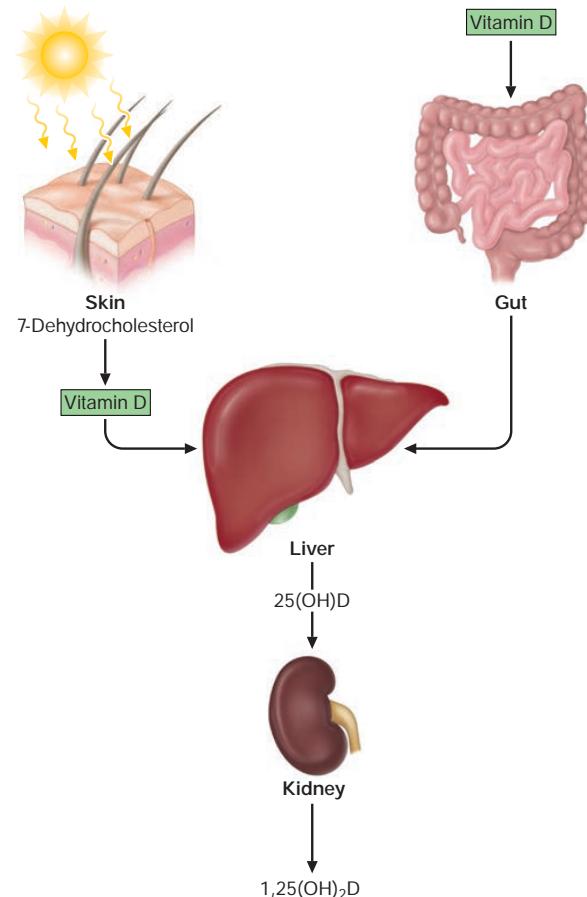
Successful treatment of hypermagnesemia generally involves identifying and interrupting the source(s) of magnesium and employing measures to increase magnesium clearance from the ECF. Use of magnesium-free cathartics or enemas may be helpful in clearing ingested magnesium from the gastrointestinal tract. Vigorous IV hydration should be attempted, if appropriate. Hemodialysis is effective and may be required in patients with significant renal insufficiency. Calcium, administered IV in doses of 100–200 mg over 1–2 h, has been reported to provide temporary improvement in signs and symptoms of hypermagnesemia.

## VITAMIN D

### SYNTHESIS AND METABOLISM

1,25-Dihydroxyvitamin D [ $1,25(OH)_2D$ ] is the major steroid hormone involved in regulation of mineral ion homeostasis. Vitamin D and its metabolites are hormones and hormone precursors rather than vitamins, since in the proper biologic setting, they can be synthesized endogenously (Fig. 409-4). In response to ultraviolet radiation of the skin, a photochemical cleavage results in the formation of vitamin D from 7-dehydrocholesterol. Cutaneous production of vitamin D is decreased by melanin and high solar protection factor sunblocks, which effectively impair skin penetration by ultraviolet light. The increased use of sunblocks in North America and Western Europe and a reduction in the magnitude of solar exposure of the general population over the past several decades has led to an increased reliance on dietary sources of vitamin D. In the United States and Canada, these sources largely consist of fortified cereals and dairy products, in addition to fish oils and egg yolks. Vitamin D from plant sources is in the form of vitamin  $D_2$ , whereas that from animal sources is vitamin  $D_3$ . These two forms have equivalent biologic activity and are activated equally well by the vitamin D hydroxylases in humans. Vitamin D enters the circulation, whether absorbed from the intestine or synthesized cutaneously, bound to vitamin D-binding protein, an  $\alpha$ -globulin synthesized in the liver. Vitamin D is subsequently 25-hydroxylated in the liver by a cytochrome P450 oxidase in the mitochondria and microsomes. The activity of this hydroxylase is not tightly regulated, and the resultant metabolite, 25-hydroxyvitamin D [25(OH)D], is the major circulating and storage form of vitamin D. Approximately 88% of 25(OH)D circulates bound to the vitamin D-binding protein, 0.03% is free, and the rest circulates bound to albumin. The half-life of 25(OH)D is  $\sim 2$ –3 weeks, with that of 25(OH)D<sub>2</sub> being shorter than that of 25(OH)D<sub>3</sub> due to a lower affinity of vitamin D-binding protein for the former. The half-life of 25(OH)D is also greatly shortened when vitamin D-binding protein levels are reduced, as can occur with increased urinary losses in the nephrotic syndrome.

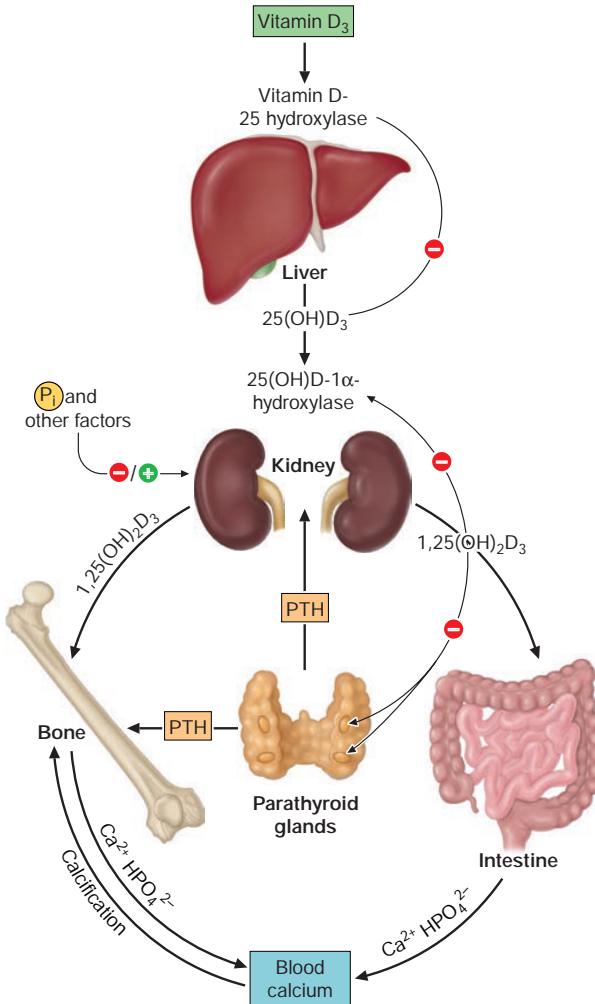
The second hydroxylation, required for the formation of the mature hormone, occurs in the kidney (Fig. 409-5). The 25-hydroxyvitamin



**FIGURE 409-4** Vitamin D synthesis and activation. Vitamin D is synthesized in the skin in response to ultraviolet radiation and also is absorbed from the diet. It is then transported to the liver, where it undergoes 25-hydroxylation. This metabolite is the major circulating form of vitamin D. The final step in hormone activation, 1 $\alpha$ -hydroxylation, occurs in the kidney.

D-1 $\alpha$ -hydroxylase (encoded by the *CYP27B1* gene) is a tightly regulated cytochrome P450-like mixed-function oxidase expressed in the proximal convoluted tubule cells of the kidney. PTH and hypophosphatemia are the major inducers of this microsomal enzyme in the kidney, whereas calcium, FGF23, and the enzyme's product, 1,25(OH)<sub>2</sub>D, repress it. The 25-hydroxyvitamin D-1 $\alpha$ -hydroxylase is also present in numerous other cell types, where it is not subject to hormonal regulation. It is expressed in epidermal keratinocytes, but keratinocyte production of 1,25(OH)<sub>2</sub>D is not thought to contribute to circulating levels of this hormone. In addition to being present in the trophoblastic layer of the placenta, the 1 $\alpha$ -hydroxylase is produced by macrophages associated with granulomas and lymphomas. In these latter pathologic states, the activity of the enzyme is induced by interferon and TNF-but is not regulated by calcium or 1,25(OH)<sub>2</sub>D; therefore, hypercalcemia, associated with elevated levels of 1,25(OH)<sub>2</sub>D, may be observed. Treatment of sarcoidosis-associated hypercalcemia with glucocorticoids, ketoconazole, or chloroquine reduces 1,25(OH)<sub>2</sub>D production and effectively lowers serum calcium. In contrast, chloroquine has not been shown to lower the elevated serum 1,25(OH)<sub>2</sub>D levels in patients with lymphoma.

The major pathway for inactivation of vitamin D metabolites is an additional hydroxylation step by the vitamin D 24-hydroxylase, an enzyme that is expressed in most tissues. 1,25(OH)<sub>2</sub>D is the major inducer of this enzyme; therefore, this hormone promotes its own inactivation, thereby limiting its biologic effects. FGF23 also induces this hydroxylase, thereby reducing circulating 1,25(OH)<sub>2</sub>D levels by increasing its inactivation, as well as by impairing its synthesis.



**FIGURE 409-5** Schematic representation of the hormonal control loop for vitamin D metabolism and function. A reduction in the serum calcium below ~2.2 mmol/L (8.8 mg/dL) prompts a proportional increase in the secretion of parathyroid hormone (PTH) and so mobilizes additional calcium from the bone. PTH promotes the synthesis of 1,25(OH)<sub>2</sub>D in the kidney, which in turn stimulates the mobilization of calcium from bone and intestine and regulates the synthesis of PTH by negative feedback.

Mutations of the gene encoding this enzyme (*CYP24A1*) can lead to infantile hypercalcemia, and in those less severely affected, long-standing hypercalcemia, nephrocalcinosis, and nephrolithiasis can occur.

Polar metabolites of 1,25(OH)<sub>2</sub>D are secreted into the bile and reabsorbed via the enterohepatic circulation. Impairment of this recirculation, which is seen with diseases of the terminal ileum, leads to accelerated losses of vitamin D metabolites.

#### ACTIONS OF 1,25(OH)<sub>2</sub>D

1,25(OH)<sub>2</sub>D mediates its biologic effects by binding to a member of the nuclear receptor superfamily, the vitamin D receptor (VDR). This receptor belongs to the subfamily that includes the thyroid hormone receptors, the retinoid receptors, and the peroxisome proliferator-activated receptors; however, in contrast to the other members of this subfamily, only one VDR isoform has been isolated. The VDR binds to target DNA sequences as a heterodimer with the retinoid X receptor, recruiting a series of coactivators that modify chromatin and approximate the VDR to the basal transcriptional apparatus, resulting in the induction of target gene expression. The mechanism of transcriptional repression by the VDR varies with different target genes but has been shown to involve either interference with the action of activating

transcription factors or the recruitment of novel proteins to the VDR complex, resulting in transcriptional repression.

The affinity of the VDR for 1,25(OH)<sub>2</sub>D is approximately three orders of magnitude higher than that for other vitamin D metabolites. In normal physiologic circumstances, these other metabolites are not thought to stimulate receptor-dependent actions. However, in states of vitamin D toxicity, the markedly elevated levels of 25(OH)D may lead to hypercalcemia by interacting directly with the VDR and by displacing 1,25(OH)<sub>2</sub>D from vitamin D-binding protein, resulting in increased bioavailability of the active hormone.

The VDR is expressed in a wide range of cells and tissues. The molecular actions of 1,25(OH)<sub>2</sub>D have been studied most extensively in tissues involved in the regulation of mineral ion homeostasis. This hormone is a major inducer of calbindin 9K, a calcium-binding protein expressed in the intestine, which is thought to play an important role in the active transport of calcium across the enterocyte. The two major calcium transporters expressed by intestinal epithelia, TRPV5 and TRPV6 (transient receptor potential vanilloid), are also vitamin D responsive. By inducing the expression of these and other genes in the small intestine, 1,25(OH)<sub>2</sub>D increases the efficiency of intestinal calcium absorption, and it also has been shown to have several important actions in the skeleton. The VDR is expressed in osteoblasts and regulates the expression of several genes in this cell. These genes include the bone matrix proteins osteocalcin and osteopontin, which are upregulated by 1,25(OH)<sub>2</sub>D, in addition to type I collagen, which is transcriptionally repressed by 1,25(OH)<sub>2</sub>D. Both 1,25(OH)<sub>2</sub>D and PTH induce the expression of RANK ligand, which promotes osteoclast differentiation and increases osteoclast activity, by binding to RANK on osteoclast progenitors and mature osteoclasts. This is the mechanism by which 1,25(OH)<sub>2</sub>D induces bone resorption. 1,25(OH)<sub>2</sub>D regulates phosphate homeostasis, primarily by inducing the expression of FGF23 in osteocytes. However, the skeletal features associated with VDR-knockout mice (rickets, osteomalacia) are largely corrected by increasing calcium and phosphorus intake, underscoring the importance of vitamin D action in the gut.

The VDR is expressed in the parathyroid gland, and 1,25(OH)<sub>2</sub>D has been shown to have antiproliferative effects on parathyroid cells and to suppress the transcription of the parathyroid hormone gene. These effects of 1,25(OH)<sub>2</sub>D on the parathyroid gland are an important part of the rationale for current therapies directed at preventing and treating hyperparathyroidism associated with renal insufficiency.

The VDR is also expressed in tissues and organs that do not play a role in mineral ion homeostasis. Notable in this respect is the observation that 1,25(OH)<sub>2</sub>D has an antiproliferative effect on several cell types, including keratinocytes, breast cancer cells, and prostate cancer cells. The effects of 1,25(OH)<sub>2</sub>D and the VDR on keratinocytes are particularly intriguing, since the VDR is primarily a transcriptional repressor in these cells. Alopecia is seen in humans and mice with mutant VDRs but is not a feature of vitamin D deficiency; thus, the effects of the VDR on the hair follicle are ligand-independent.

#### VITAMIN D DEFICIENCY

The mounting concern about the relationship between solar exposure and the development of skin cancer has led to increased reliance on dietary sources of vitamin D. Although the prevalence of vitamin D deficiency varies, the third National Health and Nutrition Examination Survey (NHANES III) revealed that vitamin D deficiency is prevalent throughout the United States, with the prevalence being >29% in obese children. The clinical syndrome of vitamin D deficiency can be a result of deficient production of vitamin D in the skin, lack of dietary intake, accelerated losses of vitamin D, impaired vitamin D activation, or resistance to the biologic effects of 1,25(OH)<sub>2</sub>D (Table 409-6). The elderly and nursing home residents are particularly at risk for vitamin D deficiency, since both the efficiency of vitamin D synthesis in the skin and the absorption of vitamin D from the intestine decline with age. The presence of terminal ileal disease also results in impaired enterohepatic circulation of vitamin D metabolites. While intestinal malabsorption of dietary fats and short bowel syndrome, including that associated with intestinal bypass surgery, lead to vitamin D deficiency, the cause

**TABLE 409-6 Causes of Impaired Vitamin D Action**

Vitamin D deficiency	Impaired 1 $\alpha$ -hydroxylation
Impaired cutaneous production	Hypoparathyroidism
Dietary absence	Ketoconazole
Malabsorption (short gut syndrome, gastric bypass)	1 $\alpha$ -hydroxylase mutation
Accelerated loss of vitamin D	FGF23 excess
Increased metabolism (barbiturates, phenytoin, rifampin)	Oncogenic osteomalacia
Impaired enterohepatic circulation	Hypophosphatemic rickets
Nephrotic syndrome	Fibrous dysplasia
CYP3A4 mutation	Chronic kidney disease
Impaired 25-hydroxylation	Target organ resistance
Liver disease, isoniazid	Vitamin D receptor mutation
25-Hydroxylase mutation	Phenytoin
	Other
	Obesity

of vitamin D deficiency in obese individuals is poorly understood. In addition to intestinal diseases, accelerated inactivation of vitamin D metabolites can be seen with drugs that induce hepatic cytochrome P450 mixed-function oxidases such as barbiturates, phenytoin, and rifampin. Gain-of-function mutations in *CYP3A4* accelerate the oxidation and inactivation of vitamin D metabolites, thus resulting in decreased serum levels of 25OHD and 1,25(OH)<sub>2</sub>D. This form of rickets is autosomal recessive and presents during early childhood and can be treated with high doses of calcitriol or vitamin D. Impaired 25-hydroxylation, associated with severe liver disease or isoniazid, is an uncommon cause of vitamin D deficiency. A mutation in the gene responsible for 25-hydroxylation has been identified in a few kindreds. Increased circulating FGF23 levels impair 1 $\alpha$ -hydroxylation, preventing the production of 1,25(OH)<sub>2</sub>D. High levels of FGF23 are seen in those with genetic disorders associated with hypophosphatemic rickets, the most common of which is X-linked hypophosphatemia, and are prevalent in populations with profound renal dysfunction. Thus, therapeutic interventions should be considered in patients whose creatinine clearance is <0.5 mL/s (30 mL/min). Mutations in the renal 1 $\alpha$ -hydroxylase are the basis for the genetic disorder pseudovitamin D-deficiency rickets (also called vitamin D-dependent rickets type I). This autosomal recessive disorder presents with the syndrome of vitamin D deficiency in the first year of life. Patients present with growth retardation, rickets, and hypocalcemic seizures. Serum 1,25(OH)<sub>2</sub>D levels are low despite normal 25(OH)D levels and elevated PTH levels. Treatment with vitamin D metabolites that do not require 1 $\alpha$ -hydroxylation for activity results in disease remission, although lifelong therapy is required. A second autosomal recessive disorder, hereditary vitamin D-resistant rickets (also called vitamin D-dependent rickets type II), a consequence of vitamin D receptor mutations, is a greater therapeutic challenge. These patients present in a similar fashion during the first year of life, but alopecia often accompanies the disorder, demonstrating a functional role of the VDR in the keratinocyte stem cell population required for hair follicle regeneration. Serum levels of 1,25(OH)<sub>2</sub>D are dramatically elevated in these individuals both because of increased production due to stimulation of 1 $\alpha$ -hydroxylase activity as a consequence of secondary hyperparathyroidism and because of impaired inactivation since induction of the 24-hydroxylase by 1,25(OH)<sub>2</sub>D requires an intact VDR. Since the receptor mutation results in hormone resistance, daily calcium and phosphate infusions may be required to bypass the defect in intestinal mineral ion absorption.

Regardless of the cause, the clinical manifestations of vitamin D deficiency are largely a consequence of impaired intestinal calcium absorption. Mild to moderate vitamin D deficiency is asymptomatic, whereas long-standing vitamin D deficiency results in hypocalcemia accompanied by secondary hyperparathyroidism, impaired mineralization of the skeleton (osteopenia on x-ray or decreased bone mineral density), and proximal myopathy. Vitamin D deficiency also has been shown to be associated with an increase in overall mortality, including

cardiovascular causes. In the absence of an intercurrent illness, the hypocalcemia associated with long-standing vitamin D deficiency rarely presents with acute symptoms of hypocalcemia such as numbness, tingling, and seizures. However, the concurrent development of hypomagnesemia, which impairs parathyroid function, or the administration of potent bisphosphonates, which impair bone resorption, can lead to acute symptomatic hypocalcemia in vitamin D-deficient individuals.

**Rickets and Osteomalacia** In children, before epiphyseal fusion, vitamin D deficiency results in growth retardation associated with an expansion of the growth plate known as *rickets*. Three layers of chondrocytes are present in the normal growth plate: the reserve zone, the proliferating zone, and the hypertrophic zone. Rickets associated with impaired vitamin D action is characterized by expansion of the hypertrophic chondrocyte layer. The expansion of the growth plate is a consequence of impaired apoptosis of the late hypertrophic chondrocytes, an event that precedes replacement of these cells by osteoblasts during endochondral bone formation. Investigations in murine models demonstrate that hypophosphatemia, which in vitamin D deficiency is a consequence of secondary hyperparathyroidism, is a key etiologic factor in the development of the rachitic growth plate. Impaired actions specific to vitamin D also contribute to the expansion of the hypertrophic layer in the rachitic growth plate.

The hypocalcemia and hypophosphatemia that accompany vitamin D deficiency result in impaired mineralization of bone matrix proteins, a condition known as *osteomalacia*. Osteomalacia is also a feature of long-standing hypophosphatemia, which may result from renal phosphate wasting, or chronic use of etidronate or phosphate-binding antacids. This hypomineralized matrix is biomechanically inferior to normal bone; as a result, patients with osteomalacia are prone to bowing of weight-bearing extremities and skeletal fractures. Vitamin D and calcium supplementation have been shown to decrease the incidence of hip fracture among ambulatory nursing home residents in France, suggesting that undermineralization of bone contributes significantly to morbidity in the elderly. Proximal myopathy is a striking feature of severe vitamin D deficiency both in children and in adults. Rapid resolution of the myopathy is observed upon vitamin D treatment.

Although vitamin D deficiency is the most common cause of rickets and osteomalacia, many disorders lead to inadequate mineralization of the growth plate and bone. Calcium deficiency without vitamin D deficiency, the disorders of vitamin D metabolism previously discussed, and hypophosphatemia can all lead to inefficient mineralization. Even in the presence of normal calcium and phosphate levels, chronic acidosis and drugs such as bisphosphonates can lead to osteomalacia. The inorganic calcium/phosphate mineral phase of bone cannot form at low pH. Bisphosphonates bind to and prevent hydroxyapatite crystal growth. Since alkaline phosphatase is necessary for normal mineral deposition, probably because the enzyme can hydrolyze inhibitors of mineralization such as inorganic pyrophosphate, genetic inactivation of the alkaline phosphatase gene (hereditary hypophosphatasia) also can lead to osteomalacia in the setting of normal calcium and phosphate levels.

### Diagnosis of Vitamin D Deficiency, Rickets, and Osteomalacia

The most specific screening test for vitamin D deficiency in otherwise healthy individuals is a serum 25(OH)D level. Although the normal ranges vary, levels of 25(OH)D <37 nmol/L (<15 ng/mL) are associated with increasing PTH levels and lower bone density. The National Academy of Medicine has defined vitamin D sufficiency as a vitamin D level >50 nmol/L (>20 ng/mL), although higher levels may be required to optimize intestinal calcium absorption in the elderly and those with underlying disease states, including obesity. Vitamin D deficiency leads to impaired intestinal absorption of calcium, resulting in decreased serum total and ionized calcium values. This hypocalcemia results in secondary hyperparathyroidism, a homeostatic response that initially maintains serum calcium levels at the expense of the skeleton. Due to the PTH-induced increase in bone turnover, alkaline phosphatase levels are often increased. In addition to increasing bone resorption,

PTH decreases urinary calcium excretion while promoting phosphaturia. This results in hypophosphatemia, which exacerbates the mineralization defect in the skeleton. With prolonged vitamin D deficiency resulting in osteomalacia, calcium stores in the skeleton become relatively inaccessible, since osteoclasts cannot resorb unmineralized osteoid, and frank hypocalcemia ensues. Since PTH is a major stimulus for the renal 25(OH)D<sub>1</sub>-hydroxylase, there is increased synthesis of the active hormone, 1,25(OH)<sub>2</sub>D. Paradoxically, levels of this hormone are often normal in severe vitamin D deficiency. Therefore, measurements of 1,25(OH)<sub>2</sub>D are not accurate reflections of vitamin D stores and should not be used to diagnose vitamin D deficiency in patients with normal renal function.

Radiologic features of vitamin D deficiency in children include a widened, expanded growth plate that is characteristic of rickets. These findings not only are apparent in the long bones but also are present at the costochondral junction, where the expansion of the growth plate leads to swellings known as the "rachitic rosary." Impairment of intramembranous bone mineralization leads to delayed fusion of the calvarial sutures and a decrease in the radiopacity of cortical bone in the long bones. If vitamin D deficiency occurs after epiphyseal fusion, the main radiologic finding is a decrease in cortical thickness and relative radiolucency of the skeleton. A specific radiologic feature of osteomalacia, whether associated with phosphate wasting or vitamin D deficiency, is pseudofractures, or Looser's zones. These are radiolucent lines that occur where large arteries are in contact with the underlying skeletal elements; it is thought that the arterial pulsations lead to the radiolucencies. As a result, these pseudofractures are usually a few millimeters wide, are several centimeters long, and are seen particularly in the scapula, the pelvis, and the femoral neck.

## TREATMENT

### Vitamin D Deficiency

Based on the National Academy of Medicine 2010 report, the recommended daily intake of vitamin D is 600 IU from 1 to 70 years of age, and 800 IU for those >70. Based on the observation that 800 IU of vitamin D, with calcium supplementation, decreases the risk of hip fractures in elderly women, this higher dose is thought to be an appropriate daily intake for prevention of vitamin D deficiency in adults. The Vitamin D and Omega-3 Trial (VITAL) revealed that supplementation of vitamin D in people >50 years of age with normal vitamin D levels, at doses above the recommended daily intake, does not further improve bone mineral density or skeletal microarchitecture and does not prevent falls. The safety margin for vitamin D is large, and vitamin D toxicity usually is observed only in patients taking doses in the range of 40,000 IU daily. Treatment of vitamin D deficiency should be directed at the underlying disorder, if possible, and also should be tailored to the severity of the condition. Vitamin D should always be repleted in conjunction with calcium supplementation since most of the consequences of vitamin D deficiency are a result of impaired mineral ion homeostasis. In patients in whom 1-hydroxylation is impaired, metabolites that do not require this activation step are the treatment of choice. They include 1,25(OH)<sub>2</sub>D<sub>3</sub> (calcitriol [Rocaltrol], 0.25–0.5 µg/d) and 1-hydroxyvitamin D<sub>2</sub> (doxercalciferol [Hectorol], 2.5–5 µg/d). Outside the United States, 1-hydroxyvitamin D<sub>3</sub> (alfacalcidol [One-Alpha]) (0.25–1.0 µg/d) is also used. If the pathway required for activation of vitamin D is intact, severe vitamin D deficiency can be treated with pharmacologic repletion initially (50,000 IU weekly for 3–12 weeks), followed by maintenance therapy (800 IU daily). Pharmacologic doses may be required for maintenance therapy in patients who are taking medications such as barbiturates or phenytoin that accelerate metabolism of or cause resistance to 1,25(OH)<sub>2</sub>D. Polymorphisms in the 25-hydroxylase and the 24-hydroxylase genes can also lead to different responses to the normal recommended daily intake of vitamin D. The hepatic enzyme cytochrome P450 3A4 (CYP3A4) is a strong inducer of the catabolism of vitamin D metabolites. Polymorphisms of the

CYP3A4 gene and certain drugs, such as phenytoin and rifampin, lead to strong induction of this enzyme; thus, those affected may also require higher doses of vitamin D supplementation. Calcium supplementation should include 1.5–2 g/d of elemental calcium. Normocalcemia is usually observed within 1 week of the institution of therapy, although increases in PTH and alkaline phosphatase levels may persist for 3–6 months. The most efficacious methods to monitor treatment and resolution of vitamin D deficiency are serum and urinary calcium measurements. In patients who are vitamin D replete and are taking adequate calcium supplementation, the 24-h urinary calcium excretion should be in the range of 100–250 mg/24 h. Lower levels suggest problems with adherence to the treatment regimen or with absorption of calcium or vitamin D supplements. Levels >250 mg/24 h predispose to nephrolithiasis and should lead to a reduction in vitamin D dosage and/or calcium supplementation.

### Acknowledgement

The authors would like to acknowledge the advice of Marie Demay (also a former author of this chapter), Michael Mannstadt, and Marc Wein, who helped us put this chapter together.

### FURTHER READING

- Amrein K et al: Vitamin D deficiency 2.0: An update on the current status worldwide. *Eur J Clin Nutr* 74:1498, 2020.
- Bikle D et al: Vitamin D metabolites in captivity? Should we measure free or total 25(OH)D to assess vitamin D status? *J Steroid Biochem Mol Biol* 173:105, 2017.
- Carpenter TO et al: Burosomab therapy in children with X-linked hypophosphatemia. *N Engl J Med* 378:1987, 2018.
- Christakos S et al: Vitamin D: Metabolism, molecular mechanism of action, and pleiotropic effects. *Physiol Rev* 96:365, 2016.
- De Baaj JH et al: Magnesium in man: Implications for health and disease. *Physiol Rev* 95:1, 2015.
- Institute of Medicine (IOM): *Dietary Reference Intakes for Calcium and Vitamin D*. Washington, DC, National Academies Press, 2011.
- Kim JM et al: Osteoblast-osteoclast communication and bone homeostasis. *Cells* 10:2073, 2020.
- Robling AG, Bonevald LF: The osteocyte: New insights. *Ann Rev Physiol* 82:485, 2020.
- Siddiqui JA, Partridge NC: Physiological bone remodeling: Systemic regulation and growth factor involvement. *Physiology* 31:233, 2016.
- Zofkova I: Hypercalcemia. Pathophysiological aspects. *Phys Res* 65:1, 2016.

410

## Disorders of the Parathyroid Gland and Calcium Homeostasis

John T. Potts, Jr., Harald Jüppner



### PARATHYROID GLAND DISORDERS

#### INTRODUCTION

Four parathyroid glands are located posterior to the thyroid gland. They produce parathyroid hormone (PTH), which is the primary regulator of calcium physiology. PTH acts directly on bone, where it induces calcium (and phosphate) release, and on the kidney, where it enhances calcium reabsorption in the distal tubules. In the proximal renal tubules, PTH increases excretion of phosphate and the synthesis of 1,25-dihydroxyvitamin D (1,25[OH]<sub>2</sub>D), a hormone that increases

gastrointestinal calcium absorption. Serum PTH levels are tightly regulated by a negative feedback loop. Calcium, acting through the calcium-sensing receptor, and vitamin D, acting through its nuclear receptor, reduce PTH release and synthesis. Additional evidence indicates that fibroblast growth factor 23 (FGF23), a phosphaturic hormone, can suppress PTH secretion. Understanding the hormonal pathways that regulate calcium and phosphate levels as well as bone metabolism is essential for effective diagnosis and management of a wide array of hyper- and hypocalcemic disorders.

Hyperparathyroidism, characterized by excess production of PTH, is a common cause of hypercalcemia and is usually the result of autonomously functioning adenomas or hyperplasia. Surgery for this disorder is highly effective and has been shown to reverse some of the deleterious effects of long-standing PTH excess on bone density. Humoral hypercalcemia of malignancy (HJM) is also a common cause of hypercalcemia, which is usually due to the overproduction of parathyroid hormone-related peptide (PTHrP) by cancer cells. The similarities in the biochemical characteristics of hyperparathyroidism and HJM, first noted by Albright in 1941, are now known to reflect the actions of PTH and PTHrP through the same G protein-coupled PTH/PTHrP receptor (PTHR1). The converse, namely hypocalcemia, can be caused by the lack of functional PTH, i.e., hypoparathyroidism, or by reduced PTH responsiveness of the proximal renal tubules, i.e., pseudohypoparathyroidism (PHP).

The genetic basis of hyperparathyroidism, multiple endocrine neoplasia (MEN) types 1 and 2, familial hypocalciuric hypercalcemia (FHH), Jansen's syndrome, different forms of hypoparathyroidism and PHP, disorders of excess urinary phosphate excretion and of vitamin D synthesis, action, and metabolism, and the molecular events associated with parathyroid gland neoplasia has provided new insights into the regulation of calcium and phosphate homeostasis. In addition, PTH and possibly some of its analogues are promising therapeutic agents for the treatment of postmenopausal or senile osteoporosis, and calcimimetic agents, which activate the calcium-sensing receptor, have provided new approaches for PTH suppression.

## PARATHYROID GLAND DISORDERS

### PTH

**Structure and Physiology** PTH is an 84-amino-acid single-chain peptide. The amino-terminal portion, PTH(1–34), is highly conserved and is critical for the biologic actions of the molecule. Modified synthetic fragments of the amino-terminal sequence as small as PTH(1–11) are sufficient to activate the PTH/PTHrP receptor, if provided at high enough concentrations (see below). The carboxyl-terminal portions of full-length PTH(1–84) also can bind to a separate binding protein/receptor; however, its properties and biologic role(s), if any, remain undefined.

The primary function of PTH is to maintain the extracellular fluid (ECF) calcium concentration within a narrow normal range. The hormone acts directly on bone and kidney and indirectly on the intestine through its effects on synthesis of  $1,25(\text{OH})_2\text{D}$  to increase serum calcium concentrations; in turn, PTH production is closely regulated by the concentration of serum ionized calcium. This feedback system is the critical homeostatic mechanism for maintenance of ECF calcium. Any tendency toward hypocalcemia, as might be induced by calcium- or vitamin D-deficient diets, is counteracted by an increased secretion of PTH. This in turn (1) increases the rate of dissolution of bone mineral, thereby increasing the flow of calcium (and phosphate) from bone into blood; (2) reduces the renal clearance of calcium, returning more of the calcium and phosphate filtered at the glomerulus into ECF; and (3) increases the efficiency of calcium absorption in the intestine by stimulating the production of  $1,25(\text{OH})_2\text{D}$ . Immediate control of blood calcium is due to PTH effects on bone and, to a lesser extent, on renal calcium clearance. Maintenance of steady-state calcium balance, on the other hand, probably results from the effects of  $1,25(\text{OH})_2\text{D}$  on calcium absorption (Chap. 409). The renal actions of PTH are exerted at multiple sites and include an increase in urinary phosphate

excretion (proximal tubule), augmentation of calcium reabsorption (distal tubule), and stimulation of the renal 25(OH)D-1-hydroxylase. As much as 12 mmol (500 mg) of calcium is transferred between the ECF and bone each day (a large amount in relation to the total ECF calcium pool), and PTH has a major effect on this transfer. The homeostatic role of the hormone can preserve calcium concentration in blood at the cost of bone demineralization.

PTH has multiple actions on bone, some direct and some indirect. PTH-mediated changes in bone calcium release can be seen within minutes. The chronic effects of PTH are to increase the number of bone cells, both osteoblasts and osteoclasts, and to increase the remodeling of bone; these effects are apparent within hours after the hormone is given and persist for hours after PTH is withdrawn. Continuous exposure to elevated PTH (as in hyperparathyroidism or long-term infusions in animals) leads to increased osteoclast-mediated bone resorption. However, the intermittent administration of small amounts of PTH, elevating hormone levels for 1–2 h each day, leads to a net increase of bone mass rather than bone loss. Striking increases, especially in trabecular bone in the spine and hip, have been reported with the use of PTH. PTH(1–34) as monotherapy caused a highly significant reduction in fracture incidence in a worldwide placebo-controlled trial.

Osteoblasts (or their stromal cell precursors), which have PTH/PTHrP receptors, are crucial to this bone-forming effect of PTH. When PTH activates PTH/PTHrP receptors on osteocytes, release of calcium from the matrix surrounding these cells is enhanced; osteoclasts, which mediate bone breakdown, lack such receptors. PTH-mediated stimulation of osteoclasts is indirect, acting in part through cytokine released from osteoblasts to activate osteoclasts; in experimental studies of bone resorption in vitro, osteoblasts must be present for PTH to activate osteoclasts to resorb bone (Chap. 409).

**Synthesis, Secretion, and Metabolism** • **SYNTHESIS** Parathyroid cells have multiple methods of adapting to increased needs for PTH production. Most rapid (within minutes) is secretion of preformed hormone in response to hypocalcemia. Second, within hours, PTH mRNA expression is induced by sustained hypocalcemia. Finally, protracted challenge leads within days to cellular replication to increase parathyroid gland mass.

PTH is initially synthesized as a larger molecule (preproparathyroid hormone, consisting of 115 amino acids). After a first cleavage step to remove the "pre" sequence of 25 amino acid residues, a second cleavage step removes the "pro" sequence of 6 amino acid residues before secretion of the mature peptide comprising 84 residues. Mutations in the prepro-region of the gene can cause hypoparathyroidism by interfering with hormone synthesis, transport, or secretion.

Transcriptional suppression of the PTH gene by calcium is nearly maximal at physiologic calcium concentrations. Hypocalcemia increases transcriptional activity within hours.  $1,25(\text{OH})_2\text{D}$  strongly suppresses PTH gene transcription. In patients with chronic kidney disease (CKD), IV administration of supraphysiologic levels of  $1,25(\text{OH})_2\text{D}$  or analogues of this active metabolite can dramatically suppress PTH overproduction and is thus used clinically to control severe secondary hyperparathyroidism. Regulation of proteolytic destruction of preformed hormone (posttranslational regulation of hormone production) is an important mechanism for mediating rapid (within minutes) changes in hormone availability. High calcium increases and low calcium inhibits the proteolytic destruction of stored hormone.

**REGULATION OF PTH SECRETION** PTH secretion increases steeply to a maximum value of about five times the basal rate of secretion as the calcium concentration falls from normal to 1.9–2.0 mmol/L (7.6–8.0 mg/dL; measured as total calcium). However, the ionized fraction of blood calcium is the important determinant of hormone secretion. Severe intracellular magnesium deficiency impairs PTH secretion (see below).

ECF calcium controls PTH secretion by interaction with a calcium-sensing receptor (CaSR), a G protein-coupled receptor (GPCR) for which  $\text{Ca}^{2+}$  ions act as the primary ligand (see below). This receptor, which also has phosphate binding sites, is a member of a distinctive

subgroup of the GPCR superfamily that mediates its actions through two closely related alpha-subunits of signaling G proteins, namely G<sub>q</sub> and G<sub>11</sub>, and is characterized by a large extracellular domain suitable for “clamping” the small-molecule ligand. Stimulation of the CaSR by high calcium levels suppresses PTH secretion. The CaSR is present in parathyroid glands and the calcitonin-secreting cells of the thyroid (C cells), as well as in multiple other sites, including brain and kidney. Genetic evidence has revealed a key biologic role for the CaSR in parathyroid gland responsiveness to calcium and in renal calcium clearance. Heterozygous loss-of-function mutations in CaSR cause the syndrome of FHH, in which the blood calcium abnormality resembles that observed in hyperparathyroidism but with hypocalciuria; two more recently defined variants of FHH, namely FHH2 and FHH3, are caused either by heterozygous loss-of-function mutations in G<sub>11</sub>, the alpha-subunit of one of the signaling proteins downstream of the CaSR, or by heterozygous mutations in AP2A1. Homozygous loss-of-function mutations in the CaSR are the cause of severe neonatal hyperparathyroidism, a disorder that is typically lethal if not treated within the first days of life. On the other hand, heterozygous gain-of-function mutations cause a form of hypocalcemia resembling hypoparathyroidism (see below).

**METABOLISM** The secreted form of PTH is indistinguishable by immunologic criteria and by molecular size from the 84-amino-acid peptide (PTH[1–84]) extracted from glands. However, much of the immunoreactive material found in the circulation is smaller than the extracted or secreted hormone. The principal circulating fragments of immunoreactive hormone lack a portion of the critical amino-terminal sequence required for biologic activity and, hence, are biologically inactive fragments (so-called middle and carboxyl-terminal fragments). Much of the proteolysis of hormone occurs in the liver and kidney. Peripheral metabolism of PTH does not appear to be regulated by physiologic states (high vs low calcium, etc.); hence, peripheral metabolism of hormone, although responsible for rapid clearance of secreted hormone, appears to be a high-capacity, metabolically invariant catabolic process.

The rate of clearance of the secreted 84-amino-acid peptide from blood is more rapid than the rate of clearance of the biologically inactive fragment(s) corresponding to the middle and carboxyl-terminal regions of PTH. Consequently, the interpretation of results obtained with earlier PTH radioimmunoassays was influenced by the nature of the peptide fragments detected by the anti-PTH antibodies used in these assays.

Although the problems inherent in PTH measurements have been largely circumvented by use of double-antibody immunometric assays, it is now known that some of these assays detect, besides the intact molecule, large amino-terminally truncated forms of PTH, which are present in normal and uremic individuals in addition to PTH(1–84). The concentration of these fragments relative to that of full-length PTH(1–84) is higher with induced hypercalcemia than in eucalcemic or hypocalcemic conditions and is higher in patients with impaired renal function. PTH(7–84) has been identified as a major component of these amino-terminally truncated fragments. Some evidence suggests that the PTH(7–84) (and probably related amino-terminally

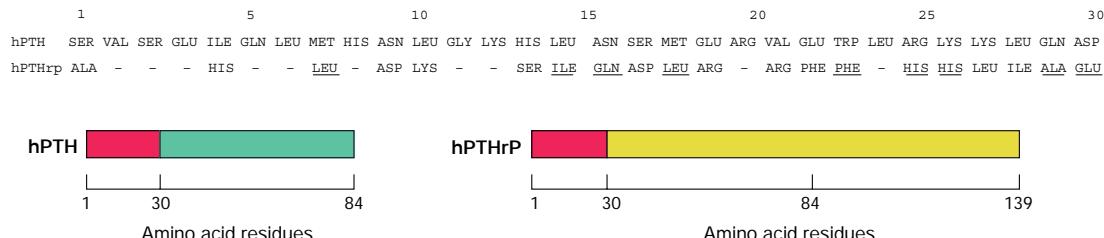
truncated fragments) can act, through yet undefined mechanisms, as an inhibitor of PTH action and may therefore be of clinical significance, particularly in patients with CKD. In this group of patients, efforts to prevent secondary hyperparathyroidism by a variety of measures (vitamin D analogues, higher calcium intake, higher dialysate calcium, phosphate-lowering strategies, and calcimimetic drugs) can lead to oversuppression of the parathyroid glands since some amino-terminally truncated PTH fragments, such as PTH(7–84), react in many immunometric PTH assays (now termed *second-generation* assays; see below under “Diagnosis”), thus overestimating the levels of biologically active, intact PTH. Excessive parathyroid gland suppression in CKD can lead to adynamic bone disease (see below), which has been associated in children with further impaired growth and increased bone fracture rates in adults and can furthermore lead to significant hypercalcemia. The measurement of PTH with newer third-generation immunometric assays, which use detection antibodies directed against extreme amino-terminal PTH epitopes and thus detect only full-length PTH(1–84), may provide some advantage to prevent bone disease in CKD.

### PTHRP

**Structure and Physiology** PTHrP is responsible for most instances of HHM (Chap. 93), a syndrome that resembles primary hyperparathyroidism but without elevated PTH levels. Most cell types normally produce PTHrP, including brain, pancreas, heart, lung, mammary tissue, placenta, endothelial cells, and smooth muscle. In fetal animals, PTHrP directs transplacental calcium transfer, and high concentrations of PTHrP are produced in mammary tissue and secreted into milk, but the biologic significance of this hormone in breast milk is unknown. PTHrP also plays an essential role in endochondral bone formation and in branching morphogenesis of the breast and possibly in uterine contraction and other biologic functions.

PTH and PTHrP, although products of different genes, exhibit considerable functional and structural homology (Fig. 410-1) and have evolved from a shared ancestral gene. The structure of the gene encoding human PTHrP, however, is more complex than that of PTH, containing multiple additional exons, which can undergo alternate splicing patterns during formation of the mature mRNA. Protein products of 139, 141, and 173 amino acids are produced, and other molecular forms may result from tissue-specific degradation at accessible internal cleavage sites. The biologic roles of these various molecular species and the nature of the circulating forms of PTHrP are unclear. In fact, it is uncertain whether PTHrP circulates at any significant level in healthy children and adults. As a paracrine factor, PTHrP may be produced, act, and be destroyed locally within tissues. In adults, PTHrP appears to have little influence on calcium homeostasis, except in disease states, when large tumors, especially of the squamous cell type as well as renal cell carcinomas, lead to massive overproduction of the hormone and hypercalcemia.

Both PTH and PTHrP bind to and activate the PTH/PTHrP receptor. The PTH/PTHrP receptor (also known as the PTH-1 receptor, PTHR1) belongs to a subfamily of GPCRs that includes the receptors for calcitonin, glucagon, secretin, vasoactive intestinal peptide, and a



**FIGURE 410-1** Schematic diagram to illustrate similarities and differences in structure of human parathyroid hormone (hPTH) and human PTH-related peptide (hPTHrP). Close structural (and functional) homology exists between the first 30 amino acids of hPTH and hPTHrP. The PTHrP sequence may be 139 amino acid residues in length. PTH is only 84 residues long; after residue 30, there is little structural homology between the two. Dashed lines in the PTHrP sequence indicate identity; underlined residues, although different from those of PTH, still represent conservative changes (charge or polarity preserved). Ten amino acids are identical, and a total of 20 of 30 are homologous.

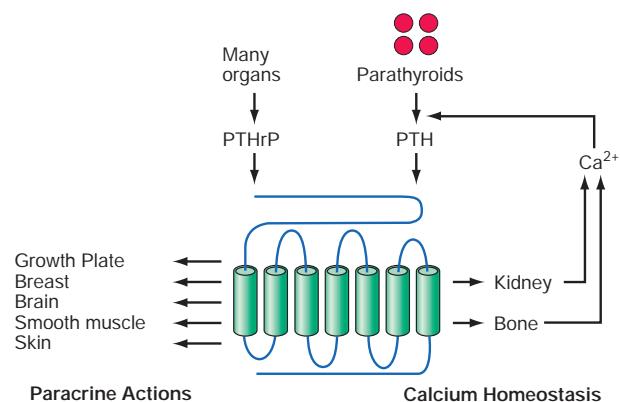
**3172** few other peptides. Although both ligands activate the PTHR1, the two peptides induce distinct responses in the receptor, which explains how a single receptor without isoforms can serve different biologic roles. The extracellular regions of the receptor are involved in hormone binding, and the intracellular domains, after hormone activation, bind G protein subunits to transduce hormone signaling into cellular responses through the stimulation of second messenger formation. A second receptor that binds PTH, originally termed the PTH-2 receptor (PTH2R), is primarily expressed in brain, pancreas, and testis. Different mammalian PTHR1s respond equivalently to PTH and PTHrP, at least when tested with traditional assays, whereas the human PTH2R responds efficiently only to PTH, but not to PTHrP. PTH2Rs from other species show little or no stimulation of second-messenger formation in response to PTH or PTHrP. In fact, the endogenous ligand of the PTH2R was shown to be a hypothalamic peptide referred to as tubular infundibular peptide of 39 residues, TIP39, that is distantly related to PTH and PTHrP. The PTHR1 and the PTH2R can be traced backward in evolutionary time to fish. Furthermore, the zebrafish genome contains, in addition to the PTHR1 and the PTH2R orthologs, a third receptor, the PTH3R, that is more closely related to the fish PTHR1 than to the fish PTH2R. The evolutionary conservation of structure and function suggests important biologic roles for these receptors, even in fish, which lack discrete parathyroid glands but produce two molecules that are closely related to mammalian PTH.

Studies using the cloned PTHR1 confirm that it can be coupled to more than one G protein and second-messenger pathway, apparently explaining the multiplicity of pathways stimulated by PTH. Activation of protein kinases (A and C) and calcium transport channels is associated with a variety of hormone-specific tissue responses. These responses include inhibition of phosphate and bicarbonate transport, stimulation of calcium transport, and activation of renal 1-hydroxylase in the kidney. The responses in bone include effects on collagen synthesis, alkaline phosphatase, ornithine decarboxylase, citrate decarboxylase, and glucose-6-phosphate dehydrogenase activities; phospholipid synthesis; and calcium and phosphate transport. Ultimately, these biochemical events lead to an integrated hormonal response in bone turnover and calcium homeostasis. PTH also activates  $\text{Na}^+/\text{Ca}^{2+}$  exchangers at renal distal tubular sites and stimulates translocation of preformed calcium transport channels, moving them from the interior to the apical surface to increase tubular uptake of calcium. PTH-dependent stimulation of phosphate excretion involves reduced expression of two sodium-dependent phosphate co-transporters, NPT2a and NPT2c, at the apical membrane, thereby reducing phosphate reabsorption in the proximal renal tubules. Similar mechanisms may be involved in other renal tubular transporters that are influenced by PTH. Recent studies reaffirm the critical linkage of blood phosphate lowering to net calcium entry into blood by PTH action and emphasize the participation of bone cells other than osteoclasts in the rapid calcium-elevating actions of PTH.

PTHrP exerts important developmental influences on fetal bone development and in adult physiology. Homozygous ablation of the gene encoding PTHrP (or disruption of the PTHR1 gene) in mice causes a lethal phenotype in which animals are born with pronounced acceleration of chondrocyte maturation that resembles a lethal form of chondrodysplasia in humans that is caused by homozygous or compound heterozygous, inactivating PTHR1 mutations (Fig. 410-2). Heterozygous inactivating PTHR1 mutations in humans furthermore can be a cause of delayed tooth eruption, while heterozygous inactivating PTHrP mutations lead to premature growth plate closure and reduced adult heights.

## CALCITONIN

(See also Chap. 388) Calcitonin is a peptide hormone with hypocalcemic properties that in several mammalian species acts as an indirect antagonist to the calcemic actions of PTH. Calcitonin seems to be of limited physiologic significance in humans, at least with regard to calcium homeostasis. It is of medical significance because of its role as a tumor marker in sporadic and hereditary cases of medullary thyroid carcinoma and its medical use as an adjunctive treatment in severe



**FIGURE 410-2** Dual role for the actions of the PTH/PTHrP receptor (PTHR1). Parathyroid hormone (PTH; endocrine–calcium homeostasis) and PTH-related peptide (PTHrP; paracrine–multiple tissue actions including growth plate cartilage in developing bone) use the single receptor for their disparate functions mediated by the amino-terminal 34 residues of either peptide. Other regions of both ligands interact with other receptors (not shown).

hypercalcemia and in Paget's disease of bone. Levels can also be elevated in patients with PHP.

The hypocalcemic activity of calcitonin is accounted for primarily by inhibition of osteoclast-mediated bone resorption and secondarily by stimulation of renal calcium clearance. These effects are mediated by receptors on osteoclasts and renal tubular cells. Calcitonin exerts additional effects through receptors present in the brain, the gastrointestinal tract, and the immune system. The hormone, for example, exerts analgesic effects directly on cells in the hypothalamus and related structures, possibly by interacting with receptors for related peptide hormones such as calcitonin gene-related peptide (CGRP) or amylin. Both of these ligands have specific high-affinity receptors that share considerable structural similarity with the PTHR1 and can also bind to and activate calcitonin receptors. The calcitonin receptor shares considerable structural similarity with the PTHR1.

The naturally occurring calcitonins consist of a peptide chain of 32 amino acids. There is considerable sequence variability among species. Calcitonin from salmon, which is used therapeutically, is 10–100 times more potent than mammalian forms in lowering serum calcium.

The circulating level of calcitonin in humans is lower than that in many other species. In humans, even extreme variations in calcitonin production do not change calcium and phosphate metabolism; no definite effects are attributable to calcitonin deficiency (totally thyroidectomized patients receiving only replacement thyroxine) or excess (patients with medullary carcinoma of the thyroid, a calcitonin-secreting tumor) (Chap. 388). Calcitonin has been a useful pharmacologic agent to suppress bone resorption in Paget's disease (Chap. 412) and osteoporosis (Chap. 411) and in the treatment of hypercalcemia of malignancy (see below). However, bisphosphonates are usually more effective, and the physiologic role, if any, of calcitonin in humans is uncertain. On the other hand, ablation of the calcitonin gene (combined because of the close proximity with ablation of the CGRP gene) in mice leads to reduced bone mineral density, suggesting that its biologic role in mammals is still not fully understood.

## HYPERCALCEMIA

**Introduction** (See also Chap. 54) Hypercalcemia can be a manifestation of a serious illness such as malignancy or can be detected coincidentally by laboratory testing in a patient with no obvious illness. The number of patients recognized with asymptomatic hypercalcemia, usually hyperparathyroidism, increased in the late twentieth century when wider testing became readily available.

Whenever hypercalcemia is confirmed, a definitive diagnosis must be established. Although hyperparathyroidism, a frequent cause of asymptomatic hypercalcemia, is a chronic disorder in which manifestations, if any, may be expressed only after months or years,

**TABLE 410-1 Classification of Causes of Hypercalcemia**

<b>I. Parathyroid-Related</b>
A. Primary hyperparathyroidism
1. Adenoma(s)
2. Multiple endocrine neoplasia
3. Carcinoma
B. Lithium therapy
C. Familial hypocalciuric hypercalcemia
<b>II. Malignancy-Related</b>
A. Solid tumor with metastases (breast)
B. Solid tumor with humoral mediation of hypercalcemia (lung, kidney)
C. Hematologic malignancies (multiple myeloma, lymphoma, leukemia)
<b>III. Vitamin D-Related</b>
A. Vitamin D intoxication
B. ↑ 1,25(OH) <sub>2</sub> D: sarcoidosis and other granulomatous diseases, lymphoma
C. ↑ 1,25(OH) <sub>2</sub> D: impaired 1,25(OH) <sub>2</sub> D metabolism due to 24-hydroxylase deficiency or increased 1,25(OH) <sub>2</sub> D synthesis due to inactivating mutations involving the sodium-dependent phosphate co-transporters
<b>IV. Associated with High Bone Turnover</b>
A. Hyperthyroidism
B. Immobilization
C. Thiazides
D. Vitamin A intoxication
E. Fat necrosis
<b>V. Associated with Renal Failure</b>
A. Severe secondary hyperparathyroidism
B. Aluminum intoxication and adynamic bone disease
C. Milk-alkali syndrome

hypercalcemia can also be the earliest manifestation of malignancy, the second most common cause of hypercalcemia in the adult. The causes of hypercalcemia are numerous (Table 410-1), but hyperparathyroidism and cancer account for 90% of all cases.

Before undertaking a diagnostic workup, it is essential to be sure that true hypercalcemia, not a false-positive laboratory test, is present. A false-positive diagnosis of hypercalcemia is usually the result of inadvertent hemoconcentration during blood collection or elevation in serum proteins such as albumin. Hypercalcemia is in most cases a chronic problem, and it is cost-effective to obtain several serum calcium measurements; these tests need not be in the fasting state.

Clinical features are helpful in differential diagnosis. Hypercalcemia in an adult who is asymptomatic is usually due to primary hyperparathyroidism. In malignancy-associated hypercalcemia, the disease is usually not occult; rather, symptoms of malignancy bring the patient to the physician, and hypercalcemia is discovered during the evaluation. In such patients, the interval between detection of hypercalcemia and death, especially without vigorous treatment, is often <6 months. Accordingly, if an asymptomatic individual has had hypercalcemia or some manifestation of hypercalcemia such as kidney stones for >1 or 2 years, it is unlikely that malignancy is the cause. Nevertheless, differentiating primary hyperparathyroidism from occult malignancy can occasionally be difficult, and careful evaluation is required, particularly when the duration of the hypercalcemia is unknown. Hypercalcemia not due to hyperparathyroidism or malignancy can result from excessive vitamin D action, impaired metabolism of 1,25(OH)<sub>2</sub>D, high bone turnover from any of several causes, or renal failure (Table 410-1). Dietary history and a history of ingestion of vitamins or drugs are often helpful in diagnosing some of the less frequent causes. Immunometric PTH assays serve as the principal laboratory test in establishing the diagnosis.

Hypercalcemia from any cause can result in fatigue, depression, mental confusion, anorexia, nausea, vomiting, constipation, reversible renal tubular defects, increased urine output, a short QT interval in the electrocardiogram, and, in some patients, cardiac arrhythmias.

There is a variable relation from one patient to the next between the severity of hypercalcemia and the symptoms. Generally, symptoms are more common at calcium levels >2.9–3.0 mmol/L (11.6–12.0 mg/dL), but some patients, even at this level, are asymptomatic. When the calcium level is >3.2 mmol/L (12.8 mg/dL), calcification in kidneys, skin, vessels, lungs, heart, and stomach occurs, and renal insufficiency may develop, particularly if blood phosphate levels are normal or elevated due to impaired renal excretion. Severe hypercalcemia, usually defined as 3.7–4.5 mmol/L (14.8–18.0 mg/dL), can be a medical emergency; coma and cardiac arrest can occur.

Acute management of the hypercalcemia is usually successful. The type of treatment is based on the severity of the hypercalcemia and the nature of associated symptoms, as outlined below.

## PRIMARY HYPERPARATHYROIDISM

**Pathophysiology** • **NATURAL HISTORY AND INCIDENCE** Primary hyperparathyroidism is a generalized disorder of calcium, phosphate, and bone metabolism due to an increased secretion of PTH. The elevation of circulating hormone usually leads to hypercalcemia and hypophosphatemia. There is great variation in the manifestations. Patients may present with multiple signs and symptoms, including recurrent nephrolithiasis, peptic ulcers, mental changes, and, less frequently, extensive bone resorption. However, with greater awareness of the disease and wider use of multiphasic screening tests, including measurements of blood calcium, the diagnosis is frequently made in patients who have no symptoms and minimal, if any, signs of the disease other than hypercalcemia and elevated levels of PTH. The manifestations may be subtle, and the disease may have a benign course for many years or a lifetime. This milder form of the disease is usually termed *asymptomatic hyperparathyroidism*. Rarely, hyperparathyroidism develops or worsens abruptly and causes severe complications such as marked dehydration and coma, so-called hypercalcemic parathyroid crisis.

The annual incidence of the disease is calculated to be as high as 0.2% in patients >60 years old, with an estimated prevalence, including undiscovered asymptomatic patients, of 1%; some reports suggest the incidence may be declining. If confirmed, these changing estimates may reflect less frequent routine testing of serum calcium in recent years, earlier overestimates in incidence, or unknown factors. The disease has a peak incidence between the third and fifth decades but occurs in young children and the elderly.

**ETIOLOGY** Parathyroid tumors are most often encountered as isolated adenomas without other endocrinopathy. They may also arise in hereditary syndromes such as MEN syndromes. As many as 10% of patients with hyperparathyroidism are found to have mutations in 1 of 11 genes (see below). Parathyroid tumors may also arise as secondary to underlying disease (excessive stimulation in secondary hyperparathyroidism, especially chronic renal failure), or after other forms of excessive stimulation such as lithium therapy. These etiologies are discussed below.

**Solitary Adenomas** A single abnormal gland is the cause in ~80% of patients; the abnormality in the gland is usually a benign neoplasm or adenoma and rarely a parathyroid carcinoma. Some surgeons and pathologists report that the enlargement of multiple glands is common; double adenomas are reported. In ~15% of patients, all glands are hyperfunctioning; *chief cell parathyroid hyperplasia* is usually hereditary and frequently associated with other endocrine abnormalities.

**Hereditary Syndromes and Multiple Parathyroid Tumors** Hereditary hyperparathyroidism can occur without other endocrine abnormalities but is usually part of a MEN syndrome (Chap. 388). MEN 1 (Wermer's syndrome) consists of hyperparathyroidism and tumors of the pituitary and pancreas, often associated with gastric hypersecretion and peptic ulcer disease (Zollinger-Ellison syndrome). MEN 2A is characterized by pheochromocytoma and medullary carcinoma of the thyroid, as well as hyperparathyroidism; MEN 2B has additional associated features such as multiple neuromas but usually lacks hyperparathyroidism. Each of these MEN syndromes is transmitted in an apparent

3174 autosomal dominant manner, although, as noted below, the genetic basis of MEN 1 involves biallelic loss of a tumor suppressor.

The *hyperparathyroidism jaw tumor* (HPT-JT) syndrome occurs in families with parathyroid tumors (sometimes carcinomas) in association with benign jaw tumors. This disorder is caused by mutations in *CDC73* (*HRPT2*), and mutations in this gene are also observed in sporadic parathyroid cancers. Some kindreds exhibit hereditary hyperparathyroidism without other endocrinopathies, which has been referred to as *nonsyndromic familial isolated hyperparathyroidism* (FIHP). In some of these familial cases, the disease co-segregates with heterozygous mutations in *GCM2*. Inactivating or activating mutations in this parathyroid-specific transcription factor had initially been identified in familial forms of hypoparathyroidism. However, identification of a *GCM2* mutation in a parathyroid adenoma raised the possibility that mutations in this gene could also be causing some forms of FIHP. Furthermore, there is speculation that some FIHP cases may be examples of variable expression of the other syndromes such as MEN 1, MEN 2, or the HPT-JT syndrome, but they may also have distinctive, still unidentified genetic causes.

**Genetic Defects Associated with Hyperparathyroidism** As in many other types of neoplasia, two fundamental types of genetic defects have been identified in parathyroid gland tumors: (1) overactivity of protooncogenes and (2) loss of function of tumor-suppressor genes. The former, by definition, can lead to uncontrolled cellular growth and function by activation (gain-of-function mutation) of a single allele of the responsible gene, whereas the latter requires loss of function of both allelic copies. Biallelic loss of function of a tumor-suppressor gene is usually characterized by a germline defect (all cells) and an additional somatic deletion/mutation in the tumor (Fig. 410-3).

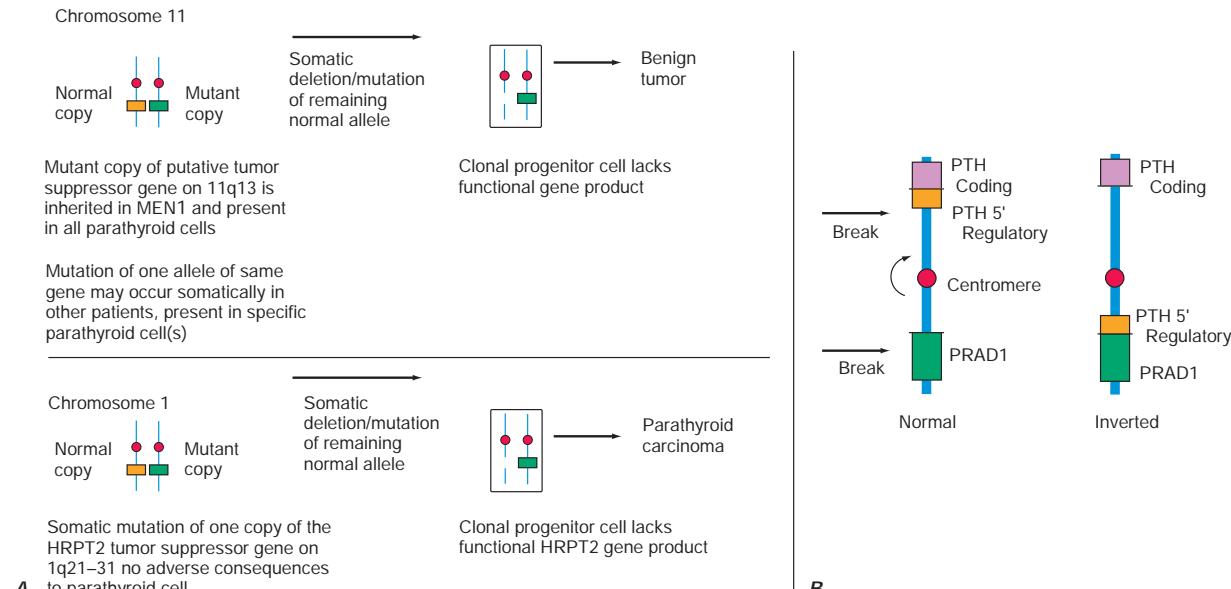
Mutations in the *MEN1* gene locus, encoding the protein MENIN, on chromosome 11q13 are responsible for causing MEN 1; the normal allele of this gene fits the definition of a tumor-suppressor gene. Inheritance of one mutated allele in this hereditary syndrome, followed by

loss of the other allele via somatic cell mutation, leads to monoclonal expansion and tumor development. Also, in ~15–20% of sporadic parathyroid adenomas, both alleles of the *MEN1* locus on chromosome 11 are somatically deleted, implying that the same defect responsible for MEN 1 can also cause the sporadic disease (Fig. 410-3A). Consistent with the Knudson hypothesis for two-step neoplasia in certain inherited cancer syndromes (Chap. 71), the earlier onset of hyperparathyroidism in the hereditary syndromes reflects the need for only one mutational event to trigger the monoclonal outgrowth. In sporadic adenomas, typically occurring later in life, two different somatic events must occur before the *MEN1* gene is silenced.

Other presumptive anti-oncogenes involved in hyperparathyroidism include a still unidentified gene mapped to chromosome 1p seen in 40% of sporadic parathyroid adenomas and a gene mapped to chromosome Xp11 in patients with secondary hyperparathyroidism and renal failure, who progressed to “tertiary” hyperparathyroidism, now known to reflect monoclonal outgrowths within previously hyperplastic glands.

A more complex pattern, still incompletely resolved, arises with genetic defects and carcinoma of the parathyroids. This appears to be due to biallelic loss of a functioning copy of a gene, *HRPT2* (or *CDC73*), originally identified as the cause of the HPT-JT syndrome. Several inactivating mutations have been identified in *HRPT2* (located on chromosome 1q21–31), which encodes a 531-amino-acid protein called parafibromin. The responsible genetic mutations in *HRPT2* (or another gene) appear to be necessary, but not sufficient, for parathyroid cancer.

In general, the detection of additional genetic defects in these parathyroid tumor-related syndromes and the variations seen in phenotypic expression/penetrance indicate the multiplicity of the genetic factors responsible. Nonetheless, the ability to detect the presence of the major genetic contributors has greatly aided a more informed management of family members of patients identified in the hereditary syndromes such as MEN 1, MEN 2, and HPT-JT.



**FIGURE 410-3** **A.** Schematic diagram indicating molecular events in tumor susceptibility. The patient with the hereditary abnormality (multiple endocrine neoplasia [MEN]) is envisioned as having one defective gene inherited from the affected parent on chromosome 11, but one copy of the normal gene is present from the other parent. In the monoclonal tumor (benign tumor), a somatic event, here partial chromosomal deletion, removes the remaining normal gene from a cell. In nonhereditary tumors, two successive somatic mutations must occur, a process that takes a longer time. By either pathway, the cell, deprived of growth-regulating influence from this gene, has unregulated growth and becomes a tumor. A different genetic locus also involving loss of a tumor-suppressor gene termed *HRPT2* is involved in the pathogenesis of parathyroid carcinoma. **B.** Schematic illustration of the mechanism and consequences of gene rearrangement and overexpression of the *PRAD1* protooncogene (pericentromeric inversion of chromosome 11) in parathyroid adenomas. The excessive expression of *PRAD1* (a cell cycle control protein, cyclin D1) by the highly active PTH gene promoter in the parathyroid cell contributes to excess cellular proliferation. (Image A reproduced with permission from A Arnold: *Genetic basis of endocrine disease 5. Molecular genetics of parathyroid gland neoplasia*. J Clin Endocrinol Metab 77:1108, 993. Image B reproduced with permission from J Habener, in L DeGroot, JL Jameson (eds): *Endocrinology*, 4th ed. Philadelphia, PA: Saunders; 2001.)

An important contribution from studies on the genetic origin of parathyroid carcinoma has been the realization that the mutations involve a different pathway than that involved with the benign gland enlargements. Unlike the pathogenesis of genetic alterations seen in colon cancer, where lesions evolve from benign adenomas to malignant disease by progressive genetic changes, the alterations commonly seen in most parathyroid cancers (*HRPT2* mutations) are infrequently seen in sporadic parathyroid adenomas.

Abnormalities at the *Rb* gene were the first to be noted in parathyroid cancer. The *Rb* gene, a tumor-suppressor gene located on chromosome 13q14, was initially associated with retinoblastoma but has since been implicated in other neoplasias, including parathyroid carcinoma. Early studies implicated allelic deletions of the *Rb* gene in many parathyroid carcinomas and decreased or absent expression of the *Rb* protein. However, because there are often large deletions in chromosome 13 that include many genes in addition to the *Rb* locus (with similar findings in some pituitary carcinomas), it remains possible that other tumor-suppressor genes on chromosome 13 may be playing a role in parathyroid carcinoma.

Study of the parathyroid cancers found in some patients with the HPT-JT syndrome has led to identification of a much larger role for mutations in the *HRPT2* gene in most parathyroid carcinomas, including those that arise sporadically, without apparent association with the HPT-JT syndrome. Mutations in the coding region have been identified in 75–80% of all parathyroid cancers analyzed, leading to the conclusion that, with addition of presumed mutations in the noncoding regions, this genetic defect may be seen in essentially all parathyroid carcinomas. Of special importance was the discovery that, in some sporadic parathyroid cancers, germline mutations have been found; this, in turn, has led to careful investigation of the families of these patients and a new clinical indication for genetic testing in this setting.

Hypercalcemia occurring in family members (who are also found to have the germline mutations) can lead to the finding, at parathyroid surgery, of premalignant parathyroid tumors.

Overall, it seems there are multiple factors in parathyroid cancer, in addition to the *HRPT2* and *Rb* gene, although the *HRPT2* gene mutation is the most invariant abnormality. *RET* encodes a tyrosine kinase type receptor; specific inherited germline mutations lead to a constitutive activation of the receptor, thereby explaining the autosomal dominant mode of transmission and the relatively early onset of neoplasia. In the MEN 2 syndrome, the *RET* protooncogene may be responsible for the earliest disorder detected, the polyclonal disorder (C-cell hyperplasia, which then is transformed into a clonal outgrowth—a medullary carcinoma with the participation of other, still uncharacterized genetic defects).

In some parathyroid adenomas, activation of a protooncogene has been identified (Fig. 410-3B). A reciprocal translocation involving chromosome 11 has been identified that juxtaposes the *PTH* gene promoter upstream of *CCND1*, encoding a cyclin D protein that plays a key role in normal cell division. This translocation plus other mechanisms that cause an equivalent overexpression of cyclin D1 are found in 20–40% of parathyroid adenomas.

Mouse models have confirmed the role of several of the major identified genetic defects in parathyroid disease and the MEN syndromes. Loss of the *MEN1* gene locus and overexpression of the *CCND1* protooncogene or the mutated *RET* protooncogene have been analyzed by genetic manipulation in mice, with the expected onset of parathyroid tumors or medullary carcinoma, respectively.

**Pathology** Adenomas are most often located in the inferior parathyroid glands, but in 6–10% of patients, parathyroid adenomas may be located in the thymus, the thyroid, the pericardium, or behind the esophagus. Adenomas are usually 0.5–5 g in size but may be as large as 10–20 g (normal glands weigh 25 mg on average). Chief cells are predominant in both hyperplasia and adenoma. With chief cell hyperplasia, the enlargement may be so asymmetric that some involved glands appear grossly normal. If generalized hyperplasia is present, however, histologic examination reveals a uniform pattern of chief cells and disappearance of fat even in the absence of an increase in gland weight.

Thus, microscopic examination of biopsy specimens of several glands is essential to interpret findings at surgery.

Parathyroid carcinoma is often not aggressive. Long-term survival without recurrence is common if at initial surgery the entire gland is removed without rupture of the capsule. Recurrent parathyroid carcinoma is usually slow growing with local spread in the neck, and surgical correction of recurrent disease may be feasible. Occasionally, however, parathyroid carcinoma is more aggressive, with distant metastases (lung, liver, and bone) found at the time of initial operation. It may be difficult to appreciate initially that a primary tumor is carcinoma; increased numbers of mitotic figures and increased fibrosis of the gland stroma may precede invasion. The diagnosis of carcinoma is often made in retrospect. Hyperparathyroidism from a parathyroid carcinoma may be indistinguishable from other forms of primary hyperparathyroidism but is usually more severe clinically. A potential clue to the diagnosis is offered by the degree of calcium elevation. Calcium values of 3.5–3.7 mmol/L (14–15 mg/dL) are frequent with carcinoma and may alert the surgeon to remove the abnormal gland with care to avoid capsular rupture. Recent findings concerning the genetic basis of some patients with parathyroid carcinoma (distinct from that of benign adenomas) indicate the need, in these kindreds, for family screening (see below).

**Signs and Symptoms** Many patients with hyperparathyroidism are asymptomatic. Manifestations of hyperparathyroidism involve primarily the kidneys and the skeletal system. Kidney involvement, due either to deposition of calcium in the renal parenchyma or to recurrent nephrolithiasis, was present in 60–70% of patients prior to 1970. With earlier detection, renal complications occur in <20% of patients in many large series. Renal stones are usually composed of either calcium oxalate or calcium phosphate. In occasional patients, repeated episodes of nephrolithiasis or the formation of large calculi may lead to urinary tract obstruction, infection, and loss of renal function. Nephrocalcinosis may also cause decreased renal function and phosphate retention.

The distinctive bone manifestation of hyperparathyroidism is *osteitis fibrosa cystica*, which occurred in 10–25% of patients in series reported 50 years ago. Histologically, the pathognomonic features are an increase in the giant multinucleated osteoclasts in scalloped areas on the surface of the bone (Howship's lacunae) and a replacement of the normal cellular and marrow elements by fibrous tissue. X-ray changes include resorption of the phalangeal tufts and replacement of the usually sharp cortical outline of the bone in the digits by an irregular outline (subperiosteal resorption). In recent years, *osteitis fibrosa cystica* is very rare in primary hyperparathyroidism, probably due to the earlier detection of the disease.

Dual-energy x-ray absorptiometry of the spine provides reproducible quantitative estimates (within a few percent) of spinal bone density. Similarly, bone density in the extremities can be quantified by densitometry of the hip or of the distal radius at a site chosen to be primarily cortical. CT is a very sensitive technique for estimating spinal bone density, but reproducibility of standard CT is no better than 5%. Newer CT techniques (spiral, "extreme" CT) are more reproducible but are currently available in a limited number of medical centers. Cortical bone density is reduced, while cancellous bone density, especially in the spine, is relatively preserved.

In symptomatic patients, dysfunctions of the central nervous system (CNS), peripheral nerve and muscle, gastrointestinal tract, and joints also occur. It has been reported that severe neuropsychiatric manifestations may be reversed by parathyroidectomy. When present in symptomatic patients, neuromuscular manifestations may include proximal muscle weakness, easy fatigability, and atrophy of muscles and may be so striking as to suggest a primary neuromuscular disorder. The distinguishing feature is the complete regression of neuromuscular disease after surgical correction of the hyperparathyroidism.

Gastrointestinal manifestations are sometimes subtle and include vague abdominal complaints and disorders of the stomach and pancreas. Again, cause and effect are unclear. In MEN 1 patients with hyperparathyroidism, duodenal ulcer may be the result of associated pancreatic tumors that secrete excessive quantities of gastrin

(Zollinger-Ellison syndrome). Pancreatitis has been reported in association with hyperparathyroidism, but the incidence and the mechanism are not established.

Much attention has been paid in recent years to the manifestations of and optimum management strategies for asymptomatic hyperparathyroidism. This is now the most prevalent form of the disease. *Asymptomatic primary hyperparathyroidism* is defined as biochemically confirmed hyperparathyroidism (elevated or inappropriately normal PTH levels despite hypercalcemia) with the absence of signs and symptoms typically associated with more severe hyperparathyroidism such as features of renal or bone disease.

Four conferences on the topic have been held in the United States over the past two decades, with the most recent in 2014. The published proceedings include discussion of more subtle manifestations of disease, its natural history (without parathyroidectomy), and guidelines both for indications for surgery and medical monitoring in nonoperated patients.

Issues of concern include the potential for cardiovascular deterioration, the presence of subtle neuropsychiatric symptoms, and the longer-term status of skeletal integrity in patients not treated surgically. The current consensus is that medical monitoring rather than surgical correction of hyperparathyroidism may be justified in certain patients. The current recommendation is that patients who show mild disease, as defined by the meeting guidelines (Table 410-2), can be safely followed under management guidelines (Table 410-3). There is, however, growing uncertainty about subtle disease manifestations and whether surgery is therefore indicated in most patients. Among the issues is the evidence of eventual (>8 years) deterioration in bone mineral density after a decade of relative stability. There is concern that this late-onset deterioration in bone density in nonoperated patients could contribute significantly to the well-known age-dependent fracture risk (osteoporosis). Significant and sustained improvements in bone mineral density are seen after successful parathyroidectomy, and there is some evidence for reduction in fractures.

Cardiovascular disease including left ventricular hypertrophy, cardiac functional defects, and endothelial dysfunction have been reported as reversible in European patients with more severe symptomatic disease after surgery, leading to numerous studies of these cardiovascular features in those with milder disease. There are reports of endothelial dysfunction in patients with mild asymptomatic hyperparathyroidism, but the expert panels concluded that more observation is needed, especially regarding whether there is reversibility with surgery.

A topic of considerable interest and some debate is assessment of neuropsychiatric status and health-related quality of life status in hyperparathyroid patients both before surgery and in response to parathyroidectomy. Several observational studies have suggested improvements in symptom score after surgery. Randomized studies of surgery versus observation, however, have yielded inconclusive

**TABLE 410-3 Guidelines for Monitoring in Asymptomatic Primary Hyperparathyroidism**

PARAMETER	GUIDELINE
Serum calcium	Annually
Renal	eGFR, annually; serum creatinine, annually. If renal stones suspected, 24-h biochemical stone profile, renal imaging by x-ray, ultrasound, or CT
Serum creatinine	Annually
Skeletal	Every 1–2 years (3 sites), x-ray or VFA of spine if clinically indicated (e.g., height loss, back pain)

Abbreviations: CT, computed tomography; eGFR, estimated glomerular filtration rate; VFA, vertebral fracture assessment.

Source: Data from JP Bilezikian et al: Guidelines for the management of asymptomatic primary hyperparathyroidism: Summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab* 99:3561, 2014.

results, especially regarding benefits of surgery. Many studies report that hyperparathyroidism is associated with increased neuropsychiatric symptoms, but it is not possible at present to determine which patients might improve after surgery.

**Diagnosis** The diagnosis is typically made by detecting an elevated immunoreactive PTH level in a patient with asymptomatic hypercalcemia (Fig. 410-4) (see “Differential Diagnosis: Special Tests,” below). Serum phosphate is usually low but may be normal, especially if renal failure has developed.

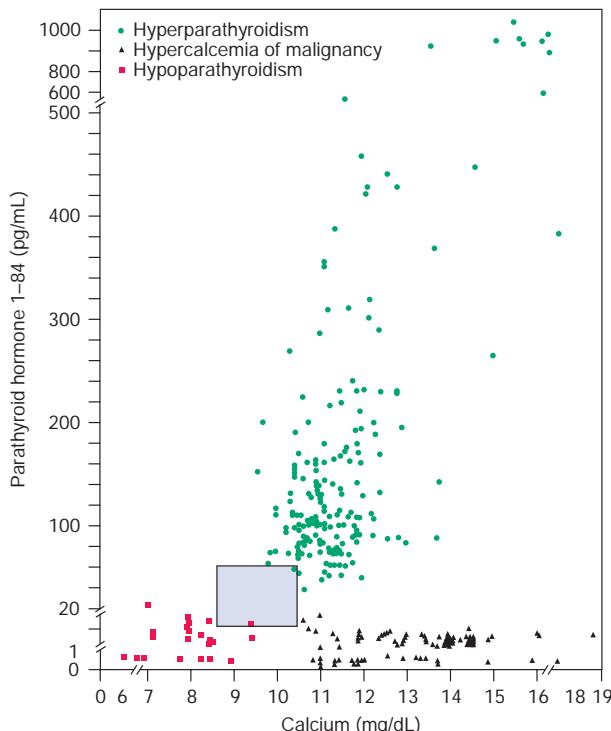
Several modifications in PTH assays have been introduced in efforts to improve their utility in light of information about metabolism of PTH (as discussed above). First-generation assays were based on displacement of radiolabeled PTH from antibodies that reacted with PTH (often also PTH fragments). Double-antibody or immunoassays (one antibody that is usually directed against the carboxyl-terminal portion of intact PTH to capture the hormone and

**TABLE 410-2 Guidelines for Surgery in Asymptomatic Primary Hyperparathyroidism<sup>a</sup>**

PARAMETER	GUIDELINE
Serum calcium (above normal)	>1 mg/dL
Renal	Creatinine clearance <60 mL/min 24-h urine for calcium >400 mg/d and increased stone risk by biochemical stone risk analysis Presence of nephrolithiasis or nephrocalcinosis by x-ray, ultrasound, or CT
Skeletal	BMD by DXA: T score <-2.5 at lumbar spine, total hip, femoral neck, or distal one-third radius Vertebral fracture by x-ray, CT, MRI, or VFA
Age	<50

Abbreviations: BMD, bone mineral density; CT, computed tomography; DXA, dual-energy x-ray absorptiometry; MRI, magnetic resonance imaging; VFA, vertebral fracture assessment.

<sup>a</sup>Data from JP Bilezikian et al: Guidelines for the management of asymptomatic primary hyperparathyroidism: Summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab* 99:3561, 2014.



**FIGURE 410-4** Levels of immunoreactive parathyroid hormone (PTH) detected in patients with primary hyperparathyroidism, hypercalcemia of malignancy, and hypoparathyroidism. Boxed area represents the upper and normal limits of blood calcium and/or immunoreactive PTH. (Reproduced with permission from SR Nussbaum et al (eds): *Endocrinology*, 4th ed. Philadelphia, PA: Saunders; 2001.)

a second radio- or enzyme-labeled antibody that is usually directed against the amino-terminal portion of intact PTH) greatly improved the diagnostic discrimination of the tests by eliminating interference from circulating biologically inactive fragments, detected by the original first-generation assays. Double-antibody assays are now referred to as second-generation. Such PTH assays have, in some centers and testing laboratories, been replaced by third-generation assays after it was discovered that large PTH fragments, devoid of only the extreme amino-terminal portion of the PTH molecule, are also present in blood and are detected, incorrectly as full-length PTH. These amino-terminally truncated PTH fragments were prevented from registering in the newer third-generation assays by use of a detection antibody directed against the extreme amino-terminal epitope. These assays may be useful for clinical research studies as in management of chronic renal disease, but the consensus is that either second- or third-generation assays are useful in the diagnosis of primary hyperparathyroidism and for the diagnosis of high-turnover bone disease in CKD.

## TREATMENT

### Hyperparathyroidism

Surgical excision of the abnormal parathyroid tissue is the definitive therapy for this disease. As noted above, medical surveillance without operation for patients with mild, asymptomatic disease is, however, still preferred by some physicians and patients, particularly when the patients are more elderly. Evidence favoring surgery, if medically feasible, is growing because of concerns about skeletal, cardiovascular, and neuropsychiatric disease, even in mild hyperparathyroidism.

Two surgical approaches are generally practiced. The conventional parathyroidectomy procedure was neck exploration with general anesthesia; this procedure is being replaced in many centers, whenever feasible, by an outpatient procedure with local anesthesia, termed *minimally invasive parathyroidectomy*.

Parathyroid exploration is challenging and should be undertaken by an experienced surgeon. Certain features help in predicting the pathology (e.g., multiple abnormal glands in familial cases). However, some critical decisions regarding management can be made only during the operation.

With conventional surgery, one approach is still based on the view that typically only one gland (the adenoma) is abnormal. If an enlarged gland is found, a normal gland should be sought. In this view, if a biopsy of a normal-sized second gland confirms its histologic (and presumed functional) normality, no further exploration, biopsy, or excision is needed. At the other extreme is the minority viewpoint that all four glands be sought and that most of the total parathyroid tissue mass be removed. The concern with the former approach is that the recurrence rate of hyperparathyroidism may be high if a second abnormal gland is missed; the latter approach could involve unnecessary surgery and an unacceptable rate of hypoparathyroidism. When normal glands are found in association with one enlarged gland, excision of the single adenoma usually leads to cure or at least years free of symptoms. Long-term follow-up studies to establish true rates of recurrence are limited.

Recently, there has been growing experience with new surgical strategies that feature a minimally invasive approach guided by improved preoperative localization and intraoperative monitoring by PTH assays. Preoperative  $^{99m}$ Tc sestamibi scans with single-photon emission CT (SPECT) are used to predict the location of an abnormal gland and intraoperative sampling of PTH before and at 5-min intervals after removal of a suspected adenoma to confirm a rapid fall (>50%) to normal levels of PTH. In several centers, a combination of preoperative four-dimensional (4D) CT and ultrasound examination and, less frequently, sestamibi imaging, cervical block anesthesia, minimal surgical incision, and intraoperative PTH measurements has allowed successful outpatient surgical management with a clear-cut cost benefit compared to general anesthesia and more extensive neck surgery. The use of these minimally invasive

approaches requires clinical judgment to select patients unlikely to have multiple-gland disease (e.g., MEN or secondary hyperparathyroidism). The growing acceptance of the technique and its relative ease for the patient has lowered the threshold for surgery.

Severe hypercalcemia may provide a preoperative clue to the presence of parathyroid carcinoma. In such cases, when neck exploration is undertaken, the tissue should be widely excised; care is taken to avoid rupture of the capsule to prevent local seeding of tumor cells.

Multiple-gland hyperplasia, as predicted in familial cases, poses more difficult questions of surgical management. Once a diagnosis of hyperplasia is established, all the glands must be identified. Two schemes have been proposed for surgical management. One is to totally remove three glands with partial excision of the fourth gland; care is taken to leave a good blood supply for the remaining gland. Other surgeons advocate total parathyroidectomy with immediate transplantation of a portion of a removed, minced parathyroid gland into the muscles of the forearm, with the view that surgical excision is easier from the ectopic site in the arm if there is recurrent hyperfunction.

In a minority of cases, if no abnormal parathyroid glands are found in the neck, the issue of further exploration must be decided. There are documented cases of five or six parathyroid glands and of unusual locations for adenomas such as in the mediastinum.

When a second parathyroid exploration is indicated, the minimally invasive techniques for preoperative localization such as ultrasound, CT scan, and isotope scanning are combined with venous sampling and/or selective digital arteriography in one of the centers specializing in these procedures. Intraoperative monitoring of PTH levels by rapid PTH immunoassays may be useful in guiding the surgery. At one center, long-term cures have been achieved with selective embolization or injection of large amounts of contrast material into the end-arterial circulation feeding the parathyroid tumor.

A decline in serum calcium occurs within 24 h after successful surgery; usually, blood calcium falls to low-normal values for 3–5 days until the remaining parathyroid tissue resumes full hormone secretion. Acute postoperative hypocalcemia is likely only if severe bone mineral deficits are present or if injury to all the normal parathyroid glands occurs during surgery. In general, there are few problems encountered in patients with uncomplicated disease such as a single adenoma (the clear majority), who do not have symptomatic bone disease or a large deficit in bone mineral, who are vitamin D and magnesium sufficient, and who have good renal and gastrointestinal function. The extent of postoperative hypocalcemia varies with the surgical approach. If all glands are biopsied, hypocalcemia may be transiently symptomatic and more prolonged. Hypocalcemia is more likely to be symptomatic after second parathyroid explorations, particularly when normal parathyroid tissue was removed at the initial operation and when the manipulation and/or biopsy of the remaining normal glands are more extensive in the search for the missing adenoma.

Patients with hyperparathyroidism have efficient intestinal calcium absorption due to the increased levels of  $1,25(\text{OH})_2\text{D}$  stimulated by PTH excess. Once hypocalcemia signifies successful surgery, patients can be put on a high-calcium intake or be given oral calcium supplements. Despite mild hypocalcemia, most patients do not require parenteral therapy. If the serum calcium falls to <2 mmol/L (8 mg/dL), and if the phosphate level rises simultaneously, the possibility that surgery has caused hypoparathyroidism must be considered. With unexpected hypocalcemia, coexistent hypomagnesemia should be considered, as it interferes with PTH secretion and causes functional hypoparathyroidism (Chap. 409).

Signs of hypocalcemia include symptoms such as muscle twitching, a general sense of anxiety, and positive Chvostek's and Troussseau's signs coupled with serum calcium consistently <2 mmol/L (8 mg/dL). Parenteral calcium replacement at a low level should be instituted when hypocalcemia is symptomatic. The rate and duration of IV therapy are determined by the severity of the symptoms

and the response of the serum calcium to treatment. An infusion of 0.5–2 mg/kg per hour or 30–100 mL/h of a 1-mg/mL solution usually suffices to relieve symptoms. Usually, parenteral therapy is required for only a few days. If symptoms worsen or if parenteral calcium is needed for >2–3 days, therapy with a vitamin D analogue and/or oral calcium (2–4 g/d) should be started (see below). It is cost-effective to use calcitriol (doses of 0.5–1 µg/d) because of the rapidity of onset of effect and prompt cessation of action when stopped, in comparison to other forms of vitamin D. A rise in blood calcium after several months of vitamin D replacement may indicate restoration of parathyroid function to normal. It is also appropriate to monitor serum PTH serially to estimate gland function in such patients.

If magnesium deficiency is present, it can complicate the postoperative course since magnesium deficiency impairs the secretion of PTH. Hypomagnesemia should be corrected whenever detected. Magnesium replacement can be effective orally (e.g., MgCl<sub>2</sub>, MgOH<sub>2</sub>), but parenteral repletion is usual to ensure postoperative recovery, if magnesium deficiency is suspected due to low blood magnesium levels. Because the depressant effect of magnesium on central and peripheral nerve functions does not occur at levels <2 mmol/L (normal range 0.8–1.2 mmol/L), parenteral replacement can be given rapidly. A cumulative dose as great as 0.5–1 mmol/kg of body weight can be administered if severe hypomagnesemia is present; often, however, total doses of 20–40 mmol are sufficient.

#### MEDICAL MANAGEMENT

The guidelines for recommending surgical intervention, if feasible (Table 410-2), as well as for monitoring patients with asymptomatic hyperparathyroidism who elect not to undergo parathyroidectomy (Table 410-3), reflect the changes over time since the first conference on the topic in 1990. Medical monitoring rather than corrective surgery is still acceptable, but it is clear that surgical intervention is the more frequently recommended option for the reasons noted above. Tightened guidelines favoring surgery include lowering the recommended level of serum calcium elevation, more careful attention to skeletal integrity through reference to peak skeletal mass at baseline (T scores) rather than age-adjusted bone density (Z scores), as well as the presence of any fragility fracture. The other changes noted in the two guidelines (Tables 410-2 and 410-3) reflect accumulated experience and practical consideration, such as a difficulty in quantity of urine collections. Despite the usefulness of the guidelines, the importance of individual patient and physician judgment and preference is clear in all recommendations.

When surgery is not selected or not medically feasible, there is interest in the potential value of specific medical therapies. There is no long-term experience regarding specific clinical outcomes such as fracture prevention, but it has been established that bisphosphonates increase bone mineral density significantly without changing serum calcium (as does estrogen, but the latter is not favored because of reported adverse effects in other organ systems). Calcimimetics that lower PTH secretion lower calcium but do not affect bone mineral density (BMD).

#### OTHER PARATHYROID CAUSES OF HYPERCALCEMIA

**Lithium Therapy** Lithium, used in the management of bipolar depression and other psychiatric disorders, causes hypercalcemia in ~10% of treated patients. The hypercalcemia is dependent on continued lithium treatment, remitting and recurring when lithium is stopped and restarted. The parathyroid adenomas reported in some hypercalcemic patients with lithium therapy may reflect the presence of an independently occurring parathyroid tumor; a permanent effect of lithium on parathyroid gland growth need not be implicated as most patients have complete reversal of hypercalcemia when lithium is stopped. However, long-standing stimulation of parathyroid cell replication by lithium may predispose to development of adenomas (as is documented in secondary hyperparathyroidism and renal failure).

At the levels achieved in blood in treated patients, lithium can be shown in vitro to shift the PTH secretion curve to the right in response to calcium; i.e., higher calcium levels are required to lower PTH secretion, probably acting at the calcium sensor (see below). This effect can cause elevated PTH levels and consequent hypercalcemia in otherwise normal individuals. Fortunately, there are usually alternative medications for the underlying psychiatric illness. Parathyroid surgery should not be recommended unless hypercalcemia and elevated PTH levels persist after lithium is discontinued.

#### GENETIC DISORDERS CAUSING HYPERPARATHYROIDISM-LIKE SYNDROMES

**Familial Hypocalciuric Hypercalcemia** FHH (also called *familial benign hypercalcemia*) is inherited as an autosomal dominant trait. Affected individuals are discovered because of asymptomatic hypercalcemia. Most cases of FHH (FHH1) are caused by an inactivating mutation in a single allele of the CaSR (see below), leading to inappropriately normal or even increased secretion of PTH, whereas another hypercalcemic disorder, namely the exceedingly rare Jansen's disease, is caused by a constitutively active PTH/PTHrP receptor in target tissues. Neither FHH1 nor Jansen's disease, however, is a growth disorder of the parathyroids. Other forms of FHH are caused either by heterozygous mutations in GNA11 (encoding G 11), one of the signaling proteins at the CaSR (FHH2), or by heterozygous mutations in APIA1 (FHH3).

The pathophysiology of FHH1 is now understood. The primary defect is abnormal sensing of the blood calcium by the parathyroid gland and renal tubule, causing inappropriate secretion of PTH and excessive reabsorption of calcium in the distal renal tubules. The CaSR is a member of the third family of GPCRs (type C or type III). The receptor responds to increased ECF calcium concentration by suppressing PTH secretion through second-messenger signaling at the CaSR involving the G proteins, G 11 and G q, thereby providing negative-feedback regulation of PTH secretion. Many different inactivating CaSR mutations have been identified in patients with FHH1. These mutations lower the capacity of the sensor to bind calcium, and the mutant receptors function as though blood calcium levels were low; excessive secretion of PTH occurs from an otherwise normal gland. Approximately two-thirds of patients with FHH have mutations within the protein-coding region of the CaSR gene. The remaining one-third of kindreds may have mutations in the promoter/introns of the CaSR gene or are caused by mutations in other genes.

Even before elucidation of the pathophysiology of FHH, abundant clinical evidence served to separate the disorder from primary hyperparathyroidism; these clinical features are still useful in differential diagnosis. Patients with primary hyperparathyroidism have <99% renal calcium reabsorption, whereas most patients with FHH have >99% reabsorption. The hypercalcemia in FHH is often detectable in affected members of the kindreds in the first decade of life, whereas hypercalcemia rarely occurs in patients with primary hyperparathyroidism or the MEN syndromes who are aged <10 years. PTH may be elevated in the different forms of FHH, but the values are usually normal or lower for the same degree of calcium elevation than is observed in patients with primary hyperparathyroidism. Parathyroid surgery performed in a few patients with FHH before the nature of the syndrome was understood led to permanent hypoparathyroidism; nevertheless, hypocalciuria persisted, establishing that hypocalciuria is not PTH-dependent (now known to be due to the abnormal CaSR in the kidney).

Few clinical signs or symptoms are present in patients with FHH, while other endocrine abnormalities are not. Most patients are detected as a result of family screening after hypercalcemia is detected in a proband. In those patients inadvertently operated upon for primary hyperparathyroidism, the parathyroids appeared normal or moderately hyperplastic. Parathyroid surgery is not appropriate, nor, in view of the lack of symptoms, does medical treatment seem needed to lower the calcium. One striking exception to the rule against parathyroid surgery in this syndrome is the occurrence, usually in consanguineous marriages (due to the rarity of the gene mutation), of a homozygous or compound heterozygote state, resulting in severe impairment of

CaSR function. In this condition, neonatal severe hypercalcemia, total parathyroidectomy is mandatory, but calcimimetics have been used as a temporary measure. Rare but well-documented cases of acquired hypocalciuric hypercalcemia are reported due to antibodies against the CaSR. They appear to be a complication of an underlying autoimmune disorder and respond to therapies directed against the underlying disorder.

**Jansen's Disease** Activating mutations in the PTH/PTHrP receptor (PTHR1) have been identified as the cause of this rare autosomal dominant syndrome. Because the mutations lead to constitutive activation of receptor function, one abnormal copy of the mutant receptor is sufficient to cause the disease, thereby accounting for its dominant mode of transmission. Besides often severe hypercalcemia, patients affected by Jansen's disease have short-limbed dwarfism due to abnormal regulation of chondrocyte maturation in the growth plates of the bone that are formed through the endochondral process. In adult life, there are numerous abnormalities in bone, including multiple cystic resorptive areas resembling those seen in severe hyperparathyroidism. Hypercalcemia and hypophosphatemia with undetectable or low PTH levels are typically observed. The pathogenesis of the growth plate abnormalities in Jansen's disease has been confirmed by transgenic experiments in which targeted expression of the mutant PTH/PTHrP receptor to the proliferating chondrocyte layer of growth plate emulated several features of the human disorder. Other genetic mutations in the parathyroid gland or PTH target cells that affect  $\text{Ca}^{2+}$  metabolism are illustrated in Figure 410-5.

## MALIGNANCY-RELATED HYPERCALCEMIA

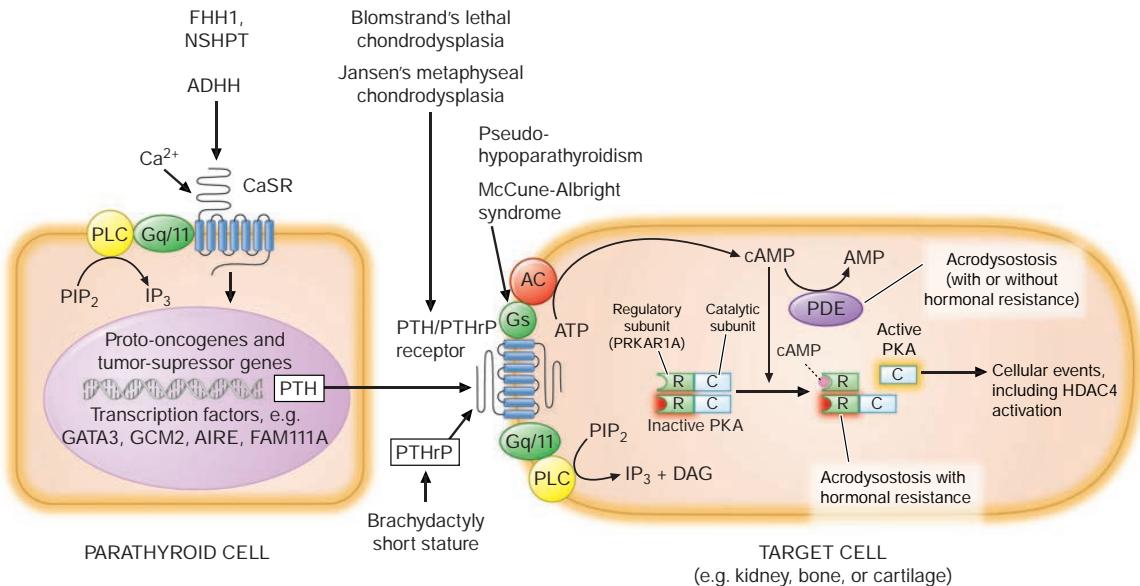
### Clinical Syndromes and Mechanisms of Hypercalcemia

Hypercalcemia due to malignancy is common (occurring in as many as 20% of cancer patients, especially with certain types of tumors such

as lung carcinoma), often severe and difficult to manage, and, on rare occasions, difficult to distinguish from primary hyperparathyroidism. Although malignancy is usually clinically obvious or readily detectable by medical history, hypercalcemia can occasionally be due to an occult tumor. Previously, hypercalcemia associated with malignancy was thought to be due to local invasion and destruction of bone by tumor cells; many cases are now known to result from the elaboration by the malignant cells of humoral mediators of hypercalcemia. PTHrP is the responsible humoral agent in most solid tumors that cause hypercalcemia.

The histologic character of the tumor is more important than the extent of skeletal metastases in predicting hypercalcemia. Small-cell carcinoma (oat cell) and adenocarcinoma of the lung, although the most common lung tumors associated with skeletal metastases, rarely cause hypercalcemia. By contrast, many patients with squamous cell carcinoma of the lung develop hypercalcemia. Histologic studies of bone in patients with squamous cell or epidermoid carcinoma of the lung, in sites invaded by tumor as well as areas remote from tumor invasion, reveal increased bone resorption.

Two main mechanisms of hypercalcemia are operative in cancer hypercalcemia. Many solid tumors associated with hypercalcemia, particularly squamous cell and renal tumors, produce and secrete PTHrP that causes increased bone resorption and mediates the hypercalcemia through systemic actions on the skeleton. Alternatively, direct bone marrow invasion occurs with hematologic malignancies such as leukemia, lymphoma, and multiple myeloma. Lymphokines and cytokines (including PTHrP) produced by cells involved in the marrow response to the tumors promote resorption of bone through local destruction. Several hormones, hormone analogues, cytokines, and growth factors have been implicated as the result of clinical assays, in vitro tests, or chemical isolation. The etiologic factor produced by activated normal lymphocytes and by myeloma and lymphoma cells, originally termed *osteoclast activation factor*, now appears to represent the biologic action



**FIGURE 410-5** Illustration of some genetic mutations that alter calcium metabolism by effects on the parathyroid cell or target cells of parathyroid hormone (PTH) action. Alterations in PTH production by the parathyroid cell can be caused by changes in the response to extracellular fluid calcium ( $\text{Ca}^{2+}$ ) that are detected by the calcium-sensing receptor (CaSR). Furthermore, PTH (or PTH-related peptide [PTHrP]) can show altered efficacy in target cells such as in proximal tubular cells, by altered function of its receptor (PTH/PTHrP receptor) or the signal transduction proteins, G proteins such as  $G_{\alpha q}$  that is linked to adenylate cyclase (AC), the enzyme responsible for producing cyclic AMP (cAMP) (also illustrated are  $G_{\alpha q}/G_{\alpha 11}$ , which activate an alternate pathway of receptor signal transmission involving the generation of inositol triphosphate [ $\text{IP}_3$ ] or diacylglycerol [DAG]). Heterozygous loss-of-function mutations in the CaSR cause familial benign hypocalciuric hypercalcemia (FHH) and homozygous mutations (both alleles mutated) and neonatal severe hyperparathyroidism (NSHPT); heterozygous gain-of-function causes autosomal dominant hypocalciuric hypocalcemia (ADHH). Other defects in parathyroid cell function that occur at the level of gene regulation (oncogenes or tumor-suppressor genes) or transcription factors are discussed in the text. Blomstrand's lethal chondrodysplasia is due to homozygous or compound heterozygous loss-of-function mutations in the PTH/PTHrP receptor, a neonatally lethal disorder, while pseudohypoparathyroidism involves inactivation at the level of the G proteins, specifically mutations that eliminate or reduce  $G_{\alpha q}$  activity in the kidney (see text for details). Acrodysostosis can occur with (mutant regulatory subunit of PKA) or without hormonal resistance (mutant  $PDE4D$  or  $PDE3A$ ). Jansen's metaphyseal chondrodysplasia and McCune-Albright syndrome represent gain-of-function mutations in the PTH/PTHrP receptor and  $G_{\alpha q}$  protein, respectively.

**3180** of several different cytokines, probably interleukin 1 and lymphotoxin or tumor necrosis factor (TNF). In some lymphomas, there is a third mechanism, caused by an increased blood level of  $1,25(\text{OH})_2\text{D}$ , produced by the abnormal lymphocytes or adjacent macrophages.

In the more common mechanism, usually termed *humoral hypercalcemia of malignancy* (HHM), solid tumors (cancers of the lung and kidney, in particular), in which bone metastases are absent, minimal, or not detectable clinically, secrete PTHrP measurable by immunoassay. Secretion by the tumors of the PTH-like factor, PTHrP, activates the PTHR1, resulting in a pathophysiology closely resembling hyperparathyroidism, but with normal or suppressed PTH levels. The clinical picture resembles primary hyperparathyroidism (hypophosphatemia accompanies hypercalcemia), and elimination or regression of the primary tumor leads to disappearance of the hypercalcemia.

As in hyperparathyroidism, patients with the HHM have elevated urinary nephrogenous cyclic AMP excretion, hypophosphatemia, and increased urinary phosphate clearance. However, in HHM, immunoreactive PTH is undetectable or suppressed, making the differential diagnosis easier. Other features of the disorder differ from those of true hyperparathyroidism. Although the biologic actions of PTH and PTHrP are exerted through the same receptor, subtle differences in receptor activation by the two ligands must account for some of the discordance in pathophysiology, when an excess of one or the other peptide occurs. Other cytokines elaborated by the malignancy may contribute to the variations from hyperparathyroidism in these patients as well. Patients with HHM may have low to normal levels of  $1,25(\text{OH})_2\text{D}$ , instead of elevated levels as in true hyperparathyroidism. In some patients with the HHM, osteoclastic resorption is unaccompanied by an osteoblastic or bone-forming response, implying inhibition of the normal coupling of bone formation and resorption.

Several different assays (single- or double-antibody, different epitopes) have been developed to detect PTHrP. Most data indicate that circulating PTHrP levels are undetectable or low in normal individuals except perhaps in pregnancy (high in human milk) and elevated in most cancer patients with the humoral syndrome. The etiologic mechanisms in cancer hypercalcemia may be multiple in the same patient. For example, in breast carcinoma (metastatic to bone) and in a distinctive type of T-cell lymphoma/leukemia initiated by human T-cell lymphotropic virus I, hypercalcemia is caused by direct local lysis of bone as well as by a humoral mechanism involving excess production of PTHrP. Hyperparathyroidism has been reported to coexist with the humoral cancer syndrome, and rarely, ectopic hyperparathyroidism due to tumor elaboration of true PTH is reported.

**Diagnostic Issues** Levels of PTH measured by the double-antibody technique are undetectable or extremely low in tumor hypercalcemia, as would be expected with the mediation of the hypercalcemia by a factor other than PTH (the hypercalcemia suppresses the normal parathyroid glands). In a patient with minimal symptoms referred for hypercalcemia, low or undetectable PTH levels would focus attention on a possible occult malignancy (except for very rare cases of ectopic hyperparathyroidism).

Ordinarily, the diagnosis of cancer hypercalcemia is not difficult because tumor symptoms are prominent when hypercalcemia is detected. Indeed, hypercalcemia may be noted incidentally during the workup of a patient with known or suspected malignancy. Clinical suspicion that malignancy is the cause of the hypercalcemia is heightened when there are other signs or symptoms of a paraneoplastic process such as weight loss, fatigue, muscle weakness, or unexplained skin rash, or when symptoms specific for a particular tumor are present. Squamous cell tumors are most frequently associated with hypercalcemia, particularly tumors of the lung, kidney, head and neck, and urogenital tract. Radiologic examinations can focus on these areas when clinical evidence is unclear. Bone scans with technetium-labeled bisphosphonate are useful for detection of osteolytic metastases; the sensitivity is high, but specificity is low; results must be confirmed by conventional x-rays to be certain that areas of increased uptake are due to osteolytic metastases per se. Bone marrow biopsies are helpful in patients with anemia or abnormal peripheral blood smears.

## TREATMENT

### Malignancy-Related Hypercalcemia

Treatment of the hypercalcemia of malignancy is first directed to control of tumor; reduction of tumor mass usually corrects hypercalcemia. If a patient has severe hypercalcemia yet has a good chance for effective tumor therapy, treatment of the hypercalcemia should be vigorous while awaiting the results of definitive therapy (see "General Approach to Hypercalcemic States" below). If hypercalcemia occurs in the late stages of a tumor that is resistant to antitumor therapy, the treatment of the hypercalcemia should be judicious as high calcium levels can have a mild sedating effect. Standard therapies for hypercalcemia (discussed below) are applicable to patients with malignancy.

### VITAMIN D-RELATED HYPERCALCEMIA

Vitamin D-mediated hypercalcemia can be due to excessive ingestion of vitamin D analogs or abnormal metabolism of the vitamin. Abnormal metabolism of the vitamin is usually acquired in association with a widespread granulomatous disorder. Vitamin D metabolism is carefully regulated, particularly the activity of renal 1'-hydroxylase, the enzyme responsible for the production of  $1,25(\text{OH})_2\text{D}$  (**Chap. 409**). The regulation of 1'-hydroxylase and the normal feedback suppression by  $1,25(\text{OH})_2\text{D}$  seem to work less well in infants than in adults and to operate poorly, if at all, in sites other than the renal tubule; these phenomena may explain the occurrence of hypercalcemia secondary to excessive  $1,25(\text{OH})_2\text{D}$  production in infants with Williams' syndrome (see below) and in adults with sarcoidosis or lymphoma.

**Vitamin D Intoxication** Chronic ingestion of 40–100 times the normal physiologic requirement of vitamin D (amounts >40,000–100,000 U/d) is usually required to produce significant hypercalcemia in otherwise healthy individuals. The stated upper limit of safe dietary intake is 2000 U/d (50 µg/d) in adults because of concerns about potential toxic effects of cumulative supraphysiologic doses. These recommendations are now regarded as too restrictive, since some estimates are that in elderly individuals in northern latitudes, 2000 U/d may be necessary to avoid vitamin D insufficiency.

Hypercalcemia in vitamin D intoxication is due to an excessive biologic action of the vitamin, perhaps the consequence of increased levels of  $25(\text{OH})\text{D}$  rather than merely increased levels of the active metabolite  $1,25(\text{OH})_2\text{D}$  (the latter may not be elevated in vitamin D intoxication). These actions lead to both increased intestinal absorption of calcium and increased release of calcium from bone.  $25(\text{OH})\text{D}$  has definite, if low, biologic activity in the intestine and bone. The production of  $25(\text{OH})\text{D}$  is less tightly regulated than is the production of  $1,25(\text{OH})_2\text{D}$ . Hence concentrations of  $25(\text{OH})\text{D}$  are elevated several-fold in patients with excess vitamin D intake.

The diagnosis is substantiated by documenting elevated levels of  $25(\text{OH})\text{D} > 100 \text{ ng/mL}$ . Hypercalcemia is usually controlled by restriction of dietary calcium intake and appropriate attention to hydration. These measures, plus discontinuation of vitamin D, usually lead to resolution of hypercalcemia. However, because of the increased bone resorption caused by high levels of vitamin D, simple cessation of calcium intake is often insufficient therapy. Further, vitamin D stores in fat may be substantial, and vitamin D intoxication may persist for weeks after vitamin D ingestion is terminated. Such patients are responsive to glucocorticoids, which in doses of 40–100 mg/d of prednisone or its equivalent usually return serum calcium levels to normal over several days; severe intoxication may require intensive therapy.

**Sarcoidosis and Other Granulomatous Diseases** In patients with sarcoidosis and other granulomatous diseases, such as tuberculosis and fungal infections, excess  $1,25(\text{OH})_2\text{D}$  is synthesized in macrophages or other cells in the granulomas. Indeed, increased  $1,25(\text{OH})_2\text{D}$  levels have been reported in anephric patients with sarcoidosis and hypercalcemia. Macrophages obtained from granulomatous tissue convert  $25(\text{OH})\text{D}$  to  $1,25(\text{OH})_2\text{D}$  at an increased rate. There is

a positive correlation in patients with sarcoidosis between 25(OH)D levels (reflecting vitamin D intake) and the circulating concentrations of 1,25(OH)<sub>2</sub>D, whereas normally, there is no increase in 1,25(OH)<sub>2</sub>D with increasing 25(OH)D levels due to multiple feedback controls on renal 1'-hydroxylase (**Chap. 409**). The usual regulation of active metabolite production by calcium and phosphate or by PTH does not operate in these patients. Instead, macrophages increase their production of the vitamin D receptor and of the 1'-hydroxylase in response to tumor necrosis factor and other inflammatory stimuli. Clearance of 1,25(OH)<sub>2</sub>D from blood may be decreased in sarcoidosis as well. PTH levels are usually low and 1,25(OH)<sub>2</sub>D levels are elevated, but primary hyperparathyroidism and sarcoidosis may coexist in some patients.

Management of the hypercalcemia can often be accomplished by avoiding excessive sunlight exposure and limiting vitamin D and calcium intake. Presumably, however, the abnormal sensitivity to vitamin D and abnormal regulation of 1,25(OH)<sub>2</sub>D synthesis will persist as long as the disease is active. Alternatively, glucocorticoids in the equivalent of 100 mg/d of hydrocortisone or equivalent doses of glucocorticoids may help control hypercalcemia. Glucocorticoids appear to act by blocking excessive production of 1,25(OH)<sub>2</sub>D, as well as the response to it in target organs.

**Hypercalcemia of Infancy** Several variants of this rare abnormality of calcium homeostasis are now known. For example, *Williams' syndrome* is an autosomal dominant disorder characterized by multiple congenital development defects, including supravalvular aortic stenosis, intellectual disability, and an elfin facies, in association with hypercalcemia due to abnormal sensitivity to vitamin D. The hypercalcemia associated with the syndrome was first recognized in England, where it was thought, incorrectly, to be caused by the fortification of milk with vitamin D. The cardiac and developmental abnormalities were independently described, but the connection between these defects and hypercalcemia was not described until later. Levels of 1,25(OH)<sub>2</sub>D can be elevated, ranging from 46 to 120 nmol/L (150–500 pg/mL). The mechanism of the abnormal sensitivity to vitamin D and of the increased circulating levels of 1,25(OH)<sub>2</sub>D is still unclear. Studies suggest that genetic mutations involving microdeletions at the elastin locus and perhaps other genes on chromosome 7 may play a role in the pathogenesis. Other more recently defined causes of hypercalcemia in infants and young children can be 24-hydroxylase deficiency that impairs metabolism of 1,25(OH)<sub>2</sub>D or homozygous mutations involving the sodium-dependent phosphate transporters (*NPT2a* mutations lead to more severe hypercalcemia than *NPT2c* mutations).

## HIGH-BONE-TURNOVER STATES

**Hyperthyroidism** As many as 20% of hyperthyroid patients have high-normal or mildly elevated serum calcium concentrations; hypercalciuria is even more common. The hypercalcemia is due to increased bone turnover, with bone resorption exceeding bone formation. Severe calcium elevations are not typical, and the presence of such suggests a concomitant disease such as hyperparathyroidism. Usually, the diagnosis is obvious, but signs of hyperthyroidism may occasionally be occult, particularly in the elderly (**Chap. 384**). Hypercalcemia is managed by treatment of the hyperthyroidism. Reports that thyroid-stimulating hormone (TSH) itself normally has a bone-protective effect suggest that suppressed TSH levels also play a role in hypercalcemia.

**Immobilization** Immobilization is a rare cause of hypercalcemia in adults in the absence of an associated disease but may cause hypercalcemia in children and adolescents, particularly after spinal cord injury and paraplegia or quadriplegia. With resumption of ambulation, the hypercalcemia in children usually returns to normal.

The mechanism appears to involve a disproportion between bone formation and bone resorption; the former decreased and the latter increased. Hypercalciuria and increased mobilization of skeletal calcium can develop in normal volunteers subjected to extensive bed rest, although hypercalcemia is unusual. Immobilization of an adult with a disease associated with high bone turnover, however, such as Paget's disease, may cause hypercalcemia.

**Thiazides** Administration of benzothiadiazines (thiazides) can cause hypercalcemia in patients with high rates of bone turnover. Commonly, thiazides are associated with aggravation of hypercalcemia in primary hyperparathyroidism, but this effect can be seen in other high-bone-turnover states as well. The mechanism of thiazide action is complex. Chronic thiazide administration leads to reduction in urinary calcium; the hypocalciuric effect appears to reflect the enhancement of proximal tubular resorption of sodium and calcium in response to sodium depletion. Some of this renal effect is due to augmentation of PTH action and is more pronounced in individuals with intact PTH secretion. However, thiazides cause hypocalciuria in hypoparathyroid patients on high-dose vitamin D and oral calcium replacement if sodium intake is restricted. This finding is the rationale for the use of thiazides as an adjunct to therapy in hypoparathyroid patients, as discussed below. Thiazide administration to normal individuals causes a transient increase in blood calcium (usually within the high-normal range) that reverts to preexisting levels after a week or more of continued administration. If hormonal function and calcium and bone metabolism are normal, homeostatic controls are reset to counteract the calcium-elevating effect of the thiazides. In the presence of hyperparathyroidism or increased bone turnover from another cause, homeostatic mechanisms are ineffective. The abnormal effects of the thiazide on calcium metabolism disappear within days of cessation of the drug.

**Vitamin A Intoxication** Vitamin A intoxication is a rare cause of hypercalcemia and is most commonly a side effect of dietary faddism (**Chap. 333**). Calcium levels can be elevated into the 3–3.5-mmol/L (12–14 mg/dL) range after the ingestion of 50,000–100,000 units of vitamin A daily (10–20 times the minimum daily requirement). Typical features of severe hypercalcemia include fatigue, anorexia, and, in some, severe muscle and bone pain. Excess vitamin A intake is presumed to increase bone resorption.

The diagnosis can be established by history and by measurement of vitamin A levels in serum. Occasionally, skeletal x-rays reveal periosteal calcifications, particularly in the hands. Withdrawal of the vitamin is usually associated with prompt disappearance of the hypercalcemia and reversal of the skeletal changes. As in vitamin D intoxication, administration of 100 mg/d hydrocortisone or its equivalent leads to a rapid return of the serum calcium to normal.

## RENAL FAILURE-ASSOCIATED HYPERCALCEMIA

**Severe Secondary Hyperparathyroidism** The pathogenesis of secondary hyperparathyroidism in CKD is incompletely understood. Resistance to the normal level of PTH is a major factor contributing to the development of hypocalcemia, which, in turn, is a stimulus to parathyroid gland enlargement. Recent findings have indicated that an increase of FGF23 production by osteocytes (and possibly osteoblasts) in bone occurs well before an elevation in PTH is detected. FGF23 is a potent inhibitor of the renal 1'-hydroxylase, and the FGF23-dependent reduction in 1,25(OH)<sub>2</sub>D thus seems to be an important stimulus for the development of secondary hyperparathyroidism.

Secondary hyperparathyroidism occurs not only in patients with renal failure but also in those with osteomalacia due to multiple causes (**Chap. 409**), including deficiency of vitamin D action and PHP (deficient response to PTH downstream of PTHR1 in the proximal renal tubules). For both disorders, hypocalcemia seems to be the common denominator in initiating the development of secondary hyperparathyroidism. Primary and secondary hyperparathyroidism can be distinguished conceptually by the autonomous growth of the parathyroid glands in primary hyperparathyroidism (presumably irreversible) and the adaptive response of the parathyroids in secondary hyperparathyroidism (typically reversible). In fact, reversal over weeks from an abnormal pattern of secretion, presumably accompanied by involution of parathyroid gland mass to normal, occurs in patients with osteomalacia who have been treated effectively with calcium and vitamin D. However, it is now recognized that a true clonal outgrowth (irreversible) can arise in long-standing, inadequately treated CKD (e.g., tertiary hyperparathyroidism; see below).

Patients with secondary hyperparathyroidism may develop bone pain, ectopic calcification, and pruritus. The bone disease seen in patients with secondary hyperparathyroidism and CKD is termed *renal osteodystrophy* and affects primarily bone turnover. However, osteomalacia is frequently encountered as well and may be related to the circulating levels of FGF23.

Two other skeletal disorders have been frequently associated in the past with CKD patients treated by long-term dialysis, who received aluminum-containing phosphate binders. Aluminum deposition in bone (see below) leads to an osteomalacia-like picture. The other entity is a low-turnover bone disease termed “aplastic” or “adynamic” bone disease; PTH levels are lower than typically observed in CKD patients with secondary hyperparathyroidism. It is believed that the condition is caused, at least in part, by excessive PTH suppression, which may be even greater than previously appreciated in light of evidence that some of the immunoreactive PTH detected by most commercially available PTH assays is not the full-length biologically active molecule (as discussed above) but may consist of amino-terminally truncated fragments that do not activate the PTHR1. The low PTH level is thought to contribute to low bone formation and consequent decreased ability of the skeleton to incorporate circulating calcium into bone matrix.

## TREATMENT

### Hypercalcemia in Secondary Hyperparathyroidism

Medical therapy to reverse secondary hyperparathyroidism in CKD includes reduction of excessive blood phosphate by restriction of dietary phosphate, the use of nonabsorbable phosphate binders, and careful, selective addition of calcitriol (0.25–2 µg/d) or related analogues. Calcium carbonate became preferred over aluminum-containing antacids to prevent aluminum-induced bone disease. However, synthetic gels that also bind phosphate (such as sevelamer; **Chap. 311**) are now widely used, with the advantage of avoiding not only aluminum retention but also excess calcium loading, which may contribute to cardiovascular calcifications. Intravenous calcitriol (or related analogues), administered as several pulses each week, helps control secondary hyperparathyroidism. Aggressive but carefully administered medical therapy can often, but not always, reverse hyperparathyroidism and its symptoms and manifestations.

Occasional patients develop severe manifestations of secondary hyperparathyroidism, including hypercalcemia, pruritus, extraskelatal calcifications, and painful bones, despite aggressive medical efforts to suppress the hyperparathyroidism. PTH hypersecretion no longer responsive to medical therapy, a state of severe hyperparathyroidism in patients with CKD that requires surgery, has been referred to as *tertiary hyperparathyroidism*. Parathyroid surgery is necessary to control this condition. Based on genetic evidence from examination of tumor samples in these patients, the emergence of autonomous parathyroid function is due to a monoclonal outgrowth of one or more previously hyperplastic parathyroid glands. The adaptive response has become an independent contributor to disease; this finding seems to emphasize the importance of optimal medical management to reduce the proliferative response of the parathyroid cells that enables the irreversible genetic change.

## OTHER CAUSES OF HYPERCALCEMIA

**Aluminum Intoxication** Aluminum intoxication (and often hypercalcemia as a complication of medical treatment) in the past occurred in patients on chronic dialysis; manifestations included acute dementia and unresponsive and severe osteomalacia. Bone pain, multiple nonhealing fractures, particularly of the ribs and pelvis, and a proximal myopathy occur. Hypercalcemia develops when these patients are treated with vitamin D or calcitriol because of impaired skeletal responsiveness. Aluminum is present at the site of osteoid mineralization, osteoblastic activity is minimal, and calcium incorporation into the skeleton is impaired. The disorder is now rare because of the

avoidance of aluminum-containing antacids or aluminum excess in the dialysis regimen.

**Milk-Alkali Syndrome** The milk-alkali syndrome is due to excessive ingestion of calcium and absorbable antacids such as milk or calcium carbonate. It is much less frequent since proton pump inhibitors and other treatments became available for peptic ulcer disease. For a time, the increased use of calcium carbonate in the management of secondary hyperparathyroidism led to reappearance of the syndrome. Several clinical presentations—acute, subacute, and chronic—have been described, all of which feature hypercalcemia, alkalosis, and renal failure. The chronic form of the disease, termed *Burnett's syndrome*, is associated with irreversible renal damage. The acute syndromes reverse if the excess calcium and absorbable alkali are stopped.

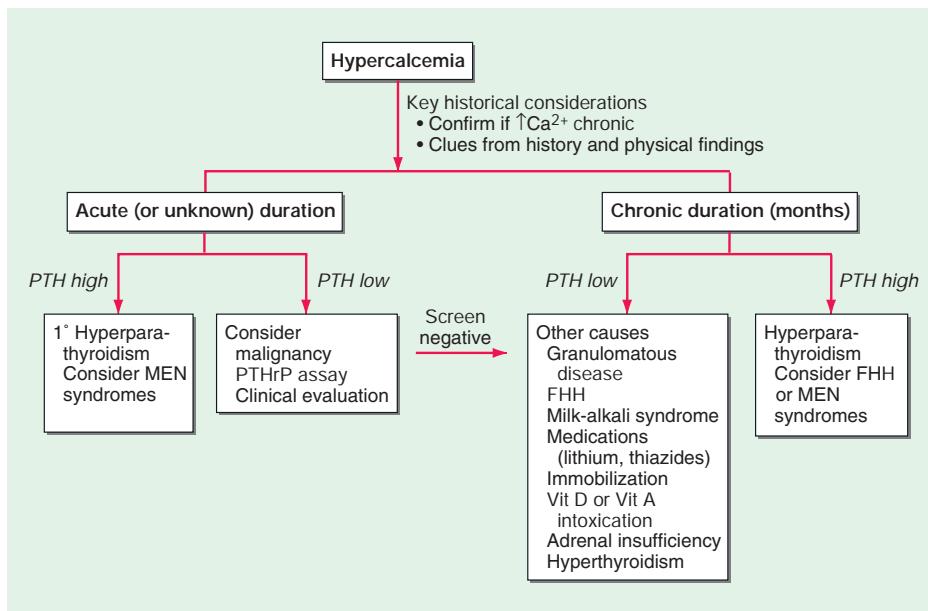
Individual susceptibility is important in the pathogenesis, as some patients are treated with calcium carbonate and alkali regimens without developing the syndrome. One variable is the fractional calcium absorption as a function of calcium intake. Some individuals absorb a high fraction of calcium, even with intakes 2 g of elemental calcium per day, instead of reducing calcium absorption with high intake, as occurs in most normal individuals. Resultant mild hypercalcemia after meals in such patients is postulated to contribute to the generation of alkalosis. Development of hypercalcemia causes increased sodium excretion and some depletion of total-body water. These phenomena and perhaps some suppression of endogenous PTH secretion due to mild hypercalcemia lead to increased bicarbonate resorption and to alkalosis in the face of continued calcium carbonate ingestion. Alkalosis per se selectively enhances calcium resorption in the distal nephron, thus aggravating the hypercalcemia. The cycle of mild hypercalcemia bicarbonate retention alkalosis renal calcium retention severe hypercalcemia perpetuates and aggravates hypercalcemia and alkalosis as long as calcium and absorbable alkali are ingested.

## DIFFERENTIAL DIAGNOSIS OF HYPERCALCEMIA

Differential diagnosis of hypercalcemia is best achieved by using clinical criteria, but immunometric assays to measure PTH are especially useful in distinguishing among major causes (**Fig. 410-6**). The clinical features that deserve emphasis are the presence or absence of symptoms or signs of disease and evidence of chronicity. If one discounts fatigue or depression, >90% of patients with primary hyperparathyroidism have *asymptomatic hypercalcemia*; symptoms of malignancy are usually present in cancer-associated hypercalcemia. Disorders other than hyperparathyroidism and malignancy cause <10% of cases of hypercalcemia, and some of the nonparathyroid causes are associated with clear-cut manifestations such as renal failure.

Hyperparathyroidism is the likely diagnosis in patients with *chronic hypercalcemia*. If hypercalcemia has been manifest for >1 year, malignancy can usually be excluded as the cause. A striking feature of malignancy-associated hypercalcemia is the rapidity of the course, whereby signs and symptoms of the underlying malignancy are evident within months of the detection of hypercalcemia. Although clinical considerations are helpful in arriving at the correct diagnosis of the cause of hypercalcemia, appropriate laboratory testing is essential for definitive diagnosis. The immunoassay for PTH usually separates hyperparathyroidism from all other causes of hypercalcemia (exceptions are very rare reports of ectopic production of excess PTH by non-parathyroid tumors). Patients with hyperparathyroidism have elevated PTH levels despite hypercalcemia, whereas patients with malignancy and the other causes of hypercalcemia (except for disorders mediated by PTH such as lithium-induced hypercalcemia) have levels of hormone below normal or undetectable. Assays based on the double-antibody method for PTH exhibit very high sensitivity (especially if serum calcium is simultaneously evaluated) and specificity for the diagnosis of primary hyperparathyroidism (**Fig. 410-4**).

In summary, PTH values are elevated in >90% of parathyroid-related causes of hypercalcemia, undetectable or low in malignancy-related hypercalcemia, and undetectable or normal in vitamin D-related and high-bone-turnover causes of hypercalcemia. In view of the specificity



**FIGURE 410-6 Algorithm for the evaluation of patients with hypercalcemia.** PTH levels (high or low) should be interpreted in the context of serum calcium levels, as they may be inappropriately high or low for the level of serum calcium. See text for details. FHH, familial hypocalciuric hypercalcemia; MEN, multiple endocrine neoplasia; PTH, parathyroid hormone; PTHrP, parathyroid hormone-related peptide; Vit, vitamin.

of the PTH immunoassay and the high frequency of hyperparathyroidism in hypercalcemic patients, it is cost-effective to measure the PTH level in all hypercalcemic patients unless malignancy or a specific nonparathyroid disease is obvious. False-positive PTH assay results are rare but can be due to heterotopic antibodies. Immunoassays for PTHrP are helpful in diagnosing certain types of malignancy-associated hypercalcemia. Although FHH is parathyroid-related, the disease should be managed distinctively from hyperparathyroidism. Clinical features and the low urinary calcium excretion can help make the distinction. Because the incidence of malignancy and hyperparathyroidism both increase with age, they can coexist as two independent causes of hypercalcemia.

1,25(OH)<sub>2</sub>D levels are elevated in many (but not all) patients with primary hyperparathyroidism. In other disorders associated with hypercalcemia, concentrations of 1,25(OH)<sub>2</sub>D are low or, at the most, normal. However, this test is of low specificity and is not cost-effective, as not all patients with hyperparathyroidism have elevated 1,25(OH)<sub>2</sub>D levels and not all nonparathyroid hypercalcemic patients have suppressed 1,25(OH)<sub>2</sub>D. Measurement of 1,25(OH)<sub>2</sub>D is, however, critically valuable in establishing the cause of hypercalcemia in sarcoidosis and certain lymphomas.

A useful general approach is outlined in Fig. 410-6. If the patient is *asymptomatic* and there is evidence of *chronicity* to the hypercalcemia, hyperparathyroidism is almost certainly the cause. If PTH levels (usually measured at least twice) are elevated, the clinical impression is confirmed and little additional evaluation is necessary. If there is only a short history or no data as to the duration of the hypercalcemia, *occult malignancy* must be considered; if the PTH levels are not elevated, then a thorough workup must be undertaken for malignancy, including chest x-ray, CT of chest and abdomen, and bone scan. Immunoassays for PTHrP may be especially useful in such situations. Attention should also be paid to clues for underlying hematologic disorders such as anemia, increased plasma globulin, and abnormal serum immunoelectrophoresis; bone scans can be negative in some patients with metastases such as in multiple myeloma. Finally, if a patient with chronic hypercalcemia is asymptomatic and malignancy therefore seems unlikely on clinical grounds, but PTH values are not elevated, it is useful to search for other chronic causes of hypercalcemia such as occult sarcoidosis. A careful history of dietary supplements and drug use may suggest intoxication with vitamin D or vitamin A or the use of thiazides.

## TREATMENT

### General Approach to Hypercalcemic States

The approach to medical treatment of hypercalcemia varies with its severity. Mild hypercalcemia, <3 mmol/L (12 mg/dL), can be managed by hydration. More severe hypercalcemia (levels of 3.2–3.7 mmol/L [13–15 mg/dL]) must be managed aggressively; above that level, hypercalcemia can be life-threatening and requires emergency measures (Table 410-4). By using a combination of approaches in severe hypercalcemia, the serum calcium concentration can be decreased within 24–48 h in most patients, enough to relieve acute symptoms, prevent death from hypercalcemic crisis, and permit diagnostic evaluation. Therapy can then be directed at the underlying disorder—the second priority.

Hypercalcemia develops because of excessive skeletal calcium release, increased intestinal calcium absorption, or inadequate renal calcium excretion. Understanding the particular pathogenesis helps guide therapy. For example, hypercalcemia in patients with malignancy is primarily due to excessive skeletal calcium release and is, therefore, minimally improved by restriction of dietary calcium. On the other hand, patients with vitamin D hypersensitivity or vitamin D intoxication have excessive intestinal calcium absorption, and restriction of dietary calcium is beneficial. Decreased renal function or ECF depletion decreases urinary calcium excretion. In such situations, rehydration may rapidly reduce or reverse the hypercalcemia, even though increased bone resorption persists. As outlined below, the more severe the hypercalcemia, the greater is the number of combined therapies that should be used. Rapid-acting (hours) approaches—rehydration, forced diuresis, and calcitonin—can be used with the most effective antiresorptive agents such as bisphosphonates (since severe hypercalcemia usually involves excessive bone resorption).

### HYDRATION, INCREASED SALT INTAKE, AND MILD AND FORCED DIURESIS

The first principle of treatment is to restore normal hydration. Many hypercalcemic patients are dehydrated because of vomiting, inanition, and/or hypercalcemia-induced defects in urinary concentrating ability. The resultant drop in glomerular filtration rate is

TABLE 410-4 Therapies for Severe Hypercalcemia

TREATMENT	ONSET OF ACTION	DURATION OF ACTION	ADVANTAGES	DISADVANTAGES
<b>Most Useful Therapies</b>				
Hydration with normal saline	Hours	During infusion	Rehydration invariably needed	Volume overload
Forced diuresis; normal saline plus loop diuretic	Hours	During treatment	Rapid action	Volume overload, cardiac decompensation, intensive monitoring, electrolyte disturbance, inconvenience
Pamidronate	1–2 days	10–14 days to weeks	High potency; intermediate onset of action	Fever in 20%, hypophosphatemia, hypocalcemia, hypomagnesemia, rarely jaw necrosis
Zoledronate	1–2 days	>3 weeks	Same as for pamidronate (lasts longer)	Same as pamidronate above
Denosumab	1–2 days	>3 weeks	Strongest antiresorptive	Occasional severe hypocalcemia, rarely jaw necrosis, skin infections
<b>Special Use Therapies</b>				
Calcitonin	Hours	1–2 days	Rapid onset of action; useful as adjunct in severe hypercalcemia	Rapid tachyphylaxis
Phosphate oral	24 h	During use	Chronic management (with hypophosphatemia); low toxicity if P < 4 mg/dL	Limited use except as adjuvant or chronic therapy
Glucocorticoids	Days	Days, weeks	Oral therapy, antitumor agent	Active only in certain malignancies, vitamin D excess, and sarcoidosis; glucocorticoid side effects
Dialysis	Hours	During use and 24–48 h afterward	Useful in renal failure; onset of effect in hours; can immediately reverse life-threatening hypercalcemia	Complex procedure, reserved for extreme or special circumstances

Source: Data from JP Bilezikian et al: Guidelines for the management of asymptomatic primary hyperparathyroidism: Summary statement from the Fourth International Workshop. *J Clin Endocrinol Metab* 99:3561, 2014.)

accompanied by an additional decrease in renal tubular sodium and calcium clearance. Restoring a normal ECF volume corrects these abnormalities and increases urine calcium excretion by 2.5–7.5 mmol/d (100–300 mg/d). Increasing urinary sodium excretion to 400–500 mmol/d increases urinary calcium excretion even further than simple rehydration. After rehydration has been achieved, saline can be administered or furosemide or ethacrynic acid can be given twice daily to depress the tubular reabsorptive mechanism for calcium (care must be taken to prevent dehydration). The combined use of these therapies can increase urinary calcium excretion to 12.5 mmol/d (500 mg/d) in most hypercalcemic patients. Since this is a substantial percentage of the exchangeable calcium pool, the serum calcium concentration usually falls 0.25–0.75 mmol/L (1–3 mg/dL) within 24 h. Precautions should be taken to prevent potassium and magnesium depletion; calcium-containing renal calculi are a potential complication.

Under life-threatening circumstances, the preceding approach can be pursued more aggressively, but the availability of effective agents to block bone resorption (such as bisphosphonates) has reduced the need for extreme diuresis regimens (Table 410-4). Depletion of potassium and magnesium is inevitable unless replacements are given; pulmonary edema can be precipitated. The potential complications can be reduced by careful monitoring of central venous pressure and plasma or urine electrolytes; catheterization of the bladder may be necessary. Dialysis treatment may be needed when renal function is compromised.

### BISPHOSPHONATES

The bisphosphonates are analogues of pyrophosphate, with high affinity for bone, especially in areas of increased bone turnover, where they are powerful inhibitors of bone resorption. These bone-seeking compounds are stable in vivo because phosphatase enzymes cannot hydrolyze the central carbon-phosphorus-carbon bond. The bisphosphonates are concentrated in areas of high bone turnover and are taken up by and inhibit osteoclast action; the mechanism of action is complex. The bisphosphonate molecules that contain amino groups in the side chain structure (see below) interfere with prenylation of proteins and can lead to cellular apoptosis. The

highly active nonamino group-containing bisphosphonates are also metabolized to cytotoxic products.

A number of second- or third-generation compounds have become the mainstays of antiresorptive therapy for treatment of hypercalcemia and osteoporosis. The newer bisphosphonates have a highly favorable ratio of blocking resorption versus inhibiting bone formation; they inhibit osteoclast-mediated skeletal resorption yet do not cause mineralization defects at ordinary doses. Though the bisphosphonates have similar structures, the routes of administration, efficacy, toxicity, and side effects vary. The potency of the compounds for inhibition of bone resorption varies >10,000-fold, increasing in the order of etidronate, tiludronate, pamidronate, alendronate, risedronate, and zoledronate. The IV use of pamidronate and zoledronate is approved for the treatment of hypercalcemia; between 30 and 90 mg pamidronate, given as a single IV dose over a few hours, returns serum calcium to normal within 24–48 h with an effect that lasts for weeks in 80–100% of patients. Zoledronate given in doses of 4 or 8 mg per 5-min infusion has a more rapid and more sustained effect than pamidronate in direct comparison.

These drugs are used extensively in cancer patients. Absolute survival improvements are noted with pamidronate and zoledronate in multiple myeloma, for example. However, though still rare, there are increasing reports of jaw necrosis, especially after dental surgery, mainly in cancer patients treated with multiple doses of the more potent bisphosphonates.

### DENOSUMAB

Denosumab is the most recent antiresorptive therapy to be approved for the treatment of hypercalcemia, a monoclonal antibody that binds to RANK ligand (RANKL) and prevents it from binding to the receptor RANK on osteoclast precursors and mature osteoclasts. The inhibition of differentiation, activation, and function of osteoclasts leads to a reduction in bone resorption. It has a profound suppressive effect on biochemical markers of bone resorption and is the most powerful antiresorptive agent currently available. Repeated doses of denosumab, 120 mg given subcutaneously, may be effective in patients with hypercalcemia of malignancy who have lost responsiveness to bisphosphonates.

## OTHER THERAPIES

*Calcitonin* acts within a few hours of its administration, principally through receptors on osteoclasts, to block bone resorption. *Calcitonin*, after 24 h of use, is no longer effective in lowering calcium. Tac-hypophysis, a known phenomenon with this drug, seems to explain the results since the drug is initially often effective. Therefore, in life-threatening hypercalcemia, calcitonin can be used effectively within the first 24 h in combination with rehydration and saline diuresis while waiting for more sustained effects from a simultaneously administered bisphosphonate such as pamidronate. Usual doses of calcitonin are 2–8 U/kg of body weight IV, SC, or IM every 6–12 h. *Plicamycin* (formerly mithramycin), which inhibits bone resorption, and *gallium nitrate*, which exerts a hypocalcemic action also by inhibiting bone resorption, are no longer used because of superior alternatives such as bisphosphonates.

*Glucocorticoids* have utility, especially in hypercalcemia complicating certain malignancies. They increase urinary calcium excretion and decrease intestinal calcium absorption when given in pharmacologic doses, but they also cause negative skeletal calcium balance. In normal individuals and in patients with primary hyperparathyroidism, glucocorticoids neither increase nor decrease the serum calcium concentration. In patients with hypercalcemia due to certain osteolytic malignancies, however, glucocorticoids may be effective as a result of antitumor effects. The malignancies in which hypercalcemia responds to glucocorticoids include multiple myeloma, leukemia, Hodgkin's disease, other lymphomas, and carcinoma of the breast, at least early in the course of the disease. Glucocorticoids are also effective in treating hypercalcemia due to vitamin D intoxication and sarcoidosis. Glucocorticoids are also useful in the rare form of hypercalcemia, now recognized in certain autoimmune disorders in which inactivating antibodies against the receptor imitate FHH. Elevated PTH and calcium levels are effectively lowered by the glucocorticoids. In all the preceding situations, the hypocalcemic effect develops over several days, and the usual glucocorticoid dosage is 40–100 mg prednisone (or its equivalent) daily in four divided doses. The side effects of chronic glucocorticoid therapy may be acceptable in some circumstances.

*Dialysis* is often the treatment of choice for severe hypercalcemia complicated by renal failure, which is difficult to manage medically. Peritoneal dialysis with calcium-free dialysis fluid can remove 5–12.5 mmol (200–500 mg) of calcium in 24–48 h and lower the serum calcium concentration by 0.7–2.2 mmol/L (3–9 mg/dL). Large quantities of phosphate are lost during dialysis, and serum inorganic phosphate concentration usually falls, potentially aggravating hypercalcemia. Therefore, the serum inorganic phosphate concentration should be measured after dialysis, and phosphate supplements should be added to the diet or to dialysis fluids if necessary.

*Phosphate therapy*, PO or IV, has a limited role in certain circumstances (**Chap. 409**). Correcting hypophosphatemia lowers the serum calcium concentration by several mechanisms, including bone/calcium exchange. The usual oral treatment is 1–1.5 g phosphorus per day for several days, given in divided doses. It is generally believed, but not established, that toxicity does not occur if therapy is limited to restoring serum inorganic phosphate concentrations to normal.

Raising the serum inorganic phosphate concentration above normal decreases serum calcium levels, sometimes strikingly. Intravenous phosphate is one of the most dramatically effective treatments available for severe hypercalcemia but is toxic and even dangerous (fatal hypocalcemia). For these reasons, it is used rarely and only in severely hypercalcemic patients with cardiac or renal failure where dialysis, the preferable alternative, is not feasible or is unavailable.

## SUMMARY

The various therapies for hypercalcemia are listed in Table 410-4. The choice depends on the underlying disease, the severity of the hypercalcemia, the serum inorganic phosphate level, and the renal, hepatic, and bone marrow function. Mild hypercalcemia

( $3 \text{ mmol/L}$  [ $12 \text{ mg/dL}$ ]) can usually be managed by hydration. Severe hypercalcemia ( $3.7 \text{ mmol/L}$  [ $15 \text{ mg/dL}$ ]) requires rapid correction. IV pamidronate or zoledronate or subcutaneous denosumab should be administered. In addition, for the first 24–48 h, aggressive sodium-calcium diuresis with IV saline should be given and, following rehydration, large doses of furosemide or ethacrynic acid, but only if appropriate monitoring is available and cardiac and renal function are adequate. Intermediate degrees of hypercalcemia between  $3$  and  $3.7 \text{ mmol/L}$  ( $12$  and  $15 \text{ mg/dL}$ ) should be approached with vigorous hydration and then the most appropriate selection for the patient of the combinations used with severe hypercalcemia.

## HYPOCALCEMIA

(See also Chap. 54)

**Pathophysiology** *Chronic hypocalcemia* is less common than hypercalcemia; causes include CKD, hereditary and acquired hypoparathyroidism, vitamin D deficiency, PTH resistance, and hypomagnesemia.

Acute rather than chronic hypocalcemia is seen in critically ill patients or as a consequence of certain medications and often does not require specific treatment. Transient hypocalcemia is seen with severe sepsis, burns, acute kidney injury, and extensive transfusions with citrated blood. Although as many as one-half of patients in an intensive care setting are reported to have calcium concentrations of  $<2.1 \text{ mmol/L}$  ( $8.5 \text{ mg/dL}$ ), most do not have a reduction in ionized calcium. Patients with severe sepsis may have a decrease in ionized calcium (true hypocalcemia), but in other severely ill individuals, hypoalbuminemia is the primary cause of the reduced total calcium concentration. Alkalosis increases calcium binding to proteins.

Medications such as protamine, heparin, and glucagon may cause transient hypocalcemia. These forms of hypocalcemia are usually not associated with tetany and resolve with improvement in the overall medical condition. The hypocalcemia after repeated transfusions of citrated blood usually resolves quickly.

Patients with *acute pancreatitis* have hypocalcemia that persists during the acute inflammation and varies in degree with disease severity. The cause of hypocalcemia remains unclear. PTH values are reported to be low, normal, or elevated, and both resistance to PTH and impaired PTH secretion have been postulated. Occasionally, a chronic low total calcium and low ionized calcium concentration are detected in an elderly patient without obvious cause and with a paucity of symptoms; the pathogenesis is unclear.

Chronic hypocalcemia, however, is usually symptomatic and requires treatment. Neuromuscular and neurologic manifestations of chronic hypocalcemia include muscle spasms, carpopedal spasm, facial grimacing, and, in extreme cases, laryngeal spasm and convulsions. Respiratory arrest may occur. Increased intracranial pressure occurs in some patients with long-standing hypocalcemia, often in association with papilledema. Mental changes include irritability, depression, and psychosis. The QT interval on the electrocardiogram is prolonged, in contrast to its shortening with hypercalcemia. Arrhythmias occur, and digitalis effectiveness may be reduced. Intestinal cramps and chronic malabsorption may occur. Chvostek's or Troussseau's sign can be used to confirm latent tetany.

**Classification of Hypocalcemia** The classification of hypocalcemia shown in **Table 410-5** is based on an organizationally useful premise that PTH is responsible for minute-to-minute regulation of plasma calcium concentration and, therefore, that the occurrence of hypocalcemia must mean a failure of the homeostatic action of PTH. Failure of the PTH response can occur if there is hereditary or acquired parathyroid gland failure, if a mutant PTH is secreted, if PTH is ineffective in target organs, or if the action of the hormone is overwhelmed by the loss of calcium from the ECF at a rate faster than it can be replaced.

**PTH Absent** Hereditary or acquired forms of hypoparathyroidism have a number of common components. The disease is rare with estimates from all causes to be ~25–35 patients/100,000 of the population

**TABLE 410-5 Functional Classification of Hypocalcemia (Excluding Neonatal Conditions)**

PTH Absent	
Hereditary hypoparathyroidism	Hypomagnesemia
Acquired hypoparathyroidism	
PTH Ineffective	
Chronic kidney disease	Active vitamin D ineffective
Active vitamin D lacking	Intestinal malabsorption
↓ Dietary intake or sunlight	Vitamin D-dependent rickets type II
Defective metabolism:	
Anticonvulsant therapy	Pseudohypoparathyroidism
Vitamin D-dependent rickets type I	Mutant, less active PTH
PTH Overwhelmed	
Severe, acute hyperphosphatemia	Osteitis fibrosa after parathyroideectomy
Tumor lysis	
Acute kidney injury	
Rhabdomyolysis	

Abbreviation: PTH, parathyroid hormone.

(based on U.S. and Danish estimates). Symptoms of untreated hypocalcemia are shared by both types of hypoparathyroidism, although the onset of hereditary hypoparathyroidism can be more gradual and associated with other developmental defects. Basal ganglia calcification and extrapyramidal syndromes are more common and earlier in onset in hereditary hypoparathyroidism. Acquired hypoparathyroidism secondary to surgery in the neck is more common than hereditary hypoparathyroidism, but the frequency of surgically induced parathyroid failure has diminished as a result of improved surgical techniques that spare the parathyroid glands and increased use of nonsurgical therapy for hyperthyroidism. PHP, an example of resistance to PTH action rather than a failure of parathyroid gland production, may share several features with hypoparathyroidism, including extrasseous calcification and extrapyramidal manifestations such as choreoathetotic movements and dystonia.

Papilledema, raised intracranial pressure, and lenticular cataracts may occur in both hereditary and acquired hypoparathyroidism, as do chronic changes in fingernails and hair, the latter usually reversible with treatment of hypocalcemia. Certain skin manifestations, including alopecia and candidiasis, are characteristic of hereditary hypoparathyroidism associated with autoimmune polyglandular failure (Chap. 388).

Hypocalcemia associated with hypomagnesemia is associated with both deficient PTH release and impaired responsiveness to the hormone. Patients with hypocalcemia secondary to hypomagnesemia have low levels of circulating PTH, indicative of diminished hormone release despite a maximum physiologic stimulus by hypocalcemia. Hypoparathyroidism can be due to hereditary or acquired causes or acute but reversible gland dysfunction (hypomagnesemia).

**GENETIC CAUSES** Hereditary hypoparathyroidism can occur as an isolated entity without other endocrine or dermatologic manifestations or in association with other abnormalities (Chap. 388).

**Hypoparathyroidism Associated with Other Abnormalities** Hypoparathyroidism associated with defective development of both the thymus and the parathyroid glands is termed *DiGeorge syndrome*, or the *velocardiofacial syndrome*. Congenital cardiovascular, facial, and other developmental defects are present, and patients may die in early childhood with severe infections, hypocalcemia and seizures, or cardiovascular complications. Patients can survive into adulthood, and milder, incomplete forms may become manifest in childhood or adolescence. Most cases are sporadic, but autosomal dominant forms involving microdeletions of chromosome 22q11.2 or point mutations in the transcription factor TBX1 in that chromosomal region exist. Another autosomal dominant developmental defect with hypoparathyroidism, deafness,

and renal dysplasia (HDR) is caused by mutations in the transcription factor GATA3 (chromosome 10p14), which is important in embryonic development and is expressed in developing kidney, ear structures, and the parathyroids. Autosomal recessive disorders comprising hypoparathyroidism include *Kenney-Caffey syndrome type 1*, which also features short stature, osteosclerosis, and thick cortical bones, and the related *Sanjad-Sakati syndrome*, which also exhibits growth failure and other dysmorphic features. Both syndromes involve mutations in a chaperone protein called *TBCE* (chromosome 1q42-q43), which is relevant to tubulin function. FAM111A defects (chromosome 11q12.1) were identified as the cause of *Kenney-Caffey syndrome type 2*.

Hypoparathyroidism that can occur in association with a complex hereditary autoimmune syndrome involving failure of the adrenals, the ovaries, the immune system, and the parathyroids in association with recurrent mucocutaneous candidiasis, alopecia, vitiligo, and pernicious anemia is commonly referred to as *polyglandular autoimmune type 1 deficiency* (Chap. 388). This disorder is caused by mutations in the *AIRE* gene (chromosome 21q22.3). A stop codon mutation occurs in many Finnish families with the disorder, while another mutation (Y85C) is typically observed in Jews of Iraqi and Iranian descent.

Hypoparathyroidism is also seen in two disorders associated with mitochondrial dysfunction and myopathy, one termed *Kearns-Sayre syndrome* (KSS), with ophthalmoplegia and pigmentary retinopathy, and the other termed *MELAS syndrome* (mitochondrial encephalopathy, lactic acidosis, and stroke-like episodes). Mutations or deletions in mitochondrial genes have been identified.

**Isolated Hypoparathyroidism** Several forms of hypoparathyroidism, each rare in frequency, are seen as isolated defects; the genetic mechanisms are varied. The inheritance includes autosomal dominant, autosomal recessive, and X-linked modes.

**PTH Mutations** Three separate autosomal defects involving the prepro sequence of PTH have been recognized. The dominant forms are caused by point mutations in a critical region involved in intracellular transport of the hormone precursor. For example, an Arg for Cys mutation interferes with processing of the precursor and is believed to trigger an apoptotic cellular response, hence acting as a dominant negative. The two recessive forms require both PTH alleles encoding the prepro sequence to be mutated. Only one homozygous mutation affecting the secreted PTH (Arg25>Cys25) has been described thus far that leads to an autosomal recessive form of hypoparathyroidism. The defect for an X-linked recessive form of hypoparathyroidism has been localized to chromosome Xq26-q27, perhaps involving the *SOX3* gene.

**CaSR Mutations** Abnormalities in the CaSR are detected in three distinctive hypocalcemic disorders. All are rare, but several different gain-of-function mutations have been found in one form of hypocalcemia termed *autosomal dominant hypocalcemic hypercalcioruria (ADHH)*. The receptor senses the ambient calcium level as excessive and suppresses PTH secretion, leading to hypocalcemia. The hypocalcemia is aggravated by constitutive receptor activity in the renal tubule causing excretion of inappropriate amounts of calcium. Recognition of the syndrome is important because efforts to treat the hypocalcemia with vitamin D analogues and increased oral calcium exacerbate the already excessive urinary calcium excretion (several grams or more per 24 h), leading to irreversible renal damage from stones and ectopic calcification.

**Other Causes of Isolated Hypoparathyroidism** These include homozygous, inactivating mutations in the parathyroid-specific transcription factor *GCM2* or heterozygous point mutations in this protein, which have a dominant-negative effect on the wild-type protein and thus lead to an autosomal dominant form of hypoparathyroidism. Furthermore, heterozygous mutations in *G 11*, one of the two signaling proteins downstream of the CaSR, have been identified as a cause of autosomal dominant hypoparathyroidism. The *Bartter syndrome* is a group of disorders associated with disturbances in electrolyte and acid-base balance, sometimes with nephrocalcinosis and other features. Several types of ion channels or transporters are involved. Curiously, *Bartter syndrome type V* has electrolyte and pH disturbances but is caused by a gain-of-function mutation in the CaSR. The defect may be more

severe than in ADHH and explains the additional features seen beyond hypocalcemia and hypercalcuria. As with autoimmune disorders that block the CaSR (discussed above under hypercalcemic conditions), there are autoantibodies that at least transiently activate the CaSR, leading to suppressed PTH secretion and hypocalcemia.

**ACQUIRED HYPOPARATHYROIDISM** *Acquired chronic hypoparathyroidism* is usually the result of inadvertent surgical removal of all the parathyroid glands; in some instances, not all the tissue is removed, but the remainder undergoes vascular supply compromise secondary to fibrotic changes in the neck after surgery. In the past, the most frequent cause of acquired hypoparathyroidism was surgery for hyperthyroidism. Hypoparathyroidism now usually occurs after surgery for hyperparathyroidism when the surgeon, facing the dilemma of removing too little tissue and thus not curing the hyperparathyroidism, removes too much. Parathyroid function may not be totally absent in all patients with postoperative hypoparathyroidism.

Rare causes of acquired chronic hypoparathyroidism include radiation-induced damage subsequent to radioiodine therapy of hyperthyroidism and glandular damage in patients with hemochromatosis or hemosiderosis after repeated blood transfusions. Infection may involve one or more of the parathyroids but usually does not cause hypoparathyroidism because all four glands are rarely involved.

*Transient hypoparathyroidism* is frequent following surgery for hyperparathyroidism. After a variable period of hypoparathyroidism, normal parathyroid function may return due to hyperplasia or recovery of remaining tissue. Occasionally, recovery occurs months after surgery.

## TREATMENT

### Acquired and Hereditary Hypoparathyroidism

Conventional treatment has involved replacement with vitamin D and  $1,25(\text{OH})_2\text{D}$  (calcitriol) combined with a high oral calcium intake. In most patients, blood calcium and phosphate levels are maintained satisfactorily, but some patients show a tendency to alternate between hypocalcemia and hypercalcemia, thus requiring close monitoring of each patient. Compared to typical daily requirements in euparathyroid patients ( $200\text{--}1000 \text{ }\mu\text{g/d}$ ), much higher doses of vitamin D are needed for the treatment of hypoparathyroid patients (as much as 100-fold higher), which reflects the reduced conversion of vitamin D to  $1,25(\text{OH})_2\text{D}$ . Thus, treatment with  $1,25(\text{OH})_2\text{D}$  ( $0.5\text{--}1 \text{ }\mu\text{g/d}$  of calcitriol) is frequently preferred, particularly since calcitriol is cleared much more rapidly from the circulation than vitamin D.

Oral calcium and vitamin D restore the overall calcium-phosphate balance but do not reverse the lowered urinary calcium reabsorption typical of hypoparathyroidism. Therefore, blood calcium levels should be maintained in these patients at the lower end of the normal range in order to avoid excessive urinary calcium excretion; otherwise, nephrocalcinosis and kidney stones can develop, and the risk of CKD is increased. Thiazide diuretics lower urine calcium by as much as  $100 \text{ mg/d}$  in hypoparathyroid patients on vitamin D, provided they are maintained on a low-sodium diet. Use of thiazides seems to be of benefit in mitigating hypercalcuria and easing the daily management of these patients.

Until recently, hypoparathyroidism has been the only endocrine disorders not being treated with the missing hormone. After the initial experimental use of PTH(1-34), the synthetic PTH fragment used in treatment of osteoporosis, showed promise, full-length PTH(1-84) has been shown to be effective and is now approved by the U.S. Food and Drug Administration for therapy of hypoparathyroidism. Published reports illustrate its use substantially reduced the requirements for supplemental calcium and active vitamin to maintain serum calcium but did not prevent, throughout the day, excessive urinary calcium losses.

**HYPOMAGNESEMIA** Severe hypomagnesemia ( $<0.4 \text{ mmol/L}$ ;  $<0.8 \text{ meq/L}$ ) is associated with hypocalcemia (Chap. 409). Restoration of the

total-body magnesium deficit leads to rapid reversal of hypocalcemia. There are at least two causes of the hypocalcemia—impaired PTH secretion and reduced responsiveness to PTH. **For further discussion of causes and treatment of hypomagnesemia, see Chap. 409.**

PTH levels are undetectable or inappropriately low in severe hypomagnesemia despite the stimulus of severe hypocalcemia, and acute repletion of magnesium leads to a rapid increase in PTH level. Serum phosphate levels are often not elevated, in contrast to the situation with acquired or idiopathic hypoparathyroidism, probably because phosphate deficiency is often seen in hypomagnesemia. In addition to diminished PTH secretion, some patients with low calcium and magnesium levels show a blunted peripheral response to exogenous PTH as documented by subnormal response in urinary phosphorus and urinary cyclic AMP excretion.

## TREATMENT

### Hypomagnesemia

Repletion of magnesium cures the condition. Repletion should be parenteral. Attention must be given to restoring the intracellular deficit, which may be considerable. After IV magnesium administration, serum magnesium may return transiently to the normal range, but unless replacement therapy is adequate, serum magnesium will again fall. If the cause of the hypomagnesemia is renal magnesium wasting, magnesium may have to be given long-term to prevent recurrence (Chap. 409).

**PTH Ineffective** PTH is ineffective when the PTHR1-signaling protein complex is defective (as in the different forms of PHP, discussed below) or in CKD in which the calcium-elevating action of PTH is impaired.

Typically, hypophosphatemia is more severe than hypocalcemia in vitamin D deficiency states because the increased PTH levels, although only partly effective in elevating blood calcium, are readily capable of promoting urinary phosphate excretion.

PHP, on the other hand, has a pathophysiology that is different from the other disorders of ineffective PTH action. PHP resembles hypoparathyroidism (in which PTH synthesis is deficient) and is manifested by hypocalcemia and hyperphosphatemia yet elevated PTH levels. The cause of the disorder is defective PTH-dependent activation of the stimulatory G protein complex or the downstream effector protein kinase A, resulting in failure of PTH to increase intracellular cyclic AMP or to respond to elevated cyclic AMP levels (see below).

**CKD** Improved medical management of CKD allows many patients to survive for decades and, hence, provides time enough to develop features of renal osteodystrophy, which must be controlled to avoid additional morbidity. Impaired production of  $1,25(\text{OH})_2\text{D}$  is a principal factor that causes calcium deficiency, secondary hyperparathyroidism, and bone disease; hyperphosphatemia, which lowers further blood calcium levels, typically occurs only in the later stages of the disease. Low levels of  $1,25(\text{OH})_2\text{D}$  due to increased FGF23 production in bone (and possibly other tissues) are critical in the development of hypocalcemia. It is notable that FGF23 levels are often dramatically elevated in end-stage kidney disease (ESKD). The uremic state also causes impairment of intestinal absorption by mechanisms other than defects in vitamin D metabolism. Nonetheless, treatment with supraphysiologic amounts of vitamin D or calcitriol can correct impaired calcium absorption. Increased FGF23 levels are seen already during the early CKD stages and have been reported to correlate with kidney disease progression, increased mortality, and left ventricular hypertrophy. Strategies involving different oral phosphate binders have therefore been pursued to lower intestinal phosphate absorption early during the course of kidney disease and to thereby lower FGF23 levels. However, these approaches have been largely disappointing. Furthermore, there is concern as to whether supplementation with activated vitamin D analogues increases further the circulating FGF23 levels and their “off-target” effects in CKD patients.

## Chronic Kidney Disease

Therapy of CKD ([Chap. 311](#)) involves appropriate management of patients prior to dialysis and adjustment of regimens once dialysis is initiated. Attention should be paid to restriction of phosphate in the diet; avoidance of aluminum-containing phosphate-binding antacids; provision of an adequate calcium intake by mouth, usually 1–2 g/d; and supplementation with 0.25–1 µg/d calcitriol or other activated forms of vitamin D. The aim of therapy is to restore normal calcium balance to prevent osteomalacia and severe secondary hyperparathyroidism (it is usually recommended to maintain PTH levels between 100 and 300 pg/mL) and, in light of evidence of genetic changes and monoclonal outgrowths of parathyroid glands in CKD patients, to prevent secondary hyperparathyroidism from becoming autonomous hyperparathyroidism. Reduction of hyperphosphatemia and restoration of normal intestinal calcium absorption by calcitriol can improve blood calcium levels and reduce the manifestations of secondary hyperparathyroidism. Since adynamic bone disease can occur in association with low PTH levels, it is important to avoid excessive suppression of the parathyroid glands while recognizing the beneficial effects of controlling the secondary hyperparathyroidism. These patients should be closely monitored with PTH assays that detect only the full-length or biologically active PTH(1–84) to ensure that inactive, inhibitory PTH fragments are not measured. Use of oral phosphate-binding agents such as sevelamer lower blood phosphate levels in ESKD, but their use in earlier CKD stages does not seem to be beneficial in lowering blood phosphate levels and to prevent the rise in FGF23.

**VITAMIN D DEFICIENCY DUE TO INADEQUATE DIET AND/OR SUNLIGHT** Vitamin D deficiency due to inadequate intake of dairy products enriched with vitamin D, lack of vitamin supplementation, and reduced sunlight exposure in the elderly, particularly during winter in northern latitudes, is more common in the United States than previously recognized. Biopsies of bone in elderly patients with hip fracture (documenting osteomalacia) and abnormal levels of vitamin D metabolites, PTH, calcium, and phosphate indicate that vitamin D deficiency may occur in as many as 25% of elderly patients, particularly in northern latitudes in the United States. Concentrations of 25(OH)D are low or low-normal in these patients. Quantitative histomorphometric analysis of bone biopsy specimens from such individuals reveals widened osteoid seams consistent with osteomalacia ([Chap. 409](#)). PTH hypersecretion compensates for the tendency for the blood calcium to fall but also increases renal phosphate excretion and thus causes osteomalacia.

Treatment involves adequate replacement with vitamin D and calcium until the deficiencies are corrected. Severe hypocalcemia rarely occurs in moderately severe vitamin D deficiency of the elderly, but vitamin D deficiency must be considered in the differential diagnosis of mild hypocalcemia.

Mild hypocalcemia, secondary hyperparathyroidism, severe hypophosphatemia, and a variety of nutritional deficiencies occur with gastrointestinal diseases. Hepatocellular dysfunction can lead to reduction in 25(OH)D levels, as in portal or biliary cirrhosis of the liver, and malabsorption of vitamin D and its metabolites, including 1,25(OH)<sub>2</sub>D, may occur in a variety of bowel diseases, hereditary or acquired. Hypocalcemia itself can lead to steatorrhea, due to deficient production of pancreatic enzymes and bile salts. Depending on the disorder, vitamin D or its metabolites can be given parenterally, guaranteeing adequate blood levels of active metabolites.

**DEFECTIVE VITAMIN D METABOLISM • Anticonvulsant Therapy** Anticonvulsant therapy with any of several agents induces acquired vitamin D deficiency by increasing the conversion of vitamin D to inactive compounds and/or causing resistance to its action. The more marginal the vitamin D intake in the diet, the more likely that anticonvulsant therapy will lead to abnormal mineral and bone metabolism.

**Vitamin D-Dependent Rickets Type I** Vitamin D-dependent rickets type I, previously termed *pseudo-vitamin D-resistant rickets*, is caused by homozygous or compound heterozygous mutations in the gene encoding 25(OH)D-1'-hydroxylase. It differs from true vitamin D-resistant rickets (vitamin D-dependent rickets type II, see below) in that it is typically less severe and the biochemical and radiographic abnormalities can be readily reversed with physiologic doses of the vitamin's active metabolite, 1,25(OH)<sub>2</sub>D ([Chap. 409](#)). Clinical features include hypocalcemia, often with tetany or convulsions; hypophosphatemia due to secondary hyperparathyroidism; and thus, osteomalacia and increased levels of alkaline phosphatase.

**Vitamin D-Dependent Rickets Type II** Vitamin D-dependent rickets type II results from end-organ resistance to the active metabolite 1,25(OH)<sub>2</sub>D. The clinical features resemble those of the type I disorder and include hypocalcemia, hypophosphatemia, secondary hyperparathyroidism, and rickets but also partial or total alopecia. Plasma levels of 1,25(OH)<sub>2</sub>D are elevated, in keeping with the refractoriness of the end-organs. This disorder is caused by homozygous or compound heterozygous mutations in the gene encoding the vitamin D receptor; treatment requires regular, usually nocturnal calcium infusions, which normalize PTH levels, thus reducing urinary phosphate excretion and thereby improving rickets and thus growth, but do not restore hair growth ([Chap. 409](#)).

**PSEUDOHYPOPARTHYROIDISM** PHP refers to a group of distinct inherited disorders. Patients affected by PHP type Ia (PHP1A) develop symptoms and signs of hypocalcemia in association with distinctive skeletal and developmental defects, referred to as Albright's hereditary osteodystrophy (AHO). The hypocalcemia is due to a deficient PTH response in the proximal renal tubules, probably leading to insufficient 1,25(OH)<sub>2</sub>D production and thus impaired intestinal calcium absorption. Furthermore, PTH resistance in this portion of the kidney impairs urinary phosphate excretion, thus leading to elevated serum phosphate levels. Patients affected by PHP type Ib (PHP1B) also present with hypocalcemia and hyperphosphatemia but less frequently with obvious AHO features. In response to the hypocalcemia observed in either disorder, PTH levels increase, leading to parathyroid hyperplasia and, in some cases, to autonomous PTH secretion. Studies, both clinical and basic, have clarified some aspects of these disorders, including the variable clinical spectrum, the pathophysiology, the genetic defects, and their mode of inheritance.

A working classification of the various PHP forms is given in [Table 410-6](#). The classification scheme is based on the signs of ineffective PTH action (low calcium and high phosphate), low or normal urinary cyclic AMP response to exogenous PTH, the presence or absence of AHO, and assays to measure the concentration of the G<sub>s</sub> subunit of the adenylate cyclase enzyme. Using these criteria, there are four types: PHP types Ia and Ib (PHP1A and PHP1B); pseudopseudohypoparathyroidism (PPHP), and PHP type II (PHP2). Another classification has been proposed recently, which is being debated.

**PHP1A and PHP1B** Individuals with PHP type I (PHP1), the most common of the disorders, show deficient urinary cyclic AMP excretion in response to administration of exogenous PTH. Patients with PHP1 are divided into PHP1A and PHP1B. Patients with PHP1A show evidence for AHO and reduced amounts of G<sub>s</sub> protein/activity, as determined in readily accessible tissues such as erythrocytes, lymphocytes, or fibroblasts. Only some PHP1B patients show typically AHO features, but they usually have normal G<sub>s</sub> activity. PHP1C, sometimes listed as a third form of PHP1, is really a variant of PHP1A, although the mutant G<sub>s</sub> shows normal activity in certain *in vitro* assays.

Most patients who have PHP1A reveal characteristic features of AHO, which consist of short stature, early-onset obesity, round face, obesity, skeletal anomalies (brachydactyly), intellectual impairment, and/or heterotopic calcifications. Patients have low calcium and high phosphate levels, as with true hypoparathyroidism. PTH levels, however, are elevated, reflecting resistance to hormone action. In addition, hormonal resistance is observed at other G<sub>s</sub>-coupled receptors, particularly at the TSH receptor, leading to elevated levels of this hormone.

**TABLE 410-6 Classification of Pseudohypoparathyroidism (PHP) and Pseudopseudohypoparathyroidism (PPHP)**

TYPE	HYPOCALCEMIA, HYPERPHOSPHATEMIA	RESPONSE OF URINARY cAMP TO PTH	SERUM PTH	G <sub>s</sub> SUBUNIT DEFICIENCY	AHO	RESISTANCE TO HORMONES OTHER THAN PTH
PHP1A	Yes	↓	↑	Yes	Yes	Yes
PPHP	No	Normal	Normal	Yes	Yes	No
PHP1B	Yes	↓	↑	No	Yes (less frequently and usually less severe)	Yes (in some patients)
PHP2	Yes	Normal	↑	No	No	No
Acrodyostosis due to <i>PRKAR1A</i> mutations with hormonal resistance	Yes	Normal (but ↓ phosphaturic response)	↑	No	Yes	Yes

Abbreviations: ↓, decreased; ↑, increased; AHO, Albright's hereditary osteodystrophy; cAMP, cyclic adenosine monophosphate; PTH, parathyroid hormone.

Amorphous deposits of calcium and phosphate are found in the basal ganglia. The typical shortening of metacarpal and metatarsal bones is caused by premature closing of the epiphyses and is probably a particularly sensitive sign of overall advanced skeletal maturation resulting in adult short stature.

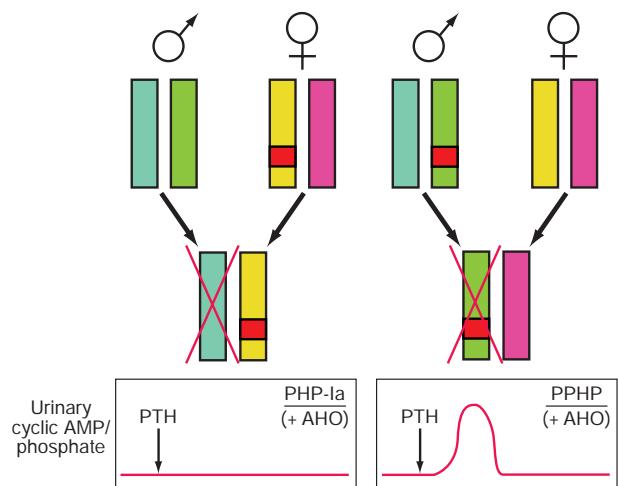
**INHERITANCE AND GENETIC DEFECTS** Multiple defects at the *GNAS* locus have now been identified in PHP1A, PHP1B, and PPHP patients. This gene, which is located on chromosome 20q13.3, encodes the  $\alpha$ -subunit of the stimulatory G protein ( $G_s$ ), among other products (see below). Mutations involving the *GNAS* exons encoding  $G_s$ , which are the cause of PHP1A and PPHP, include abnormalities at splice junctions, point mutations, insertions, and/or deletions that all result in a  $G_s$  protein with defective function, resulting in a 50% reduction of *in vitro*  $G_s$  activity in erythrocytes or other cells. While PHP1A is caused by inactivating  $G_s$  mutations on the maternal allele, PPHP is caused by the same or similar mutations on the paternal *GNAS* allele (Fig. 410-7). The  $G_s$  transcript is biallelically expressed in most tissues; however, expression from paternal allele is silenced through as-of-yet-unknown mechanisms in some tissues, including proximal renal tubules, thyroid, and pituitary. Consequently, inheritance of a molecular defect involving the paternal exons encoding  $G_s$  has no implications with regard to hormone function, while inactivating  $G_s$  mutations involving the maternal *GNAS* allele lead to little or no  $G_s$  protein in these tissues (Chap. 466). Thus, females affected by either PHP1A or PPHP will have offspring with PHP1A, if these children inherit the allele carrying the *GNAS* mutation; in contrast, if the mutant allele is inherited from a male affected by either disorder, the offspring will exhibit PPHP. However, patients affected by either disorder develop some but not all AHO features, making it likely that  $G_s$  haploinsufficiency occurs during embryonic or postnatal development.

The complex mechanisms that control the *GNAS* gene contributed particularly to challenges involved in unraveling the pathogenesis of PHP1B. Analysis of families in which multiple members are affected by PHP1B, as well as studies of the complex parent-specific methylation of four regions within the complex *GNAS* locus, revealed that the autosomal dominant forms of PHP1B (AD-PHP1B) are caused either by microdeletions, duplications, or inversions within or upstream of the *GNAS* locus. These genetic mutations are associated with a loss of DNA methylation at one or several loci on the maternal *GNAS* allele (Table 410-6). These abnormalities in methylation silence maternal  $G_s$  expression, thus leading in the proximal renal tubules—where  $G_s$  appears to be expressed predominantly from the maternal allele—to PTH resistance. While most cases of AD-PHP1B are by now resolved at the molecular level, the genetic defect responsible for the sporadic variant of PHP1B (sporPHP1B), the most frequent form of PHP1B, remains to be defined, except for those sporPHP1B cases that are caused by paternal uniparental isodisomy/heterodisomy of chromosome 20q (patUPD20q).

PHP1B patients, who rarely develop an AHO phenotype as severe as in PHP1A, develop hypocalcemia and hyperphosphatemia caused by PTH resistance and thus elevated PTH levels. The previously used

Ellsworth-Howard test to assess the presence or absence of hormone resistance is used much less frequently, largely because of routinely available sensitive PTH assays (Table 410-6). As for PHP1A, these endocrine abnormalities become apparent only if disease-causing mutations are inherited maternally. Bone responsiveness may be excessive rather than blunted in PHP1B (and in PHP1A) patients, based on case reports that have emphasized an osteitis fibrosa-like pattern in several PHP1B patients. Some patients present with PTH resistance in the absence of AHO features and without *GNAS* methylation changes; it remains unclear why this PHP variant readily resolves upon treatment with vitamin D supplements.

PHP2 refers to patients with hypocalcemia and hyperphosphatemia, who have normal urinary cyclic AMP excretion, but an impaired urinary phosphaturic response to PTH. In one PHP2 variant, referred to as acrodyostosis with hormonal resistance, patients have a heterozygous



**FIGURE 410-7** Paternal imprinting of renal parathyroid hormone (PTH) resistance (*GNAS* gene for  $G_s$  subunit) in pseudohypoparathyroidism (PHP1A and PHP1B). An impaired excretion of urinary cyclic AMP and phosphate is observed in patients with PHP type I. In the renal cortex, there is selective silencing of paternal  $G_{\alpha}$  expression; consequently, mutations involving the maternal *GNAS* exons encoding  $G_{\alpha}$  or loss of methylation at *GNAS* exon A/B leads to reduced or completely absent  $G_{\alpha}$  protein in this portion of the kidney. The disease becomes manifest only in patients who inherit the defective gene from an obligate female carrier (left). If a genetic defect involving *GNAS* exons encoding  $G_{\alpha}$  is inherited from an obligate male carrier of the mutation (PHP1A or PPHP patient), no biochemical abnormality is encountered, and the administration of PTH causes an appropriate increase in the urinary cyclic AMP and phosphate concentration (pseudoPHP [PPHP]; right). Both patterns of inheritance lead to some but not all features of Albright's hereditary osteodystrophy (AHO), most likely because of haploinsufficiency; for example,  $G_{\alpha}$  protein derived from both parental *GNAS* alleles must be active for normal bone development. Maternal inheritance of a mutation (deletion, duplication, or inversion within or upstream of the *GNAS* locus) causes AD-PHP1B, while paternal inheritance does not lead to any detectable abnormality.

**3190** defect in the regulatory subunit of PKA (PRKAR1A) that mediates the response to PTH distal to cyclic AMP production. Acrodysostosis without or with only mild hormonal resistance can be caused by heterozygous mutations in the cyclic AMP-selective phosphodiesterase 4D. In patients with one variant of acrodysostosis that is associated with hypertension, it was shown to be caused by heterozygous phosphodiesterase 3A mutations.

The diagnosis of these hormone-resistant states can usually be made when there is a positive family history for signs and symptoms of hypocalcemia with or without AHO features. In both categories—PHP1A and PHP1B—serum PTH levels are elevated, particularly when patients start to experience hypocalcemia during childhood. However, patients with PHP1B or PHP2 without skeletal findings present only with hypocalcemia and high PTH levels, as evidence for hormone resistance. In PHP1A and PHP1B, the response of urinary cyclic AMP to the administration of exogenous PTH is blunted. The diagnosis of PHP2, in the absence of acrodysostosis, is more complex, and vitamin D deficiency must be excluded before such a diagnosis can be entertained.

## TREATMENT

### Pseudohypoparathyroidism

Treatment of PHP is similar to that of hypoparathyroidism, except that calcium and activated vitamin D analogues are usually given at higher doses to maintain blood calcium levels within the normal range and PTH levels in the upper end of normal or slightly elevated. Patients with PHP1 show no PTH resistance in the distal tubules—hence, urinary calcium clearance is typically reduced, and these individuals are not at risk of developing nephrocalcinosis, as are patients with hypoparathyroidism, unless overtreatment occurs, for example, after the completion of pubertal development and skeletal maturation, when calcium and  $1,25(\text{OH})_2\text{D}$  treatment should be reduced. Variability in response makes it necessary to establish the optimal regimen for each patient.

**PTH Overwhelmed** Occasionally, loss of calcium from the ECF is so severe that PTH cannot compensate. Such situations include acute pancreatitis and severe, acute hyperphosphatemia, often in association with renal failure, conditions in which there is rapid efflux of calcium from the ECF. Severe hypocalcemia can occur quickly; PTH rises in response to hypocalcemia but does not return blood calcium to normal.

**SEVERE, ACUTE HYPERPHOSPHATEMIA** Severe hyperphosphatemia is associated with extensive tissue damage or cell destruction (*Chap. 409*). The combination of increased release of phosphate from muscle and impaired ability to excrete phosphorus because of renal failure causes moderate to severe hyperphosphatemia, the latter causing calcium loss from the blood and mild to moderate hypocalcemia. Hypocalcemia is usually reversed with tissue repair and restoration of renal function as phosphorus and creatinine values return to normal. There may even be a mild hypercalcemic period in the oliguric phase of renal function recovery. This sequence, severe hypocalcemia followed by mild hypercalcemia, reflects widespread deposition of calcium in muscle and subsequent redistribution of some of the calcium to the ECF after phosphate levels return to normal.

Other causes of hyperphosphatemia include hypothermia, massive hepatic failure, and hematologic malignancies, either because of high cell turnover of malignancy or because of cell destruction by chemotherapy.

## TREATMENT

### Severe, Acute Hyperphosphatemia

Treatment is directed toward lowering of blood phosphate by the administration of phosphate-binding antacids or dialysis. Although

calcium replacement may be necessary if hypocalcemia is severe and symptomatic, calcium administration during the hyperphosphatemic period tends to increase extraosseous calcium deposition and aggravate tissue damage. The levels of  $1,25(\text{OH})_2\text{D}$  may be low during the hyperphosphatemic phase and return to normal during the oliguric phase of recovery.

**OSTEITIS FIBROSA AFTER PARATHYROIDECTOMY** Severe hypocalcemia after parathyroid surgery is rare now that osteitis fibrosa cystica is an infrequent manifestation of hyperparathyroidism. When osteitis fibrosa cystica is severe, however, bone mineral deficits can be large. After parathyroidectomy, hypocalcemia can persist for days if calcium replacement is inadequate. Treatment may require parenteral administration of calcium; addition of calcitriol and oral calcium supplementation is sometimes needed for weeks to a month or two until bone defects are filled (which, of course, is of therapeutic benefit in the skeleton), making it possible to discontinue parenteral calcium and/or reduce the amount.

**Differential Diagnosis** Care must be taken to ensure that true hypocalcemia is present; in addition, acute transient hypocalcemia can be a manifestation of a variety of severe, acute illnesses, as discussed above. *Chronic hypocalcemia*, however, can usually be ascribed to a few disorders associated with absent or ineffective PTH. Important clinical criteria include the duration of the illness, signs or symptoms of associated disorders, and the presence of features that suggest a hereditary abnormality. A nutritional history can be helpful in recognizing a low intake of vitamin D and calcium in the elderly, and a history of excessive alcohol intake may suggest magnesium deficiency.

Hypoparathyroidism and PHP are typically lifelong illnesses, usually (but not always) appearing by adolescence; hence, a recent onset of hypocalcemia in an adult is more likely due to nutritional deficiencies, CKD, or intestinal disorders that result in deficient or ineffective vitamin D. Neck surgery, even long past, however, can be associated with a delayed onset of postoperative hypoparathyroidism. A history of seizure disorder raises the issue of anticonvulsive medication. Developmental defects may point to the diagnosis of PHP1A. Rickets and a variety of neuromuscular syndromes and deformities may indicate ineffective vitamin D action, either due to defects in vitamin D metabolism or to vitamin D deficiency.

A pattern of *low calcium with high phosphorus* in the absence of renal failure or massive tissue destruction almost invariably means hypoparathyroidism or PHP. A *low calcium and low phosphorus* pattern points to absent or ineffective vitamin D, thereby impairing the action of PTH on calcium metabolism (but not phosphate clearance). The relative ineffectiveness of PTH in calcium homeostasis in vitamin D deficiency, anticonvulsant therapy, gastrointestinal disorders, and hereditary defects in vitamin D metabolism leads to secondary hyperparathyroidism as a compensation. The excess PTH on renal tubule phosphate transport accounts for renal phosphate wasting and hypophosphatemia.

## FURTHER READING

Bastepe M, Jüppner H: Pseudohypoparathyroidism, Albright's hereditary osteodystrophy, and progressive osseous heteroplasia: Disorders caused by inactivating GNAS mutations, in *Endocrinology*, 6th ed, in *Endocrinology*, JL Jameson, LJ DeGroot (eds). Philadelphia, W.B. Saunders Company, 2016.

Bilezikian JP et al: Guidelines for the management of asymptomatic primary hyperparathyroidism: Summary statement from the fourth international workshop. *J Clin Endocrinol Metab* 99:3561, 2014.

Thakker RV et al: Genetic disorders of calcium homeostasis caused by abnormal regulation of parathyroid hormone secretion or responsiveness, in *Endocrinology*, 6th ed, in *Endocrinology*, JL Jameson, LJ DeGroot (eds). Philadelphia, W.B. Saunders Company, 2016.



Osteoporosis, a condition characterized by decreased bone strength, is prevalent among postmenopausal women but also occurs in both women and men as a function of age and with underlying conditions or major risk factors associated with loss of bone mass. Its chief clinical manifestations are vertebral and hip fractures, although fractures can occur at almost any skeletal site. Osteoporosis affects >10 million individuals in the United States, but only a proportion are diagnosed and treated.

## DEFINITION

*Osteoporosis* is defined as a reduction in the strength of bone that leads to an increased risk of fractures. Loss of bone tissue causes deterioration in skeletal microarchitecture, and thus, the process of bone loss causes a greater detriment to bone strength than might be appreciated from the simple measure of bone “density.” The World Health Organization (WHO) operationally defined osteoporosis as a bone density that falls 2.5 standard deviations (SDs) or more below the mean for young healthy adults of the same sex and race—also referred to as a *T-score* of  $-2.5$ . Postmenopausal women at the lower end of the young normal range (*T-score*  $<-1.0$ ) are defined as having low bone density and may be at increased risk of osteoporosis. Although fracture risk is lower in this group, >50% of fractures among postmenopausal women, including hip fractures, occur in those with low bone density as the numerical size of that population is larger than the group with bone density osteoporosis. As a consequence, clinical assessment has evolved to include an estimate of the risk of fracture, incorporating bone mineral density (BMD) with age, gender, and other clinical risk factors to allow a calculated 10-year risk of hip or major osteoporosis-related fracture. This has evolved into a second definition of osteoporosis with cut points for intervention that are variable across different geographies.

Osteoporosis-related fractures are defined as fractures of any bone in adults that occur in the setting of trauma less than or equal to a fall from standing height, with the exceptions of fingers, toes, face, and skull. However, in individuals thought to be at risk of osteoporosis, any traumatic fracture must be regarded as possibly indicative of an underlying skeletal problem, raising consideration of further evaluation.

## EPIDEMIOLOGY

In the United States, as many as 10.8 million women and 2.5 million men have osteoporosis (BMD *T-score*  $<-2.5$  at lumbar spine, total hip, or femoral neck). This does not include additional people who present with an osteoporosis-related fracture but with low bone mass (*T-score*  $<-1$  to  $-2.5$ ). It is estimated that 2 million osteoporosis-related fractures occur each year in the United States at a cost of \$19 billion, a problem that will increase as the population ages with an estimate of 3 million fractures and \$25 billion in costs by 2025. The failure to identify the first fracture and intervene is estimated to cost \$6 billion to Medicare alone for secondary fractures. About 40 million individuals have low bone mass (*T-score*  $<-1$  to  $-2.5$ ) that potentially puts them at increased of fracture and of developing osteoporosis. Osteoporosis is mostly age related, as bone tissue is lost progressively. In women, the loss of ovarian function at menopause (typically around age 50) precipitates rapid bone loss such that most women meet the diagnostic criterion for osteoporosis by age 70–80. As the population ages, the number of individuals with osteoporosis and fractures rises. As many of the fractures defined as related to osteoporosis occur in individuals with low bone mass, identification of those at high risk of fracture and their evaluation and treatment have become important issues in clinical management.

The epidemiology of fractures follows the trend for loss of bone density, with most fractures, especially those of the hip and vertebrae,

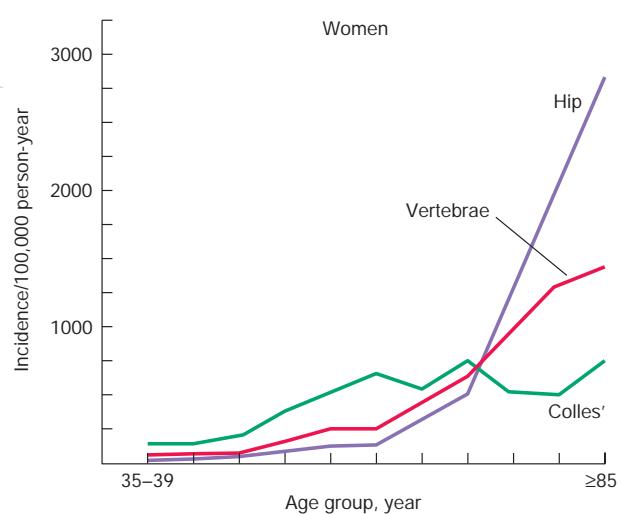


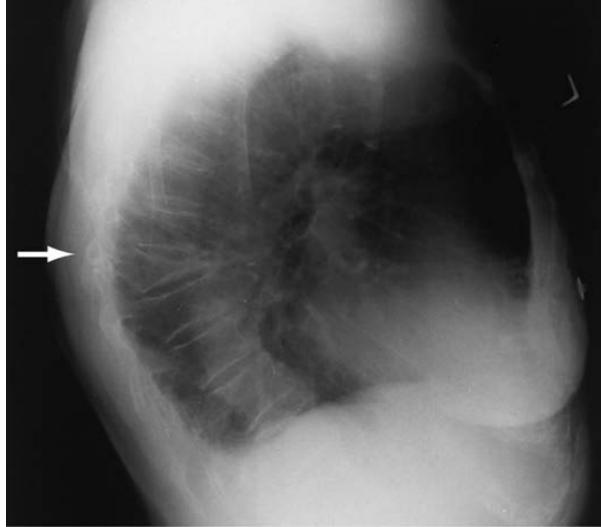
FIGURE 411-1 Epidemiology of vertebral, hip, and Colles' fractures with age. (Reproduced with permission from C Cooper, LJ Melton 3rd: Epidemiology of osteoporosis. *Trends Endocrinol Metab* 3:224, 1992.)

showing exponential increases with advancing age (Fig. 411-1). Lifetime osteoporotic fracture risk for a Caucasian woman who reaches the age of 50 is ~50%, and corresponding risk for a 50-year-old man is ~25%. Recent data suggest that fractures are increasing despite the availability of effective medications. This may be related to the failure to evaluate patients who fall into a high-risk group for underlying skeletal problems leading to fractures.

About 300,000 hip fractures occur each year in the United States, almost all requiring hospital admission and emergency surgical intervention. The lifetime probability that a 50-year-old white individual will have a hip fracture is 14% for women and 5% for men; the risk for African Americans is about half of those rates, and the risk for Asians and nonblack Hispanics appears similar to that for Caucasians. Surgical intervention for hip fractures is associated with a high incidence of mortality and morbidity, with 20–25% of patients dying in the year following the injury with higher mortality rates among males and African Americans. About 30% of survivors require long-term care (at least temporarily), and many never regaining the independence that they had prior to the fracture.

There are ~500,000 symptomatic vertebral fractures per year in the United States, but >1,000,000 vertebral fractures may actually occur yearly since only about one-third are recognized clinically at the time of the event. Many of these initially “silent” vertebral fractures are identified incidentally during radiography for other purposes (Fig. 411-2). Even when asymptomatic, these vertebral fractures are a major sign of skeletal fragility and may carry the same predictive value for subsequent fracture. Vertebral fractures rarely require hospitalization but are associated with long-term morbidity and an increase in mortality. The occurrence of the first fracture increases the risk of further fractures, especially in the first year after clinically evident fractures. The consequence is height loss (often of several inches), kyphosis, and secondary pain and discomfort related to altered biomechanics of the back. Thoracic fractures can be associated with restrictive lung disease, whereas lumbar fractures are associated with abdominal symptoms that include distention, early satiety, and constipation.

Approximately 400,000 wrist fractures occur in the United States each year. Fractures of other bones (including ~150,000 pelvic fractures and >100,000 proximal humerus fractures) also occur with osteoporosis. Although some fractures result from major trauma, the threshold for fracture is reduced in osteoporotic bone (Fig. 411-3). The occurrence of a traumatic fracture in someone at risk of osteoporosis in the skeleton necessitates evaluation for reduced bone mass and, if appropriate, intervention to reduce future fracture risk. Fewer than 10% of these patients are currently investigated for osteoporosis within 6 months of a new fracture. In addition to reduced bone density



**FIGURE 411-2** Lateral spine x-ray showing severe osteopenia and a severe wedge-type deformity (severe anterior compression).

with advancing age, there are a number of risk factors for fracture; the common ones are summarized in **Table 411-1**. Prior fractures, a family history of osteoporosis-related fractures (particularly hip fractures), low body weight, cigarette consumption, and excessive alcohol consumption are all independent predictors of fracture. Chronic diseases with inflammatory components that increase skeletal remodeling, such as rheumatoid arthritis, increase the risk of osteoporosis, as do diseases associated with malabsorption. Chronic diseases that increase the risk of falling or frailty, including dementia, Parkinson's disease, and multiple sclerosis, also increase fracture risk (Table 411-1). Many other risk factors for osteoporosis have been described including air pollution, triclosan, gastric bypass surgery, diabetes, cerebrovascular accidents, dementia (including Alzheimer's), the death of a spouse, and depression and its treatment, to name a few.

The increasing frailty with age is a potent risk factor for fracture, as is sensory inattention (e.g., walking while looking at mobile phone).

In the United States and Europe, osteoporosis-related fractures are more common among women than men, presumably due to a lower peak bone mass as well as postmenopausal bone loss in women. However, this gender difference in bone density and age-related increase in hip fractures is not as apparent in some other cultures, possibly due to genetics, physical activity level, or diet.

Fractures are themselves risk factors for future fractures (Table 411-1). Vertebral fractures increase the risk of other vertebral fractures as well as fractures of the peripheral skeleton such as the hip and wrist. Wrist fractures also increase the risk of vertebral and hip fractures. Among individuals aged >50, any fracture except those of the fingers, toes, face, and skull should be considered as potentially related to osteoporosis

**TABLE 411-1** Risk Factors for Osteoporosis Fracture

NONMODIFIABLE	POTENTIALLY MODIFIABLE
Personal history of fracture as an adult	Current cigarette smoking
History of fracture in first-degree relative	Estrogen deficiency
Female gender	Early menopause (<45 years) or bilateral ovariectomy
Advanced age	Prolonged premenstrual amenorrhea (>1 year)
White race	Poor nutrition especially low calcium and vitamin D intake
Dementia	Alcoholism
	Impaired eyesight despite adequate correction
	Recurrent falls
	Inadequate physical activity
	Poor health/frailty

regardless of the specific circumstances of the fracture. Osteoporotic bone is more likely to fracture than is normal bone at any level of trauma, and a fracture in a person aged >50 should trigger evaluation for osteoporosis. This often does not occur since postfracture care is fragmented. Recent attempts to coordinate care using a fracture liaison health care provider to guide patients through the system and ensure their evaluation and treatment for osteoporosis may improve care but is more difficult to do in the open medical care systems in the United States. In countries with single-payer systems, that approach does seem to be effective, as is also the case in closed health care systems in the United States.

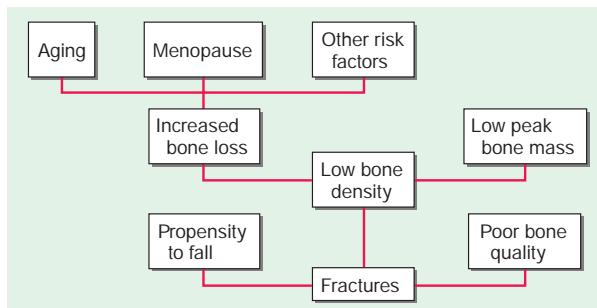
The risk for future fracture after a first fracture is not linear. Highest risk occurs within the first 2 years after the first fracture. A recent large Medicare database study indicated that almost 20% of women will have a second fracture within 2 years after the first. Risk diminishes to less than half of that rate in the subsequent 3 years and declines to baseline thereafter for most fracture types, although risk after a vertebral or hip fracture may persist.

## PATHOPHYSIOLOGY

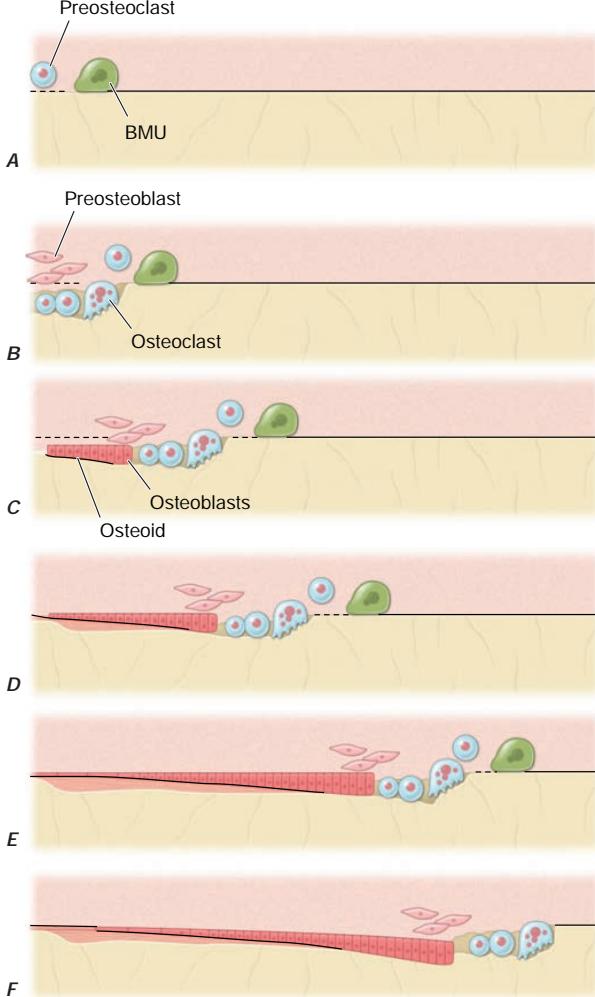
### BONE REMODELING

Osteoporosis results from bone loss due to age-related changes in bone remodeling as well as extrinsic and intrinsic factors that exaggerate this process. These changes may be superimposed on a low peak bone mass. Consequently, understanding the bone remodeling process is fundamental to understanding both the pathophysiology of osteoporosis (**Chap. 409**) and the effects of pharmacologic intervention. During growth, the skeleton increases in size by linear growth and by apposition of new bone tissue on the outer surfaces of the cortex (**Fig. 411-4**). The latter process is called *modeling*, a process that also allows the long bones to adapt in shape to the stresses placed on them. Increased sex hormone production at puberty is required for skeletal maturation, which reaches maximum mass and density in early adulthood. Recent data suggest that delayed puberty may be associated with low bone mass that persists into adulthood. The sexual dimorphism in skeletal size becomes obvious after puberty, although true bone density remains similar between the sexes. Nutrition and lifestyle also play an important role in growth, although genetic factors primarily determine peak skeletal mass and density.

Numerous genes control skeletal growth, peak bone mass, and body size, as well as skeletal structure and density. Heritability estimates of 50–80% for bone density and size have been derived on the basis of twin studies. Though peak bone mass is often lower among individuals with a family history of osteoporosis, association studies of candidate genes (vitamin D receptors; type I collagen, estrogen receptors [ERs], and interleukin 6 [IL-6]; and insulin-like growth factor I [IGF-I]) and bone mass, bone turnover, and fracture prevalence have been inconsistent. Linkage studies suggest that a genetic locus on chromosome 11 is associated with high bone mass. Families with high bone mass and without much apparent age-related bone loss have been



**FIGURE 411-3** Factors leading to osteoporotic fractures.



**FIGURE 411-4 Mechanism of bone remodeling.** The basic molecular unit (BMU) moves along the trabecular surface at a rate of ~10 µm/d. The figure depicts remodeling over ~120 days. **A.** Origination of BMU-lining cells contracts to expose collagen and attract preosteoclasts. **B.** Osteoclasts fuse into multinucleated cells that resorb a cavity. Mononuclear cells continue resorption, and preosteoblasts are stimulated to proliferate. **C.** Osteoblasts align at bottom of cavity and start forming osteoid (black). **D.** Osteoblasts continue formation and mineralization. Previous osteoid starts to mineralize (horizontal lines). **E.** Osteoblasts begin to flatten. **F.** Osteoblasts turn into lining cells; bone remodeling at initial surface (left of drawing) is now complete, but BMU is still advancing (to the right). (Adapted with permission from SM Ott, in JP Bilezikian, LG Raisz, GA Rodan: *Principles of Bone Biology*, vol. 18. San Diego, CA: Academic Press; 1996.)

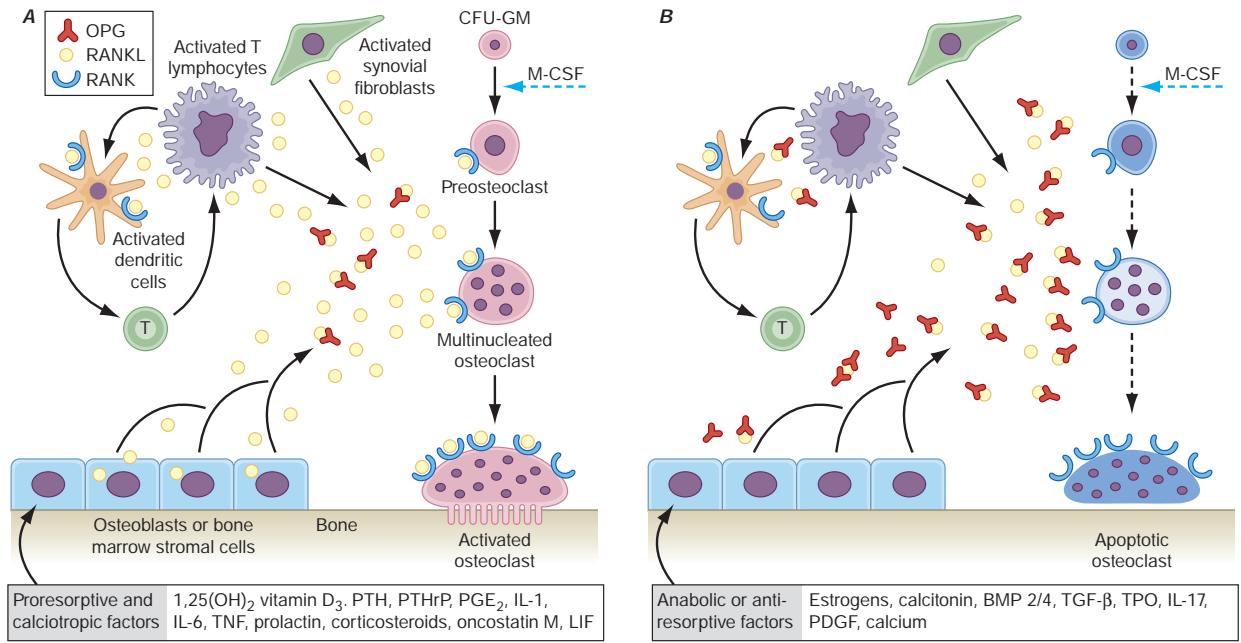
shown to have a point mutation in LRP5, a low-density lipoprotein receptor-related protein. The role of this gene in the general population is not clear, although a nonfunctional mutation results in osteoporosis-pseudoglioma syndrome, and LRP5 signaling appears to be important in controlling bone formation. Genome-wide scans for low bone mass suggest multiple genes are involved, many of which are also implicated also in control of body size.

In adults, bone *remodeling*, not modeling, is the principal metabolic skeletal process. Bone remodeling has two primary functions: (1) to repair microdamage within the skeleton to maintain skeletal strength and ensure the relative youth of the skeleton and (2) to supply calcium when needed from the skeleton to maintain serum calcium. Remodeling may be activated by microdamage to bone as a result of excessive or accumulated stress. Acute demands for calcium involve osteoclast-mediated resorption as well as calcium transport by osteocytes. Chronic demands for calcium can result in secondary hyperparathyroidism,

increased bone remodeling, and overall loss of bone tissue. Bone remodeling occurs through well-coordinated activity of osteocytes, osteoblasts, and osteoclasts. Osteocytes are the terminal-differentiated cells derived from osteoblasts after incorporation into newly formed bone tissue. Osteoblasts derive from mesenchymal cell lineage and osteoclasts from monocyte/macrophage lineage. Remodeling sites are discrete units with osteoclasts initiating the process by removal of bone tissue and osteoblasts synthesizing new organic bone that becomes gradually mineralized.

Bone remodeling also is regulated by multiple hormones, including estrogens (in both genders), androgens, vitamin D, and parathyroid hormone (PTH), as well as locally produced growth factors, such as IGF-I, transforming growth factor (TGF-), PTH-related peptide (PTHrP), interleukins (ILs), prostaglandins, and members of the tumor necrosis factor (TNF) superfamily. These factors primarily modulate the rate at which new remodeling sites are activated, a process that results initially in bone resorption by osteoclasts, followed by a period of repair during which new bone tissue is synthesized by osteoblasts (**Chap. 409**). The cytokine responsible for communication between the osteoblasts, other marrow cells, and osteoclasts is receptor activator of nuclear factor- B (RANK) ligand (RANKL). RANKL, a member of the TNF family, is secreted by osteocytes, osteoblasts, and certain cells of the immune system. The osteoclast receptor for this protein is referred to as RANK. Activation of RANK by RANKL is a final common path in osteoclast development and activation. A humoral decoy for RANKL, also secreted by osteoblasts, is referred to as *osteoprotegerin* (**Fig. 411-5**). Modulation of osteoclast recruitment and activity appears to be related to the interplay among these three factors. Additional influences include nutrition (particularly calcium intake) and physical activity level. RANKL production is in part regulated by the canonical Wnt signaling pathway. Wnt activation through mechanical loading or by hormonal or cytokine factors stimulates bone formation by increasing formation and activity of osteoblasts and decreases RANKL secretion, which inhibits production and activity of osteoclasts. Sclerostin, also an osteocyte protein, is a major inhibitor of Wnt activation and bone formation. Both the RANKL and Wnt pathways have become major targets for pharmacologic treatment of osteoporosis (see below).

In young adults, resorbed bone is replaced by an equal amount of new bone tissue. Thus, the mass of the skeleton remains constant after peak bone mass is achieved by the age of ~20 years. After age 30–45, however, the resorption and formation processes become imbalanced, and resorption exceeds formation. This imbalance may begin at different ages, varies at different skeletal sites, and becomes exaggerated in women after menopause or any other cause of estrogen deficiency. Excessive bone loss can be due to an increase in osteoclastic activity and/or a decrease in osteoblastic activity. In addition, an increase in remodeling activation frequency, and thus the number of remodeling sites, can magnify the small imbalance seen at each remodeling unit. Increased recruitment of bone remodeling sites produces a reversible reduction in bone tissue but also can result in permanent loss of tissue and disrupted skeletal architecture, with the imbalance between resorption and formation within each cycle. In trabecular bone, if the osteoclasts penetrate trabeculae, they leave no template for new bone formation to occur, and consequently, rapid bone loss ensues and cancellous connectivity becomes impaired. A higher number of remodeling sites increases the likelihood of this event. In cortical bone, increased activation of remodeling creates more porous bone. The effect of this increased porosity on cortical bone strength may be modest if the overall diameter of the bone is not changed. However, decreased apposition of new bone on the periosteal surface coupled with increased endocortical resorption of bone decreases the biomechanical strength of long bones. Even a slight exaggeration in normal bone loss increases the risk of osteoporosis-related fractures because of the architectural changes that occur, and osteoporosis is largely a disease of disordered skeletal architecture, although currently the only clinical tool generally available (dual-energy x-ray absorptiometry [DXA]) measures mass (an estimate of the mineral in bone) not architecture. Several tools are becoming available that may give more



**FIGURE 411-5** Hormonal control of bone resorption. **A.** Proresorptive and calcitropic factors. **B.** Anabolic and antiosteoclastic factors. RANKL expression is induced in osteoblasts, activated T cells, synovial fibroblasts, and bone marrow stromal cells. It binds to membrane-bound receptor RANK to promote osteoclast differentiation, activation, and survival. Conversely, osteoprotegerin (OPG) expression is induced by factors that block bone catabolism and promote anabolic effects. OPG binds and neutralizes RANKL, leading to a block in osteoclastogenesis and decreased survival of preexisting osteoclasts. CFU-GM, colony-forming units, granulocyte macrophage; IL, interleukin; LIF, leukemia inhibitory factor; M-CSF, macrophage colony-stimulating factor; OPG-L, osteoprotegerin ligand; PDGF, platelet-derived growth factor; PGE<sub>2</sub>, prostaglandin E<sub>2</sub>; PTH, parathyroid hormone; RANKL, receptor activator of nuclear factor-κB; TGF-β, transforming growth factor β; TNF, tumor necrosis factor; TPO, thrombospondin. (Reproduced with permission from WJ Boyle et al: Osteoclast differentiation and activation. *Nature* 423:337, 2003.)

insight into the architecture of the skeleton (including the trabecular bone score, a noninvasive addition to DXA).

## CALCIUM NUTRITION

Peak bone mass may be impaired by inadequate calcium intake during growth among other nutritional factors (calories, protein, and other minerals), leading to increased risk of osteoporosis later in life. During the adult phase of life, insufficient calcium intake contributes to secondary hyperparathyroidism and an increase in the rate of bone remodeling to assist in maintaining normal serum calcium levels. PTH stimulates the hydroxylation of vitamin D in the kidney, leading to increased levels of 1,25-dihydroxyvitamin D [1,25(OH)<sub>2</sub>D] and enhanced gastrointestinal calcium absorption. PTH also reduces renal calcium loss. Although these are appropriate compensatory homeostatic responses for adjusting calcium economy, the long-term effects are detrimental to the skeleton because the increased remodeling rates and the ongoing imbalance between resorption and formation at remodeling sites combine to accelerate loss of bone tissue.

Total daily calcium intakes <400 mg are detrimental to the skeleton, and intakes in the range of 600–800 mg, which is about the average intake among adults in the United States, are also probably suboptimal. The recommended daily required intake of 1000–1200 mg for adults accommodates population heterogeneity in controlling calcium balance (Chap. 332). Such intakes should preferentially come from dietary sources, with supplements used only when dietary intakes fall short and cannot be modified easily. The supplement should be enough to bring total intake to ~1200 mg/d. Recent studies have suggested that there may be differences in safety based on calcium source; high intakes primarily from supplement sources appear to result in a greater risk of renal stones and perhaps cardiovascular calcifications (although the literature is inconsistent and controversial). Increasing calcium intake above this level does not improve calcium homeostasis or bone formation. Increasing calcium intake by itself will not prevent bone loss due to other factors (e.g., postmenopausal status)

## VITAMIN D

(See also Chap. 409) Severe vitamin D deficiency causes rickets in children and osteomalacia in adults. However, vitamin D insufficiency (circulating levels of 25-hydroxyvitamin D [25(OH)D] that may be inadequate but above the level that results in rickets) may be more prevalent than previously thought, particularly among individuals at increased risk such as the elderly; those living in northern latitudes; and individuals with poor nutrition, obesity, malabsorption, or chronic liver or renal disease. Dark-skinned individuals are also at high risk of vitamin D in the insufficiency range or lower, but African Americans have a low risk of osteoporosis, with better calcium homeostasis than Caucasians.

Although there is considerable controversy about overall optimal health targets for serum 25(OH)D, there is evidence that for optimal skeletal health, serum 25(OH)D should be >75 nmol/L (30 ng/mL). To achieve this level for most adults requires skin exposure to sunlight (estimated to be exposure of face and arms for at least one-half hour each day) or an intake of at least 800–1000 units/d, or even higher in individuals with risk factors (as above).

Vitamin D insufficiency leads to compensatory secondary hyperparathyroidism and is an important risk factor for osteoporosis and fractures. Some studies have shown that >50% of inpatients on a general medical service exhibit biochemical features of vitamin D deficiency, including increased levels of PTH and alkaline phosphatase and lower levels of ionized calcium. Among those living in northern latitudes, vitamin D levels decline during the winter months without supplementation. This is associated with seasonal bone loss, reflecting increased bone turnover. Even among healthy ambulatory individuals, mild vitamin D deficiency is increasing in prevalence. In part, this is due to decreased exposure to sunlight coupled with increased use of potent sunscreens, although not all studies suggest that sunscreens inhibit D synthesis in the skin. Treatment with vitamin D can return levels to normal (>75 μmol/L [30 ng/mL]) and prevent the associated increase in bone remodeling, bone loss, and fractures. Reduced falls and fracture rates also have been documented among individuals in

northern latitudes who have greater vitamin D intake and have higher 25(OH)D levels (though one study suggested an increased fall risk with 25(OH)D levels >70 ng/mL). Although vitamin D levels are suspected to affect risk and/or severity of other diseases, including cancers (colorectal, prostate, and breast), autoimmune diseases, multiple sclerosis, cardiovascular disease, and diabetes, most controlled clinical trials have not confirmed these effects. For most adults in the United States, supplements of 1000–2000 IU/d are adequate and safe. Recent data suggesting that those with low vitamin D levels have a more severe clinical course than those with normal vitamin D levels have added impetus to ensuring that vitamin D levels are normal in all adults, even though a cause-and-effect relationship has not been demonstrated.

### ESTROGEN STATUS

Estrogen deficiency causes bone loss by two distinct but interrelated mechanisms: (1) activation of new bone remodeling sites and (2) initiation or exaggeration of an imbalance between bone formation and resorption, in favor of the latter. The change in activation frequency causes a transient bone loss until a new steady state between resorption and formation is achieved. The remodeling imbalance, however, results in a permanent decrement in mass. In addition, the very presence of more remodeling sites in the skeleton increases the probability that trabeculae will be penetrated, eliminating the template on which new bone can be formed and accelerating the loss of bony tissue. The consequence is loss of skeletal architecture, particularly in trabecular bone, and it is possible that at any given bone density the risk of a fracture is likely to be greater in those who have experienced bone loss than in those for whom that level of bone mass represents normal. Recent addition of the trabecular bone score in DXA measurements is an attempt to capture these architectural changes.

The most common estrogen-deficient state is the cessation of ovarian function at the time of menopause, which occurs on average at age 51 (Chap. 395). Thus, with current life expectancy, an average woman will spend ~30 years without an ovarian supply of estrogen. Breast cancer treatment with aromatase inhibitors is an increasingly common cause of even more severe estrogen deficiency. The mechanism by which estrogen deficiency causes bone loss is summarized in Fig. 411-5. Marrow cells (macrophages, monocytes, osteoclast precursors, mast cells) as well as bone cells (osteoblasts, osteocytes, osteoclasts) express both ERs (α and β). Loss of estrogen increases production of RANKL and reduces production of osteoprotegerin, increasing osteoclast formation and recruitment. Estrogen also may play a role in determining the life span of bone cells by controlling the rate of apoptosis. Thus, in situations of estrogen deprivation, the life span of osteoblasts may be decreased, whereas the longevity and activity of osteoclasts are increased. The rate and duration of bone loss after menopause are heterogeneous and unpredictable. Once surfaces are lost in cancellous bone, the rate of bone loss declines. In cortical bone, loss is slower but may continue for a longer time period.

Since remodeling is initiated at the surface of bone, it follows that trabecular bone—which has a considerably larger surface area (80% of the total) than cortical bone—will be affected preferentially by estrogen deficiency. Fractures occur earliest at sites where trabecular bone contributes most to bone strength; consequently, vertebral fractures are the most common early skeletal consequence of estrogen deficiency.

In males, estrogen may have an important role in regulation of bone remodeling. In an experiment in which males were rendered estrogen and androgen deficient, restoring estrogen supply reduced remodeling rate more than restoring androgen.

### PHYSICAL ACTIVITY

Inactivity, such as prolonged bed rest or paralysis, results in significant bone loss. Concordantly, athletes have higher bone mass than non-athletes. These changes in skeletal mass are most marked when the stimulus begins during growth and before the age of puberty. Adults are less capable than children of increasing bone mass after restoration of physical activity. Epidemiologic data support the beneficial effects on the skeleton of chronic high levels of physical activity. Fracture risk is lower in rural communities and in countries where physical activity

is maintained into old age. However, when exercise is initiated during adult life, the effects of moderate exercise on the skeleton are modest, with a bone mass increase of 1–2% in short-term studies of <2 years' duration. It is argued that more active individuals are less likely to fall and are more capable of protecting themselves upon falling, thereby reducing fracture risk. Continuing physical activity into the later years appears to slow cognitive decline, another major reason for including exercise programs for the aging population.

### CHRONIC DISEASES

Various genetic and acquired diseases are associated with an increase in the risk of osteoporosis (Table 411-2). Mechanisms that contribute to bone loss are unique for each disease and typically result from multiple factors, including nutrition, reduced physical activity levels, and factors that affect rates of bone remodeling. In most, but not all circumstances, the primary diagnosis is made before osteoporosis presents clinically. Both type 1 and type 2 diabetes mellitus are associated with an increased fracture risk, with increased risk at higher bone density than in the nondiabetic population. This may be due to differences in the chemical composition of bone tissue that is more brittle than normal, a predilection for conversion of precursors to adipose cells rather than osteoblasts, and the sequelae of diabetes that increase the risk of falls and injury.

Severe bone loss occurs in quadriplegic and paraplegic individuals below the level of the injury. The combination of loss of muscle function and innervation of both muscle and bone contributes to failure to recover mobility, which leads to a high fracture risk in those attempting to pursue athletic activities despite their primary diagnosis (e.g., wheelchair athletes). Bone loss also follows a stroke and is again dependent on the severity of the paralysis. The risk of fracture can be predicted by the FRAX (Fracture Risk Assessment) score and seems highest in the first year after stroke diagnosis. The increasing prevalence of transgender and gender nonconforming individuals has prompted a guideline for evaluation of bone density in that population by the International Society of Clinical Densitometry published in 2019.

**TABLE 411-2 Diseases Associated with an Increased Risk of Generalized Osteoporosis in Adults**

Hypogonadal states	Hematologic disorders/malignancy
Turner's syndrome	Multiple myeloma
Klinefelter's syndrome	Lymphoma and leukemia
Anorexia nervosa	Malignancy-associated parathyroid hormone-related peptide (PTHrP) production
Hypothalamic amenorrhea	Mastocytosis
Hyperprolactinemia	Hemophilia
Other primary or secondary hypogonadal states	Thalassemia
Endocrine disorders	Selected inherited disorders
Cushing's syndrome	Osteogenesis imperfecta
Hyperparathyroidism	Marfan's syndrome
Thyrotoxicosis	Hemochromatosis
Diabetes mellitus (both type 1 and 2)	Hypophosphatasia
Acronegaly	Glycogen storage diseases
Adrenal insufficiency	Homocystinuria
Nutritional and gastrointestinal disorders	Ehlers-Danlos syndrome
Malnutrition	Porphyria
Parenteral nutrition	Menkes' syndrome
Malabsorption syndromes	Epidermolysis bullosa
Gastrectomy	Other disorders
Severe liver disease, especially biliary cirrhosis	Immobilization
Pernicious anemia	Chronic obstructive pulmonary disease
Rheumatologic disorders	Pregnancy and lactation
Rheumatoid arthritis	Scoliosis
Ankylosing spondylitis	Multiple sclerosis
	Sarcoidosis
	Amyloidosis

**TABLE 411-3 Drugs Associated with an Increased Risk of Generalized Osteoporosis in Adults**

Glucocorticoids	Excessive thyroxine
Cyclosporine	Aluminum
Cytotoxic drugs	Gonadotropin-releasing hormone agonists
Anticonvulsants	Heparin
Aromatase inhibitors	Lithium
Selective serotonin reuptake inhibitors	Protein pump inhibitors
	Thiazolidinediones
	Androgen deprivation therapies

## MEDICATIONS

A large number of medications used in clinical practice have potentially detrimental effects on the skeleton (**Table 411-3**). *Glucocorticoids* are the most common cause of medication-induced osteoporosis. It is often not possible to determine the extent to which osteoporosis is related to glucocorticoid treatment or to other factors, as the effects of medication are superimposed on the effects of the primary disease, which in itself may be associated with bone loss (e.g., rheumatoid arthritis). Excessive doses of thyroid hormone can accelerate bone remodeling and result in bone loss.

Other medications have less detrimental effects on the skeleton than pharmacologic doses of glucocorticoids. *Anticonvulsants* are thought to increase the risk of osteoporosis, although many affected individuals have concomitant insufficiency of  $1,25(\text{OH})_2\text{D}$ , as some anticonvulsants induce the cytochrome P450 system and vitamin D metabolism. Patients undergoing transplantation are at high risk for rapid bone loss and fracture not only from glucocorticoids but also from treatment with other *immunosuppressants* such as cyclosporine and tacrolimus (FK506). In addition, these patients often have underlying metabolic abnormalities such as hepatic or renal failure that predispose to bone loss. Recently, long-term use of proton pump inhibitors has been shown in observational studies to be associated with a higher risk of fracture. Given their widespread and frequent long-term use, the skeletal effect is important from a public health perspective and when reviewing risk for fracture in individuals.

*Aromatase inhibitors*, which potently block the aromatase enzyme that converts androgens and other adrenal precursors to estrogen, reduce circulating postmenopausal estrogen supply dramatically. These agents, which are used in various stages for breast cancer treatment, also have been shown to have a detrimental effect on bone density and risk of fracture. Androgen deprivation therapies, used to treat men with prostate cancer, also result in rapid loss of bone and increased fracture risk. Various diabetes medications, including but not limited to thiazolidinediones, and antidepressants, including the selective serotonin reuptake inhibitors, increase risk of osteoporosis and fracture. It is difficult in some cases to separate the risk accrued by the underlying disease from that attributable to the medication. Thus, both depression and diabetes are risk factors for fracture by themselves.

## SMOKING

Smoking produces detrimental effects on bone mass mediated directly by toxic effects on osteoblasts or indirectly by modifying estrogen metabolism. On average, cigarette smokers reach menopause 1–2 years earlier than the general population. Cigarette smoking also produces secondary effects that can modulate skeletal status, including intercurrent respiratory and other illnesses, frailty, decreased exercise, poor nutrition, and the need for additional medications (e.g., glucocorticoids for lung disease).

## OTHER POTENTIAL FACTORS

In the past few years, a large number of potential risk factors for fracture have been identified. These include excessive alcohol intake and other drugs of abuse, pollution, use of triclosan, chronic obstructive pulmonary disease, excess vitamin B, and hormonal therapies utilized among the transgender population.

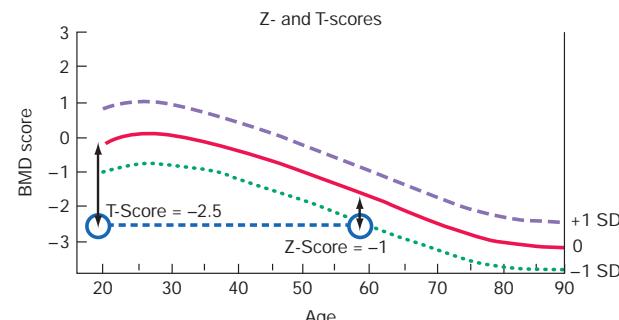
## DIAGNOSIS

### MEASUREMENT OF BONE MASS

Several noninvasive techniques are available for estimating skeletal mass or BMD. They include single-energy x-ray absorptiometry (SXA), DXA, quantitative computed tomography (CT), and ultrasound. DXA is a highly accurate x-ray technique that has become the standard for measuring bone density. Though it can be used for measurement in any skeletal site, clinical determinations usually are made of the lumbar spine and hip. DXA also can be used to measure the wrist total body bone mass and body composition. Two x-ray energies are used to estimate mineralized tissue, allowing for correction for attenuation through soft tissue. The mineral content is divided by bone area, which partially corrects for body and bone size. However, this correction is only partial since DXA is a two-dimensional scanning technique and cannot estimate the depth or posteroanterior length of the bone. Thus, small slim people tend to have lower than average BMD, a feature that is important in interpreting BMD measurements. Bone spurs, which are common in osteoarthritis, tend to falsely increase bone density mostly of the spine and are a particular problem in measuring spine BMD in older individuals. Because DXA measurement devices are provided by two different manufacturers, the output varies in absolute terms. Consequently, it has become standard practice to relate the results to “normal” values by using T-scores (a T-score of 1 equals 1 SD), which compare individual results to those in a young adult population that is matched for race and sex. The mean value is given a score of zero and the range +2.5 to -2.5 (i.e., 2.5 SDs above or below the mean). Z-scores (also SDs) compare individual results to those of an age and gender-matched reference population. Thus, a 60-year-old woman with a Z-score of -1 (1 SD below mean for age) has a T-score of -2.5 (2.5 SD below mean for a young control group) (**Fig. 411-6**). A T-score <-2.5 in the lumbar spine, femoral neck, or total hip has been defined as osteoporosis. Although the outputs from different instruments and, more importantly, different manufacturers correlate well, different machines used over time may show changes in BMD that may be attributed to biological changes or simply the result of differences between machines. This is particularly true with measurements of the hip. Consequently, it is recommended that serial measurements be performed on the same machine and preferably by the same technician.

As noted above, since >50% of fractures occur in individuals with low bone mass (i.e., a T-score between -1.0 and -2.5), it is usual to report fracture risk in addition to BMD. To that end the absolute fracture risk assessment tool FRAX often accompanies the report of bone density. FRAX estimates include age, gender, height, weight, fracture history, hip fracture in a parent, steroid use, rheumatoid arthritis, other secondary causes, and bone density of the femoral neck. The program then calculates the estimated risk over a 10-year time frame for major osteoporosis-related fractures (clinical spine, hip, wrist, and proximal humerus) as well as hip fracture.

CT can also be used to measure the spine and hip but is rarely used clinically, in part because the radiation exposure and cost are both



**FIGURE 411-6** Relationship between Z-scores and T-scores in a 60-year-old woman. BMD, bone mineral density; SD, standard deviation.

much higher than with DXA. High-resolution peripheral quantitative computed tomography (HR-pQCT) can be used to measure bone in the forearm or tibia and is a research tool that provides information on skeletal architecture noninvasively. Magnetic resonance imaging (MRI) can also be used to obtain some architectural information on the forearm and perhaps the hip but, again, is primarily a research tool at present.

Ultrasound can be used to measure bone mass by calculating the attenuation of the signal as it passes through bone or the speed with which it traverses the bone. Although the ultrasound technique was purported to assess properties of bone other than mass (e.g., quality), this has not been confirmed. Because of its relatively low cost and mobility, ultrasound bone density measurement is amenable for use as a screening procedure in stores or health fairs.

All these techniques for measuring BMD have been approved by the U.S. Food and Drug Administration (FDA) on the basis of their capacity to predict fracture risk. The hip is the preferred site of measurement in most individuals, since it allows prediction of hip fracture risk, the most important consequence of osteoporosis, better than any other bone density measurement site. When hip measurements are performed by DXA, the spine is usually measured at the same time. In younger individuals such as perimenopausal or early postmenopausal women, spine measurements may be a more sensitive indicator of bone loss. When the spine or hip is not measurable due to severe degenerative spine disease or scoliosis or prior spine or hip surgery, BMD of the wrist is often measured.

### INDICATIONS FOR BONE MASS MEASUREMENT

Several clinical guidelines have been developed for the use of bone densitometry in clinical practice (**Table 411-4**). The National Osteoporosis Foundation (NOF) guidelines recommend bone mass measurements in postmenopausal women who have one or more risk factors for osteoporosis in addition to age, sex, and estrogen deficiency. The guidelines further recommend that bone mass measurement be considered in all women by age 65, a position ratified by the U.S. Preventive Health Services Task Force. In males, the use of bone density determination is not recommended until the age of 70 years in the absence of multiple risk factors or the occurrence of an osteoporosis-related fracture.

The FRAX score incorporates risk factors (age, prior fracture, family history of hip fracture, low body weight, cigarette consumption, excessive alcohol use, steroid use, and rheumatoid arthritis) with BMD to assess the 10-year fracture probabilities. Fracture risk probability calculators are available as part of the report from all DXA machines and also available online (<https://www.sheffield.ac.uk/FRAX/>) (**Fig. 411-7**). In the United States, it has been determined to be cost effective to treat if the 10-year major osteoporotic fracture risk from FRAX is 20% and/or the 10-year risk of hip fracture is 3%. FRAX is an imperfect tool, as it does not include any assessment of fall risk, and secondary causes are excluded when BMD is entered. More importantly, it does not distinguish the contribution toward future fracture probability from an acute recent fracture versus the lesser importance of the more remote fracture. Moreover, there is no mandate for vertebral fracture diagnosis and no additional fracture probability estimated for patients who have had multiple fractures. Nonetheless, it is useful as an educational tool for patients, particularly for those who are excessively worried about BMD levels despite relative youth and health.

**TABLE 411-4 Indications for Bone Mineral Density Testing**

- Women aged 65 and men aged 70; regardless of clinical risk factors
- Younger postmenopausal women, women in the menopausal transition, and men aged from 50 to 69 with clinical risk factors for fracture
- Adults who have a fracture at or after age 50
- Adults with a condition (e.g., rheumatoid arthritis) or taking a medication (e.g., glucocorticoids at a daily dose >5 mg prednisone or equivalent for >3 months associated with low bone mass or bone loss

### VERTEBRAL IMAGING

DXA equipment can also be used to obtain lateral images of the thoracic and lumbar spine, a technique called vertebral fracture assessment (VFA). While not as definitive as a radiograph, it is an excellent screening tool for vertebral abnormality in both women and men based on age and BMD even in the absence of any specific symptoms since the majority of vertebral fractures are asymptomatic for a long time. Furthermore, the VFA can be used to evaluate vertebral abnormality as a cause of height loss or back pain that suggests the possibility of an undiagnosed vertebral fracture.

Because vertebral fractures are often asymptomatic when they first occur, the diagnosis of vertebral fracture is rarely made at the time of the fracture occurrence. Since vertebral fractures, whether symptomatic or asymptomatic, are associated with the same clinical sequelae, it is critical that patients with these fractures are identified. Vertebral fracture prevalence in the United States based on the National Health and Nutrition Evaluation Studies (NHANES) population appears to be ~10% in the 1970s and 20% in the 1980s, when the strictest criteria for diagnosis are utilized. The NOF and other organizations have recommended that women by the age of 65 and men by the age of 70 undergo vertebral imaging if a T-score is -1.5 at the spine, hip, or femoral neck. Vertebral imaging is also recommended for women by the age of 70 and men by the age of 80 if a T-score is <-1.0. For younger individuals, vertebral imaging is recommended for those with an osteoporosis related fracture, height loss, or glucocorticoid use. (**See Table 411-5.**)

### APPROACH TO THE PATIENT

#### Osteoporosis

The development of underlying skeletal changes is a gradual process occurring under a variety of influences throughout adult life. Recognition of these influences allows intervention at several points, although the need for aggressive management is clearly dependent on a careful evaluation of each individual patient.

The menopausal transition affects all women by their late 50s and represents an opportunity to initiate a discussion about bone loss, the role of estrogen loss, and other risk factors that might exacerbate it. Assessment of fracture risk using a tool such as FRAX (with or without bone density) provides a 10-year estimate of hip and major osteoporosis-related fracture risk and opens the discussion about preventive steps including, if required, the use of medication. If risk is low, then nutrition and lifestyle are the focus. If menopausal symptoms are prominent and estrogen intervention is needed, then the added protection against bone loss should be emphasized.

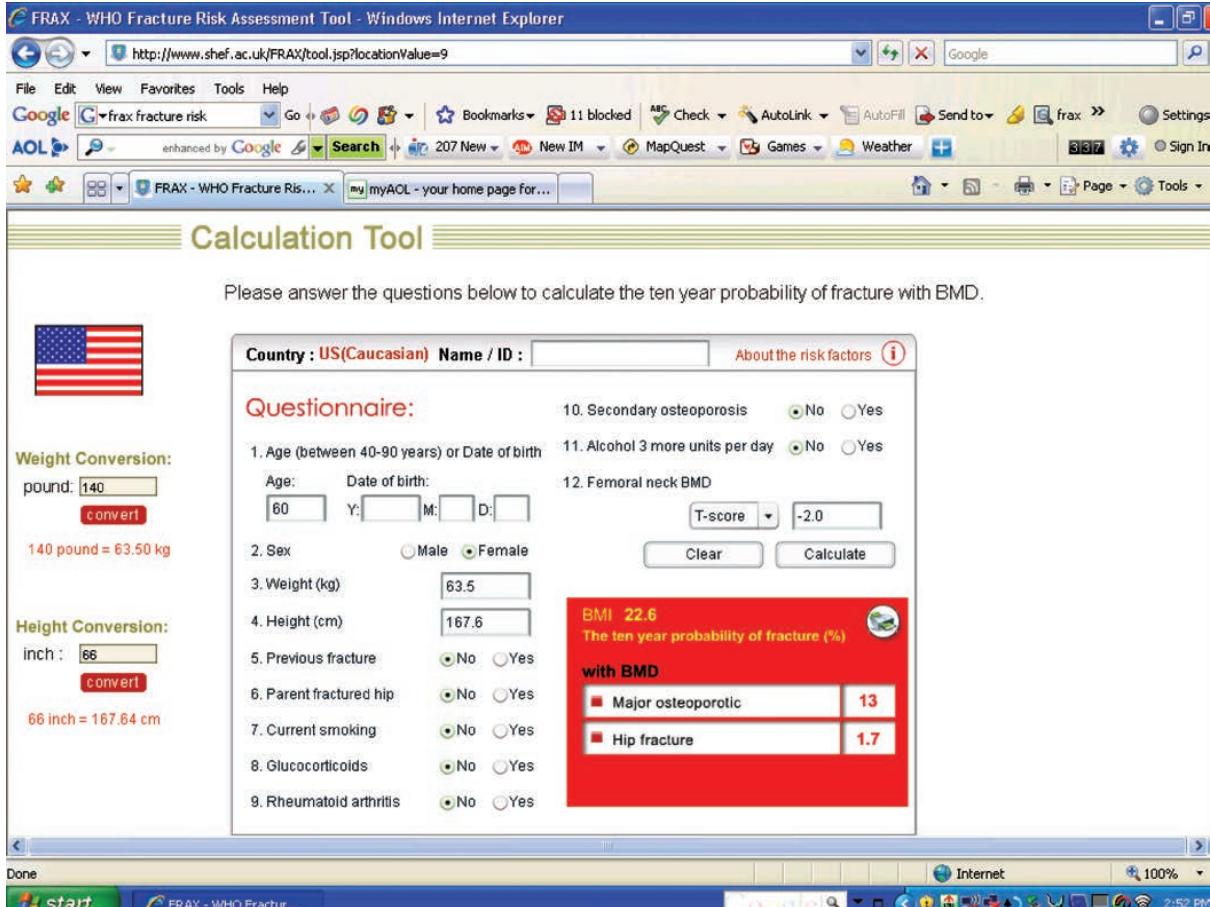
Among older women, the occurrence of a fracture should precipitate an evaluation of skeletal status including bone density testing. In this case, any fracture, whether traumatic or not, should trigger the assessment. Although osteoporosis is associated with a risk of fracture on minimal trauma, individuals with osteoporosis are consequently more likely to fracture at greater levels of trauma, and such individuals should not be excluded from osteoporosis evaluation simply because of the level of trauma. This concept, while obvious, still needs emphasis with individual patients, physicians, and payors.

Patients who present with hip or spine fractures by definition have osteoporosis and will require treatment for both the fracture itself and the underlying skeletal disorder. Other long bone fractures (e.g., distal radius) are triggers for evaluation of the skeleton upon which treatment decisions can be based.

In all individuals presenting with a fracture as a result of a fall, fall prevention strategies are an important adjunct to other lifestyle and nutritional interventions that must be reviewed with all patients.

#### ROUTINE LABORATORY EVALUATION

There is no established algorithm for the evaluation of women who present with osteoporosis. A general evaluation that includes



**FIGURE 411-7** FRAX calculation tool. When the answers to the indicated questions are filled in, the calculator can be used to assess the 10-year probability of fracture. The calculator (available online at <http://www.shef.ac.uk/FRAX/tool.jsp?locationValue=9>) also can risk adjust for various ethnic groups.

complete blood count, serum and 24-h urine calcium, and renal and hepatic function tests is useful for identifying selected secondary causes of low bone mass, particularly for women with fractures or unexpectedly low Z-scores. An elevated serum calcium level suggests hyperparathyroidism or malignancy, whereas a reduced serum calcium level may reflect malnutrition or a malabsorption disease, such as celiac disease. In the presence of hypercalcemia, a serum PTH level differentiates between hyperparathyroidism (PTH↑) and malignancy (PTH↓), and a high PTHrP level can help document the presence of humoral hypercalcemia of malignancy (Chap. 410). A low urine calcium (<50 mg/24 h) suggests malnutrition, or

malabsorption; a high urine calcium (>300 mg/24 h) during normal calcium intake (excluding calcium supplements for at least a week before the urine collection) is indicative of hypercalciuria. Hypercalciuria occurs primarily in three situations: (1) a renal calcium leak, which is more common in males with osteoporosis; (2) absorptive hypercalciuria, which can be idiopathic or associated with increased 1,25(OH)<sub>2</sub>D in granulomatous disease; or (3) hematologic malignancies or conditions associated with excessive bone turnover such as Paget's disease, hyperparathyroidism, and hyperthyroidism. Renal hypercalciuria is treated with thiazide diuretics, which lower urine calcium and help improve calcium economy. In this setting, thiazides alone can improve bone mass and possibly reduce risk of fracture. They might also reduce renal stone risk.

Individuals who have osteoporosis-related fractures or bone density in the osteoporotic range should have a measurement of serum 25(OH)D level since the intake of vitamin D required to achieve a target level >30 ng/mL is highly variable. Hyperthyroidism should be evaluated by measuring thyroid-stimulating hormone (TSH).

When there is clinical suspicion of Cushing's syndrome, urinary free cortisol levels or a fasting serum cortisol should be measured after overnight dexamethasone. When bowel disease, malabsorption, or malnutrition is suspected, serum albumin, cholesterol, and a complete blood count should be checked. Asymptomatic malabsorption may be heralded by anemia (macrocytic—vitamin B<sub>12</sub> or folate deficiency; microcytic—iron deficiency) or low serum cholesterol or urinary calcium levels. If these or other features suggest malabsorption, further evaluation is required. Asymptomatic celiac

#### TABLE 411-5 Indications for Vertebral Testing

Consider vertebral imaging tests for the following individuals<sup>a</sup>

- All women aged 70 and all men aged 80 if bone mineral density (BMD) T-score at the spine, total hip, or femoral neck is <1.0
- Women aged from 65 to 69 and men aged from 70 to 79 if BMD T-score at the spine, total hip, or femoral neck is <1.5
- Postmenopausal women and men aged 50 with specific risk factors:
  - Low-trauma fracture during adulthood (aged 50)
  - Historical height loss of 1.5 in. (4 cm)<sup>b</sup>
  - Prospective height loss of 0.8 in. (2 cm)<sup>c</sup>
  - Recent or ongoing long-term glucocorticoid treatment

<sup>a</sup>If bone density testing is not available, vertebral imaging may be considered based on age alone. <sup>b</sup>Current height compared to peak height during childhood. <sup>c</sup>Cumulative height loss measured during interval medical assessment.

disease with selective malabsorption is being found with increasing frequency; the diagnosis can be made by testing for transglutaminase IgA antibodies but may require confirmation by endoscopic biopsy. A trial of a gluten-free diet can also be confirmatory (*Chap. 325*). When osteoporosis is found associated with symptoms of rash, multiple allergies, diarrhea, or flushing, mastocytosis should be considered and excluded by using 24-h urine histamine collection or serum tryptase.

Myeloma can masquerade as generalized osteoporosis, although it more commonly presents with bone pain and characteristic “punched-out” lesions on radiography. Serum and urine electrophoresis and/or evaluation for serum free light chains in urine are required to exclude this diagnosis. More commonly, a monoclonal gammopathy of undetermined significance (MGUS) is found, and the patient is subsequently monitored to ensure that this is not an incipient myeloma. MGUS itself may be associated with an increased risk of osteoporosis. A bone marrow biopsy may be required to rule out myeloma (in patients with equivocal electrophoretic results) and also can be used to exclude mastocytosis, leukemia, and other marrow infiltrative disorders such as Gaucher's disease.

An important cause of fracture among the aging population is diabetes, both type 1 and type 2. Patients with diabetes appear, at any given bone density, to be at higher risk of fracture than nondiabetics. The reasons include the effects on muscle and nerve that increase the risk of falls, but also the possibility that there is an underlying skeletal fragility as part of the metabolic consequences of diabetes itself.

#### BONE BIOPSY

Tetracycline labeling of the skeleton allows determination of the rate of remodeling as well as evaluation for other metabolic bone diseases. The current use of BMD tests, in combination with hormonal evaluation and biochemical markers of bone remodeling, has largely replaced the clinical use of bone biopsy, although it remains an important tool in the diagnosis of chronic kidney disease–mineral bone disease (CKD-MBD), in evaluating the mechanism of action of osteoporosis pharmacologies, and in clinical research.

#### BIOCHEMICAL MARKERS

Several biochemical tests are available that provide an index of the overall rate of bone remodeling (*Table 411-6*). Biochemical markers usually are characterized as those related primarily to *bone formation* or *bone resorption*. These tests measure the overall state of bone remodeling at a single point in time. Clinical use of these tests has been hampered by biologic variability (in part related to circadian rhythm) as well as analytic variability, although the latter is improving.

For the most part, remodeling markers do not predict rates of bone loss well enough in individuals to make accurate assessment of potential future changes in bone density. However, they do provide adjunct information that assists in both evaluation of the patient and in assessment of treatment response. Markers of bone resorption may help in the prediction of fracture risk, independently of bone density, particularly in older individuals. In women

65 years, when bone density results are greater than the usual treatment thresholds noted above, a high level of bone resorption

should prompt consideration of treatment. The primary use of biochemical markers is for monitoring the response to treatment. With the introduction of antiresorptive therapeutic agents, bone remodeling declines rapidly, with the fall in resorption occurring earlier than the fall in formation. Inhibition of bone resorption is maximal within 3 months or so. Thus, measurement of bone resorption (serum C-terminal telopeptide measured in a fasting specimen is the preferred marker) before initiating therapy and 2–6 months after starting therapy provides an earlier estimate of patient response than does bone densitometry. A decline in resorative markers can be ascertained after treatment with bisphosphonates, denosumab, or estrogen; this effect is less marked after treatment with weaker agents such as raloxifene or calcitonin. Bone turnover markers are also useful in monitoring the effects of 1–34hPTH, or teriparatide, which rapidly increases bone formation (P1NP is the most sensitive, but osteocalcin is also a very good formation marker) and later bone resorption. The recent suggestion of “drug holidays” (see below) has opened another use for biochemical markers, allowing evaluation of the off-effect of drugs such as bisphosphonates.

## TREATMENT

### Osteoporosis

#### MANAGEMENT OF PATIENTS WITH FRACTURES

Treatment of a patient with osteoporosis frequently involves management of acute fractures as well as treatment of the underlying disease. Hip fractures almost always require surgical repair if the patient is to become ambulatory again. Depending on the location and severity of the fracture, condition of the neighboring joint, and general status of the patient, procedures may include open reduction and internal fixation with pins and plates, hemiarthroplasties, and total arthroplasties. These surgical procedures are followed by intense rehabilitation in an attempt to return patients to their pre-fracture functional level. Long bone fractures often require either external or internal fixation. Other fractures (e.g., vertebral, rib, and pelvic fractures) can often be managed with supportive care, requiring no specific orthopedic treatment.

Only ~25–30% of vertebral compression fractures present with sudden-onset back pain. For acutely symptomatic fractures, treatment with analgesics is required, including nonsteroidal anti-inflammatory agents and/or acetaminophen, sometimes with the addition of a narcotic agent. (A few small, randomized clinical trials suggest that calcitonin may reduce pain related to acute vertebral compression fracture). A technique that involves percutaneous injection of artificial cement (polymethylmethacrylate) into the vertebral body (vertebroplasty or kyphoplasty) may offer significant pain relief in some patients; however, controlled trials of these procedures have provided some doubt of their efficacy in the longer term. Furthermore, risks include acute extravasation of cement outside of the vertebral body with neurologic impairment and possibly an increased risk of vertebral fracture in adjacent vertebrae due to increased rigidity of the treated vertebral body. Short periods of bed rest may be helpful for pain management, but in general, early mobilization is recommended as it helps prevent further bone loss associated with immobilization. Occasionally, use of a soft elastic-style brace may facilitate earlier mobilization. Muscle spasms often occur with acute compression fractures and can be treated with muscle relaxants and heat treatments. Severe pain usually resolves within 6–10 weeks. More chronic severe pain might suggest the possibility of multiple myeloma or other underlying conditions.

Vertebral fractures cause height loss because of the loss of vertebral body height during compression of the vertebral body. These fractures can produce kyphotic posture, particularly when wedge shaped, or just loss of thoracic height. Chronic pain following vertebral fracture is probably not bony in origin; instead,

**TABLE 411-6 Biochemical Markers of Bone Metabolism in Clinical Use**

#### Bone formation

Serum bone-specific alkaline phosphatase

Serum osteocalcin

Serum propeptide of type I procollagen

#### Bone resorption

Urine and serum cross-linked N-telopeptide

Urine and serum cross-linked C-telopeptide

it is related to abnormal strain on muscles, ligaments, and tendons and to secondary facet-joint arthritis associated with alterations in thoracic and/or abdominal shape. Chronic pain may also be the result of ribs sitting right on top of the iliac crest bones, particularly in patients who have had multiple vertebral compression fractures. Chronic pain is difficult to treat effectively and may require analgesics, sometimes including narcotic analgesics with the attendant risk of addiction. Frequent intermittent rest in a supine or semireclining position is often required to allow the soft tissues, which are under tension, to relax. Back and core-strengthening exercises may be beneficial. Heat treatments help relax muscles and reduce the muscular component of discomfort. Various physical modalities, such as ultrasound and transcutaneous nerve stimulation, may be beneficial in some patients. Pain also occurs in the neck region, not as a result of compression fractures (which almost never occur in the cervical spine as a result of osteoporosis) but because of chronic strain associated with trying to elevate the head in a person with a significant thoracic kyphosis.

Multiple vertebral fractures often are associated with psychological symptoms; this is not always appreciated. The changes in body configuration and back pain can lead to marked loss of self-image and a secondary depression. Altered balance, precipitated by the kyphosis and the anterior movement of the body's center of gravity, leads to a fear of falling, a consequent tendency to remain indoors, and the onset of social isolation. These symptoms sometimes can be alleviated by family support and/or psychotherapy. Medication may be necessary when depressive features are present.

Multiple studies show that patients presenting with fractures after age 50 years (even fractures traditionally linked to osteoporosis) are largely not screened or treated for osteoporosis. Estimates suggest that <25% of fracture patients receive follow-up care. Recently, several studies have demonstrated the effectiveness of a relatively simple and inexpensive program that reduces the risk of subsequent fractures. In the Kaiser system, it is estimated that a 20% decline in hip fracture occurrence was seen with the introduction of a fracture liaison service. This approach has also been successful in other non-U.S. health systems. This involves a health care professional (usually a nurse or physician's assistant) whose job is to educate patients and coordinate evaluation and osteoporosis treatment as patients move through the emergency room, inpatient care in an acute care hospital, rehabilitation hospital care, and/or orthopedic practice to outpatient management. If the Kaiser experience can be repeated, there would not only be significant savings of health care dollars but also a dramatic drop in hip fracture incidence and a marked improvement in morbidity and mortality among the aging population.

#### MANAGEMENT OF THE UNDERLYING DISEASE

**Risk Factor Reduction** After risk assessment, patients should be thoroughly educated to reduce the impact of modifiable risk factors associated with bone loss and falling. Medications should be reviewed to ensure that all are necessary and taken at the lowest required dose. Glucocorticoid medication, if present, should be evaluated to determine that it is truly indicated and is being given in doses that are as low as possible. For those on thyroid hormone replacement, TSH testing should be performed to determine that an excessive dose is not being used, as iatrogenic thyrotoxicosis can be associated with increased bone loss. In patients who smoke, efforts should be made to facilitate smoking cessation. Reducing risk factors for falling also includes alcohol abuse treatment and a review of the medical regimen for any drugs that might be associated with orthostatic hypotension and/or sedation, including hypnotics and anxiolytics. If nocturia occurs, the frequency should be reduced, if possible (e.g., by decreasing or modifying diuretic use), as arising in the middle of sleep is a common precipitant of a fall. Patients should be instructed about environmental safety with regard to eliminating exposed wires, curtain strings, slippery rugs, and mobile tables. Avoiding stocking feet on wood floors, checking

TABLE 411-7 Adequate Calcium Intake

LIFE STAGE GROUP	ESTIMATED ADEQUATE DAILY CALCIUM INTAKE, mg/d
Young children (1–3 years)	500
Older children (4–8 years)	800
Adolescents and young adults (9–18 years)	1300
Men and women (19–50 years)	1000
Men and women (51 years and older)	1200

Note: Pregnancy and lactation needs are the same as for nonpregnant women (e.g., 1300 mg/d for adolescents/young adults and 1000 mg/d for those  $\geq$  19 years old).

Source: Data from Institute of Medicine. Dietary Reference Intakes for Calcium, Phosphorus, Magnesium, Vitamin D, and Fluoride. Washington, DC: The National Academies Press; 1997.

carpet condition (particularly on stairs), and providing good light in paths to bathrooms and outside the home are important preventive measures. Treatment for impaired vision is recommended, particularly a problem with depth perception, which is specifically associated with increased falling risk. Elderly patients with neurologic impairment (e.g., stroke, Parkinson's disease, Alzheimer's disease) are particularly at risk of falling and require specialized supervision and care. In patients with risk factors for falls, especially those who live alone or spend significant time alone, medical alert systems should be prescribed.

**Nutritional Recommendations • Calcium** A large body of data indicates that less than optimal calcium intake results in bone loss. Consequently, an adequate intake suppresses bone turnover. Recommended intakes from an Institute of Medicine report are shown in Table 411-7. The NHANES have consistently documented that average calcium intakes fall considerably short of these recommendations. The preferred source of calcium is diet, but many patients require calcium supplementation to bring intake to  $\sim$ 1000 mg/d. Best sources of calcium include dairy products (milk, yogurt, and cheese), nondairy milks (almond, rice, soy), and fortified foods such as certain cereals, waffles, snacks, juices, and crackers. Some of these fortified foods contain as much calcium per serving as milk. Various vegetables and fruits, such as kale, broccoli, and dried figs, contain reasonably high calcium content, although some of it may not be fully bioavailable. Calcium intake calculators are available at NOF.org or NYSOPEP.org and will give a rough idea of total calcium intake.

If calcium supplements are required, they should be taken in doses sufficient to bring total intake to the required level ( $\sim$ 1000 mg/d). Doses of supplements should be 600 mg per single dose, as the calcium absorption fraction decreases at higher doses. Calcium supplements should be calculated on the basis of the elemental calcium content of the supplement, not the weight of the calcium salt (Table 411-8). Calcium supplements containing carbonate are best taken with food since they require acid for solubility. Calcium citrate supplements can be taken at any time. To confirm

TABLE 411-8 Elemental Calcium Content of Various Oral Calcium Preparations

CALCIUM PREPARATION	ELEMENTAL CALCIUM CONTENT
Calcium citrate	60 mg/300 mg
Calcium lactate	80 mg/600 mg
Calcium gluconate	40 mg/500 mg
Calcium carbonate	400 mg/g
Calcium carbonate + 5 µg vitamin D <sub>3</sub> (OsCal 250)	250 mg/tablet
Calcium carbonate (Tums 500)	500 mg/tablet

Source: Adapted with permission from SM Krane, MF Holick, in *Harrison's Principles of Internal Medicine*, 14th ed. New York, NY: McGraw Hill; 1998.

bioavailability, calcium supplements can be placed in distilled vinegar. They should dissolve within 30 min.

Several controlled clinical trials of calcium, mostly with accompanying vitamin D, have confirmed reductions in clinical fractures, including fractures of the hip (~20–30% risk reduction), particularly in elderly individuals who are more likely to be dietarily deficient. All recent studies of pharmacologic agents have been conducted in the context of calcium replacement ( $\pm$  vitamin D). Thus, it is standard practice to ensure an adequate calcium and vitamin D intake in patients with osteoporosis whether they are receiving additional pharmacologic therapy or not. A systematic review confirmed a greater BMD response to antiresorptive therapy when calcium intake was adequate.

Although side effects from supplemental calcium are minimal (eructation and constipation mostly with carbonate salts), individuals with a history of kidney stones should have a 24-h urine calcium determination before starting increased calcium to avoid exacerbating hypercalcaturia. A recent analysis of published data has suggested that high intakes of calcium from supplements are associated with an increase in the risk of renal stones, calcification in arteries, and potentially an increased risk of heart disease and stroke. This is an evolving story with data both confirming and refuting the finding. Since high calcium intakes also increase the risk of renal stones and confer no extra benefit to the skeleton, the recommendation that total intakes should be between 1000 and 1500 mg/d seems reasonable.

**Vitamin D** Diet alone rarely contains sufficient vitamin D to maintain target circulating levels (serum 25[OH]D consistently  $>75 \mu\text{mol/L}$  [30 ng/mL]). Vitamin D is synthesized from a precursor in the skin under the influence of heat and ultraviolet light (Chap. 409). Production is blocked by sunscreen and sun avoidance. However, large segments of the population do not obtain sufficient vitamin D from either skin production or dietary sources. Since vitamin D supplementation at doses that would achieve these serum levels is safe and inexpensive, the National Academy of Medicine (formerly, Institute of Medicine [IOM]) recommends daily intakes of 200 IU for adults  $<50$  years of age, 400 IU for those 50–70 years, and 600 IU for those  $>70$  years (based on obtaining a serum level of 20 ng/mL, lower than the level recommended by most other guidelines). Multivitamin tablets usually contain 400 IU, and many calcium supplements also contain vitamin D. Some data suggest that higher doses (1000 IU) may be required in the elderly and chronically ill. The IOM report suggests that it is safe to take up to 4000 IU/d. For those with osteoporosis or those at risk of osteoporosis, 1000–2000 IU/d can usually maintain serum 25(OH)D above 30 ng/mL. Vitamin D supplementation by itself does not appear to reduce fracture risk, but the combination of adequate calcium intake and vitamin D does decrease fracture risk. Low vitamin D levels appear associated with more serious outcomes in response to COVID-19. Whether this is a cause-and-effect relationship or a chance occurrence is not known, but it certainly argues for ensuring normal circulation levels of vitamin D.

**Other Nutrients** Other nutrients such as salt, high animal protein intakes, and caffeine may have modest effects on calcium excretion or absorption. Adequate vitamin K status is required for optimal carboxylation of osteocalcin. States in which vitamin K nutrition or metabolism is impaired, such as with long-term warfarin therapy, have been associated with reduced bone mass. Research concerning cola-based soda beverage intake is controversial but suggests a possible link to reduced bone mass through factors that appear independent of caffeine.

Magnesium is abundant in foods, and magnesium deficiency is quite rare in the absence of a serious chronic disease. Magnesium supplementation may be warranted in patients with inflammatory bowel disease, celiac disease, chemotherapy, severe diarrhea, malnutrition, or alcoholism. Dietary phytoestrogens, which are derived primarily from soy products and legumes (e.g., garbanzo beans

[chickpeas] and lentils), exert some estrogenic activity but are insufficiently potent to justify their use in place of a pharmacologic agent in the treatment of osteoporosis.

Patients with hip fractures are often frail and relatively malnourished. Some data suggest an improved outcome in such patients when they are provided calorie and protein supplementation. Excessive protein intake can increase renal calcium excretion, but this can be corrected by an adequate calcium intake. Strontium as a dietary mineral has also been implicated with strontium ranelate approved in some countries for treatment of osteoporosis. No evidence suggests that strontium in doses used in supplements can reduce fracture risk, but by virtue of replacing calcium in bone with the larger strontium atom, this supplement can produce an increase in bone density of questionable significance.

**Exercise** Exercise in young individuals increases the likelihood that they will attain the maximal genetically determined peak bone mass. Meta-analyses of studies performed in postmenopausal women indicate that weight-bearing exercise helps prevent bone loss but does not appear to result in substantial gain of bone mass. This beneficial effect wanes if exercise is discontinued. Most of the studies are short term, and a more substantial effect on bone mass is likely if exercise is continued over a long period. Exercise also has beneficial effects on neuromuscular function, and it improves coordination, balance, and strength, thereby reducing the risk of falling. A walking program is a practical way to start. Other activities such as dancing, racquet sports, cross-country skiing, and use of gym equipment, are also recommended, depending on the patient's personal preference and general condition. Even women who cannot walk benefit from swimming or water exercises, not so much for the effects on bone, which are quite minimal, but because of effects on muscle. Exercise habits should be consistent, optimally at least three times a week. For most patients, we suggest participation in exercise regimens that the patient enjoys in order to improve adherence. We also emphasize the importance of making exercise a social activity, again to improve adherence. Many individuals experience a fear of falling that can lead to social isolation and depression. Group exercises can help to alleviate this problem by providing sense of social connectivity among participants.

Tai chi is a traditional Chinese martial art that utilizes a series of gentle, flowing movements to promote and maintain flexibility, balance, endurance, proprioception, and strength. It involves constant movement through all three spatial dimensions. As a three-dimensional exercise, tai chi may be incorporated as part of balance training program for individuals with osteoporosis. Tai chi is generally considered a safe activity. The evidence for fall prevention in tai chi has been evaluated in randomized controlled trials. Results have been controversial; however, recent systematic reviews have provided a growing body of evidence to indicate that participating in tai chi can significantly reduce the risk of falls in older adults. The benefit appears to be greater when tai chi is practiced with increased frequency.

It is recommended that individuals with osteoporosis or osteoporotic vertebral fractures participate in exercise programs that involve both resistance and balance training. Slow, controlled movements are recommended in order to avoid injuries. Exercise modifications and avoidance of certain postures (such as spinal flexion) may be advised for those with prior injuries and back or joint pain. Caution should be taken to avoid activities that can lead to potential fractures, such as performing activities on slippery surfaces or twisting or bending the spine quickly while transitioning between different positions. Precautions should be taken to avoid injury when exercising with loads and performing exercises that challenge balance. For individuals with osteoporosis who have a high risk for fracture, vertebral fractures, sedentary lifestyle, or comorbid conditions that impact exercise tolerance, consultation with physical therapists to learn safe exercise practices is recommended.

Several guidelines for the treatment of osteoporosis have been published over the past few years. Patients presenting with fractures of the hip and spine should be evaluated for treatment. Patients presenting with low-trauma fractures in the setting of a BMD in the low bone mass or osteoporosis range should be treated with pharmacologic agents. Most guidelines also suggest that patients be considered for treatment when BMD T-score is  $-2.5$ , a level consistent with the diagnosis of osteoporosis. Treatment should also be considered in postmenopausal women with fracture or multiple risk factors even if BMD is not in the osteoporosis range. Treatment thresholds depend on cost-effectiveness analyses but, in the United States, are a  $>20\%$  10-year major fracture probability and  $>3\%$  10-year hip fracture probability. It must be emphasized, however, that as with other diseases, risk assessment is an inexact science when applied to individual patients. Fractures are chance occurrences that can happen to anyone and do! Patients often accept risks that are higher than the physician might like out of concern for the (usually considerably lower) risks of adverse events of drugs.

Pharmacologic therapies for osteoporosis are either antiresorptive or anabolic. The antiresorptive agents include medications that have broad effects such as hormone/estrogen therapy and selective estrogen receptor modulators (SERMS) as well as those agents that are specific for the treatment of osteoporosis (bisphosphonates, denosumab, and calcitonin). The anabolic agents are teriparatide, abaloparatide, and romosozumab. Denosumab, considered as an antiresorptive, allows bone formation to continue, and thus, there is an increase in bone density beyond that occurring with agents that inhibit resorption directly leading to a reduction in bone formation.

**Antiresorptive Agents** • **Estrogens** A large body of clinical trial data indicates that various types of estrogens (conjugated equine estrogens, estradiol, estrone, esterified estrogens, ethinyl estradiol, and mestranol) reduce bone turnover, prevent bone loss, and induce small increases in bone mass of the spine, hip, and total body. The effects of estrogen are seen in women with natural or surgical menopause and in late postmenopausal women with or without established osteoporosis. Estrogens are efficacious when administered orally, transdermally, or by subcutaneous implant. For both oral and transdermal routes of administration, combined estrogen/progestin preparations are now available in many countries, obviating the problem of taking two tablets or using a patch and oral progestin.

For oral estrogens, the standard recommended doses have been  $0.3\text{ mg/d}$  for esterified estrogens,  $0.625\text{ mg/d}$  for conjugated equine estrogens, and  $5\text{ }\mu\text{g/d}$  for ethinyl estradiol. For transdermal estrogen, the commonly used dose supplies  $50\text{ }\mu\text{g}$  of estradiol per day, but a lower dose may be appropriate for some individuals. Dose-response data for conjugated equine estrogens indicate that lower doses ( $0.3$  and  $0.45\text{ mg/d}$ ) are effective. Doses even lower have also been shown to slow bone loss.

**Fracture Data** Epidemiologic databases indicate that women who take estrogen replacement have a  $50\%$  reduction, on average, of osteoporosis-related fractures, including hip fractures. The beneficial effect of estrogen is greatest among those who start replacement early and continue the treatment; the benefit declines after discontinuation to the extent that there is no residual protective effect against fracture by 10 years after discontinuation. The first clinical trial evaluating fractures as secondary outcomes, the Heart and Estrogen-Progestin Replacement Study (HERS) trial, showed no effect of hormone therapy on hip or other clinical fractures in women with established coronary artery disease. These data made the results of the Women's Health Initiative (WHI) exceedingly important (Chap. 395). The estrogen-progestin arm of the WHI in  $>16,000$  postmenopausal healthy women indicated that hormone therapy reduces the risk of hip and clinical spine fracture by  $34\%$  and that of all clinical fractures by  $24\%$ .

A few smaller clinical trials have evaluated spine fracture occurrence as an outcome with estrogen therapy. They have consistently

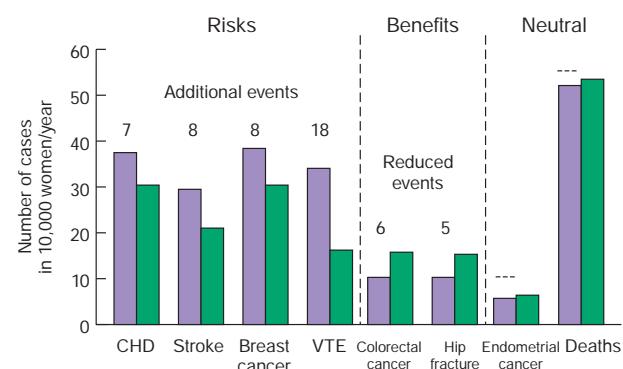


FIGURE 411-8 Effects of hormone therapy on event rates: green, placebo; purple, estrogen and progestin. CHD, coronary heart disease; VTE, venous thromboembolic events. (Adapted with permission from Women's Health Initiative. WHI HRT Update.)

shown that estrogen treatment reduces the incidence of vertebral compression fracture.

The WHI has provided a vast amount of data on the multisystemic effects of hormone therapy. Although earlier observational studies suggested that estrogen replacement might reduce heart disease, the WHI showed that combined estrogen-progestin treatment increased risk of fatal and nonfatal myocardial infarction by  $\sim 29\%$ , confirming data from the HERS study. Other important relative risks included a  $40\%$  increase in stroke, a  $100\%$  increase in venous thromboembolic disease, and a  $26\%$  increase in risk of breast cancer. Subsequent analyses have confirmed the increased risk of stroke and, in a substudy, showed a twofold increase in dementia. Benefits other than the fracture reductions noted above included a  $37\%$  reduction in the risk of colon cancer. These relative risks have to be interpreted in light of absolute risk (Fig. 411-8). For example, out of 10,000 women treated with estrogen-progestin for 1 year, there will be 8 excess heart attacks, 8 excess breast cancers, 18 excess venous thromboembolic events, 5 fewer hip fractures, 44 fewer clinical fractures, and 6 fewer colorectal cancers. These numbers must be multiplied by the number of years of hormone treatment. There was no effect of combined hormone treatment on the risk of uterine cancer or total mortality.

It is important to note that these WHI findings apply specifically to hormone treatment in the form of conjugated equine estrogen plus medroxyprogesterone acetate. The relative benefits and risks of unopposed estrogen in women who had hysterectomies vary somewhat. They still show benefits against fracture occurrence and increased risk of venous thrombosis and stroke, similar in magnitude to the risks for combined hormone therapy. In contrast, though, the estrogen-only arm of WHI indicated no increased risk of heart attack or breast cancer. The data suggest that at least some of the detrimental effects of combined therapy are related to the progestin component. In addition, there is the possibility, suggested by primate data, that the risk accrues mainly to women who have some years of estrogen deficiency before initiating treatment. Nonetheless, there is marked reluctance among women for estrogen therapy/hormone therapy, and the U.S. Preventive Services Task Force has specifically suggested that estrogen therapy/hormone therapy not be used for disease prevention.

**Mode of Action** Two subtypes of ERs,  $\alpha$  and  $\beta$ , have been identified in bone and other tissues. Cells of monocyte lineage express both ER $\alpha$  and ER $\beta$ , as do osteoblasts. Estrogen-mediated effects vary with the receptor type. Using ER knockout mouse models, elimination of ER $\alpha$  produces a modest reduction in bone mass, whereas mutation of ER $\beta$  has less of an effect on bone. A male patient with a homozygous mutation of ER $\beta$  had markedly decreased bone density as well as abnormalities in epiphyseal closure, confirming the important role of ER $\beta$  in bone biology.

The mechanism of estrogen action in bone is an area of active investigation (Fig. 411-5). Although data are conflicting, estrogens may inhibit osteoclasts directly. However, the majority of estrogen (and androgen) effects on bone resorption are mediated indirectly through paracrine factors produced by osteoblasts. These actions include (1) increasing osteoprotegerin production by osteoblasts, (2) increasing IGF-I and TGF-, and (3) suppressing IL-1 ( and ), IL-6, TNF-, and osteocalcin synthesis. The indirect estrogen actions primarily decrease bone resorption.

**Progestins** In women with a uterus, daily progestin or cyclical progestins at least 12 days per month are prescribed in combination with estrogens to reduce the risk of uterine cancer. Medroxyprogesterone acetate and norethindrone acetate blunt the high-density lipoprotein response to estrogen, but micronized progesterone does not. Neither medroxyprogesterone acetate nor micronized progesterone appears to have an independent effect on bone; at lower doses of estrogen, norethindrone acetate may have an additive benefit. In breast tissue, progestins may account for the increased risk of breast cancer with combination treatment.

**SERMs** Two SERMs are used currently in postmenopausal women: raloxifene, which is approved by the FDA for the prevention and treatment of osteoporosis as well as the prevention of breast cancer, and tamoxifen, which is approved for the prevention and treatment of breast cancer. A third SERM, bazedoxifene, is marketed in combination with conjugated estrogen for treatment of menopausal symptoms and prevention of bone loss. Bazedoxifene protects the uterus and breast from effects of estrogen and makes the use of progestin unnecessary.

*Tamoxifen* reduces bone turnover and bone loss in postmenopausal women compared with placebo groups. These findings support the concept that tamoxifen acts as an estrogenic agent in bone. There are limited data on the effect of tamoxifen on fracture risk, but the Breast Cancer Prevention study indicated a possible reduction in clinical vertebral, hip, and Colles' fractures. Tamoxifen is not FDA approved for prevention or treatment of osteoporosis. The major benefit of tamoxifen is on breast cancer occurrence and recurrence in women with ER-positive tumors. The breast cancer prevention trial indicated that tamoxifen administration over 4–5 years reduced the incidence of new invasive and noninvasive breast cancer by ~45% in women at increased risk of breast cancer. The incidence of ER-positive breast cancers was reduced by 65%. Tamoxifen increases the risk of uterine cancer in postmenopausal women, limiting its use for breast cancer prevention in women at low or moderate risk.

*Raloxifene* (60 mg/d) has effects on bone turnover and bone mass that are very similar to those of tamoxifen, indicating that this agent is also estrogenic on the skeleton. The effect of raloxifene on bone density (+1.4–2.8% vs placebo in the spine, hip, and total body) is somewhat less than that seen with standard doses of estrogens. Raloxifene reduces the occurrence of vertebral fracture by 30–50%, depending on the population; however, there are no data confirming that raloxifene can reduce the risk of nonvertebral fractures after 8 years of observation.

Raloxifene, like tamoxifen and estrogen, has effects in other organ systems. The most beneficial effect appears to be a reduction in invasive breast cancer (mainly decreased ER-positive) occurrence of ~65% in women who take raloxifene compared to placebo. In a head-to-head study, raloxifene was as effective as tamoxifen in preventing breast cancer in high-risk women, and raloxifene is FDA approved for this indication. In a further study, raloxifene had no effect on heart disease in women with increased risk for this outcome. In contrast to tamoxifen, raloxifene is not associated with an increase in the risk of uterine cancer or benign uterine disease. Raloxifene increases the occurrence of hot flashes but reduces serum total and low-density lipoprotein cholesterol, lipoprotein(a), and fibrinogen. Raloxifene, with its positive effects on breast cancer and vertebral fractures, has become a useful agent for the treatment of the younger asymptomatic postmenopausal woman. In some

women, a recurrence of menopausal symptoms may occur. Usually this is evanescent but occasionally is sufficiently impactful on daily life and sleep that the drug must be withdrawn. Raloxifene increases the risk of deep-vein thrombosis and may increase the risk of death from stroke among older women. Consequently, it is not usually recommended for women over age 70 years.

### MODE OF ACTION OF SERMS

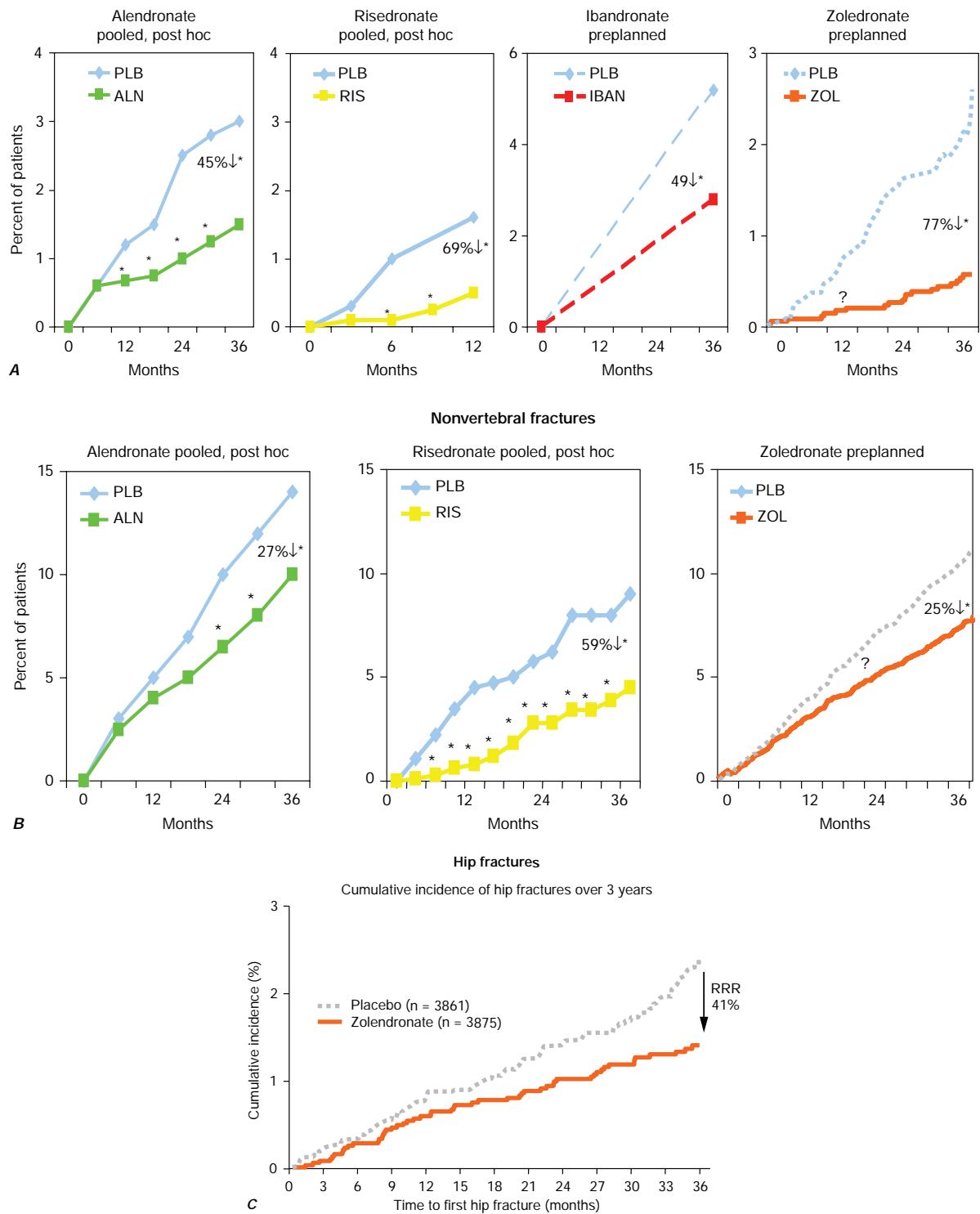
All SERMs bind to the ER, but each agent produces a unique receptor-drug conformation. As a result, specific coactivator or corepressor proteins are bound to the receptor (Chap. 377), resulting in differential effects on gene transcription that vary depending on other transcription factors present in the cell. Another aspect of selectivity is the affinity of each SERM for the different ER and ER subtypes, which are expressed differentially in various tissues. These tissue-selective effects of SERMs offer the possibility of tailoring estrogen therapy to best meet the needs and risk factor profile of an individual patient.

**Bisphosphonates** Bisphosphonates have become the mainstay of osteoporosis treatment, in part related to cost as they become generic. Alendronate, risedronate, ibandronate, and zoledronic acid are approved for the prevention and treatment of postmenopausal osteoporosis. Alendronate, risedronate, and zoledronic acid are also approved for the treatment of steroid-induced osteoporosis, and risedronate and zoledronic acid are approved for prevention of steroid-induced osteoporosis. Alendronate, risedronate, and zoledronic acid are also approved for treatment of osteoporosis in men.

*Alendronate* decreases bone turnover and increases bone mass in the spine by up to 8% versus placebo and by 6% versus placebo in the hip. Multiple trials have evaluated its effect on fracture occurrence. The Fracture Intervention Trial provided evidence in >2000 women with prevalent vertebral fractures that daily alendronate treatment (5 mg/d for 2 years and 10 mg/d for 9 months afterward) reduces vertebral fracture risk by ~50%, multiple vertebral fractures by up to 90%, and hip fractures by up to 50%. Several subsequent trials have confirmed these findings (Fig. 411-9). For example, in a study of >1900 women with low bone mass treated with alendronate (10 mg/d) versus placebo, the incidence of all nonvertebral fractures was reduced by ~47% after only 1 year. In the United States, the 70-mg weekly dose is approved for treatment of osteoporosis and the dose of 35 mg per week is approved for prevention, with those doses showing equivalence to daily dosing based on bone turnover and bone mass response.

Consequently, once-weekly therapy generally is preferred because of lower incidence of gastrointestinal side effects and ease of administration. Alendronate should be taken with a full glass of water before breakfast after an overnight fast, as bisphosphonates are poorly absorbed. Because of the potential for esophageal irritation, alendronate is contraindicated in patients who have stricture or inadequate emptying of the esophagus. It is recommended that patients remain upright (standing or sitting) for at least 30 minutes after taking the medication to avoid esophageal irritation and that food and fluids (other than water) be avoided for the same duration. In clinical trials, overall gastrointestinal symptomatology was no different with alendronate than with placebo, but in practice, all oral bisphosphonates have been associated with esophageal irritation and inflammation.

*Risedronate* also reduces bone turnover and increases bone mass. Controlled clinical trials have demonstrated 40–50% reduction in vertebral fracture risk over 3 years, accompanied by a 40% reduction in clinical nonspine fractures. The only clinical trial specifically designed to evaluate hip fracture outcome (HIP) indicated that risedronate reduced hip fracture risk in women in their seventies with confirmed osteoporosis by 40%. In contrast, risedronate was not effective at reducing hip fracture occurrence in older women (80+ years) without proven osteoporosis. Studies have shown that 35 mg of risedronate administered once weekly is therapeutically equivalent to 5 mg/d. The instructions for oral administration



**FIGURE 411-9** Effects of various bisphosphonates on fractures. **A.** Clinical vertebral fractures. **B.** Nonvertebral fractures. **C.** Hip fractures. PLB, placebo; RRR, relative risk reduction. (Data from DM Black et al: *J Clin Endocrinol Metab* 85:4118, 2000; C Roux et al: *Curr Med Res Opin* 4:433, 2004; CH Chesnut et al: *J Bone Miner Res* 19: 1241, 2004; DM Black et al: *N Engl J Med* 356:1809, 2007; JT Harrington et al: *Calcif Tissue Int* 74:129, 2003.)

noted for alendronate apply to all three oral bisphosphonates. There is also a preparation of risedronate (35 mg) that can be taken after breakfast. Risedronate is the only bisphosphonate that has this dosing flexibility.

*Ibandronate* is the third amino-bisphosphonate approved in the United States. Ibandronate (2.5 mg/d) has been shown in clinical trials to reduce vertebral fracture risk by ~40% but with no overall effect on nonvertebral fractures. In a post hoc analysis of subjects with a femoral neck T-score of -3, ibandronate reduced the risk of nonvertebral fractures by ~60%. In clinical trials, ibandronate doses of 150 mg/month PO or 3 mg every 3 months IV had greater effects on turnover and bone mass than did 2.5 mg/d. Patients should take oral ibandronate in the same way as other bisphosphonates, but with 1 h elapsing before other food or drink (other than plain water).

*Zoledronic acid* is a potent bisphosphonate with a unique administration regimen (5 mg by 30-min IV infusion at most annually). Zoledronic acid data confirm that it is highly effective in fracture risk reduction. In a study of >7000 women followed for 3 years, zoledronic acid 5 mg IV annually) reduced the risk of vertebral fractures by 70%, nonvertebral fractures by 25%, and hip fractures by 40%. These results were associated with less height loss and disability. In the treated population, there was an increased risk of almost 25% of an acute phase reaction in patients with no prior bisphosphonate exposure (fever, myalgias, headache, malaise), but effects were short-lived (2–3 days). Detailed evaluation of all bisphosphonates failed to confirm a risk of atrial fibrillation. Zoledronic acid has also been studied in a placebo-controlled trial of women and men within 3 months of an acute hip fracture. The risk of recurrent fracture was reduced by 35%, and there was a 28% reduction in mortality that was greater than might be expected by the reduction in hip fracture alone.

**Common Bisphosphonate Adverse Events** All bisphosphonates have been associated with some musculoskeletal and joint pains of unclear etiology, which are occasionally severe. There is potential for renal toxicity, and bisphosphonates are contraindicated in those with an estimated glomerular filtration rate <30–35 mL/min. Hypocalcemia can occur.

There has been concern about two potential side effects associated with bisphosphonate use. The first is osteonecrosis of the jaw (ONJ). ONJ usually follows a dental procedure in which bone is exposed (dental extractions and implants). It is presumed that the exposed bone becomes infected and dies. ONJ is more common among cancer patients receiving high doses of bisphosphonates for skeletal metastases. It is rare among persons with osteoporosis on usual doses of bisphosphonates. Oral antibiotic rinses and oral systemic antibiotics may be useful to prevent this rare adverse event if risk is perceived to be particularly high. The second is called atypical femoral fracture. These are unusual fractures that occur in the subtrochanteric femoral region or across the femoral shaft distal to the lesser trochanter. They are often preceded by pain in the lateral thigh or groin that can be present for weeks, months, or even years before the fracture. The fractures occur on trivial trauma, are horizontal with a medial beak, and are noncomminuted. A committee put together by the American Society for Bone and Mineral Research described the major and minor criteria for these fractures, which appear to be related to duration of bisphosphonate therapy. The overall risk appears quite low, especially when compared to the number of hip fractures saved by these therapies, but they often require surgical fixation and are difficult to heal. Some evidence suggests that if the fractures are found early, when there is evidence of periosteal stress reaction or stress fracture, prior to the occurrence of overt fracture, that teriparatide can help heal the fracture and preclude the need for surgical repair. We routinely inform patients initiating bisphosphonates that if they develop thigh or groin pain they should inform us. Routine x-rays will sometimes detect cortical thickening or even a stress fracture, but more commonly, MRI or technetium bone scan is required. The presence of an abnormality requires, at minimum, a period of

modified weight bearing and may need prophylactic rodding of the femur. It is important to realize that these may be bilateral (~50% of the time), and when an abnormality is found, the other femur should be checked. It is unknown whether patients who have these atypical femur fractures can ever receive antiresorptive therapies again in the future, but it seems prudent to avoid their use for the majority of these individuals.

**Mode of Action** Bisphosphonates are structurally related to pyrophosphates, compounds that are incorporated into bone matrix. Bisphosphonates specifically impair osteoclast function and reduce osteoclast number, in part by inducing apoptosis. Recent evidence suggests that the nitrogen-containing bisphosphonates also inhibit protein prenylation, one of the end products in the mevalonic acid pathway, by inhibiting the enzyme farnesyl pyrophosphate synthase. This effect disrupts intracellular protein trafficking and ultimately may lead to apoptosis. Some bisphosphonates have very long retention in the skeleton and may exert long-term effects. The consequences of this, if any, are unknown.

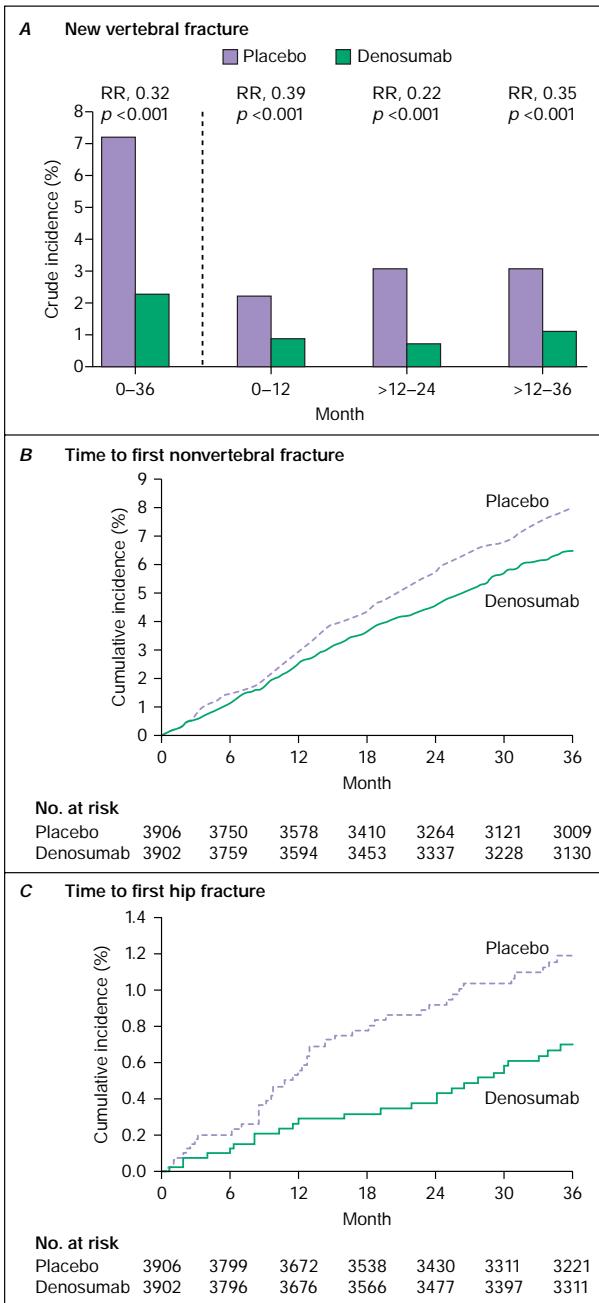
**Calcitonin** Calcitonin is a polypeptide hormone produced in the thyroid gland ([Chap. 410](#)). Its physiologic role is unclear as no skeletal disease has been described in association with calcitonin deficiency or excess. Calcitonin preparations are approved by the FDA for Paget's disease, hypercalcemia, and osteoporosis in women >5 years past menopause.

Injectable calcitonin produces small increments in bone mass of the lumbar spine. However, difficulty of administration and frequent reactions, including nausea and facial flushing, make general use limited. A nasal spray containing calcitonin (200 IU/d) is available for treatment of osteoporosis in postmenopausal women. One study suggests that nasal calcitonin produces small increments in bone mass and a small reduction in new vertebral fractures in calcitonin-treated patients (at one dose) versus those on calcium alone. There has been no proven effectiveness against nonvertebral fractures. Calcitonin is not indicated for prevention of osteoporosis and is not sufficiently potent to prevent bone loss in early postmenopausal women. Calcitonin might have an analgesic effect on bone pain, both in the subcutaneous and possibly in the nasal form. Concerns have been raised about an increase in the incidence of cancer associated with calcitonin use. Initially, the cancer noted was of the prostate, but an analysis of all data suggested a more general increase in cancer risk. In Europe, the European Medicines Agency has removed the osteoporosis indication, and an FDA Advisory Committee has voted for a similar change in the United States.

**Mode of Action** Calcitonin suppresses osteoclast activity by direct action on the osteoclast calcitonin receptor. Osteoclasts exposed to calcitonin cannot maintain their active ruffled border, which normally maintains close contact with underlying bone.

**Denosumab** Denosumab is a novel agent that, given twice yearly by subcutaneous administration in a randomized controlled trial in postmenopausal women with osteoporosis, has been shown to increase BMD in the spine, hip, and forearm and reduce vertebral, hip, and nonvertebral fractures over a 3-year period by 70, 40, and 20%, respectively ([Fig. 411-10](#)). Other clinical trials indicate ability to increase bone mass in postmenopausal women with low bone mass (above osteoporosis range) and in postmenopausal women with breast cancer treated with aromatase inhibitor therapies. In the oncology literature, denosumab reduces the risk of fractures in women on aromatase inhibitors. In a study of men with prostate cancer treated with androgen deprivation therapy, denosumab increased bone mass and reduced vertebral fracture occurrence. An analysis of five placebo-controlled studies has suggested reduced risk of falls in patients with osteoporosis treated with denosumab.

Denosumab was approved by the FDA in 2010 for the treatment of postmenopausal women who have a high risk for osteoporotic fractures, including those with a history of fracture or multiple risk



**FIGURE 411-10** Effects of denosumab on the following: **A**, new vertebral fractures; and **B**, and **C**, times to nonvertebral and hip fracture. RR, relative risk. (Reproduced with permission from SR Cummings et al: Denosumab for prevention of fractures in postmenopausal women with osteoporosis. *N Engl J Med* 361:756, 2009.)

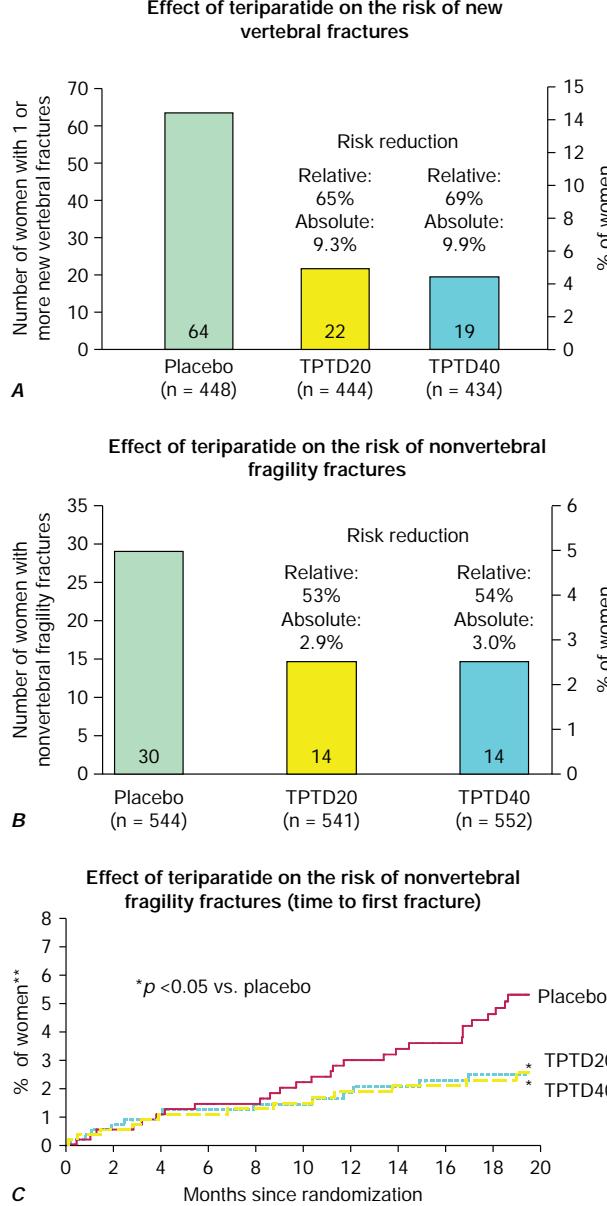
factors for fracture, and those who have failed or are intolerant to other osteoporosis therapy. Denosumab is also approved for the treatment of osteoporosis in men at high risk for fracture, women with breast cancer on aromatase inhibitors, and men with prostate cancer on androgen deprivation treatment. A long-term observational extension of the pivotal trial in postmenopausal women has provided evidence that BMD continues to increase in both the spine and hip with 3–10 years of denosumab treatment, with fracture rates that are at least as low as those seen with denosumab during the active placebo-controlled portion of the trial.

Denosumab may increase the risk of ONJ and atypical femur fractures similarly to bisphosphonates. Estimated incidence is 5/10,000 patient-years for ONJ and 1/10,000 patient-years for atypical femur fractures. Denosumab can cause hypersensitivity reactions, hypocalcemia, and skin reactions including dermatitis, rash, and eczema. Early concerns about an imbalance in infections with denosumab have largely been allayed.

When denosumab is discontinued, there is a rebound increase in bone turnover and an apparent acceleration of bone loss. This likely reflects the maturation of osteoclast precursors that have accumulated in marrow when the drug was administered and can become mature bone resorbing cells once the drug is withdrawn. The consequences of this rebound increase in remodeling-associated bone loss are a rapid increase in the risk of fracture, particularly vertebral fracture, and a specific increase in the occurrence of multiple vertebral fractures. In patients who need to stop denosumab or in patients in whom BMD and fracture risk reduction goals have been met, temporary use of bisphosphonate treatment may prevent the rebound increase in remodeling and rapid bone loss. In clinical practice, a single infusion of zoledronic acid seems to maintain BMD for 1–2 years but may need to be repeated. Oral bisphosphonates can also be prescribed. In both cases, the required duration of bisphosphonate use to eliminate the rebound effect is not clear and may vary considerably among patients.

**Mode of Action** Denosumab is a fully human monoclonal antibody to RANKL, the final common effector of osteoclast formation, activity, and survival. Denosumab binds to RANKL, inhibiting its ability to initiate formation of mature osteoclasts from osteoclast precursors and to bring mature osteoclasts to the bone surface and initiate bone resorption. Denosumab also plays a role in reducing the survival of the osteoclast. Through these actions on the osteoclast, denosumab induces potent antiresorptive action, as assessed biochemically and histomorphometrically.

**Anabolic Agents** • **Parathyroid Hormone** Endogenous PTH is an 84-amino-acid peptide that is largely responsible for calcium homeostasis (**Chap. 410**). Although chronic elevation of PTH, as occurs in hyperparathyroidism, is associated with bone loss (particularly cortical bone), PTH also can exert anabolic effects on bone. Consistent with this, some observational studies have indicated that mild endogenous hyperparathyroidism is associated with maintenance of trabecular bone mass but loss of cortical bone. On the basis of these findings, early small-scale observational studies showed that PTH analogues could augment trabecular BMD. Subsequent controlled clinical trials have confirmed that PTH can increase bone mass and reduce fracture occurrence. The first randomized controlled trial in postmenopausal women showed that PTH (1–34) (teriparatide), when superimposed on ongoing estrogen therapy, produced substantial increments in bone mass (13% over a 3-year period compared with estrogen alone) and reduced the risk of vertebral compression deformity. In the pivotal study (median, 19 months' duration), 20 µg PTH (1–34) daily by subcutaneous injection (with no additional therapy) reduced vertebral fractures by 65% and nonvertebral fractures by 40–50% (**Fig. 411-11**). Teriparatide produces rapid and robust increases in bone formation and then bone remodeling overall, resulting in substantial increases in bone mass and improvements in microarchitecture, including cancellous connectivity and cortical width. The BMD effects, particularly in the hip, are lower when patients switch from bisphosphonates to teriparatide, possibly in proportion to the potency of the antiresorptive agent. The hip BMD effect is particularly impaired when patients switch from denosumab to teriparatide. In patients on denosumab who need teriparatide treatment, there may be a role for combination therapy. In previously untreated women, teriparatide is administered as monotherapy and followed by a potent antiresorptive agent such as denosumab or a bisphosphonate. Combination therapy is generally avoided



**FIGURE 411-11** Effects of teriparatide (TPT) on the following: **A**, new vertebral fractures; and **B**, and **C**, nonvertebral fragility fractures. (**A** and **B** are data from RM Neer et al: Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med* May 344:1434, 2001. **C** reproduced with permission from RM Neer et al: Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N Engl J Med* May 344:1434, 2001.)

because of cost and potential inhibition of the anabolic activity of teriparatide.

In women with painful acute osteoporotic vertebral fractures, teriparatide reduced subsequent vertebral fractures by ~50% compared with risedronate. There was no difference in nonvertebral fracture outcome between the two medications. A study comparing teriparatide with risedronate in patients with prevalent vertebral fractures showed significant benefit for teriparatide against vertebral fractures and nearly significant benefit for teriparatide against nonvertebral fractures.

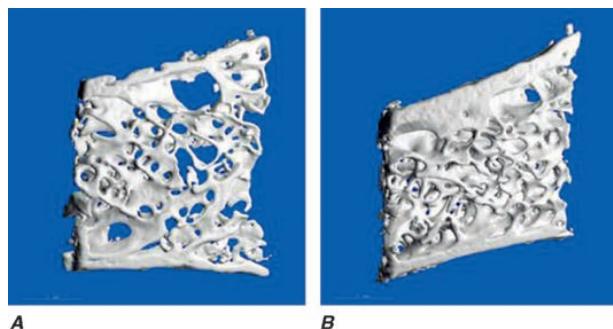
Side effects of teriparatide are generally mild and can include muscle pain, weakness, dizziness, headache, and nausea. Rodents given prolonged treatment with PTH in high doses (3–60 times the human dose) developed osteogenic sarcomas after ~18 months of treatment. Rare cases of osteosarcoma have been described in patients treated with teriparatide consistent with the background incidence of osteosarcoma in adults. Long-term surveillance studies of a high proportion of patients diagnosed with osteosarcoma as adults in both the United States and Scandinavia reveal no prior exposure to teriparatide in any of the cases.

Teriparatide use may be limited by cost and its mode of administration (daily subcutaneous injection). Alternative modes of delivery have been investigated, but none have proven successful. Because of the rodent osteosarcoma data and the maximum duration of teriparatide in the pivotal trial of 2 years, the FDA has limited teriparatide treatment to 2 years, with that becoming the lifetime maximal use. As a result, consideration is often given to restricting initial use to 1 year (using bone density response at 1 year as a guide) and saving the second year for future use if necessary.

**Mode of Action** Exogenously administered PTH appears to have direct actions on osteoblast activity, with biochemical and histomorphometric evidence of de novo bone formation within a week or two in response to teriparatide. There is subsequently resorption. Subsequently, teriparatide activates bone remodeling but still appears to favor bone formation over bone resorption. Teriparatide given by daily injection stimulates osteoblast recruitment and activity through activation of Wnt signaling. Teriparatide produces a true increase in bone tissue and an apparent restoration of bone microarchitecture (Fig. 411-12).

**Abaloparatide** Abaloparatide is a synthetic analogue of human PTHrP, which has significant homology to PTH and also binds the PTH type 1 receptor. Abaloparatide and teriparatide exert different binding affinities to the two different receptor conformations, R<sup>0</sup> and RG. Compared to teriparatide, abaloparatide binds with similar high affinity to the RG conformation but with much lesser affinity to the R<sup>0</sup> conformation. These differences appear to result in a similar bone formation stimulus but lesser bone resorption stimulus, and abaloparatide was specifically chosen for development among a large number of PTH and PTHrP analogues for what appeared to be an optimized anabolic profile.

In the phase 3 Abaloparatide Comparator Trial in Vertebral Endpoints (ACTIVE) study, 2463 postmenopausal women with osteoporosis were randomized to blinded daily subcutaneous abaloparatide versus placebo or open-label teriparatide. At 18 months, spine BMD increase was similar with abaloparatide and teriparatide (11.2% abaloparatide and 10.5% teriparatide); in the total hip, BMD



**FIGURE 411-12** Effect of parathyroid hormone (PTH) treatment on bone microarchitecture. Paired biopsy specimens from a 64-year-old woman before (**A**) and after (**B**) treatment with PTH. (Reproduced with permission from DW Dempster et al: Effects of daily treatment with parathyroid hormone on bone microarchitecture and turnover in patients with osteoporosis: A paired biopsy study. *J Bone Miner Res* 16:1846, 2001.)

increments were slightly larger with abaloparatide (4.2 vs 3.3%). New vertebral fracture incidence was reduced by 86% with abaloparatide and 80% with teriparatide compared with placebo (both  $p < .001$ ). The hazard ratio for abaloparatide versus teriparatide was not quoted. Nonvertebral fractures were reduced by 43% with abaloparatide ( $p = .05$ ) and by 28% with teriparatide (not significant;  $p = .22$ ). The ACTIVE study was extended, with 92% of eligible participants from the abaloparatide and placebo arms transitioned to open-label alendronate for a total treatment period of 24 months of alendronate. Both vertebral and nonvertebral fractures were less common in the group who transitioned from abaloparatide to alendronate, suggesting that the fracture benefit of abaloparatide can be maintained with antiresorptive treatment.

**Romosozumab** Romosozumab is a humanized antibody that blocks the osteocyte production of sclerostin, resulting in an increase in bone formation and decline in bone resorption. In the pivotal trial (FRAME), 7180 postmenopausal women with osteoporosis were randomized to receive blinded monthly subcutaneous romosozumab (210 mg) or placebo for 1 year followed by transition to open-label subcutaneous denosumab (60 mg) every 6 months for an additional year. BMD increased over 13% in the spine and almost 7% in the hip in 1 year with romosozumab. At 1 year, the incidence of new vertebral fractures in the romosozumab group was significantly reduced by 73% compared with placebo. Clinical fracture risk (nonvertebral fractures and clinical vertebral fractures combined) was significantly reduced by 36%. Nonvertebral fractures were also reduced, but the difference just missed statistical significance perhaps due to geographical differences; in the high-enrolling Latin American region, there was no significant reduction in nonvertebral fractures, probably due to a very low background incidence in that region. In the rest of the world, nonvertebral fractures were significantly reduced by >40%. During the second year of the FRAME study, both groups transitioned to denosumab. Over 24 months, women who had received romosozumab during the first 12 months and then denosumab had 75% fewer new vertebral fractures than those who had received placebo for a year followed by denosumab. There were also nearly significant trends toward reduced clinical and nonvertebral fractures in the romosozumab/denosumab group. Compared with baseline, BMD increased by 17.6% in the spine and 8.8% in the total hip in the romosozumab/denosumab group. Safety and tolerability of the two drugs were similar, with a slightly higher incidence of injection site reactions in the denosumab group. The FRAME study is in an ongoing extension where all participants received continued denosumab for an additional year. A parallel trial of very-high-risk patients, all of whom have prevalent vertebral fractures, is also ongoing and is comparing romosozumab to alendronate for 1 year, followed by transition to or continuation of alendronate for 2 additional years. In one study, there was an increase in cardiovascular side effects prompting a warning on the label.

#### OTHER PHARMACOLOGIC AGENTS NOT APPROVED IN THE UNITED STATES

Odanacatib, a cathepsin K inhibitor, inhibits the osteoclast collagenase enzyme, preventing bone resorption but not affecting osteoclast viability. This agent was in late-stage drug development. In a very large controlled clinical trial (~17,000 postmenopausal women with osteoporosis), bone mass increased substantially in the spine and hip, and vertebral, hip, and all nonvertebral fractures were reduced. Unfortunately, odanacatib was associated with a significantly increased risk of stroke, and the development of this agent was aborted in September 2016.

Testosterone has been used to treat osteoporosis associated with low testosterone levels in men. There are data that indicate that testosterone can increase bone density, but there are no data indicating improvement in any fracture endpoints. Since there are many other effects of testosterone, especially in older men (including prostate hypertrophy), decisions to use it for treatment of osteoporosis have to take the multisystemic effects into account.

Sodium fluoride was tested in two large parallel clinical trials in the late 1980s. Although BMD increased substantially, the increase was in part due to fluoride incorporation in the hydroxyapatite crystal. Fracture risk was not reduced and, in fact, was increased in nonvertebral sites. Therefore, fluoride is no longer considered a viable option for osteoporosis treatment.

Strontium ranelate has never been approved for osteoporosis in the United States but is approved in Europe and some other countries outside of the United States. It increases bone mass throughout the skeleton, but much of the increase is related to strontium incorporation into hydroxyapatite. In clinical trials, the drug reduced the risk of vertebral fractures by 37% and that of nonvertebral fractures by 14%. It appears to be modestly antiresorptive while at the same time not causing as much of a decrease in bone formation (measured biochemically). In 2014, the use of strontium was restricted because of an increased risk of cardiovascular disease and severe skin reactions. Small increased risks of venous thrombosis also occur.

Several small studies of growth hormone, alone or in combination with other agents, have not shown consistent or substantial positive effects on skeletal mass.

#### NONPHARMACOLOGIC APPROACHES

Protective pads worn around the outer thigh, which cover the trochanteric region of the hip, can prevent hip fractures in elderly residents in nursing homes. The use of hip protectors is limited largely by issues of compliance and comfort, but new devices are being developed that may circumvent these problems and provide adjunctive treatments.

*Kyphoplasty* and *vertebroplasty* are also useful nonpharmacologic approaches for the treatment of painful vertebral fractures. The data do not support routine surgical intervention for vertebral fractures since, while this can reduce pain, there is concern about long-term vertebral fracture risk.

#### TREATMENT MONITORING

There are currently no well-accepted guidelines for monitoring treatment of osteoporosis. Because most osteoporosis treatments produce small or moderate bone mass increments on average, it is reasonable to consider BMD as a monitoring tool. Changes must exceed ~4% in the spine and 6% in the hip to be considered significant in any individual. The hip is the preferred site due to larger surface area and greater reproducibility. Medication-induced increments may require several years to produce changes of this magnitude (if they do at all). Consequently, it can be argued that BMD should be repeated at intervals of every 2 years. Only significant BMD reductions should prompt a change in medical regimen, as it is expected that many individuals will not show responses greater than the detection limits of the current measurement techniques.

Biochemical markers of bone turnover can help in treatment monitoring, with significant changes seen within 3 months of initiating treatment with approved medications and the possible benefit of improving adherence. It remains unclear which endpoint is most useful. If bone turnover markers are used, a determination should be made before therapy is started and repeated 3–4 months after therapy is initiated. In general, a change in bone turnover markers must be 30–40% lower than the baseline to be significant because of the biologic and technical variability in these tests. Because markers change more rapidly than bone density, they are often early signs of treatment effect. Currently collagen C-telopeptide measured on a fasting serum sample in the morning is the preferred marker of bone resorption, and osteocalcin or the propeptide of type 1 collagen (P1NP) is the preferred marker for formation.

#### GLUCOCORTICOID-INDUCED OSTEOPOROSIS

Osteoporotic fractures are a well-characterized consequence of the hypercortisolism associated with Cushing's syndrome. However, the therapeutic use of glucocorticoids is by far the most common form

of glucocorticoid-induced osteoporosis (GCIO). Glucocorticoids are used widely in the treatment of a variety of disorders, including chronic lung disorders, rheumatoid arthritis and other connective tissue diseases, and inflammatory bowel disease, and after transplantation. Osteoporosis and related fractures are serious side effects of chronic glucocorticoid therapy. Because the effects of glucocorticoids on the skeleton are often superimposed on the consequences of aging and menopause, it is not surprising that women and the elderly are most frequently affected. The skeletal response to steroids is remarkably heterogeneous, however, and even young, growing individuals treated with glucocorticoids can present with fractures.

The risk of fractures depends on the dose and duration of glucocorticoid therapy, although recent data suggest that there may be no completely safe dose. Bone loss is more rapid during the early months of treatment, and trabecular bone is affected more severely than cortical bone. As a result, fractures have been shown to increase within 3 months of steroid treatment. There is an increase in fracture risk in both the axial skeleton and the appendicular skeleton, including risk of hip fracture. Bone loss can occur with any route of steroid administration, including high-dose inhaled glucocorticoids and intra-articular injections. Alternate-day delivery does not appear to ameliorate the skeletal effects of glucocorticoids.

## PATHOPHYSIOLOGY

Glucocorticoids increase bone loss by multiple mechanisms, including (1) inhibition of osteoblast function and an increase in osteoblast apoptosis, resulting in impaired synthesis of new bone; (2) stimulation of bone resorption, probably as a secondary effect; (3) impairment of the absorption of calcium across the intestine, probably by a vitamin D-independent effect; (4) increase of urinary calcium loss and perhaps induction of some degree of secondary hyperparathyroidism; (5) reduction of adrenal androgens and suppression of ovarian and testicular secretion of estrogens and androgens; and (6) induction of glucocorticoid myopathy, which may exacerbate effects on skeletal and calcium homeostasis as well as increase the risk of falls.

## EVALUATION OF THE PATIENT

Because of the prevalence of GCIO, it is important to evaluate the status of the skeleton in all patients starting or already receiving long-term glucocorticoid therapy. Modifiable risk factors should be identified, including those for falls. Examination should include testing of height and muscle strength. Laboratory evaluation should include an assessment of 24-h urinary calcium. All patients on long-term (>3 months) glucocorticoids should have measurement of bone mass at both the spine and the hip using DXA. If only one skeletal site can be measured, it is best to assess the spine in individuals <60 years and the hip in those >60 years.

## PREVENTION

Bone loss caused by glucocorticoids can be prevented, and the risk of fractures significantly reduced. Strategies must include using the lowest dose of glucocorticoid for disease management. Topical and inhaled routes of administration are preferred, where appropriate. Risk factor reduction is important, including smoking cessation, limitation of alcohol consumption, and participation in weight-bearing and resistance exercise, when appropriate. All patients should receive an adequate calcium and vitamin D intake from the diet or from supplements.

## TREATMENT

### Glucocorticoid-Induced Osteoporosis

Several bisphosphonates (alendronate, risedronate, and zoledronic acid) have been demonstrated in large clinical trials to reduce the risk of fractures in patients being treated with glucocorticoids and are FDA approved for the treatment of GCIO. Teriparatide is also approved for treatment of GCIO. In one trial comparing teriparatide to alendronate, BMD increases were much greater and vertebral fracture risk reduction far more substantial with teriparatide compared to alendronate. A study of denosumab indicates greater

efficacy of denosumab compared with risedronate for treatment of GCIO. The American College of Rheumatology has published guidelines for the management of GCIO.

## FURTHER READING

- Black DM, Rosen CJ: Postmenopausal osteoporosis. *N Engl J Med* 374:2096, 2016.
- Black DM et al: Atypical femur fracture risk versus fragility fracture prevention with bisphosphonates. *N Engl J Med* 383:743, 2020.
- Compston J: Glucocorticoid-induced osteoporosis: An update. *Endocrine* 61:7, 2018.
- Cosman F et al: Spine fracture prevalence in a nationally representative sample of US women and men aged >/=40 years: results from the National Health and Nutrition Examination Survey (NHANES) 2013–2014. *Osteoporos Int* 28:2319, 2017.
- Cosman F et al: Treatment sequence matters: Anabolic and antiresorptive therapy for osteoporosis. *J Bone Miner Res* 32:198, 2017.
- Khosla S, Hofbauer LC: Osteoporosis treatment: Recent developments and ongoing challenges. *Lancet Diabetes Endocrinol* 5:898, 2017.
- Reid IR: A broader strategy for osteoporosis interventions. *Nat Rev Endocrinol* 16:333, 2020.
- Roux C, Briot K: Imminent fracture risk. *Osteoporos Int* 28:1765, 2017.

**412**

## Paget's Disease and Other Dysplasias of Bone

Rajesh K. Jain, Tamara J. Vokes

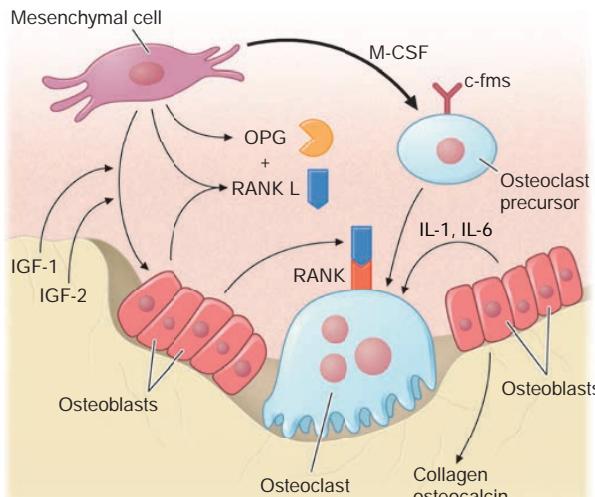
## PAGET'S DISEASE OF BONE

Paget's disease is a localized bone-remodeling disorder that affects widespread, noncontiguous areas of the skeleton. The pathologic process is initiated by overactive osteoclastic bone resorption followed by a compensatory increase in osteoblastic new bone formation, resulting in a structurally disorganized mosaic of woven and lamellar bone. Pagetic bone is expanded, less compact, and more vascular; thus, it is more susceptible to deformities and fractures. Although most patients are asymptomatic, symptoms resulting directly from bony involvement (bone pain, secondary arthritis, fractures) or secondarily from the expansion of bone causing compression of surrounding neural tissue are not uncommon.

**Epidemiology** There is a marked geographic variation in the frequency of Paget's disease, with high prevalence in Western Europe (Great Britain, France, and Germany, but not Switzerland or Scandinavia) and among those who have immigrated to Australia, New Zealand, South Africa, and North and South America. The disease is rare in native populations of the Americas, Africa, Asia, and the Middle East; when it does occur, the affected subjects usually have evidence of European ancestry, supporting the migration theory. For unclear reasons, the prevalence and severity of Paget's disease are decreasing, and the age of diagnosis is increasing.

The prevalence is greater in males and increases with age. Autopsy series reveal Paget's disease in ~3% of those over age 40. Prevalence of positive skeletal radiographs in patients aged >55 years is 2.5% for men and 1.6% for women. Elevated alkaline phosphatase (ALP) levels in asymptomatic patients have an age-adjusted incidence of 12.7 and 7 per 100,000 person-years in men and women, respectively.

**Etiology** The etiology of Paget's disease of bone remains unknown, but evidence supports both genetic and viral etiologies. A positive



**FIGURE 412-1** Diagram illustrating factors that promote differentiation and function of osteoclasts and osteoblasts and the role of the RANK pathway. Stromal bone marrow (mesenchymal) cells and differentiated osteoblasts produce multiple growth factors and cytokines, including macrophage colony-stimulating factor (M-CSF), to modulate osteoclastogenesis. RANK (receptor activator of nuclear factor- $\kappa$ B [NF- $\kappa$ B] ligand) is produced by osteoblast progenitors and mature osteoblasts and can bind to a soluble decoy receptor known as osteoprotegerin (OPG) to inhibit RANKL action. Alternatively, a cell-cell interaction between osteoblast and osteoclast progenitors allows RANKL to bind to its membrane-bound receptor, RANK, thereby stimulating osteoclast differentiation and function. RANK binds intracellular proteins called tumor necrosis factor receptor-associated factors (TRAFs) that mediate receptor signaling through transcription factors such as NF- $\kappa$ B. M-CSF binds to its receptor, c-fms, which is the cellular homologue of the fms oncogene. See text for the potential role of these pathways in disorders of osteoclast function such as Paget's disease and osteopetrosis. IGF, insulin-like growth factor; IL, interleukin.

family history is found in 15–25% of patients and, when present, raises the prevalence of the disease seven- to tenfold among first-degree relatives.

A clear genetic basis has been established for several rare familial bone disorders that clinically and radiographically resemble Paget's disease but have more severe presentation and earlier onset. A homozygous deletion of the *TNFRSF11B* gene, which encodes osteoprotegerin (Fig. 412-1), causes *juvenile Paget's disease*, also known as *familial idiopathic hyperphosphatasia*, a disorder characterized by uncontrolled osteoclastic differentiation and resorption. Familial patterns of disease in several large kindred are consistent with an autosomal dominant pattern of inheritance with variable penetrance. *Familial expansile osteolysis, expansile skeletal hyperphosphatasia, and early-onset Paget's disease* are associated with mutations in the *TNFRSF11A* gene, which encodes RANK (receptor activator of nuclear factor- B), a member of the tumor necrosis factor superfamily critical for osteoclast differentiation (Fig. 412-1). A mutation in profilin 1, a small actin protein that acts as a tumor suppressor, also causes early-onset Paget's disease with a predisposition for the development of osteosarcoma. Finally, mutations in the gene for valosin-containing protein cause a rare syndrome with autosomal dominant inheritance and variable penetrance known as *inclusion body myopathy with Paget's disease and frontotemporal dementia (IBMPFD)*. The role of genetic factors is less clear in the more common form of late-onset Paget's disease. The most common mutations identified in familial and sporadic cases of Paget's disease have been in the *SQSTM1* gene (sequestosome-1 or p62 protein) in the C-terminal ubiquitin-binding domain. The other candidate genes include *CSF1* (1p13), which encodes macrophage colony-stimulating factor (M-CSF), a cytokine that is required for osteoclast differentiation; *RIN3* (14q32), which encodes a guanine exchange factor called Rab and Ras interactor 3; *OPTN* (10p13), which is involved in regulating nuclear factor (NF)- B; *TNFRSF11A* (18q21), which encodes

receptor activator of NF- B (RANK), a receptor that is essential for osteoclast differentiation; and *TM7SF4*, which encodes dendritic cell-specific transmembrane protein (DC-STAMP), a molecule that is essential for fusion of the osteoclast. The phenotypic variability in patients with *SQSTM1* mutations suggests that additional factors, such as other genetic influences or viral infection, may influence clinical expression of the disease.

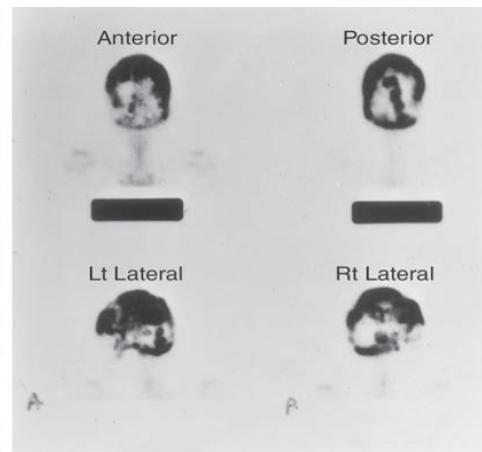
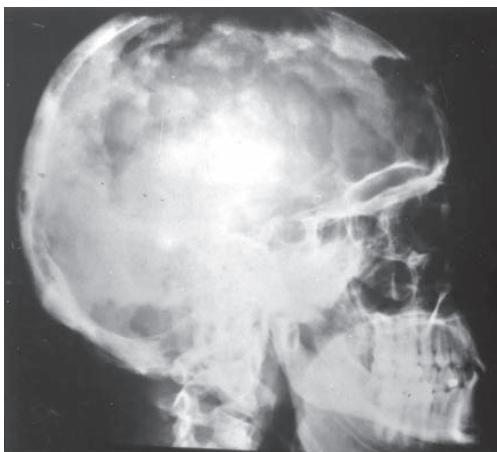
Several lines of evidence suggest that a viral infection may contribute to the clinical manifestations of Paget's disease, including (1) the presence of cytoplasmic and nuclear inclusions resembling paramyxoviruses (measles and respiratory syncytial virus) in pagetic osteoclasts and (2) viral mRNA in precursor and mature osteoclasts. The viral etiology is further supported by conversion of osteoclast precursors to pagetic-like osteoclasts by vectors containing the measles virus nucleocapsid or matrix genes. The decline in the incidence of Paget's disease coincides with the widespread vaccination against measles, also consistent with the potential role of virus in the development of the disease. However, the viral etiology has been questioned by the inability to culture a live virus from pagetic bone and by failure to clone the full-length viral genes from material obtained from patients with Paget's disease. Furthermore, patients with Paget's disease do not have higher antibody levels against paramyxoviruses or measles as compared to controls, nor do antibody levels correlate with disease severity in those with Paget's disease.

**Pathophysiology** The principal abnormality in Paget's disease is the increased number and activity of osteoclasts. Pagetic osteoclasts are large, increased 10- to 100-fold in number, and have a greater number of nuclei (as many as 100 compared to 3–5 nuclei in the normal osteoclast). The overactive osteoclasts may create a sevenfold increase in resorptive surfaces and an erosion rate of 9  $\mu\text{g}/\text{d}$  (normal is 1  $\mu\text{g}/\text{d}$ ). Several causes for the increased number and activity of pagetic osteoclasts have been identified: (1) osteoclastic precursors are hypersensitive to  $1,25(\text{OH})_2\text{D}_3$ ; (2) osteoclasts are hyperresponsive to RANK ligand (RANKL), the osteoclast stimulatory factor that mediates the effects of most osteotropic factors on osteoclast formation; (3) marrow stromal cells from pagetic lesions have increased RANKL expression; (4) osteoclast precursor recruitment is increased by interleukin (IL) 6, which is increased in the blood of patients with active Paget's disease and is overexpressed in pagetic osteoclasts; (5) expression of the protooncogene *c-fos*, which increases osteoclastic activity, is increased; and (6) the antiapoptotic oncogene *Bcl-2* in pagetic bone is overexpressed. Numerous osteoblasts are recruited to active resorption sites and produce large amounts of new bone matrix. As a result, bone turnover is high, and bone mass is normal or increased, not reduced, unless there is concomitant deficiency of calcium and/or vitamin D.

The characteristic feature of Paget's disease is increased bone resorption accompanied by accelerated bone formation. An initial osteolytic phase involves prominent bone resorption and marked hypervascularization. Radiographically, this manifests as an advancing lytic wedge, or "blade of grass" lesion. The second phase is a period of very active bone formation and resorption that replaces normal lamellar bone with haphazard (woven) bone. Fibrous connective tissue may replace normal bone marrow. In the final sclerotic phase, bone resorption declines progressively and leads to a hard, dense, less vascular pagetic or mosaic bone, which represents the so-called burned-out phase of Paget's disease. All three phases may be present at the same time at different skeletal sites.

**Clinical Manifestations** Diagnosis is often made in asymptomatic patients because they have elevated ALP levels on routine blood chemistry testing or an abnormality on a skeletal radiograph obtained for another indication. The skeletal sites most commonly involved are the pelvis, vertebral bodies, skull, femur, and tibia. Familial cases with an early presentation often have numerous active sites of skeletal involvement.

The most common presenting symptom is pain, which may result from increased bony vascularity, expanding lytic lesions, fractures, bowing, or other deformities. Bowing of the femur or tibia causes



**FIGURE 412-2** A 48-year-old woman with Paget's disease of the skull. *Left.* Lateral radiograph showing areas of both bone resorption and sclerosis. *Right.*  $^{99m}\text{Tc}$  hydroxymethylene diphosphonate (HDP) bone scan with anterior, posterior, and lateral views of the skull showing diffuse isotope uptake by the frontal, parietal, occipital, and petrous bones.

gait abnormalities and abnormal mechanical stresses with secondary osteoarthritis of the hip or knee joints. Long bone bowing also causes extremity pain by stretching the muscles attached to the bone softened by the pagetic process. Back pain results from enlarged pagetic vertebrae, vertebral compression fractures, spinal stenosis, degenerative changes of the joints, and altered body mechanics with kyphosis and forward tilt of the upper back. Rarely, spinal cord compression may result from bone enlargement or from the vascular steal syndrome. Skull involvement may cause headaches, symmetric or asymmetric enlargement of the parietal or frontal bones (frontal bossing), and increased head size. Cranial expansion may narrow cranial foramen and cause neurologic complications including hearing loss from cochlear nerve damage from temporal bone involvement, cranial nerve palsies, and softening of the base of the skull (*platybasia*) with the risk of brainstem compression. Pagetic involvement of the facial bones may cause facial deformity; loss of teeth and other dental conditions; and, rarely, airway compression.

Fractures are serious complications of Paget's disease and usually occur in long bones at areas of active or advancing lytic lesions. Common fracture sites are the femoral shaft and subtrochanteric regions. Neoplasms arising from pagetic bone are rare (<0.5%). The incidence of sarcoma appears to be decreasing, possibly because of earlier, more effective treatment with potent antiresorptive agents. The majority of tumors are osteosarcomas, which usually present with new pain in a long-standing pagetic lesion. Osteoclast-rich benign giant cell tumors may arise in areas adjacent to pagetic bone, and they respond to glucocorticoid therapy.

Cardiovascular complications may occur in patients with involvement of large (15–35%) portions of the skeleton and a high degree of disease activity (ALP four times above normal). The extensive arteriovenous shunting and marked increases in blood flow through the vascular pagetic bone lead to a high-output state and cardiac enlargement. However, high-output heart failure is relatively rare and usually develops in patients with concomitant cardiac pathology. In addition, calcific aortic stenosis and diffuse vascular calcifications have been associated with Paget's disease.

**Diagnosis** The diagnosis may be suggested on clinical examination by the presence of an enlarged skull with frontal bossing, bowing of an extremity, or short stature with simian posturing. An extremity with an area of warmth and tenderness to palpation may suggest an underlying pagetic lesion. Other findings include bony deformity of the pelvis, skull, spine, and extremities; arthritic involvement of the joints adjacent to lesions; and leg-length discrepancy resulting from deformities of the long bones.

Paget's disease is usually diagnosed from radiologic and biochemical abnormalities. Radiographic findings typical of Paget's disease

include enlargement or expansion of an entire bone or area of a long bone, cortical thickening, coarsening of trabecular markings, and typical lytic and sclerotic changes. Skull radiographs (Fig. 412-2) reveal regions of "cotton wool," or osteoporosis circumscripita, thickening of diploic areas, and enlargement and sclerosis of a portion or all of one or more skull bones. Vertebral cortical thickening of the superior and inferior end plates creates a "picture frame" vertebra. Diffuse radiodense enlargement of a vertebra is referred to as "ivory vertebra." Pelvic radiographs may demonstrate disruption or fusion of the sacroiliac joints; porotic and radiodense lesions of the ilium with whorls of coarse trabeculation; thickened and sclerotic iliopectineal line (brim sign); and softening with protrusio acetabuli, with axial migration of the hips and functional flexion contracture. Radiographs of long bones reveal bowing deformity and typical pagetic changes of cortical thickening and expansion and areas of lucency and sclerosis (Fig. 412-3). Radionuclide  $^{99m}\text{Tc}$  bone scans are less specific but are more sensitive than standard radiographs for identifying sites of active skeletal lesions. Although computed tomography (CT) and magnetic resonance imaging (MRI) studies are not necessary in most cases, CT may be useful for the assessment of possible fracture, and MRI is necessary to assess the possibility of sarcoma, giant cell tumor, or



**FIGURE 412-3** Radiograph of a 73-year-old man with Paget's disease of the right proximal femur. Note the coarsening of the trabecular pattern with marked cortical thickening and narrowing of the joint space consistent with osteoarthritis secondary to pagetic deformity of the right femur.

Biochemical evaluation is useful in the diagnosis and management of Paget's disease. The marked increase in bone turnover can be monitored using biochemical markers of bone formation and resorption. The parallel rise in markers of bone formation and resorption confirms the coupling of bone formation and resorption in Paget's disease. The degree of bone marker elevation reflects the extent and severity of the disease. For most patients, serum total ALP remains the test of choice both for diagnosis and assessing response to therapy. Occasionally, a symptomatic patient with evidence of progression at a single site may have a normal total ALP level but increased bone-specific ALP. For unclear reasons, serum osteocalcin, another marker of bone formation, is not always elevated and is not recommended for use in diagnosis or management of Paget's disease. In contrast, bone formation marker P1NP does reflect the activity of the disease and can be used instead of total ALP. Bone resorption markers (serum or urine N-telopeptide or C-telopeptide measured in the blood or urine) are also elevated in active Paget's disease and decrease more rapidly in response to therapy than does ALP.

Serum calcium and phosphate levels are normal in Paget's disease. Immobilization of a patient with active Paget's disease may rarely cause hypercalcemia and hypercalciuria and increase the risk for nephrolithiasis. However, the discovery of hypercalcemia, even in the presence of immobilization, should prompt a search for another cause of hypercalcemia. In contrast, hypocalcemia or mild secondary hyperparathyroidism may develop in Paget's patients with very active bone formation and insufficient calcium and vitamin D intake, particularly during bisphosphonate therapy when bone resorption is rapidly suppressed and active bone formation continues. Therefore, adequate calcium and vitamin D intake should be instituted prior to administration of bisphosphonates.

## TREATMENT

### Paget's Disease of Bone

The development of effective and potent pharmacologic agents (**Table 412-1**) has changed the treatment philosophy from treating only symptomatic patients to treating asymptomatic patients who are at risk for complications. According to the Endocrine Society Clinical Practice Guidelines published in 2014, pharmacologic therapy is indicated for most patients with active Paget's disease who are at risk of complications. Treatment may be initiated to control symptoms caused by metabolically active Paget's disease such as bone pain, fracture, headache, pain from pagetic radiculopathy or arthropathy, or neurologic complications; to decrease local blood flow and minimize operative blood loss in patients who need surgery at an active pagetic site; to reduce hypercalciuria that may occur during immobilization; and to decrease the risk of complications when disease activity is high (elevated ALP) and when the site of involvement involves weight-bearing bones, areas adjacent to

**TABLE 412-1** Pharmacologic Agents Approved for Treatment of Paget's Disease

NAME	DOSE AND MODE OF DELIVERY	NORMALIZATION OF ALKALINE PHOSPHATASE (ALP)
Zoledronic acid	5 mg IV over 15 min	90% of patients at 6 mo
Pamidronate	30 mg/d IV over 4 h on 3 days	-50% of patients
Risedronate	30 mg/d PO for 2 mo	73% of patients
Alendronate	40 mg/d PO for 6 mo	63% of patients
Tiludronate	800 mg/d PO for 3 mo	35% of patients
Etidronate	200–400 mg/d PO × 6 mo	15% of patients
Calcitonin (Miacalcin)	100 U SC daily for 6–18 mo (may reduce to 50 U 3× per week)	(Reduction of ALP by up to 50%)

major joints, vertebral bodies, and the skull. Whether or not early therapy prevents late complications remains to be determined. Randomized studies from the United Kingdom showed no difference in bone pain, fracture rates, quality of life, and hearing loss between patients who received pharmacologic therapy to control symptoms (bone pain) and those receiving bisphosphonates to normalize serum ALP. However, the conclusions of these studies are debatable since the majority of subjects had already received bisphosphonate therapy in the past, perhaps limiting generalizability, and because the bone deformities that occur with Paget's disease may take many years to manifest. It seems likely that the restoration of normal bone architecture following suppression of pagetic activity will prevent further deformities and complications.

Agents approved for treatment of Paget's disease suppress the very high rates of bone resorption and secondarily decrease the high rates of bone formation (Table 412-1). As a result of decreasing bone turnover, pagetic structural patterns, including areas of poorly mineralized woven bone, are replaced by more normal cancellous or lamellar bone. Reduced bone turnover can be documented by a decline in serum formation markers (ALP and P1NP) and urine or serum resorption markers (N-telopeptide, C-telopeptide).

Bisphosphonates are the mainstay of pharmacologic therapy of Paget's disease. Among them, zoledronic acid is currently recommended as the first choice, particularly for those who have severe disease or need rapid normalization of bone turnover (neurologic symptoms, severe bone pain due to a lytic lesion, risk of an impending fracture, or pretreatment prior to elective surgery in an area of active disease). Zoledronic acid normalized bone turnover faster and in a high proportion of patients (>90%) than oral bisphosphonates with the therapeutic effect persisting for months or even years. It is given at a dose of 5 mg as an intravenous infusion over 20 min, although slower rates of infusion are recommended for elderly or those with mild impairment of renal function. More significant renal impairment (glomerular filtration rate <35 mL/min) is a contraindication for use of zoledronic acid due to higher risk of further deterioration of renal function. About 20–25% of patients experience a flulike syndrome after the first infusion, which can be partly ameliorated by pretreatment with acetaminophen or nonsteroidal anti-inflammatory drugs (NSAIDs). Oral bisphosphonates, alendronate and risedronate, can be used in subjects who have mild disease or some degree of renal impairment. Oral bisphosphonates should be taken first thing in the morning on an empty stomach, followed by maintenance of upright posture with no food, drink, or other medications for 30–60 min. The first clinically useful agent, etidronate, is no longer used due to its low potency and higher risk of inducing osteomalacia. The efficacy of different agents, based on their ability to normalize or decrease ALP levels, is summarized in Table 412-1, although the response rates are not comparable because they are obtained from different studies.

The subcutaneous injectable form of salmon calcitonin is approved for the treatment of Paget's disease but is rarely used due to its low potency and should be reserved for patients who either do not tolerate bisphosphonates or have a contraindication to their use. For patients with contraindication to bisphosphonates, another alternative is denosumab, an antibody to RANKL, which has been reported to result in reduction in ALP. However, it has not been approved for this indication and has less complete and less durable effect than bisphosphonates.

## SCLEROSING BONE DISORDERS

### OSTEOPETROSIS

*Osteopetrosis* refers to a group of disorders caused by severe impairment of osteoclast-mediated bone resorption. Other terms that are often used include marble bone disease, which captures the solid x-ray appearance of the involved skeleton, and Albers-Schonberg disease, which refers to the milder, adult form of osteopetrosis also known as autosomal dominant osteopetrosis type II. The major types of

osteopetrosis include malignant (severe, infantile, autosomal recessive) osteopetrosis and benign (adult, autosomal dominant) osteopetrosis types I and II. A rare autosomal recessive intermediate form has a more benign prognosis. Autosomal recessive carbonic anhydrase (CA) II deficiency produces osteopetrosis of intermediate severity associated with renal tubular acidosis and cerebral calcification.

**Etiology and Genetics** Naturally occurring and gene-knockout animal models with phenotypes similar to those of the human disorders have been used to explore the genetic basis of osteopetrosis. The primary defect in osteopetrosis is the loss of osteoclastic bone resorption and preservation of normal osteoblastic bone formation. Osteoprotegerin (OPG) is a soluble decoy receptor that binds osteoblast-derived RANK ligand, which mediates osteoclast differentiation and activation (Fig. 412-1). Transgenic mice that overexpress OPG develop osteopetrosis, presumably by blocking RANK ligand. Mice deficient in RANK lack osteoclasts and develop severe osteopetrosis.

Recessive mutations of CA II prevent osteoclasts from generating an acid environment in the clear zone between its ruffled border and the adjacent mineral surface. Absence of CA II, therefore, impairs osteoclastic bone resorption. Other forms of human disease have less clear genetic defects. About one-half of the patients with malignant infantile osteopetrosis have a mutation in the *TCIRG1* gene encoding the osteoclast-specific subunit of the vacuolar proton pump, which mediates the acidification of the interface between bone mineral and the osteoclast ruffled border. Mutations in the *CLCN7* chloride channel gene cause autosomal dominant osteopetrosis type II.

**Clinical Presentation** The incidence of autosomal recessive severe (malignant) osteopetrosis ranges from 1 in 200,000 to 1 in 500,000 live births. As bone and cartilage fail to undergo modeling, paralysis of one or more cranial nerves may occur due to narrowing of the cranial foramen. Failure of skeletal modeling also results in inadequate marrow space, leading to extramedullary hematopoiesis with hypersplenism and pancytopenia. Hypocalcemia due to lack of osteoclastic bone resorption may occur in infants and young children. The untreated infantile disease is fatal, often before age 5.

Adult (benign) osteopetrosis is an autosomal dominant disease that is usually diagnosed by the discovery of typical skeletal changes in young adults who undergo radiologic evaluation of a fracture. The prevalence is 1 in 100,000 to 1 in 500,000 adults. The course is not always benign, because fractures may be accompanied by loss of vision, deafness, psychomotor delay, mandibular osteomyelitis, and other complications usually associated with the juvenile form. In some kindred, nonpenetrance results in skip generations, while in other families, severely affected children are born into families with benign disease. The milder form of the disease does not usually require treatment.

**Radiography** Typically, there are generalized symmetric increases in bone mass with thickening of both cortical and trabecular bone. Diaphyses and metaphyses are broadened, and alternating sclerotic and lucent bands may be seen in the iliac crests, at the ends of long bones, and in vertebral bodies. The cranium is usually thickened, particularly at the base of the skull, and the paranasal and mastoid sinuses are underpneumatized.

**Laboratory Findings** The only significant laboratory findings are elevated serum levels of osteoclast-derived tartrate-resistant acid phosphatase (TRAP) and the brain isoenzyme of creatine kinase. Serum calcium may be low in severe disease, and parathyroid hormone and 1,25-dihydroxyvitamin D levels may be elevated in response to hypocalcemia.

## TREATMENT

### Osteopetrosis

Allogeneic human leukocyte antigen (HLA)-identical bone marrow transplantation has been successful in some children. Following transplantation, the marrow contains progenitor cells and normally functioning osteoclasts. With long-term follow-up after

transplantation, radiographic improvements, such as improvements in Erlenmeyer flask deformities, are seen, although there is not complete normalization. A cure is most likely when children are transplanted before age 4. Marrow transplantation from nonidentical HLA-matched donors has a much higher failure rate. Limited studies in small numbers of patients have suggested variable benefits following treatment with interferon -1 , 1,25-dihydroxyvitamin D (which stimulates osteoclasts directly), methylprednisolone, and a low-calcium/high-phosphate diet.

Surgical intervention is indicated to decompress optic or auditory nerve compression. Orthopedic management is required for the surgical treatment of fractures and their complications, including malunion and postfracture deformity.

### PYKNODYSOSTOSIS

This is an autosomal recessive form of osteosclerosis that is believed to have affected the French impressionist painter Henri de Toulouse-Lautrec. The molecular basis involves mutations in the gene that encodes cathepsin K, a lysosomal metalloproteinase highly expressed in osteoclasts and important for bone-matrix degradation. Osteoclasts are present but do not function normally. Pyknodysostosis is a form of short-limb dwarfism that presents with frequent fractures but usually a normal life span. Clinical features include short stature; kyphoscoliosis and deformities of the chest; high arched palate; proptosis; blue sclerae; dysmorphic features including small face and chin, fronto-occipital prominence, pointed beaked nose, large cranium, and obtuse mandibular angle; and small, square hands with hypoplastic nails. Radiographs demonstrate a generalized increase in bone density, but in contrast to osteopetrosis, the long bones are normally shaped. Separated cranial sutures, including the persistent patency of the anterior fontanel, are characteristic of the disorder. There may also be hypoplasia of the sinuses, mandible, distal clavicles, and terminal phalanges. Persistence of deciduous teeth and sclerosis of the calvarium and base of the skull are also common. Histologic evaluation shows normal cortical bone architecture with decreased osteoblastic and osteoclastic activities. Serum chemistries are normal, and unlike osteopetrosis, there is no anemia. There is no known treatment for this condition, and there are no reports of attempted bone marrow transplant.

### PROGRESSIVE DIAPHYSEAL DYSPLASIA

Also known as *Camurati-Engelmann disease*, progressive diaphyseal dysplasia is an autosomal dominant disorder that is characterized radiographically by diaphyseal hyperostosis and a symmetric thickening and increased diameter of the endosteal and periosteal surfaces of the diaphyses of the long bones, particularly the femur and tibia, and, less often, the fibula, radius, and ulna. The genetic defect responsible for the disease has been localized to the area of chromosome 19q13.2 encoding tumor growth factor (TGF)- 1. The mutation promotes activation of TGF- 1. The clinical severity is variable. The most common presenting symptoms are pain and tenderness of the involved areas, fatigue, muscle wasting, and gait disturbance. The weakness may be mistaken for muscular dystrophy. Characteristic body habitus includes thin limbs with little muscle mass yet prominent and palpable bones and, when the skull is involved, large head with prominent forehead and proptosis. Patients may also display signs of cranial nerve palsies, hydrocephalus, central hypogonadism, and Raynaud's phenomenon. Radiographically, patchy progressive endosteal and periosteal new bone formation is observed along the diaphyses of the long bones. Bone scintigraphy shows increased radiotracer uptake in involved areas.

Treatment with low-dose glucocorticoids relieves bone pain and may reverse the abnormal bone formation. Intermittent bisphosphonate therapy has produced clinical improvement in a limited number of patients. Disease activity may also attenuate as patients enter adulthood.

### HYPERTOSTOSIS CORTICALIS GENERALISATA

This is also known as *van Buchem's disease*; it is an autosomal recessive disorder characterized by endosteal hyperostosis in which osteosclerosis involves the skull, mandible, clavicles, and ribs. The major

manifestations are due to narrowed cranial foramen with neural compressions that may result in optic atrophy, facial paralysis, and deafness. Adults may have an enlarged mandible. Serum ALP levels may be elevated, which reflect the uncoupled bone remodeling with high osteoblastic formation rates and low osteoclastic resorption. As a result, there is increased accumulation of normal bone. Endosteal hyperostosis with syndactyly, known as *sclerosteosis*, is a more severe form. The genetic defects for both sclerosteosis and van Buchem's disease have been associated with mutations in the *SOST* gene.

### MELORHEOSTOSIS

Melorheostosis (Greek, “flowing hyperostosis”) may occur sporadically or follow a pattern consistent with an autosomal recessive disorder. The major manifestation is progressive linear hyperostosis in one or more bones of one limb, usually a lower extremity. The name comes from the radiographic appearance of the involved bone, which resembles melted wax that has dripped down a candle. Symptoms appear during childhood as pain or stiffness in the area of sclerotic bone. There may be associated ectopic soft tissue masses, composed of cartilage or osseous tissue, and skin changes overlying the involved bone, consisting of scleroderma-like areas and hypertrichosis. The disease does not progress in adults, but pain and stiffness may persist. Laboratory tests are unremarkable. Somatic mutations in *MAP2K1*, which increases MEK1 activity downstream of the RAS pathway, and *SMAD3*, which upregulates the TGF- $\beta$ /SMAD pathway, have been identified in affected bone in patients with melorheostosis. There is no specific treatment. Surgical interventions to correct contractures are often unsuccessful.

### OSTEOPOIKILOSIS

The literal translation of osteopoikilosis is “spotted bones”; it is a benign autosomal dominant condition in which numerous small, variably shaped (usually round or oval) foci of bony sclerosis are seen in the epiphyses and adjacent metaphyses. The lesions may involve any bone except the skull, ribs, and vertebrae. They may be misidentified as metastatic lesions. The main differentiating points are that bony lesions of osteopoikilosis are stable over time and do not accumulate radionucleotide on bone scanning. In some kindred, osteopoikilosis is associated with connective tissue nevi known as *dermatofibrosis lenticularis disseminata*, also known as *Buschke-Ollendorff syndrome*. Most cases are caused by mutations in *LEMD3*, which is involved with bone morphogenetic protein (BMP) signaling. Histologic inspection reveals thickened but otherwise normal trabeculae and islands of normal cortical bone. No treatment is indicated.

### HEPATITIS C-ASSOCIATED OSTEOSCLEROSIS

Hepatitis C-associated osteosclerosis (HCAO) is a rare acquired diffuse osteosclerosis in adults with prior hepatitis C infection. After a latent period of several years, patients develop diffuse appendicular bone pain and a generalized increase in bone mass with elevated serum ALP. Bone biopsy and histomorphometry reveal increased rates of bone formation, decreased bone resorption with a marked decrease in osteoclasts, and dense lamellar bone. One patient had increased serum OPG levels, and bone biopsy showed large numbers of osteoblasts positive for OPG and reduced osteoclast number. Empirical therapy includes pain control, and there may be beneficial response to bisphosphonate. Long-term antiviral therapy may reverse the bone disease.

## DISORDERS ASSOCIATED WITH DEFECTIVE MINERALIZATION

### HYPOPHOSPHATASIA

This is a rare inherited disorder that presents as rickets in infants and children or osteomalacia in adults with paradoxically low serum levels of ALP. The frequency of the severe neonatal and infantile forms is about 1 in 100,000 live births in Canada, where the disease is most common because of its high prevalence among Mennonites and Hutterites. It is rare in African Americans. The severity of the disease is remarkably variable, ranging from intrauterine death associated with profound skeletal hypomineralization at one extreme to premature

tooth loss as the only manifestation in some adults. Severe cases are inherited in an autosomal recessive manner, but the genetic patterns are less clear for the milder forms. The disease is caused by a deficiency of tissue nonspecific (bone/liver/kidney) ALP (TNSALP), which, although ubiquitous, results only in bone abnormalities. Protein levels and functions of the other ALP isozymes (germ cell, intestinal, placental) are normal. Defective ALP permits accumulation of its major naturally occurring substrates including phosphoethanolamine (PEA), inorganic pyrophosphate (PPi), and pyridoxal 5'-phosphate (PLP). The accumulation of PPi interferes with mineralization through its action as a potent inhibitor of hydroxyapatite crystal growth.

Perinatal hypophosphatasia becomes manifest during pregnancy and is often complicated by polyhydramnios and intrauterine death. The infantile form becomes clinically apparent before the age of 6 months with failure to thrive, rachitic deformities, functional craniosynostosis despite widely open fontanels (which are actually hypomineralized areas of the calvarium), raised intracranial pressure, and flail chest with predisposition to pneumonia. Hypercalcemia and hypercalciuria are common. This form has a mortality rate of ~50%. Prognosis seems to improve for the children who survive infancy. Childhood hypophosphatasia has variable clinical presentation. Premature loss of deciduous teeth (before age 5) is the hallmark of the disease. Rickets causes delayed walking with waddling gait, short stature, and dolichocephalic skull with frontal bossing. The disease often improves during puberty but may recur in adult life. Adult hypophosphatasia presents during middle age with painful, poorly healing metatarsal stress fractures or thigh pain due to femoral pseudofractures. Presentation may be subtle with muscle pain or recurring headaches as the predominant symptoms. It is important to recognize hypophosphatasia in adults because treatment with bisphosphonates can result in increased rather than decreased bone fragility.

Laboratory investigation reveals low ALP levels and normal or elevated levels of serum calcium and phosphorus despite clinical and radiologic evidence of rickets or osteomalacia. Serum parathyroid hormone, 25-hydroxyvitamin D, and 1,25-dihydroxyvitamin D levels are normal. The elevation of PLP is specific for the disease and may even be present in asymptomatic parents of severely affected children. Because vitamin B<sub>6</sub> increases PLP levels, vitamin B<sub>6</sub> supplements should be discontinued 1 week before testing. Clinical testing is available to detect loss-of-function mutation(s) within the *ALPL* gene that encodes TNSALP.

In contrast to other forms of rickets and osteomalacia, calcium and vitamin D supplementation should be avoided because they may aggravate hypercalcemia and hypercalciuria. A low-calcium diet, glucocorticoids, and calcitonin have been used in a small number of patients with variable responses. Because fracture healing is poor, placement of intramedullary rods is best for acute fracture repair and for prophylactic prevention of fractures. In 2015, asfotase alfa, a tissue-nonspecific ALP was approved as enzyme replacement therapy for the perinatal/infantile- and juvenile-onset forms. With 7 years of therapy, children with perinatal/infantile forms showed sustained improvements in mineralization, along with improvements in other features, such as respiratory function and growth.

### AXIAL OSTEOMALACIA

This is a rare disorder characterized by defective skeletal mineralization despite normal serum calcium and phosphate levels. Clinically, the disorder presents in middle-aged or elderly men with chronic axial skeletal discomfort. Cervical spine pain may also be present. Radiographic findings are mainly osteosclerosis due to coarsened trabecular patterns typical of osteomalacia. Spine, pelvis, and ribs are most commonly affected. Histologic changes show defective mineralization and flat, inactive osteoblasts. The primary defect appears to be an acquired defect in osteoblast function. The course is benign, and there is no established treatment. Calcium and vitamin D therapies are not effective.

### FIBROGENESIS IMPERFECTA OSSUM

This is a rare condition of unknown etiology. It presents in both sexes; in middle age or later; and with progressive, intractable skeletal pain

and fractures; worsening immobilization; and a debilitating course. The only biochemical abnormality is elevated ALP. Radiographic evaluation reveals generalized osteomalacia, osteopenia, and occasional pseudofractures. Histologic features include a tangled pattern of collagen fibrils with abundant osteoblasts and osteoclasts. Use of growth hormone led to substantial short-term clinical improvement in two adult patients, but long-term outcomes are unknown. No other effective treatment is known. Spontaneous remission has been reported in a small number of patients.

## FIBROUS DYSPLASIA AND MCCUNE-ALBRIGHT SYNDROME

Fibrous dysplasia is a sporadic disorder characterized by the presence of one (monostotic) or more (polyostotic) expanding fibrous skeletal lesions composed of bone-forming mesenchyme. The association of the polyostotic form with café au lait spots and hyperfunction of an endocrine system such as pseudoprecocious puberty of ovarian origin is known as *McCune-Albright syndrome* (MAS). A spectrum of the phenotypes is caused by activating mutations in the *GNAS1* gene, which encodes the subunit of the stimulatory G protein ( $G_s$ ). As the postzygotic mutations occur at different stages of early development, the extent and type of tissue affected are variable and explain the mosaic pattern of skin and bone changes. GTP binding activates the  $G_s$  regulatory protein and mutations in regions of  $G_s$  that selectively inhibit GTPase activity, which results in constitutive stimulation of the cyclic AMP-protein kinase A signal transduction pathway. Such mutations of the  $G_s$  protein-coupled receptor may cause autonomous function in bone (parathyroid hormone receptor); skin (melanocyte-stimulating hormone receptor); and various endocrine glands including ovary (follicle-stimulating hormone receptor), thyroid (thyroid-stimulating hormone receptor), adrenal (adrenocorticotrophic hormone receptor), and pituitary (growth hormone-releasing hormone receptor). The skeletal lesions are composed largely of mesenchymal cells that do not differentiate into osteoblasts, resulting in the formation of imperfect bone. In some areas of bone, fibroblast-like cells develop features of osteoblasts in that they produce extracellular matrix that organizes into woven bone. Calcification may occur in some areas. In other areas, cells have features of chondrocytes and produce cartilage-like extracellular matrix.

**Clinical Presentation** Fibrous dysplasia occurs with equal frequency in both sexes, whereas MAS with precocious puberty is more common (10:1) in girls. The monostotic form is the most common and is usually diagnosed in patients between 20 and 30 years of age without associated skin lesions. The polyostotic form typically manifests in children <10 years old and may progress with age. Early-onset disease is generally more severe. Lesions may become quiescent in puberty and progress during pregnancy or with estrogen therapy. In polyostotic fibrous dysplasia, the lesions most commonly involve the maxilla and other craniofacial bones, ribs, and metaphyseal or diaphyseal portions of the proximal femur or tibia. Expanding bone lesions may cause pain, deformity, fractures, and nerve entrapment. Sarcomatous degeneration involving the facial bones or femur is infrequent (<1%). The risk of malignant transformation is increased by radiation, which has proven to be ineffective treatment. In rare patients with widespread lesions, renal phosphate wasting and hypophosphatemia may cause rickets or osteomalacia. Hypophosphatemia may be due to production of a phosphaturic factor by the abnormal fibrous tissue.

MAS patients may have café au lait spots, which are flat, hyperpigmented skin lesions that have rough borders ("coast of Maine") in contrast to the café au lait lesions of neurofibromatosis that have smooth borders ("coast of California"). The most common endocrinopathy is isosexual pseudoprecocious puberty in girls. Other less common endocrine disorders include thyrotoxicosis, Cushing's syndrome, acromegaly, hyperparathyroidism, hyperprolactinemia, and pseudoprecocious puberty in boys.

**Radiographic Findings** In long bones, the fibrous dysplastic lesions are typically well-defined, radiolucent areas with thin cortices and a ground-glass appearance. Lesions may be lobulated with



**FIGURE 412-4** Radiograph of a 16-year-old male with fibrous dysplasia of the right proximal femur. Note the multiple cystic lesions, including the large lucent lesion in the proximal midshaft with scalloping of the interior surface. The femoral neck contains two lucent cystic lesions.

trabeculated areas of radiolucency (Fig. 412-4). Involvement of facial bones usually presents as radiodense lesions, which may create a leonine appearance (leontiasis ossea). Expansile cranial lesions may narrow foramen and cause optic lesions, reduce hearing, and create other manifestations of cranial nerve compression.

**Laboratory Results** Serum ALP is occasionally elevated, but calcium, parathyroid hormone, 25-hydroxyvitamin D, and 1,25-dihydroxyvitamin D levels are normal. Patients with extensive polyostotic lesions may have hypophosphatemia, hyperphosphaturia, and osteomalacia. The hypophosphatemia and phosphaturia are directly related to the levels of fibroblast growth factor 23 (FGF23). Biochemical markers of bone turnover may be elevated.

## TREATMENT

### Fibrous Dysplasia and MAS

Spontaneous healing of the lesions does not occur, and there is no established effective treatment. Improvement in bone pain and partial or complete resolution of radiographic lesions have been reported after IV bisphosphonate therapy. Denosumab given every 3 months is effective in reducing bone turnover markers and could be a therapeutic option in difficult cases. Surgical stabilization is used to prevent pathologic fracture or destruction of a major joint space and to relieve nerve root or cranial nerve compression or sinus obstruction.

## OTHER DYSPLASIAS OF BONE AND CARTILAGE

### PACHYDERMOPERIOSTOSIS

Pachydermoperiostosis, or hypertrophic osteoarthropathy (primary or idiopathic), is an autosomal dominant disorder characterized by periosteal new bone formation that involves the distal extremities. The lesions present as clubbing of the digits and hyperhidrosis and thickening of the skin, primarily of the face and forehead. The changes usually appear during adolescence, progress over the next decade, and then become quiescent. During the active phase, progressive enlargement

of the hands and feet produces a paw-like appearance, which may be mistaken for acromegaly. Arthralgias, pseudogout, and limited mobility may also occur. The disorder must be differentiated from secondary hypertrophic osteopathy that develops during the course of serious pulmonary disorders. The two conditions can be differentiated by standard radiography of the digits in which secondary pachydermoperiostosis has exuberant periosteal new bone formation and a smooth and undulating surface. In contrast, primary hypertrophic osteopathy has an irregular periosteal surface.

There are no diagnostic blood or urine tests. Synovial fluid does not have an inflammatory profile. There is no specific therapy, although a limited experience with colchicine suggests some benefit in controlling the arthralgias.

### OSTEOCHONDRODYSPLASIAS

These include several hundred heritable disorders of connective tissue. These primary abnormalities of cartilage manifest as disturbances in cartilage and bone growth. Selected growth-plate chondrodysplasias are described here. **For discussion of chondrodysplasias, see Chap. 413.**

**Achondroplasia** This is a relatively common form of short-limb dwarfism that occurs in 1 in 15,000 to 1 in 40,000 live births. The disease is caused by a mutation of the fibroblast growth factor receptor 3 (*FGFR3*) gene that results in a gain-of-function state. Most cases are sporadic mutations. However, when the disorder appears in families, the inheritance pattern is consistent with an autosomal dominant disorder. The primary defect is abnormal chondrocyte proliferation at the growth plate that causes development of short, but proportionately thick, long bones. Other regions of the long bones may be relatively unaffected. The disorder is manifest by the presence of short limbs (particularly the proximal portions), normal trunk, large head, saddle nose, and an exaggerated lumbar lordosis. Severe spinal deformity may lead to cord compression. The homozygous disorder is more serious than the sporadic form and may cause neonatal death. Vosoritide, an analog of C-type natriuretic peptide, increased growth among children in phase 3 clinical trials. Treatment is controversial among patient support communities. Infigratinib, a selective FGFR1-3 tyrosine kinase inhibitor, is in phase 2 clinical trials. Pseudoachondroplasia clinically resembles achondroplasia but has no skull abnormalities.

**Enchondromatosis** This is also called *dyschondroplasia* or *Ollier's disease*; it is also a disorder of the growth plate in which the primary cartilage is not resorbed. Cartilage ossification proceeds normally, but it is not resorbed normally, leading to cartilage accumulation. The changes are most marked at the ends of long bones, where the highest growth rates occur. Chondrosarcoma develops infrequently. The association of enchondromatosis and cavernous hemangiomas of the skin and soft tissues is known as *Maffucci's syndrome*. Both Ollier's disease and Maffucci's syndrome are associated with various malignancies, including granulosa cell tumor of the ovary and cerebral glioma.

**Multiple Osteochondromas** This is also called *multiple exostoses* or *diaphyseal aclasis*; it is a genetic disorder that follows an autosomal dominant pattern of inheritance. In this condition, areas of growth plates become displaced, presumably by growing through a defect in the perichondrium. The lesion begins with vascular invasion of the growth-plate cartilage, resulting in a characteristic radiographic finding of a mass that is in direct communication with the marrow cavity of the parent bone. The underlying cortex is resorbed. The disease is caused by inactivating mutations of the *EXT1* and *EXT2* genes, whose products normally synthesize heparan sulfate chains. The resulting heparan sulfate deficiency impacts signaling pathways and leads to ectopic chondrogenesis. Solitary or multiple lesions are located in the metaphyses of long bones. Although usually asymptomatic, the lesions may interfere with joint or tendon function or compress peripheral nerves. The lesions stop growing when growth ceases but may recur during pregnancy. There is a small risk for malignant transformation into chondrosarcoma. Palovarotene, a retinoic acid receptor agonist, is in clinical trials.

### EXTRASKELETAL (ECTOPIC) CALCIFICATION AND OSSIFICATION

Deposition of calcium phosphate crystals (*calcification*) or formation of true bone (*ossification*) in nonosseous soft tissue may occur by one of three mechanisms: (1) metastatic calcification due to a supranormal calcium × phosphate concentration product in extracellular fluid; (2) dystrophic calcification due to mineral deposition into metabolically impaired or dead tissue despite normal serum levels of calcium and phosphate; and (3) ectopic ossification, or true bone formation. Disorders that may cause extraskeletal calcification or ossification are listed in **Table 412-2**.

#### METASTATIC CALCIFICATION

Soft tissue calcification may complicate diseases associated with significant hypercalcemia, hyperphosphatemia, or both. In addition, vitamin D and phosphate treatments or calcium administration in the presence of mild hyperphosphatemia, such as during hemodialysis, may induce ectopic calcification. Calcium phosphate precipitation may complicate any disorder when the serum calcium × phosphate concentration product is >75. The initial calcium phosphate deposition is in the form of small, poorly organized crystals, which subsequently organize into hydroxyapatite crystals. Calcifications that occur in hypercalcemic states with normal or low phosphate have a predilection for kidney, lungs, and gastric mucosa. Hyperphosphatemia with normal or low serum calcium may promote soft tissue calcification with predilection for the kidney and arteries. The disturbances of calcium and phosphate in renal failure and hemodialysis are common causes of soft tissue (metastatic) calcification.

#### TUMORAL CALCINOSIS

This is a rare genetic disorder characterized by masses of metastatic calcifications in soft tissues around major joints, most often shoulders, hips, and ankles. Tumoral calcinosis differs from other disorders in that the periarticular masses contain hydroxyapatite crystals or amorphous calcium phosphate complexes, whereas in fibrodysplasia ossificans progressiva (below), true bone is formed in soft tissues. About one-third of tumoral calcinosis cases are familial, with both autosomal recessive and autosomal dominant modes of inheritance reported. The disease is also associated with a variably expressed abnormality of dentition marked by short bulbous roots, pulp calcification, and radicular dentin deposited in swirls. The disorder is caused by gene mutations in *GALNT3*, *FGF23*, or *Klotho*, leading to FGF23 deficiency or resistance. The reduced activity of FGF23 leads to increased renal tubular reabsorption of phosphate, elevated serum phosphate, and spontaneous soft tissue calcification from elevated calcium-phosphate concentration product.

The disease usually presents in childhood and continues throughout the patient's life. The calcific masses are typically painless and grow at variable rates, sometimes becoming large and bulky. The masses are often

**TABLE 412-2 Diseases and Conditions Associated with Ectopic Calcification and Ossification**

Metastatic calcification	Dystrophic calcification
Hypercalcemic states	Inflammatory disorders
Primary hyperparathyroidism	Scleroderma
Sarcoidosis	Dermatomyositis
Vitamin D intoxication	Systemic lupus erythematosus
Milk-alkali syndrome	Trauma-induced
Renal failure	Ectopic ossification
Hyperphosphatemia	Myositis ossificans
Tumoral calcinosis	Postsurgery
Secondary hyperparathyroidism	Burns
Pseudohypoparathyroidism	Neurologic injury
Renal failure	Other trauma
Hemodialysis	Fibrodysplasia ossificans progressiva
Cell lysis following chemotherapy	
Therapy with vitamin D and phosphate	

located near major joints but remain extracapsular. Joint range of motion is not usually restricted unless the tumors are very large. Complications include compression of neural structures and ulceration of the overlying skin with drainage of chalky fluid and risk of secondary infection. Small deposits not detected by standard radiographs may be detected by  $^{99m}\text{Tc}$  bone scanning. The most common laboratory findings are hyperphosphatemia and elevated serum 1,25-dihydroxyvitamin D levels. Serum calcium, parathyroid hormone, and ALP levels are usually normal. Renal function is also usually normal. Urine calcium and phosphate excretions are low, and calcium and phosphate balances are positive.

An acquired form of the disease may occur with other causes of hyperphosphatemia, such as secondary hyperparathyroidism associated with hemodialysis, hypoparathyroidism, pseudohypoparathyroidism, and massive cell lysis following chemotherapy for leukemia. Tissue trauma from joint movement may contribute to the periarticular calcifications. Metastatic calcifications are also seen in conditions associated with hypercalcemia, such as in sarcoidosis, vitamin D intoxication, milk-alkali syndrome, and primary hyperparathyroidism. In these conditions, however, mineral deposits are more likely to occur in proton-transporting organs such as kidney, lungs, and gastric mucosa in which an alkaline milieu is generated by the proton pumps.

## TREATMENT

### Tumoral Calcinosis

Therapeutic successes have been achieved with surgical removal of subcutaneous calcified masses, which tend not to recur if all calcification is removed from the site. Reduction of serum phosphate by chronic phosphorus restriction may be accomplished using low dietary phosphorus intake alone or in combination with oral phosphate binders. The addition of the phosphaturic agent acetazolamide may be useful. Limited experience using the phosphaturic action of calcitonin deserves further testing.

### DYSTROPHIC CALCIFICATION

Posttraumatic calcification may occur with normal serum calcium and phosphate levels and normal ion-solubility product. The deposited mineral is either in the form of amorphous calcium phosphate or hydroxyapatite crystals. Soft tissue calcification complicating connective tissue disorders such as scleroderma, dermatomyositis, and systemic lupus erythematosus may involve localized areas of the skin or deeper subcutaneous tissue and is referred to as *calcinosis circumscripta*. Mineral deposition at sites of deeper tissue injury including periarticular sites is called *calcinosis universalis*.

### ECTOPIC OSSIFICATION

True extraskeletal bone formation that begins in areas of fasciitis following surgery, trauma, burns, or neurologic injury is referred to as *myositis ossificans*. The bone formed is organized as lamellar or trabecular, with normal osteoblasts and osteoclasts conducting active remodeling. Well-developed haversian systems and marrow elements may be present. A second cause of ectopic bone formation occurs in an inherited disorder, *fibrodysplasia ossificans progressiva*.

### FIBRODYSPLASIA OSSIFICANS PROGRESSIVA

This is also called *myositis ossificans progressiva*; it is a rare autosomal dominant disorder characterized by congenital deformities of the hands and feet and episodic soft tissue swellings that ossify. The disorder is caused by an activating mutation in activin receptor A type 1. Ectopic bone formation occurs in fascia, tendons, ligaments, and connective tissue within voluntary muscles. Tender, rubbery induration, sometimes precipitated by trauma, develops in the soft tissue and gradually calcifies. Eventually, heterotopic bone forms at these sites of soft tissue trauma. Morbidity results from heterotopic bone interfering with normal movement and function of muscle and other soft tissues. Mortality is usually related to restrictive lung disease caused by an inability of the chest to expand. Laboratory tests are unremarkable.

There is no effective medical therapy. Bisphosphonates, glucocorticoids, and a low-calcium diet have largely been ineffective in halting

progression of the ossification. Palovarotene and REGN2477 (also known as garetosmab), an anti-activin A antibody, are currently in clinical trials. Surgical removal of ectopic bone is not recommended, because the trauma of surgery may precipitate formation of new areas of heterotopic bone. Dental complications, including frozen jaw, may occur following injection of local anesthetics.

### Acknowledgment

The authors acknowledge the contribution of Dr. Murray J. Favus to this chapter in previous editions of Harrison's.

### FURTHER READING

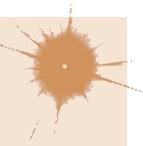
- Boyce AM, Collins MT: Fibrous dysplasia/McCune-Albright syndrome: A rare, mosaic disease of G<sub>s</sub> activation. Endocr Rev 41:345, 2020.
- Maajoor BC et al: Outcome of long-term bisphosphonate therapy in McCune-Albright syndrome and polyostotic fibrous dysplasia. J Bone Miner Res 32:264, 2017.
- Ralston SH et al: Diagnosis and management of Paget's disease of bone in adults: A clinical guideline. J Bone Miner Res 34:579, 2019.
- Reid IR et al: Treatment of Paget's disease of bone with denosumab: Case report and literature review. Calcif Tissue Int 99:322, 2016.
- Shapiro JR, Lewiecki EM: Hypophosphatasia in adults: Clinical assessment and treatment considerations. J Bone Miner Res 32:1977, 2017.
- Singer FR et al: Paget's disease of bone: An endocrine society clinical practice guideline. J Clin Endocrinol Metab 99:4408, 2014.
- Tan A et al: Long-term randomized trial of intensive versus symptomatic management in Paget's disease of the bone: The PRISM-EZ Study. J Bone Miner Res 32:1165, 2017.
- Wu CC et al: Diagnosis and management of osteopetrosis: Consensus guidelines from the osteopetrosis working group. J Clin Endocrinol Metab 102:3111, 2017.

## Section 5 Disorders of Intermediary Metabolism

413

### Heritable Disorders of Connective Tissue

Joan C. Marini, Fransiska Malfait



### CLASSIFICATION OF CONNECTIVE TISSUE DISORDERS

Some of the most common conditions that are transmitted genetically in families are disorders that produce clinically obvious changes in the bone, cartilage, skin, or relatively acellular tissues such as tendons that have been loosely defined as connective tissues. Because of their heritability, some of the disorders were recognized as potentially traceable to mutated genes soon after the principles of genetics were introduced into medicine by Garrod and others. About half a century later, McKusick emphasized the specificity of many of the diseases for selective connective tissues and suggested that they were probably caused by mutations in genes coding for the major proteins found in those tissues. In the past several decades, mutations in several hundred different genes expressed in connective tissues have been identified as the cause of many connective tissue disorders. However, classifying the disorders on the basis of either their clinical presentations or the mutations causing them continues to present a challenge for both the clinician and the molecular biologist.

Information on the disorders has continued to develop on two levels. The initial clinical classifications suggested by McKusick and

**3218** many others had to be refined as more patients were examined. For example, some patients had skin changes similar to those commonly seen in Ehlers-Danlos syndrome (EDS), but this feature was overshadowed by other features such as extreme hypotonia or sudden rupture of large blood vessels. To account for the full spectrum of presentations in patients and families, many of the disorders have been reclassified several times, dividing each into a series of subtypes.

The identification of mutations causing the diseases has developed on a parallel track. The first genes cloned for connective tissues were the two genes coding for type I collagen (*COL1A1* and *COL1A2*), the most abundant protein in bones, skin, tendons, and several other tissues. This facilitated early studies in patients with osteogenesis imperfecta (OI) that revealed mutations in type I collagen genes. Biochemical data, developed primarily with cultures of skin fibroblasts from affected individuals, demonstrated that the mutations dramatically altered the synthesis of collagen  $\alpha$ -chains or the structure of collagen fibers. The results stimulated efforts to identify additional mutations in genes coding for structural proteins. Genes for collagens provided an attractive paradigm to search for mutations, since a series of different types of collagens were found in different connective tissues and the collagen genes were readily isolated by their unique signature sequences. Also, the collagen genes were vulnerable to a large number of different mutations because of unusual structural requirements of the protein. The search for mutations in collagen genes proved fruitful in that mutations were found in most patients with OI, in many patients with hyperextensible skin and hypermobile joints, in some patients with dwarfism, and in patients with other disorders, including some such as Alport syndrome (AS) that were not initially classified as disorders of connective tissue. Also, mutations in collagen genes were found in subset of patients presenting with osteoarthritis (OA) or osteoporosis, likely representing the mildest end of the syndromic spectrum. However, the search for mutations quickly expanded to hundreds of other genes that included genes for other structural proteins, for the posttranslational modification and processing of the structural proteins, for chaperones, and for growth factors and their receptors and other genes whose functions are still not fully understood.

In many instances, the mutations helped to define the clinical subtype of the disorder, while in others, they revealed the genetic heterogeneity of the same clinical presentations. Conversely, some patients with different manifestations were found to have mutations in the same genes. In noncollagenous genes, it was sometimes difficult to establish whether a change in the structure of a gene caused the phenotypic changes in the patients or was simply a neutral polymorphism. Therefore, there has been a continuing debate as to whether the disorders should be classified by their clinical presentations or by the causative genes. As an illustration of the problems, mutations in 437 genes have been found associated with the 461 defined disorders of the skeleton. The latest nosology for the disorders remains "hybrid" in nature in the sense that the classification is not always based on the same criteria. Some diseases are grouped based on the causal gene, others are listed together, because they share common radiographic features, and still others are brought together because of a similar clinical course (lethality) or involvement of similar parts of the skeleton. A simpler system of classification proved feasible for one rare heritable disorder of skin, epidermolysis bullosa. The disorder was first defined clinically into subtypes based on the layers of the skin that were cleaved in friction-induced blisters. Most patients in each subtype were subsequently shown to have mutations in genes expressed in the corresponding layer of skin. Even with these patients, the strength of the genotype-phenotype correlation varies and mutations have not yet been found in every patient.

The best pathway through this maze of information is probably to begin by matching the signs and symptoms in a patient with the presentations that define each clinical classification. A major focus should be on the most common disorders, recognizing that the signs and symptoms may vary among different individuals and family members with the same diagnosis. Then, attempt to reach a decision, in consultation with the patient, parents, and specialist, as to whether a DNA analysis for the probable mutation is indicated. Among the considerations are the cost, the rigor with which the clinical classification has been linked

to mutated genes, the reassurance the diagnosis can bring to patients and their families, the use of the information for prenatal diagnosis, and the possibility that mutation-specific therapies may be developed in the future. For patients with the most severe forms, it is probably best to consult a specialist in the disease to determine a multidisciplinary program for management and therapy. Patient support groups have formed for many of the diseases and are an important source of information.

Patients with the most common forms of the disorders have mutations in a limited number of genes. This chapter will focus primarily on these. Also, it will provide a brief summary of biosynthesis and structure of connective tissues that may help guide the physician from the nature of the mutations to their clinical presentations.

## COMPOSITION OF CONNECTIVE TISSUES

Connective tissues such as skin, bone, cartilage, ligaments, and tendons are the critical structural frameworks of the body. They consist of a complex interacting extracellular matrix network of collagens, proteoglycans, and a large number of noncollagenous glycoproteins and proteins. While these precise combinations of up to ~500 potential extracellular matrix building blocks provide tissue-specific function, there are many overarching similarities in composition such as the role of composite collagen fibrils in providing strength and form, elastin fibrils and proteoglycans and other interacting proteins, and glycoproteins that fine-tune function (Table 413-1). The most abundant components of many connective tissues are three similar fibrillar collagens (types I, II, and III). They have a similar tensile strength that is comparable to that of steel wires. The three fibrillar collagens are distributed in a tissue-specific manner: type I collagen accounts for most of the protein of dermis, ligaments, tendons, and demineralized bone; type I and type III are the most abundant proteins of large blood vessels; and type II is the most abundant protein of cartilage.

## BIOSYNTHESIS AND TURNOVER OF CONNECTIVE TISSUES

Connective tissues are among the most stable components in living organisms, but they are not inert. During embryonic development, connective tissue membranes appear as early as the four-cell blastocyst to provide a structural scaffold for the developing embryo. With the development of blood vessels and skeleton, there is a rapid increase in the synthesis, degradation, and resynthesis of connective tissues. The turnover continues at a slower, but still rapid pace throughout postnatal development and then spikes during the growth spurt of puberty. During adulthood, the metabolic turnover of most connective tissues is slow, but it continues at a moderate pace in bone. With age, malnutrition, physical inactivity, and low gravitational stress, the rate of degradation of most connective tissues, especially in bone and skin, begins to exceed the rate of synthesis and the tissues shrink. In starvation, a large fraction of the collagen in skin and other connective tissues is degraded and provides amino acids for gluconeogenesis (Chap. 334). In both OA and rheumatoid arthritis, there is extensive degradation of articular cartilage collagen. Glucocorticoids weaken most tissues by decreasing collagen synthesis. In some pathologic states, however, collagen is deposited in excess. With most injuries to tissues, inflammatory and immune responses stimulate the deposition of collagen fibrils in the form of fibrotic scars. In humans, as distinct from many other species, the deposition of the fibrils is largely irreversible and prevents regeneration of normal tissues in diseases such as hepatic cirrhosis, pulmonary fibrosis, atherosclerosis, and nephrosclerosis.

**Structure and Biosynthesis of Fibrillar Collagens** The tensile strength of collagen fibers derives primarily from the self-assembly of protein monomers into large fibril structures in a process that resembles crystallization. The self-assembly requires monomers of highly uniform and relatively rigid structure. It also requires a complex series of posttranslational processing steps that maintain the solubility of the monomers until they are transported to the appropriate extracellular sites for fibril assembly. Because of the stringent requirements for correct self-assembly, it is not surprising that mutations in genes for fibrillar collagens cause many of the heritable diseases of connective tissues.

**TABLE 413-1 Constituents of Connective Tissues and Their Associated Heritable Conditions**

PROTEIN	TISSUE DISTRIBUTION	DISEASE	KEY MANIFESTATIONS
Collagen I	Bone, cornea, dermis, tendon	Osteogenesis imperfecta	Bone fragility with fractures and deformity; blue sclerae; dentinogenesis imperfecta; hearing loss
		EDS (various rare subtypes)	Joint hypermobility; skin hyperextensibility; skin fragility; soft connective tissue fragility
		Caffey disease	Subperiosteal new bone formation; irritability; soft tissue swelling
Collagen II	Cartilage, vitreous	Various chondrodysplasias	Skeletal dysplasia; ocular manifestations; hearing loss; orofacial findings
Collagen III	Dermis, aorta, uterus, intestine	Vascular EDS	Arterial, intestinal, and uterine fragility; thin translucent skin; easy bruising
Collagen IV	Basement membranes	Alport syndrome ( <i>COL4A3/A4/A5</i> )	Hematuria; hearing loss; ocular abnormalities
		Brain small-vessel disease ( <i>COL4A1/A2</i> )	Porencephaly; intracerebral hemorrhage; retinal arteriolar tortuosity; congenital cataract; hematuria; renal cysts; muscle cramps
Collagen V	Placental tissue, bone, dermis, cornea	Classical EDS	Joint hypermobility; skin hyperextensibility; atrophic scarring
Collagen VI	Uterus, dermis, cornea, cartilage	Bethlem myopathy and Ullrich congenital muscular dystrophy	Muscle weakness; joint contractures; joint hypermobility
Collagen VII	Skin, amniotic membrane, mucosal epithelium	Epidermolysis bullosa	Skin blistering; oral and esophageal blistering; corneal erosions
Collagen VIII	Descemet's membrane, endothelial cells	Corneal dystrophy	Corneal endothelial dystrophy; stromal edema
Collagen IX	Cartilage, vitreous	Stickler syndrome	Spondyloepiphyseal dysplasia; early-onset osteoarthritis; high myopia; vitreoretinal abnormalities; hearing loss; cleft palate; midfacial hypoplasia
Collagen X	Calcifying cartilage	Multiple epiphyseal dysplasia	Epiphyseal dysplasia; early-onset osteoarthritis
Collagen XI	Cartilage, intervertebral disk	Various chondrodysplasias	Skeletal dysplasia; ocular manifestations; hearing loss; orofacial findings
Collagen XII	Dermis, tendon, cartilage	Myopathic EDS	Joint hypermobility; congenital muscle hypotonia and/or atrophy; proximal joint contractures
Cartilage oligomeric matrix protein (COMP)	Cartilage, tendon, ligament, bone	Pseudoachondroplasia	Short-limb dwarfism; early-onset osteoarthritis
		Multiple epiphyseal dysplasia	Mildly short stature; early-onset osteoarthritis
Elastin	Dermis, arterial wall, lung	Cutis laxa	Wrinkled, redundant, sagging inelastic skin
		Williams syndrome	Cardiovascular disease (especially supravalvular aortic stenosis); orofacial features; intellectual deficit; connective tissue abnormalities; endocrine abnormalities
Fibrillin 1	Dermis, arterial wall, lung	Marfan syndrome	Aortic root aneurysm or dissection; ectopia lentis; marfanoid habitus
		Weill-Marchesani-syndrome	Short stature; joint stiffness; lens abnormalities; cardiovascular features
		Stiff skin syndrome	Progressive rock-hard skin; flexion contractures; hypertrichosis
		Geleophysic dysplasia	Short stature; joint stiffness; thickened skin; progressive cardiac valvular disease; orofacial features
Fibronectin	Dermis, tendons, ligaments	Glomerulopathy with fibronectin deposits	Glomerulopathy with fibronectin deposits
		Spondylometaphyseal dysplasia, corner fracture type	Spondylometaphyseal dysplasia characterized by flake-like, triangular, or curvilinear ossification centers at the edges of irregular metaphyses that simulate fractures; short stature
Aggrecan	Cartilage	Spondyloepiphyseal dysplasia, Kimberley type	Short stature; habitus; progressive osteoarthropathy; spondyloepiphyseal dysplasia
		Short stature; advanced bone age, with or without early-onset osteoarthritis and/or osteochondritis dissecans	Short stature and advanced bone age, with or without early-onset osteoarthritis and/or osteochondritis dissecans
		Spondylopimetaphyseal dysplasia, aggrecan type	Severe short stature; spondylopimetaphyseal dysplasia
Decorin	Dermis, tendons, ligaments, cornea	Congenital stromal corneal dystrophy	Corneal stromal opacification; visual loss; increased corneal thickness
Biglycan	Bone, cartilage, tendons	Meester-Loeys syndrome	Aortic aneurysm or dissection; orofacial features; joint hypermobility; ventricular dilatation on brain imaging; relative macrocephaly; hip dislocation; platyspondyly; phalangeal dysplasia; dysplastic epiphyses of the long bones
		X-linked spondylopimetaphyseal dysplasia	Severe short-trunked dwarfism; brachydactyly; spondylopimetaphyseal dysplasia

Abbreviation: EDS, Ehlers-Danlos syndrome.

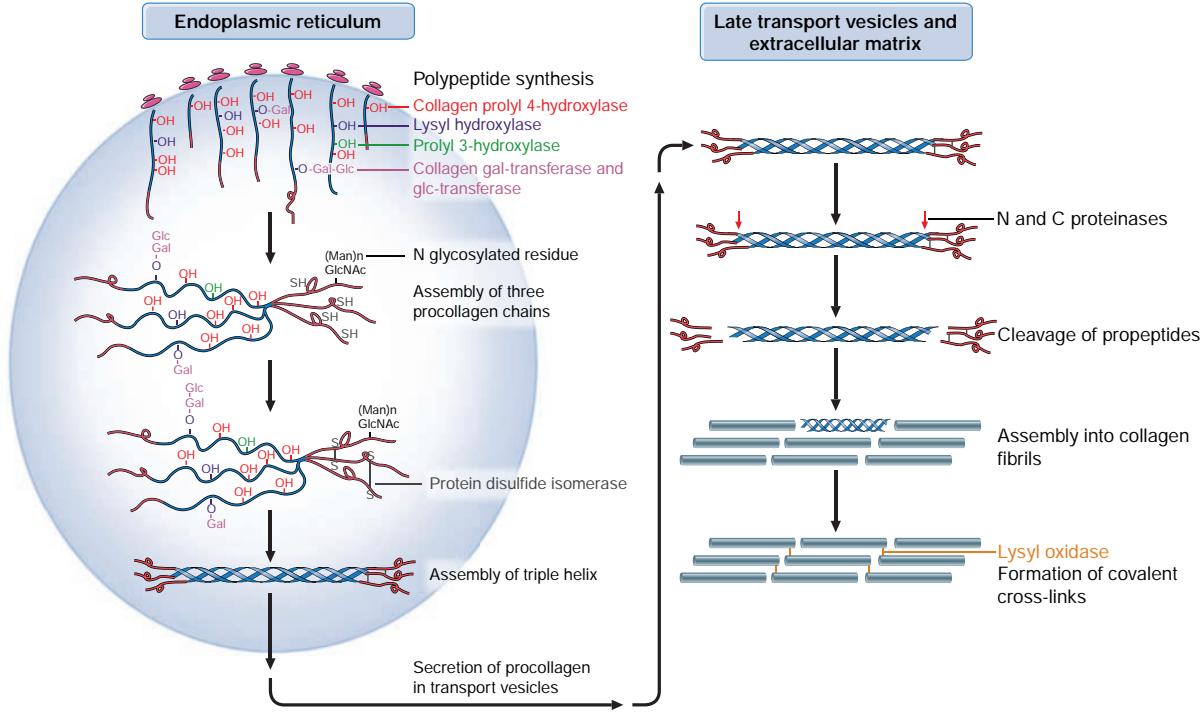


FIGURE 413-1 Schematic summary of biosynthesis of fibrillar collagens. (Reproduced with permission from J Myllyharju, KI Kivirikko: Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet* 20:33, 2004.)

The monomers of the three fibrillar collagens are formed from three polypeptide chains, called  $\alpha$  chains, that are wrapped around each other into a rope-like triple-helical conformation. The triple helix is a unique structure among proteins, and it provides rigidity to the molecule. It also orients the side chains of amino acids in an “inside out” manner relative to most other proteins so that the charged and hydrophobic residues on the surface can direct self-assembly of the monomers into fibrils. The triple-helical conformation of the monomer is generated because each of the  $\alpha$  chains has a repetitive amino acid sequence in which glycine (Gly) appears as every third amino acid. Each  $\alpha$  chain contains  $\sim 1000$  amino acids. Therefore, the sequence of each  $\alpha$  chain can be designated as  $(-\text{Gly-X-Y}-)_n$ , where X and Y represent amino acids other than glycine and n is  $>338$ . The presence of glycine, the smallest amino acid, in every third position in the sequence is critical because this residue must fit into a sterically restricted space in the interior of the helix where the three chains come together. The requirement for a glycine residue at every third position explains the significant clinical effects of mutations that convert a glycine residue to an amino acid with a bulkier side chain (see below). Many of the X- and Y-position amino acids are proline and hydroxyproline, which, because of their ring structures, provide additional rigidity to the triple helix. Other X- and Y-positions are occupied by charged or hydrophobic amino acids that precisely direct lateral and longitudinal assembly of the monomers into highly ordered fibrils. Mutations that substitute amino acids in some X- and Y-positions, particularly arginine-to-cysteine substitutions, can also produce genetic diseases.

The fibers formed by the three fibrillar collagens differ in thickness and length, but they have a similar fine structure. As viewed by electron microscopy, they all have a characteristic pattern of cross-striations that are about one-quarter the length of the monomers and reflect the precise packing into fibrils. The three fibrillar collagens, however, differ in sequences found in the X- and Y-positions of the  $\alpha$  chains and therefore in some of their physical properties. Type I collagen is a heterotrimer, composed of two identical  $\alpha_1(\text{I})$  chains and a third  $\alpha_2(\text{I})$  chain that differs slightly in its amino acid sequence. Types II and III collagen are homotrimers, each composed of three identical  $\alpha$  chains distinct to that type of collagen.

To deliver a monomer of the correct structure to the appropriate site of fibril assembly, the biosynthesis of fibrillar collagens involves a large number of unique processing steps (Fig. 413-1). The monomer, first synthesized as a soluble precursor called *procollagen*, contains an additional globular domain at each end. As the pre-pro chains of procollagen are synthesized on ribosomes, the free N-terminal ends move into the cisternae of the rough endoplasmic reticulum (ER). Signal peptides at the N-termini are cleaved, and additional posttranslational reactions begin. Proline and lysine residues in the Y-position of the Gly-X-Y repeating triplet are hydroxylated along the length of the helix by the enzymes prolyl 4-hydroxylase (P4H1) and lysyl hydroxylase (LH1), respectively. Hydroxyproline residues are essential for the three  $\alpha$  chains of the monomer to fold into a triple helix at body temperature. P4H1 requires ascorbic acid as an essential cofactor, an observation that explains why wounds fail to heal in scurvy (Chap. 333). In scurvy, some of the underhydroxylated and unfolded protein accumulates in the cisternae of the rough ER and is degraded. Many hydroxylysine residues are glycosylated with galactose or with galactose and glucose. Also, a large mannose-rich oligosaccharide is assembled on the C-terminal propeptide of each chain. The pro chains are assembled by interactions among these C-terminal propeptides that control the selection of the appropriate partner chains to form hetero- or homotrimers and provide the correct chain registration required for subsequent formation of the collagen triple helix. After the C-terminal propeptides assemble the three pro  $\alpha$  chains, a nucleus of triple helix is formed near the C-terminus, and the helical conformation is propagated toward the N-terminus in a zipper-like manner that resembles crystallization. The folding into the triple helix is spontaneous in solution, but as discussed below, identification of rare mutations causing OI demonstrated that the folding *in cellulo* is assisted by a number of ancillary proteins that also prevent collagen fibril formation within the ER. The fully folded procollagen is then transported to the Golgi via a specific COPII vesicle process. After further modifications in the Golgi stack, the procollagen is secreted into the pericellular space where distinct proteases remove the N- and C-propeptides at specific cleavage sites. The release of the propeptides decreases the solubility of the resulting collagen  $\sim 1000$ -fold. The entropic energy that is released drives the

self-assembly of the collagen into fibrils. Self-assembled collagen fibers have considerable tensile strength, but their strength is increased further by cross-linking reactions that form covalent bonds between chains in one molecule and chains in adjacent molecules. The resulting fibers, comprised of hundreds or thousands of triple-helical monomers, have some of the properties of a crystal but have innate imperfections that make them highly flexible.

Although the assembly of collagen monomers into fibers is largely a spontaneous reaction, the process in tissues is modulated by the presence of less abundant collagens (type V with type I, and type XI with type II) and by other components such as a series of small leucine-rich proteins (SLRPs). Some of the less abundant components alter the rate of fibril assembly, whereas others change the morphology of the fibers or their interactions with cells and other molecules. The presence of these other components is one explanation for why, in some tissues, the fibers are further assembled into large tendons; in others, into sheets; and in still others, into complex structures such as the hexagonal array of fibers that provide both the strength and transparency of the cornea.

Collagen fibers are resistant to most proteases, but during degradation of connective tissues, they are cleaved by specific matrix metalloproteinases (collagenases) that cause partial unfolding of the triple helices into gelatin-like structures that are further degraded by less specific proteinases.

### OTHER COLLAGENS AND RELATED MOLECULES

The unique properties of the triple helix are used to define a family of at least 28 collagens that contain repetitive -Gly-X-Y- sequences and form triple helices of varying length and complexity. The proteins are heterogeneous both in structure and function, and many are the sites of mutations causing genetic diseases. For example, the type IV collagen found in basement membranes is composed of three chains synthesized from any of six different genes; mutations in the *COL4A3*, *COL4A4*, or *COL4A5* genes cause AS.

**Fibrillin Aggregates and Elastin** In addition to tensile strength, many tissues such as the lung, large blood vessels, and ligaments require elasticity. The elasticity was originally ascribed to an amorphous rubber-like protein named elastin. Subsequent analyses, largely sparked by discoveries of mutations causing the Marfan syndrome (MFS), demonstrated that the elasticity resided in thin fibrils composed primarily of large glycoproteins named fibrillins. The fibrillins contain large numbers of epidermal growth factor-like domains interspersed with characteristic cysteine-rich domains that are also found in latent transforming growth factor (TGF- $\beta$ ) binding proteins. The fibrillins assemble into long beadlike strands that also contain numerous other components including small and variable amounts of elastin, bone morphogenic proteins (BMPs), and microfibril-associated glycoproteins (MAGPs). Besides contributing to extracellular matrix structure, a major role for fibrillins in TGF- $\beta$  signaling was emphasized by the discovery of mutations in genes coding for proteins involved in canonical TGF- $\beta$  signaling in patients with Marfan-like manifestations, including thoracic aortic aneurysm.

**Proteoglycans** The resiliency to compression of connective tissues such as cartilage or the aorta is largely explained by the presence of proteoglycans. Proteoglycans are composed of a core protein to which are attached a large series of negatively charged polymers of disaccharides (largely chondroitin sulfates). At least 30 proteoglycans have been identified. They vary in their binding to collagens and other components of matrix, but specific functions have not been assigned to most. The major proteoglycan of cartilage, called aggrecan, has a core protein of 2000 amino acids that is decorated with ~100 side chains of chondroitin sulfate and keratin sulfate. The core protein, in turn, binds to long chains of the polymeric disaccharide hyaluronan to form proteoglycan aggregates, one of the largest soluble macromolecular structures in nature. Because of its highly negative charge and extended structure, the proteoglycan aggregate binds large amounts of water and small ions to distend the three-dimensional arcade of collagen fibers found in the same tissues. It thereby makes the cartilage resilient to pressure.

## SPECIFIC DISORDERS

### OSTEOGENESIS IMPERFECTA

OI is a phenotypically and genetically heterogeneous generalized connective tissue disorder. The hallmark features of OI are increased susceptibility to skeletal fractures, bone deformity, and growth deficiency. Bone fragility is based on decreased bone mass and increased bone brittleness due to defective mineralization. Secondary features of OI are highly variable even within a type and include blue sclerae, dentinogenesis imperfecta, hearing loss, basilar invagination, pulmonary function impairment, cardiac valve abnormalities, and ligamentous laxity. Most patients have defects in the structure or quantity of type I collagen.

**Classification** OI was originally classified into *congenita* and *tarda* subtypes depending on the age of symptom onset. Sillence proposed the classification that bears his name for four types based on clinical and radiologic findings and mode of inheritance. The extension of the Sillence classification was first based on distinctive bone histology (types V and VI OI) and subsequently on the discovery of new recessive genes (types VII–XVIII). The debate between classification by phenotypic severity or gene defects has resulted in clinical and genetic classifications. The clinical classification can be useful for management but results in different type assignments in the same family or even in the same individual over their lifetime. The genetic classification (**Table 413-2**) groups patients by the causative gene. Because related causative genes were discovered close in time to each other, the genetic classification further groups types by overall mechanism and features OI as a collagen-related disorder.

Types I–IV OI are due to quantitative or structural defects in type I collagen itself. Type I is the mildest subtype, with reduced quantity of structurally normal collagen, and can produce mild or inapparent skeletal deformities. Most patients have distinctly blue sclerae. Types II, III, and IV are all caused by structural defects in one of the type I collagen chains. Type II produces bone so brittle that infants have in utero fractures of ribs and long bones and die in the perinatal period. Type III is progressively deforming with moderate to severe bone deformity, and type IV has mild to moderate bone fragility and secondary features. Subsequent rare recessive OI types are all collagen-related. Types V and VI (*ITIM5* and *SERPINF1*) particularly compromise matrix mineralization. Types VII, VIII, and IX (*CRTAP*, *P3H1*, and *PPIB*) represent defects in the components of the procollagen prolyl 3-hydroxylation complex that modifies collagen posttranslationally. Types X–XII (*SERPINH1*, *FKBP10*, and *BMP1*) have compromised procollagen processing and cross-linking. The final grouping of types XIII–XVIII (*SP7*, *TMEM38B*, *WNT1*, *CREBL1*, *SPARC*, and *MBTPS2*) alter osteoblast differentiation and impair collagen matrix quality.

The clinical heterogeneity of affected individuals within a particular OI type and even with the same mutation is not understood, with unknown modifying factors presumably involved. Among adults with OI, women are prone to fracture during pregnancy and after menopause. Some variants of mild OI are first detected perimenopausally and must be distinguished from postmenopausal osteoporosis.

**Incidence** In North America and Europe, the estimated incidence of OI is 1 per 10,000–15,000 births, based on a combination of cases recognized at birth and population surveys for milder cases. In populations with a high level of consanguinity or a founder mutation, the incidence of the rare recessive forms of OI is a significant addition to the prevalence of dominant collagen defects.

**Effects on Tissue Systems** The phenotypic features of OI are highly variable, even within the types caused by defects in type I collagen. The following section generally focuses on dominant forms comprising the majority of cases, except as specified, but the descriptions can be generalized to a large extent.

**Musculoskeletal Effects** Bone in OI is both weak and brittle. At the mildest end of the spectrum (type I OI), individuals may have only several childhood fractures and be limited only from contact sports.

**TABLE 413-2** Different Types of Osteogenesis Imperfecta (OI)

	OI TYPE	INHERITANCE	DEFECTIVE GENE	PROTEIN	OMIM	LOCUS	HYPERMINERALIZATION	DISTINGUISHING FEATURES
Defects in collagen structure and processing	I	AD	<i>COL1A1</i>	Collagen α1	166200	17q21.33	Yes	Loss of function of one of the COL1A1 alleles
	II–IV	AD	<i>COL1A1</i> , <i>COL1A2</i>	Collagen α1 or α2	166210, 259420, 166220	17q21.33, 7q21.3	Yes	Structural defects in collagen helix or C-propeptides
Procollagen processing defects	OI/EDS	AD	<i>COL1A1</i> , <i>COL1A2</i>	Procollagen α1 or α2	NA	17q21.33, 7q21.3	Yes	Defects in 90 residues at N-terminus of collagen helix that decrease pN-processing
	HBM	AD	<i>COL1A1</i> , <i>COL1A2</i>	Collagen α1 or α2	NA	17q21.33	Yes	Defects in C-propeptide cleavage site, DXA normal to increased
	XIII	AR	<i>BMP1</i>	BMP1	614856	8p21.3	Yes	Deficiency of C-propeptidase
Bone mineralization defects	V	AD	<i>IFITM5</i>	BRIL (BRIL5' MALEP)	610967	11p15.5	Yes	Calcification of interosseous membrane, dense metaphyseal band, hyperplastic callus, mesh-like pattern in lamellar bone
	Atypical VI	AD	<i>IFITM5</i>	BRIL (BRIL Ser40Leu)	610967	11p15.5	Yes	Increased osteoid, fish scale pattern in lamellar bone, increased ALP levels in childhood, symptom onset at birth
	VI	AR	<i>SERPINF1</i>	PEDF	613982	17p13.3	Yes	PEDF deficiency, increased osteoid, fish scale pattern in lamellar bone, increased ALP levels in childhood, onset after age 1 year
Defects in collagen modification	VII	AR	<i>CRTAP</i>	CRTAP	610682	3q22.3	Yes	Absent procollagen prolyl 3-hydroxylation; full OM, rhizomelia, white sclerae
	VIII	AR	<i>LERPE1</i>	P3H1	610915	1p34.2	Yes	Absent procollagen prolyl 3-hydroxylation; full OM, rhizomelia, "popcorn" metaphyses; white sclerae
	IX	AR	<i>PPIB</i>	CyPB	259440	15q22.31	Yes	Absent procollagen prolyl 3-hydroxylation; helix modification varies, without rhizomelia, white sclerae
	XIV	AR	<i>TMEM38B</i>	TRIC-B	615066	9q31.2	No	Decreased modification of collagen helix
Defects in collagen folding and cross-linking	X	AR	<i>SERPINH1</i>	HSP47	613848	11q13.5	ND	Severe skeletal deformity, blue sclerae, DI, skin abnormalities, inguinal hernias
	NA	AR	<i>KDELR2</i>	KDEL ER protein retention receptor; interacts with HSP47	619131	7p22.1	ND	Short stature, progressive skeletal deformation requiring recurrent surgical interventions
	XI	AR	<i>FKBP10</i>	FKBP65	610968	17q21.2	Yes	May have congenital contractures
	NA	AR	<i>PLOD2</i>	LH2	609220	3q24	Yes	Progressive joint contractures
Osteoblast function and differentiation	XII	AR	<i>SP7</i>	OSTERIX	613849	12q13.13	ND	Severe skeletal deformity, delayed tooth eruption, facial hypoplasia
	XV	AD/AR	<i>WNT1</i>	WNT1	615220	12q13.12	No	May have neurologic defects
	XVI	AR	<i>CREB3L1</i>	OASIS	616215	11p11.2	Yes	Defect in RIP pathway
	XVII	AR	<i>SPARC</i>	SPARC	616507	5q33.1	Yes	Progressive severe bone fragility
	XVIII	XR	<i>MBTPS2</i>	S2P	301014	Xp22.12	Yes	X-linked OI, defect in RIP pathway, rhizomelia
Unclassified disorders	NA	AR	<i>FAM46A</i>	FAM46A	617952	6q14.1	ND	Defect in BMP/TGF-β signaling pathway
	NA	AR	<i>MESD</i>	LRP chaperone MESD	618644	15q25.1	ND	Could also be classified with LRP5/6-related disorders
	NA	AR	<i>CCDC134</i>	Coiled-coil domain-containing protein 134	618788	22q13.2	ND	Could also be classified with MAPK/ERK skeletal dysplasias

*Abbreviations:* AD, autosomal dominant; ALP, alkaline phosphatase; AR, autosomal recessive; BMP, bone morphogenetic protein; DI, dentinogenesis imperfect; DXA, dual-energy x-ray absorptiometry; EDS, Ehlers-Danlos syndrome; HBM, high bone mass; NA, not applicable; ND, not determined; OM, overmodification; OMIM, Online Mendelian Inheritance in Man; TGF, transforming growth factor.

More severe forms of OI require bone to be partially unloaded with assistive devices such as walkers or canes; many severe patients use electric chairs for both the weight bearing and the normal speed of mobility. In dominant OI, fragility fractures often decrease sharply after adequate bone mass is gained at puberty. Radiographs generally show osteopenia in all types, with disordered matrix organization detected most easily in lower long bones in moderate and severe forms. In lethal OI, radiographs show continuous beading of ribs from healing fractures and crumpled and undertubulated long bones. Lateral skull radiographs may show islands of Wormian bones, even in mild forms. The appearance of "popcorn" at the metaphyses of long bones occurs in many type III and IV children and coincides with increased growth deficiency. Often these bones are so soft that normal muscle pull can produce severe deformities. Kyphoscoliosis is associated with vertebral compressions but is not prevented by bisphosphonates, suggesting a contribution from ligamentous laxity.

OI bone is weak, in that it fractures with a lower load than normal, and is brittle, in that it does not tolerate postyield displacement and snaps like chalk. The brittleness results from the paradoxical increased mineralization of OI bone. While dual-energy x-ray absorptiometry (DXA) bone density measurements uniformly return a reduced value for OI bone, it is performed with a phantom and detects mineral crystals that are in proper alignment. In contrast quantitative backscattered electron imaging or three-dimensional (3D) computed tomography (CT), which detect all mineral in 3D, reveals that both dominant and recessive (except types XIV and XV) OI bone is hypermineralized. On histomorphometry, dominant OI bone has proper formation of lamellae but increased turnover, causing decreased bone volume. Type V OI has mesh-like bone lamellae, as well as a dislocated radial head, and may have hyperplastic callus formation, while type VI OI has distinctive fish scale lamellae on polarized light microscopy.

Many OI patients across the severity spectrum have increased ligamentous laxity. Patients with defects in processing the N-terminal propeptide of type I procollagen have large and small joint hypermobility similar to EDS. Muscle weakness of unknown etiology also occurs in OI, and the weakness and ligamentous laxity contribute to delayed motor development.

**Pulmonary** The leading cause of death in OI is pulmonary disease. Young children with severe OI often have repeated pneumonia; restrictive or obstructive disease often develops in adults. Pulmonary function is impaired by marked scoliosis and chest wall deformity but also arises from intrinsic defects of lung parenchyma containing type I collagen, as shown by declining pulmonary function over time in children without scoliosis. Mice with null *CRTAP* mutations (type VII OI) have abnormal alveolar development, and patients with recessive forms also have pulmonary complications. Evaluation of even asymptomatic moderate to severe OI patients by spirometry should initiate standard pulmonary interventions.

**Cardiovascular** Cardiovascular effects of OI manifest predominantly in adults. With type I collagen as a major component of matrix in cardiac valves and aortic wall, the most frequent manifestations are valvular, especially mitral regurgitation and aortic root dilatation. Impaired mechanical properties occasionally lead to aortic dissection. Echocardiography is appropriate with heart murmurs or cardiac symptoms and every 3–5 years in asymptomatic patients.

**Dentinogenesis Imperfecta** Dentinogenesis imperfecta (DI) is associated with types III and IV OI and recessive types with collagen processing defects. Tooth agenesis, especially of premolars, is also found in types III/IV OI. Teeth with disturbed formation of dentin during development may be translucent gray or have yellowish or brownish discoloration. Defects are manifest predominantly in primary teeth; detection in secondary teeth may require radiographs to identify characteristic narrow or obliterated pulp chambers. Crumbling at the dentin-enamel junction may require capping of teeth. Hypoplastic maxilla and relative mandibular prognathism in moderate to severe OI can result in type III malocclusion and impair normal chewing, requiring surgical correction.

**Hearing Loss** About half of patients with types I, III, and IV OI develop hearing loss, but its incidence in recessive types is unknown. Hearing loss usually begins in the second decade and progresses. The initial conductive loss, based on changes in the inner ear leading to stapes footplate fixation, can evolve into a mixed conductive and sensorineural loss. Regular screening allows referral for hearing aids, stapes surgery, or cochlear implants, as appropriate.

**Other Features** A variable intensity of blue or grayish sclerae is a well-known feature of OI. The color is most striking with collagen defects, especially types I and II OI and defects that affect N-terminal procollagen processing. Blue sclerae often occur in other connective tissue disorders such as EDS or MFS and may occur in individuals without connective tissue defects. Severe neonatal OI with white sclerae should prompt consideration of recessive forms, especially prolyl 3-hydroxylation defects. Abnormalities of the skull base, such as platybasia and basilar invagination, sometimes progress to clinically devastating basilar impression. Patients with height Z-scores of <-3 should be CT scanned at 3- to 5-year intervals. Significant growth deficiency is a cardinal feature of OI, ranging from minimally shorter than siblings in mild forms to greater extents in some severe cases, with adults shorter than 5-year-old children. There is both end-organ resistance to growth hormone (GH) and defective transition to bone at the growth plate. Types I and IV OI are often responsive to recombinant GH therapy.

 **Molecular Defects** The great majority (80–85%) of cases of OI are caused by heterozygous mutations in either of the genes coding for the chains of type I procollagen, *COL1A1* or *COL1A2* (Table 413-2). Although thousands of unique mutations have been identified in type I collagen, they fall into several structural types. Null mutations in collagen chains are less detrimental than structural defects. Null mutations in *COL1A1* result in about half the normal level of collagen synthesis, but the collagen in matrix is structurally normal. These patients have mild type I OI. Null *COL1A2* mutations are rare, leading to an EDS-like condition with progressive cardiac-valvular defects.

Mutations that produce structural changes in type I collagen chains cause types II, III, and IV OI. The most common of these are mutations resulting in substitutions for glycine residues required at every third residue along the helix. In effect, any of the 338 glycine residues in the helical domain of either the pro 1 or pro 2 chain of type I procollagen is a potential site for a disease-producing mutation. Other mutations affect the splicing of the exons encoding the chains. Because each collagen exon encodes a discrete set of Gly-X-Y triplets, the abnormal splice products are most often in-frame and cause severe structural abnormalities. Use of alternative splice sites may lead to premature termination, mimicking null mutations, and a milder phenotype. Structural abnormalities in the procollagen helical region delay collagen folding and expose chains to posttranslational hydroxylation/glycosylation for a longer time. The abnormal procollagen triggers a cascade of intracellular and extracellular events including delayed collagen folding, ER stress, abnormal interaction with noncollagenous molecules, impaired osteoblast development and cross-talk with osteoclasts, and abnormal mineralization. There are some special sets of procollagen structural mutations with distinct mechanisms within types II, III, and IV. Mutations in the C-propeptide significantly delay chain assembly, and resulting procollagen is mislocalized to the ER lumen. Some of this procollagen is targeted for degradation by the ER-associated proteasomal pathway, while the secreted molecules delay pericellular processing of the C-propeptide. Mutations in the C-propeptide cleavage site itself prevent processing of the propeptide, leaving pC-collagen to be incorporated into matrix. This affects matrix mineralization, resulting in an unusual high bone mass form of OI that falls at the milder end of the type IV OI phenotype. Not surprisingly, null mutations in the C-propeptidase enzyme, BMP1, cause recessive type XII OI. Type XII OI is a severe condition because BMP1 is the cleavase for types I, II, and III procollagens and the glycoprotein decorin, which is a regulator of fibrillogenesis. Processing defects of the N-propeptide occur in the cleavage site itself or the 90 helix residues at the amino end. The persistence of the N-propeptide on a fraction of the molecules

interferes with the self-assembly of normal collagen so that thin and irregular collagen fibrils are formed. They cause extreme laxity of large and small joints, intensely blue sclerae, and an OI severity comparable to type III/IV. Rare substitutions of charged amino acids (Asp, Arg) or a branched amino acid (Val) in X- or Y-positions produce lethal phenotypes, apparently because they are located at sites for lateral assembly of the monomers or binding of other components of the matrix.

Starting in 2006, a series of noncollagenous genes have been identified that cause (mostly) recessive OI. Importantly, all the genes have encoded proteins or cellular processes related to collagen, shifting the OI paradigm to dominant OI caused by collagen defects or *IFITM5* and recessive OI caused by proteins related to collagen modification, processing, folding, and cross-linking and osteoblast differentiation. The largest group of patients with OI not caused by collagen gene mutations have types V and VI OI, affecting bone mineralization. Type V OI, with dominant inheritance, is unusual in that all patients have the same recurrent mutation at the 5'-end of *IFITM5*, which generates a novel start codon in the transmembrane protein BRIL. The gain-of-function mutation causes distinctive radiologic (ossification of interosseus membrane and dense metaphyseal band) and phenotypic findings (hypertrophic callus). Osteoblasts with type V OI have increased mineralization and differentiation in culture. Type VI OI is a recessive form caused by null mutations in *PEDF*, a collagen-interacting molecule with a known antiangiogenic effect. A connection between types V and VI OI has been revealed by a set of patients with a BRIL p.S42L substitution who have clinical, histologic, serum marker, and phenotypic features of type VI OI. Both type VI OI osteoblasts and BRIL p.S42L osteoblasts have decreased cellular mineralization and *SERPINEF1* expression, while classic type V OI osteoblasts have the opposite findings. All three types decrease collagen production.

Types VII, VIII, and IX OI are severe recessive forms caused by deficiency of one of the components of the procollagen prolyl 3-hydroxylation complex, *P3H1*, *CRTAP*, or cyclophilin B (*PPIB/CyPB*). This complex 3-hydroxylates one proline residue per chain, most critically 1(I)P986, in contrast to the proline 4-hydroxylation of multiple helical residues by P4H1. In murine models, loss of complex function results in a severe phenotype; while mutation of the P986 residue impairs collagen cross-linking and fine-tuning of collagen alignment in fibrils. The phenotype of these patients is distinctive for white sclerae, rhizomelia, and lack of relative macrocephaly; they share the bone fragility, high bone turnover, and elevated bone mineralization of classical OI.

Some recessive OI types that impair osteoblast function are caused by mutation in genes not previously understood to affect bone. Regulatory intramembrane proteolysis (RIP) is well known for its role in cholesterol synthesis, in which cells transport regulatory proteins from the ER membrane to the Golgi membrane in times of cell stress, where S1P and S2P Golgi proteases sequentially cleave the transcription factors, activating them to enter the nucleus. X-linked type XVIII OI with defective *MTPS2/S2P* and type XVI OI with deficiency of an RIP substrate Oasis, a member of the ATF6 family of stress sensors, indicate the importance of RIP for bone formation (Table 413-2).

**Inheritance and Mosaicism in Germline Cells and Somatic Cells** Types I–V OI are inherited as autosomal dominant traits, while the rare forms are mostly recessive. Many patients with mild dominant OI represent familial traits, while sporadic new mutations are often responsible for dominant severe or lethal cases. Germline mosaicism in one parent may be the etiology of a severe dominant mutation in the child; in this circumstance, a second child may be affected with the same dominant mutation from unaffected parents. Recessive mutations in genes causing the rare forms of OI lead to more severe clinical outcomes; many of these offspring do not survive childhood, but moderately to severely affected young adults show us that these conditions must also be considered.

**Diagnosis** OI is usually diagnosed on the basis of clinical and radiographic criteria. The presence of fractures together with blue sclerae, DI, or family history of the disease is usually sufficient to make

the diagnosis. X-rays reveal a decrease in bone density that can be verified by DXA bone densitometry, as well as characteristic deformities of long bones, thorax, and cranium. The differential diagnosis varies with age, including battered child syndrome, nutritional deficiencies, malignancies, and other inherited disorders such as chondrodysplasias and hypophosphatasia that can have overlapping presentations. A molecular diagnosis is now routinely obtained using targeted candidate gene sequencing, sometimes beginning with the dominant collagen and *IFITM5* panel. Although almost all cases can be diagnosed by sequencing, some may require bone histology and exome sequencing.

## TREATMENT

### Osteogenesis Imperfecta

Therapy should be directed toward maximizing the function of each individual, which includes decreasing fractures and deformity that interfere with function. Physical and occupational therapy are critical modalities. They are most commonly utilized after severe fractures or major surgery and should also be engaged consistently throughout the life span for maximizing mobility, functions of daily living, and the extent of physical conditioning possible. Water therapy is particularly useful at all ages. Diet should include adequate intake of calcium and vitamin D. Many patients are underweight for height as young children but overweight as adults, and nutritional management may be useful. Orthopedic procedures are required for deformities of long bone that interfere with standing or walking or when a bone has sustained repeated fractures. Intramedullary rods are often inserted when children are ready to stand and as needed thereafter to keep bone segments in good alignment and provide partial unloading of weight from bones. If scoliosis progresses, stabilization of the spine may be needed to maintain the curve at <60°. Medical management should also include presymptomatic screening for hearing loss, cardiac valve dysfunction, pulmonary function, and, in severe individuals, basilar invagination.

Drugs that have been developed for the therapy of postmenopausal osteoporosis are beneficial for some patients. Bisphosphonates, antiresorptive drugs that inhibit osteoclasts, increase DXA bone density and relieve vertebral compressions in most patients. They are regarded as a mainstay of care in many pediatric centers. However, several Cochran reports have not supported a clear reduction in fracture rate or bone pain from their use, and the dosing and duration of use are controversial. Currently, drugs with a bone-forming mechanism are in trials for OI, especially monoclonal antibodies to sclerostin that relieve its inhibition of osteoblast Wnt/-catenin signaling, TGF- $\beta$  inhibitors, and a PTH analogue that stimulates osteoblasts and is most beneficial for adults with milder OI. Potential therapies under investigation in animal models include chemical chaperones and mesenchymal stem cell therapy.

## EHLERS-DANLOS SYNDROMES

The Ehlers-Danlos syndromes (EDS) comprise a genetically heterogeneous group of heritable conditions that share several characteristics such as soft and hyperextensible skin, abnormal wound healing, easy bruising, and joint hypermobility. Additional clinical features that differ among the EDS subtypes include fragility of soft tissues, blood vessels, and hollow organs and involvement of the musculoskeletal system. Mutations in genes coding for fibrillar collagens (type I, III, or V) are found in many patients, but other genes are affected in rare forms.

**Classification** Several types of EDS have been defined, based on clinical characteristics, mode of inheritance, and molecular defects (Table 413-3), and the classification of these types has been a dynamic process. The current classification defines 13 clinical EDS types that are caused by alterations in 19 different genes, but a recent study described another genetically distinct EDS type, bringing the total number of EDS-associated genes to 20. The EDS classification guides the clinical diagnosis, molecular confirmation, and genetic counseling of affected individuals and their family members.

**TABLE 413-3** Different Types of Ehlers-Danlos Syndrome (EDS)

	EDS TYPE	INHERITANCE	OMIM	LOCUS	GENE	PROTEIN	KEY MANIFESTATIONS
Defects in collagen primary structure and collagen processing	Classical EDS (cEDS)	AD	130000 130010	9q34.3 2q32.2	<i>COL5A1</i> <i>COL5A2</i>	Proα1(V) Proα2(V)	Skin hyperextensibility with atrophic scarring Generalized joint hypermobility
	Classical EDS (cEDS)	AD	/	17q21.33	<i>COL1A1</i>	Proα1(I) p.Arg312Cys	Skin hyperextensibility with atrophic scarring Generalized joint hypermobility Arterial rupture at young age
	Vascular EDS (vEDS)	AD	130050	2q32.2	<i>COL3A1</i>	Proα1(III)	Arterial rupture at young age Spontaneous sigmoid colon perforation in the absence of known colon disease Uterine rupture during third trimester of pregnancy Carotid-cavernous sinus fistula (in the absence of trauma)
	Arthrochalasia EDS (aEDS)	AD	130060 130060	17q21.33 7q21.3	<i>COL1A1</i> <i>COL1A2</i>	Proα1(I) Proα2(I)	Congenital bilateral hip dislocation Severe generalized joint hypermobility with multiple dislocations Skin hyperextensibility
	Dermatosparactic EDS (dEDS)	AR	225410	5q35.3	<i>ADAMTS2</i>	ADAMTS2	Extreme skin fragility with congenital or postnatal tears Craniofacial features Progressively redundant, lax skin with excessive skinfolds Increased palmar wrinkling Severe bruising Umbilical hernia Postnatal growth retardation with short limbs Perinatal complications related to tissue fragility
	Cardiac-valvular EDS (cvEDS)	AR	225320	7q21.3	<i>COL1A2</i>	proα2(I)	Severe progressive cardiac-valvular insufficiency Skin involvement Joint hypermobility
Defects in collagen folding and collagen cross-linking	Kyphoscoliotic EDS (KEDS-PLOD1) Kyphoscoliotic EDS (KEDS-FKBP14)	AR AR	225400 614557	1p36.22 7p14.3	<i>PLOD1</i> <i>FKBP14</i>	Lysylhydroxylase 1 FKBP22	Congenital muscle hypotonia Congenital or early-onset kyphoscoliosis Generalized joint hypermobility with (sub)luxations
Defects in structure and function of myomatrix, the interface between muscle and ECM	Classical-like EDS type 1 (cIEDS1)	AR	606408	6p21.33-p21.32	<i>TNXB</i>	Tenascin XB	Skin hyperextensibility with velvety skin texture and absence of atrophic scarring Generalized joint hypermobility Easily bruised skin/spontaneous ecchymoses
	Myopathic EDS (mEDS)	AD/AR	616471	6q13-q14	<i>COL12A1</i>	Proα1(XII)	Congenital muscle hypotonia and/or muscle atrophy Joint contractures Joint hypermobility

(Continued)

**TABLE 413-3** Different Types of Ehlers-Danlos Syndrome (EDS) (Continued)

	EDS TYPE	INHERITANCE	OMIM	LOCUS	GENE	PROTEIN	KEY MANIFESTATIONS
Defects in glycosaminoglycan biosynthesis	Spondylodysplastic EDS (spEDS-B4GALT7)	AR	130070	5q35.3	<i>B4GALT7</i>	Galactosyltransferase I β4GalT7	Short stature (progressive in childhood) Muscle hypotonia (ranging from severe congenital to mild later-onset) Bowing of limbs Skeletal dysplasia
	Spondylodysplastic EDS (spEDS-B3GALT6)	AR	615349	1p36.33	<i>B3GALT6</i>	Galactosyltransferase II β3GalT6	
	Musculocontractural EDS (mcEDS-CHST14)	AR	601776	15q15.1	<i>CHST14</i>	Dermatan-4 sulfotransferase-1	Congenital multiple contractures (typically adduction/flexion contractures and talipes equinovarus) Craniofacial features Skin hyperextensibility, easy bruising, skin fragility with atrophic scars Increased palmar wrinkling
	Musculocontractural EDS (mcEDS-DSE)	AR	615539	6q22.1	<i>DSE</i>	Dermatan sulfate epimerase-1	
Defects in complement pathways	Periodontal EDS (pEDS)	AD	130080	12p13.31	<i>C1R</i> <i>C1S</i>	C1r C1s	Severe and intractable early-onset periodontitis Lack of attached gingiva Pretibial plaques
Defects in intracellular processes	Spondylodysplastic EDS (spEDS-SLC39A13)	AR	612350	11p11.2	<i>SLC39A13</i>	ZIP13	Short stature (progressive in childhood) Muscle hypotonia (ranging from severe congenital to mild later-onset) Bowing of limbs Skeletal dysplasia
	Brittle cornea syndrome (BCS)	AR	229200 614170	16q24 4q27	<i>ZNF469</i> <i>PRDM5</i>	ZNF469 PRDM5	Thin cornea with/without rupture Early-onset progressive keratoconus and/or keratoglobus Blue sclerae
Unclassified	Classical-like EDS type 2 (cIEDS2)	AR	618000	7p13	<i>AEBP1</i>	AEBP1 (ACLP)	Skin hyperextensibility with atrophic scarring Generalized joint hypermobility Foot deformities Early-onset osteopenia
Unknown	Hypermobile EDS (hEDS)	? (AD)	130020	?	?	?	Generalized joint hypermobility Systemic manifestations of generalized connective tissue fragility Musculoskeletal complaints Positive family history Exclusion of other EDS types and other joint hypermobility-associated conditions

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; ECM, extracellular matrix; OMIM, Online Mendelian Inheritance in Man.

**Incidence** An incidence of about 1 in 5000 individuals for all forms of EDS was proposed, with no apparent ethnic predisposition. The diagnosis of hypermobile EDS is more common in females than in males, but whether this is due to an increased incidence or more severe manifestation is unknown. The incidence for other types of EDS is similar in males and females. With incidences of 1 in 20,000 and 1 in 50,000–200,000 respectively, classical and vascular EDS are the most common genetically elucidated types of EDS. For the other types of EDS for which causative variants have been identified, there are no incidence estimates, but the numbers of people who have been reported worldwide with these disorders range between ~5 and ~100 individuals per EDS type. Patients with milder forms frequently do not seek medical attention.

**Skin** One of the principal features of EDS is skin hyperextensibility, that is, the skin stretches easily but snaps back after release. The skin often has a smooth, soft, or velvety feel to it and can be thin and translucent. It is fragile and tears easily, even after minor trauma, and heals slowly. Widened and thin atrophic scars are frequently observed in different types of EDS. Especially in classical EDS, atrophic scarring may be widespread, especially over pressure points and exposed areas such as the forehead, elbows, knees, and shins, with marked widening of the scars, which are covered by a very thin inelastic skin (papyraceous scars). Individuals with vascular EDS usually do not have a velvety hyperextensible skin, but skin can be thin and translucent with visible superficial veins. Easy bruising is common to most types of EDS and may manifest itself as spontaneous or recurring hematomas. These may cause discoloration of the skin due to deposition of hemosiderin, often referred to as “hemosiderotic” scars, especially in classical, vascular, and periodontic EDS.

**Ligament and Joint Changes** Joint hypermobility, another cardinal sign, is variable in severity and usually, but not always, generalized. While often an “asset” in childhood, it can become a serious burden over time, often complicated by repetitive subluxations, dislocations, sprains, and chronic joint pain that is difficult to treat. Other observed musculoskeletal features include congenital bilateral hip dislocation, spine deformities (scoliosis, kyphosis), pectus deformities (pectus carinatum, pectus excavatum), club feet and other contractures, and in some rare types, a (mild) skeletal dysplasia. Muscle hypotonia is observed in a number of EDS types and, in combination with joint laxity, may cause floppy infant syndrome or a delay in motor development.

**Other Features** Signs of more generalized connective tissue weakness and fragility can be observed in varying degrees and may help to distinguish between the different EDS types. Rupture of medium and large-sized arteries is typical of vascular EDS but has been reported in a few other types as well, i.e., classical and kyphoscoliotic type. Vascular EDS patients are also at increased risk for rupture of the gastrointestinal tract, especially the sigmoid colon, and the gravid uterus. Valvular defects and aortic root dilatation are rare and are also restricted to some of the rarer types of EDS. Obstetrical and pelvic complications such as cervical insufficiency, premature rupture of membranes, vaginal lacerations, and organ prolapses (uterus, bladder, rectum) may occur. Sclerae may be blue, and more severe ophthalmologic complications, including keratoconus, keratoglobus, and scleral or corneal rupture, may be observed in some rare types.



**Molecular Defects** Subsets of patients with different types of EDS have mutations in the structural genes for fibrillar collagen types I, III, and V (Table 413-3). About 90% of classical EDS patients harbor a heterozygous mutation in *COL5A1* or *COL5A2* coding for type V collagen, a minor collagen found in association with type I collagen. Heterozygous mutations in the *COL3A1* gene for type III collagen, which is abundant in the blood vessel wall, are responsible for vascular EDS. Arthrochalasia EDS is caused by heterozygous mutations in either *COL1A1* or *COL1A2* that make type I procollagen resistant to cleavage by procollagen N-proteinase, whereas dermatosparaxis EDS is caused by biallelic mutations in the gene that codes for the

procollagen N-proteinase itself, thereby reducing its enzyme activity. The persistence of the N-propeptide causes the formation of collagen fibrils that are thin and irregular. Other specific mutations in either *COL1A1* or *COL1A2* give rise to a few rare subtypes of EDS. These include the cardiac-valvular type, which is caused by biallelic *COL1A2* mutations, leading to a complete absence of 2(I) chains. Patients with this condition are at risk for severe, progressive cardiac-valvular disease necessitating valve replacement. A specific arginine-to-cysteine substitution in the type I collagen chain (p.Arg312Cys) is associated with an EDS phenotype that resembles that of classical EDS, but patients appear at increased risk for vascular rupture of medium-sized arteries. A few patients with a phenotype that couples EDS with signs of moderate to severe myopathy harbor heterozygous or homozygous mutations in *COL12A1*, coding for type XII collagen, a fibril-associated collagen with interrupted triple helices. Kyphoscoliotic EDS is caused by biallelic mutations either in the *PLOD1* gene, which encodes procollagen-lysine 5-dioxygenase (lysyl hydroxylase 1), an enzyme required for formation of stable cross-links in collagen fibers, or in the *FKBP14* gene, which encodes FKBP22, an endoplasmic resident molecular chaperone that acts as a quality control on the folded triple helix of type III collagen. Some patients with clinical characteristics that resemble those of classical EDS harbor biallelic mutations in either *TNXB*, encoding tenascin X, an extracellular matrix glycoprotein that appears to regulate the assembly of collagen fibers, or in *AEBP1*, which encodes the extracellular matrix-associated adipocyte enhancer-binding protein (AEBP1), which assists in collagen polymerization. Spondylyodysplastic EDS is caused by biallelic mutations in *B3GALT7*, coding for galactosyltransferase I, or in *B3GALT6*, coding for galactosyltransferase II, both key enzymes in the biosynthesis of the linker region of glycosaminoglycans. Musculocontractural EDS results from mutations in genes coding enzymes responsible for dermatan biosynthesis: *CHST14*, dermatan 4-O-sulfotransferase 1, and *DSE*, dermatan sulfate epimerase. A rare spondylyodysplastic type of EDS is caused by biallelic mutations in *SLC39A13*, encoding the intracellular zinc transporter ZIP13. Brittle cornea syndrome is caused by biallelic mutations in either *ZNF469* or *PRDM5*, both (putative) transcriptional regulators. Finally, periodontic EDS is caused by heterozygous mutations in *C1R* or *C1S*, coding for the complement pathway components C1q and C1s, respectively.

**Diagnosis** Diagnostic workup comprises clinical examination and should be followed by genetic testing in individuals who are suspected to have an EDS subtype. Genetic testing can include targeted mutation analysis in those with a family history of EDS caused by a known genetic variant or, more frequently, next-generation sequencing using multigene panels. Genetic diagnosis should lead to family testing. Of note, the genetic cause of hypermobile EDS has not been determined, and therefore, diagnosis of this condition is based on the presence of clinical manifestations. Correlations between genotype and phenotype are challenging and only starting to emerge, and as with other heritable diseases of connective tissue, there is a large degree of variability among members of the same family carrying the same mutation.

## TREATMENT

### Ehlers-Danlos Syndrome

All patients with EDS should receive multidisciplinary care and, if available, be part of a patient advocacy community. The precise treatment depends on the subtype of EDS and the clinical manifestations. Physiotherapy is essential for patients with musculoskeletal problems. Helmets and/or skin protections or joint protections, braces, or splints can be used to reduce the risk of injury in patients with skin fragility or joint hypermobility. Low-resistance exercises (such as walking or swimming) can improve joint stability, although exercises that place considerable strain on the joints (such as gymnastics or weightlifting) should be avoided. Monitoring for cardiovascular alterations using noninvasive procedures is recommended in patients at risk of adverse cardiovascular events only. Given the rarity of vascular EDS, referral to a center with EDS expertise is of vital importance. A clear protocol for emergency room evaluation

in the case of major complications should be established, and patients should carry documentation of their genetic diagnosis, such as a MedicAlert. The psychosocial impact of a vascular EDS diagnosis often requires psychological care.

## CHONDRODYSPLASIAS

**(See also Chap. 412)** Chondrodysplasias (CDs), also referred to as skeletal dysplasias or osteochondrodysplasias, encompass a heterogeneous group of disorders characterized by intrinsic abnormalities of cartilage and bone and are generally characterized by dwarfism and abnormal body proportions (disproportionate short stature). Many affected individuals develop degenerative joint changes, and mild CD in adults may be difficult to differentiate from primary generalized OA.

**Classification** The Nosology and Classification of Genetic Skeletal Disorders comprises 461 distinct disorders based on clinical, radiographic, and/or molecular phenotypes. Pathogenic variants affecting 437 different genes have currently been found in 425 of 461 (92%) of these disorders. The conditions are divided into 42 groups based on gene/protein families (e.g., the type II collagen group), phenotypic presentation (e.g., spondylometaphyseal dysplasia), and pathophysiology (i.e., lysosomal storage disorders). One gene may be responsible for more than one condition (e.g., *COL2A1* mutations may cause achondrogenesis type 2, hypochondrogenesis, spondyloepiphyseal dysplasia congenita, Kniest and Stickler syndromes), or a condition may be due to mutations in more than one gene (e.g., geleophysic dysplasia can be caused by mutations in *ADAMTSL2*, *FBNI*, and *LTBP3*).

**Incidence** The overall incidence of all forms of CD ranges from 1 per 2500 to 1 per 4000 births. Data on the frequency of individual CDs are incomplete, but the incidence of Stickler syndrome is estimated to range between 1 in 7500 and 1 in 9000. Therefore, the disease is probably among the more common heritable disorders of connective tissue. The most common form of inherited disproportionate short stature is achondroplasia, with an estimated incidence of 1 per 26,000 to 1 per 28,000 live births.

 **Molecular Defects** Mutations in the *COL2A1* gene, coding for the chain of type II collagen of cartilage, are found in a group of patients with both mild and severe CDs. For example, a mutation in *COL2A1* substituting a cysteine residue for an arginine was found in a few unrelated families with spondyloepiphyseal dysplasia (SED) and precocious generalized OA. Mutations in the gene were also found in some lethal CDs characterized by gross deformities of bones and cartilage, such as those found in SED congenita, spondyloepimetaphyseal dysplasia congenita, hypochondrogenesis/achondrogenesis type II, and Kniest syndrome. The highest incidence of *COL2A1* mutations, however, occurs in patients with the distinctive features of the Stickler syndrome, which is characterized by skeletal changes, orofacial abnormalities, and ophthalmologic and auditory abnormalities. Most of the mutations in *COL2A1* are premature stop codons that produce haploinsufficiency. In addition, some of the patients with Stickler syndrome or a closely related syndrome have mutations in two genes specific for type XI collagen (*COL11A1* and *COL11A2*), which is an unusual heterotrimer formed from chains encoded by *COL2A1*, *COL11A1*, and *COL11A2*. Mutations in the *COL11A1* gene are also found in patients with Marshall syndrome, which is similar to classic Stickler syndrome, but with more severe hearing loss and dysmorphic features, such as a flat or retracted midface with a flat nasal bridge, short nose, anteverted nostrils, long philtrum, and large-appearing eyes.

CDs are also caused by mutations in the less abundant collagens found in cartilage. For example, patients with Schmid metaphyseal CD have mutations in the gene for type X collagen, a short, network-forming collagen found in the hypertrophic zone of endochondral cartilage. The syndrome is characterized by short stature, coxa vara, flaring metaphyses, and waddling gait. As with other collagen genes, the most common mutations are of two types: nonsense mutations that lead to haploinsufficiency and structural mutations that compromise collagen assembly.

Some patients have mutations in genes for proteins that interact with collagens. Patients with pseudoachondroplasia or autosomal dominant multiple epiphyseal dysplasia have mutations in the gene for the cartilage oligomeric matrix protein (*COMP*), a protein that interacts with both collagens and proteoglycans in cartilage. However, some families with multiple epiphyseal dysplasia have a defect in one of the three genes for type IX collagen (*COL9A1*, *COL9A2*, and *COL9A3*) or in matrilin-3, another extracellular protein found in cartilage.

Some CDs are caused by mutations in genes that affect early development of cartilage and related structures. Achondroplasia is caused by mutations in the gene for a receptor for a fibroblastic growth factor (*FGFR3*). The mutations in the *FGFR3* gene causing achondroplasia are unusual in several respects. A single-base mutation in the gene that converts glycine to arginine at position 380 in the *FGFR3* gene is present in >90% of patients. Most patients harbor a sporadic new mutation, and therefore, this nucleotide change is one of the most common recurring mutations in the human genome. The mutation causes unregulated signal transduction through the receptor and inappropriate development of cartilage. Mutations that alter other domains of *FGFR3* have been found in patients with the more severe disorders of hypochondroplasia and thanatophoric dysplasia and in a few families with a variant of craniosynostosis. However, most patients with craniosynostosis appear to have mutations in the related *FGFR2* gene. The similarities between the phenotypes produced by mutations in genes for fibroblast growth factor (FGF) receptors and mutations in structural proteins of cartilage are probably explained by the observation that the activity of FGFs is regulated in part by binding of FGFs to proteins sequestered in the extracellular matrix. Therefore, the situation parallels the interactions between transforming growth factors (TGFs) and fibrillin in MFS (see below).

Other mutations involve the proteoglycans of cartilage, aggrecan (*AGC1*) and perlecan (*HSPG2*), and in the proteoglycan posttranslational sulphation pathway (*DTDST*, *PAPSS2*, and *CHST3*). Mutations in >45 other genes have been defined in CDs.

**Diagnosis** The diagnosis of CDs is made on the basis of the physical appearance, slit-lamp eye examinations, x-ray findings, histologic changes, and clinical course. Targeted gene and exome sequencing or more global sequencing strategies are used for molecular diagnosis. Given the wide spectrum of CD phenotypes, these genetic tests are becoming critical diagnostic tools. For Stickler syndrome, more precise diagnostic criteria have made it possible to identify type I variants with mutations in the *COL2A1* gene with a high degree of accuracy. It has been suggested that the type II variant with mutations in the *COL11A1* gene can be identified on the basis of a "beaded" vitreous phenotype and that the type III variant with mutations in the *COL11A2* gene can be identified on the basis of the characteristic systemic features without the ocular involvement. Prenatal diagnosis based on analysis of DNA obtained from chorionic villus or amniotic fluid is possible.

## TREATMENT

### Chondrodysplasias

The treatment of CDs is symptomatic and is directed to secondary features such as degenerative arthritis. Many patients require joint replacement surgery and corrective surgery for cleft palate. The eyes should be monitored carefully for the development of cataracts and the need for laser therapy to prevent retinal detachment. In general, patients should be advised to avoid obesity and contact sports. Counseling for the psychological problems of short stature is critical. Several clinical trials therapeutically targeting the *FGFR3* pathway in achondroplasia are underway.

## HERITABLE THORACIC AORTIC ANEURYSM DISEASE

Heritable thoracic aortic aneurysm disease (HTAD) encompasses conditions in which aortic disease has a familial occurrence, due to an underlying genetic defect. HTAD is classified as syndromic or

**TABLE 413-4** Heritable Thoracic Aortic Disease and Associated Genes and Proteins

	GENE	PROTEIN	CONDITION	OMIM	LOCUS
Extracellular matrix proteins	<i>COL3A1</i>	$\alpha 1(\text{III})$ collagen chain	Vascular EDS	130050	2q32
	<i>FBN1</i>	Fibrillin 1	Marfan Syndrome	154700	15q21.1
	<i>MFAP5</i>	Microfibrillar associated protein 5	Familial thoracic aortic aneurysm 9	616166	12p13.31
	<i>LOX</i>	Lysyl oxidase	Familial thoracic aortic aneurysm 10	617168	5q23.1
TGF- $\beta$ signaling	<i>TGFB1</i>	Transforming growth factor receptor 1	Loeys-Dietz syndrome 1	609192	9q22.33
	<i>TGFB2</i>	Transforming growth factor receptor 2	Loeys-Dietz syndrome 2	610168	3p24.1
	<i>SMAD3</i>	Mothers against decapentaplegic drosophila homolog 3	Loeys-Dietz syndrome 3	613795	15q22.33
	<i>TGFB2</i>	Transforming growth factor $\beta$ 2	Loeys-Dietz syndrome 4	614816	1q41
	<i>TGFB3</i>	Transforming growth factor $\beta$ 3	Loeys-Dietz syndrome 5	615582	14q23.3
	<i>SMAD2</i>	Mothers against decapentaplegic drosophila homolog 2	Arterial aneurysms and dissections	/	18q21.1
	<i>ACTA2</i>	Smooth muscle actin $\alpha$ 2	Familial thoracic aortic aneurysm 6	611788	10q23.31
Smooth muscle contraction	<i>MYH11</i>	Smooth muscle myosin heavy chain 11	Familial thoracic aortic aneurysm 4	132900	16p13.11
	<i>MYLK</i>	Myosin light chain kinase	Familial thoracic aortic aneurysm 7	613780	3q21.1
	<i>PRKG1</i>	Protein kinase cGMP-dependent type 1	Familial thoracic aortic aneurysm 8	615436	10q11.2-q21.1

Abbreviations: EDS, Ehlers-Danlos syndrome; OMIM, Online Mendelian Inheritance in Man; TGF, transforming growth factor.

nonsyndromic. Syndromic HTAD may associate with ocular, craniofacial, musculoskeletal, and skin features, with a recognizable, yet sometime subtle, phenotype. They are caused by mutations in genes that code for extracellular matrix proteins. Besides syndromic HTAD, there are several nonsyndromic forms of HTAD; patients with these conditions do not display an outward recognizable phenotype and are classified as having familial thoracic aortic aneurysm (FTAAs). More extensive genetic screening in cohorts of patients with thoracic aortic aneurysm is, however, slowly revealing that there is no strict boundary between syndromic and nonsyndromic HTAD entities (Table 413-4) (Chap. 280).

**Classification** The most common form of syndromic HTAD is MFS, caused by mutations in the gene for fibrillin-1 (*FBN1*). MFS was initially characterized by a triad of features: (1) skeletal changes that include long, thin extremities, frequently associated with loose joints; (2) reduced vision as the result of dislocations of the lenses (ectopia lentis); and (3) aortic aneurysms. An international panel has developed a series of revised Ghent criteria that are useful in classifying patients. Other major syndromic HTADs include different genetic variants of Loeys-Dietz syndrome (LDS) (*TGFB1*, *TGFB2*, *TGFB3*, *SMAD2*, and *SMAD3*) and vascular EDS (*COL3A1*). Rare forms of syndromic HTAD include Shprintzen-Goldberg syndrome (*SKI*), Meester-Loeys syndrome (*BGN*), and arterial tortuosity syndrome (ATS) (*SLC2A10*).

**Incidence and Inheritance** The incidence of MFS is among the highest of any heritable disorder: ~1 in 3000–5000 births in most racial and ethnic groups. The related syndromes are less common. Mutations are generally inherited as autosomal dominant traits, but about one-fourth of patients have sporadic new mutations. The LDSs are less common, but their exact incidence is currently unknown.

**Skeletal Effects** Patients with MFS typically display a marfanoid habitus with tall stature and long limbs. The ratio of the upper segment (top of the head to the top of the pubic ramus) to the lower segment (top of the pubic ramus to the floor) is usually 2 standard deviations below mean for age, race, and sex. The fingers and hands are long and slender and have a spider-like appearance (arachnodactyly). Overlapping features in MFS and LDS include scoliosis or kyphoscoliosis; anterior chest deformities, including pectus excavatum, pectus carinatum, or asymmetry; pes planus; pneumothorax; and dural ectasia. A few patients have severe joint hypermobility similar to EDS. Clubfeet, joint contractures, and cervical spine instability are more frequently observed in LDS. Patients with *SMAD3* mutations are particularly prone to premature OA.

**Cardiovascular Features** Cardiovascular abnormalities are the major source of morbidity and mortality both in MFS and LDS (Chap. 280). Patients with MFS often have mitral valve prolapse that develops early in life and that progresses to mitral valve regurgitation of increasing severity in about one-quarter of patients. Dilatation of the root of the aorta and the sinuses of Valsalva are characteristic and ominous features of MFS that can develop at any age. The rate of dilation is unpredictable, but it can lead to aortic regurgitation, dissection of the aorta, and rupture. Dilatation is probably accelerated by physical and emotional stress as well as by pregnancy. Cardiovascular features of LDS also include dilatation of the aortic root at the level of the sinus of Valsalva, which can progress to dissection or rupture when left untreated. LDS is also known for its involvement of aneurysms affecting arterial branches of head, neck, thoracic and abdominal aorta, lung, and lower extremities and for the presence or tortuosity of these vessels. In contrast to MFS, congenital heart malformations are often noted.

**Ocular Features** Myopia is the most common ocular feature of MFS and often presents in early childhood. Displacement of the lens from the center of the pupil (ectopia lentis) occurs in ~60% of MFS patients. The ocular globe is frequently elongated. Retinal detachment, early cataract formation, and glaucoma can occur. Ectopia lentis does not usually occur in LDS, but other ocular features may be present, such as blue sclerae, strabismus, amblyopia, and myopia.

**Other Features** MFS patients typically have a high arched palate. Patients with LDS characteristically display hypertelorism (widely spaced eyes) and cleft palate or bifid (split) uvula. They may also have craniostenosis. Shared mucocutaneous features include striae, typically over the shoulders and buttocks, and inguinal and incisional hernias. Patients with LDS may display more EDS-like skin features, such as thin translucent skin and widened scars.

 **Molecular Defects** More than 90% of patients clinically classified as having MFS by the Ghent criteria have a mutation in the gene for fibrillin-1 (*FBN1*). Mutations in the same gene are found in a few patients who do not meet the Ghent criteria. Most *FBN1* gene mutations are unique and are scattered throughout its 65 coding exons. Approximately 10% are recurrent new mutations that are largely located in CpG sequences known to be "hot spots." About one-third of the mutations introduce premature termination codons, and about two-thirds are missense mutations that alter calcium-binding domains in the repetitive epidermal growth factor-like domains of the protein. Rarer mutations alter the processing of the protein. As in many genetic diseases, the severity of the phenotype cannot be predicted from the

**3230** nature of the mutation. In LDS, components of the TGF- signaling pathway are mutated, including the cytokines (TGF 2, TGF 3), the receptors (TGFBR1, TGFBR2), and the downstream effectors (SMAD2, SMAD3).

The discovery that syndromes similar to MFS are caused by alterations in the TGF- signaling pathway refocused attention on structural similarity between fibrillin-1 and TGF- binding proteins that sequester TGF- in the extracellular matrix. As a result, some of the manifestations of MFS have been shown to arise from alterations in binding sites that modulate TGF- bioavailability during development of the skeleton and other tissues. In both MFS and LDS, the pathogenic mechanisms involve increased TGF- signaling, which contributes to aneurysm formation.

**Diagnosis** When HTAD is present, genetic testing can confirm the diagnosis and allow identification of at-risk individuals. Referral to a specialty genetics service is critically important, and genetic counseling before testing is recommended. In view of phenotypic overlap between the syndromic HTAD, a multigene panel (usually including genes for syndromic and nonsyndromic HTAD) is recommended. All patients with a suspected diagnosis of MFS should have a slit-lamp examination and an echocardiogram. Also, homocystinuria should be ruled out by amino acid analysis of plasma (Chap. 420). The diagnosis of MFS according to the international Ghent standards places emphasis on two cardinal features, dilation of the ascending aorta with or without dissection and ectopia lentis. Other cardiovascular and ocular manifestations and findings in other organ systems such as the skeleton, dura, skin, and lungs contribute to a systemic score that guides diagnosis when aortic disease is present but ectopia lentis is not.

## TREATMENT

### Marfan Syndrome and Loeys-Dietz Syndromes

Patients should be advised that vascular risks are increased by severe physical exertion, smoking, emotional stress, and pregnancy. Low-level moderate aerobic exercise and limits on isometric exercise are recommended. Prophylactic beta blocker and/or angiotensin II receptor blocker therapy are prescribed in normotensive individuals, and blood pressure control is important for those with hypertension. Surgical correction of the aorta, aortic valve, and mitral valve has been successful in many patients, but tissues are frequently friable. The scoliosis tends to be progressive, and surgical stabilization may be required. Dislocated lenses rarely require surgical removal, but patients should be followed closely for retinal detachment.

## ACKNOWLEDGEMENT

Darwin J. Prockop and John F. Bateman contributed to this chapter in the 20th edition, and some material from that chapter has been retained here.

## FURTHER READING

- De Backer J et al: Genetic testing for aortopathies: primer for the nongeneticist. *Curr Opin Cardiol* 34:585, 2019.
- Forlino A, Marini JC: Osteogenesis imperfecta. *Lancet* 387:1657, 2016.
- Loeys BL et al: The revised Ghent nosology for the Marfan syndrome. *J Med Genet* 47:476, 2010.
- Malrait F et al: The Ehlers-Danlos syndromes. *Nat Rev Dis Primers* 6:64, 2020.
- Marini JC et al: Osteogenesis imperfecta. *Nat Rev Dis Primers* 3:17052, 2017.
- Marzin P, Cormier-Daire V: New perspectives on the treatment of skeletal dysplasia. *Ther Adv Endocrinol Metab* 11:2042018820904016, 2020.
- Mortier GR et al: Nosology and classification of genetic skeletal disorders: 2019 revision. *Am J Med Genet A* 179:2393, 2019.
- Prockop DJ, Kivirikko, KI: Collagen: Molecular biology, diseases and potentials for therapy. *Ann Rev Biochem* 64:403, 1995.



## DEFINITION

Hemochromatosis is a relatively common inherited disorder of iron metabolism prevalent in European populations. Once thought to be a single disease entity, it is now known to be an iron-storage disorder with genetic heterogeneity but with a final common metabolic pathway resulting in the inappropriately high cellular release of iron. This leads to an increase in intestinal iron absorption and the deposition of excess iron in parenchymal cells with eventual tissue damage and organ failure. Thus, the term *hemochromatosis* now refers to a group of genetic diseases that predispose to iron overload, potentially leading to fibrosis and organ failure. Cirrhosis of the liver, diabetes mellitus, arthritis, cardiomyopathy, and hypogonadotropic hypogonadism are the major clinical manifestations.

The following terminology is widely accepted.

1. *Heredity hemochromatosis* is most often caused by a mutation in the homeostatic iron regulator (*HFE*) gene, which is tightly linked to the HLA-A locus on chromosome 6p. Persons who are homozygous for the mutation are at increased risk of iron overload and account for 80–90% of clinical hereditary hemochromatosis in persons of northern European descent. In such subjects, the presence of hepatic fibrosis, cirrhosis, arthropathy, or hepatocellular carcinoma constitutes iron overload-related disease. Rarer forms of non-*HFE* hemochromatosis are caused by mutations in other genes involved in iron metabolism (Table 414-1). The disease can be recognized during its early stages when iron overload and organ damage are minimal. At this stage, the disease is best referred to as *early hemochromatosis* or *precirrhotic hemochromatosis*.
2. *Secondary iron overload* occurs as a result of an iron-loading anemia, such as thalassemia or sideroblastic anemia, in which erythropoiesis is increased but ineffective. In the acquired iron-loading disorders, massive iron deposits in parenchymal tissues can lead to the same clinical and pathologic features as in hemochromatosis.

TABLE 414-1 Classification of Iron Overload States

### Heredity Hemochromatosis

Hemochromatosis, *HFE*-related (type 1)

C282Y homozygosity

C282Y/H63D compound heterozygosity

Hemochromatosis, non-*HFE*-related

Juvenile hemochromatosis (type 2A) (hemojuvelin mutations)

Juvenile hemochromatosis (type 2B) (hepcidin mutation)

Mutated transferrin receptor 2, *TFR2* (type 3)

Mutated ferroportin 1 gene, *SLC40A1* (type 4)

### Acquired Iron Overload

Iron-loading anemias

Thalassemia major

Sideroblastic anemia

Chronic hemolytic anemias

Transfusional and parenteral iron overload

Dietary iron overload

Chronic liver disease

Hepatitis C

Alcoholic cirrhosis, especially when advanced

Nonalcoholic steatohepatitis

Porphyria cutanea tarda

Dysmetabolic iron overload syndrome

Post-portacaval shunting

### Miscellaneous

Iron overload in sub-Saharan Africa

Neonatal iron overload

Aceruloplasminemia

Congenital atransferrinemia

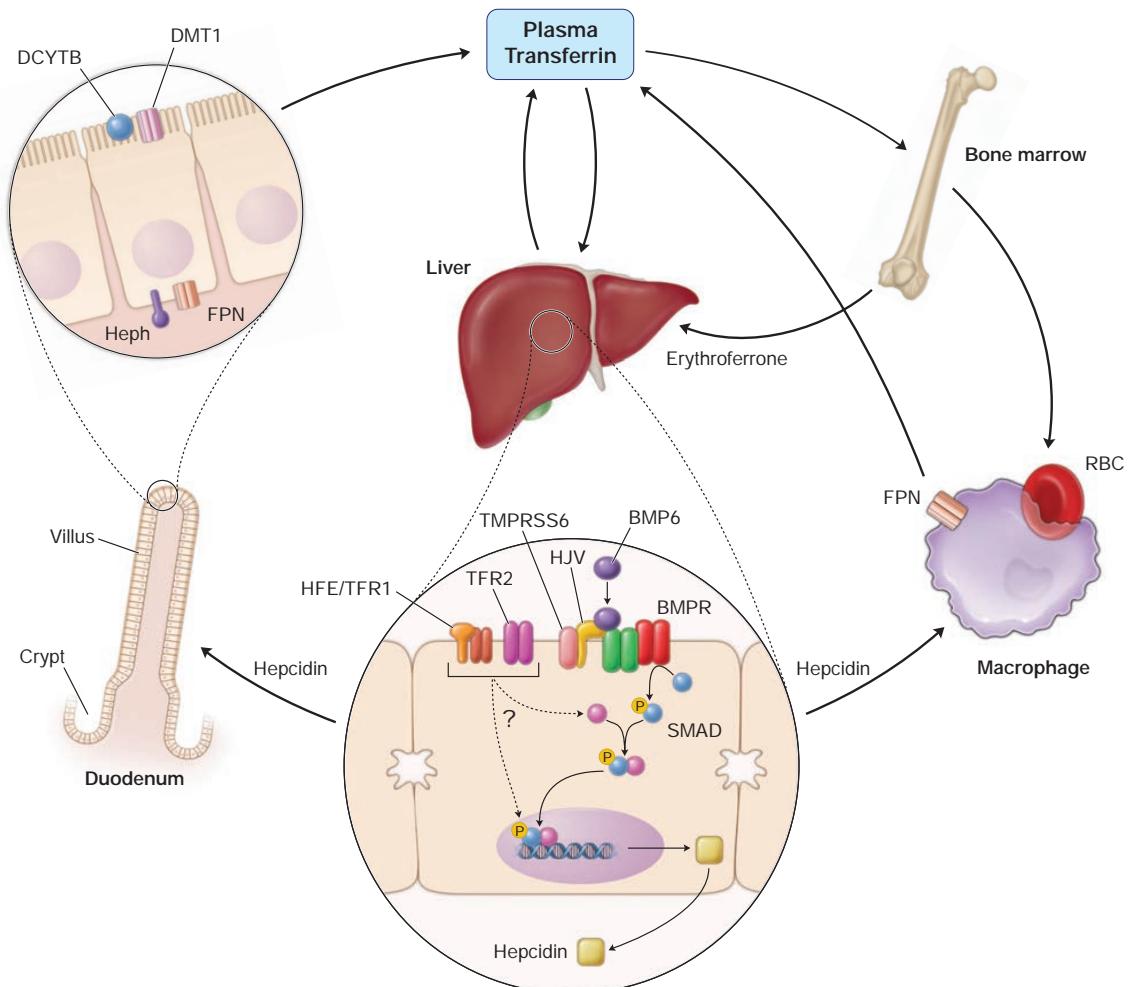
Although *HFE*-associated hemochromatosis mutations are common, the prevalence varies in different ethnic groups. It is most common in populations of northern European extraction in whom ~1 in 10 persons are heterozygous carriers and 0.3–0.5% are homozygotes, with even higher percentages in some Celtic populations such as those residing in Ireland and Brittany. However, expression of the disease is variable and modified by several factors, especially alcohol consumption, dietary iron intake, blood loss associated with menstruation and pregnancy, and blood donation. Recent population studies indicate that ~30% of homozygous men develop iron overload-related disease and about 6% develop hepatic cirrhosis. For women, iron overload-related disease is closer to 1%. In addition, there are as yet unidentified modifying genes responsible for expression. Nearly 70% of untreated patients develop the first symptoms between ages 40 and 60. The disease is rarely evident before age 20, although with family screening (see “Screening for Hemochromatosis,” below) and periodic health

examinations, asymptomatic subjects with iron overload can be identified, including young menstruating women.

In contrast to *HFE*-associated hemochromatosis, the non-*HFE*-associated forms of hemochromatosis (Table 414-1) are rare, but they affect all populations and may affect young people (juvenile hemochromatosis).

These result from mutations in one or more of the genes encoding proteins in the hepcidin pathway (Fig. 414-1), including hepcidin, hemojuvelin, and transferrin receptor 2 (TFR2). The resultant clinical disease is very similar to *HFE*-related disease because they all lead to hepcidin deficiency, which is the final common pathway (Fig. 414-1).

A rare autosomal dominant form of hemochromatosis results from two types of mutations in the gene for the iron transporter ferroportin. Loss-of-function mutations decrease the cell surface localization of ferroportin in certain tissues, thereby reducing its ability to export iron (“ferroportin disease”). A second mutation abolishes the hepcidin-induced ferroportin internalization and degradation resulting in a



**FIGURE 414-1 Pathways of normal iron homeostasis.** Dietary inorganic iron traverses the brush border membrane of duodenal enterocytes via divalent metal-ion transporter 1 (DMT1) after reduction of ferric ( $Fe^{3+}$ ) iron to the ferrous ( $Fe^{2+}$ ) state by intestinal ferrereductases such as duodenal cytochrome B (DCYTB). Iron then moves from the enterocyte to the circulation via a process requiring the basolateral iron exporter ferroportin (FPN) and the iron oxidase hephaestin (Heph). In the circulation, iron binds to plasma transferrin and is thereby distributed to sites of iron utilization and storage. Much of the diferric transferrin supplies iron to immature erythrocyte cells in the bone marrow for hemoglobin synthesis. At the end of their life, senescent red blood cells (RBCs) are phagocytosed by macrophages, and iron is returned to the circulation after export through ferroportin. The liver-derived peptide hepcidin represses basolateral iron transport in the gut as well as iron released from macrophages and other cells and serves as a central regulator of body-iron traffic. At least two separate signals regulate hepcidin production in response to changes in body-iron requirements. The first involves the detection of circulating diferric transferrin by HFE and TFR2. A second relies on hepatic iron stores activating the hemojuvelin (HJV)-dependent bone morphogenetic protein (BMP)/SMAD pathway. This pathway is modified by erythroferrone released from erythroid precursor cells, which binds to BMP6 and inhibits its function. TMPRSS6 is a protease that regulates hepcidin production, possibly by modulating HJV activity. Heme is metabolized by heme oxygenase within the enterocytes, and the released iron then follows the same pathway. Mutations in the genes encoding HFE, TFR2, HJV, and hepcidin all lead to decreased hepcidin release and increased iron absorption, resulting in hemochromatosis (Table 414-1).

## GENETIC BASIS

The most common mutation in the *HFE* gene is a homozygous G to A transition that leads to a cysteine to tyrosine substitution at position 282 (C282Y) of the HFE protein. It has been identified in 85–90% of patients with hereditary hemochromatosis in populations of northern European descent but is found in only 60% of cases from Mediterranean populations. A second, relatively common *HFE* variant (H63D) results in a substitution of aspartic acid for histidine at residue 63 of the HFE protein. Homozygosity for H63D is not associated with clinically significant iron overload. Some compound heterozygotes (i.e., one copy each of C282Y and H63D) have mild to moderately increased body-iron stores but develop clinical disease only in association with cofactors such as heavy alcohol intake or hepatic steatosis. *HFE*-associated hemochromatosis is inherited as an autosomal recessive trait, and heterozygotes have no, or minimal, increase in iron stores. However, this slight increase in hepatic iron can act as a cofactor that may modify the expression of other diseases such as porphyria cutanea tarda (PCT) or nonalcoholic steatohepatitis (NASH).

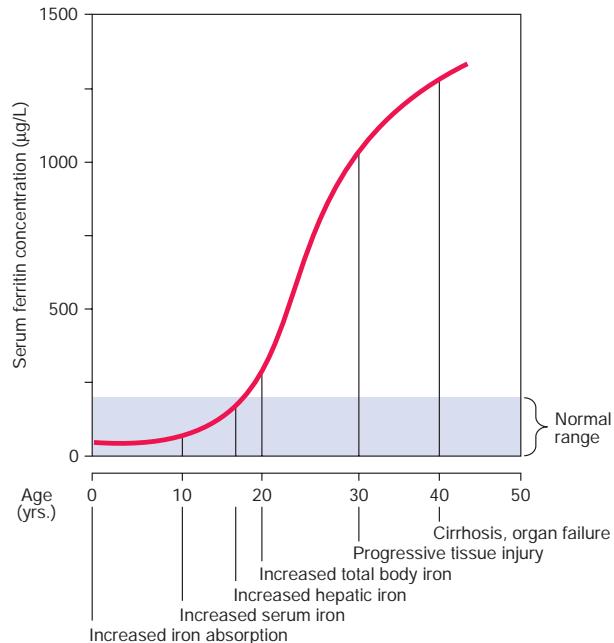
Mutations in other genes involved in iron metabolism are responsible for non-*HFE*-associated hemochromatosis, including juvenile hemochromatosis, which affects persons in the second and third decades of life (Table 414-1). Mutations in the genes encoding hepcidin, TFR2, and hemojuvelin (Fig. 414-1) result in clinicopathologic features that are indistinguishable from *HFE*-associated hemochromatosis. However, loss-of-function mutations in ferroportin, which is responsible for the efflux of iron from most cell types, result in iron loading of reticuloendothelial macrophages as well as parenchymal cells.

## PATOPHYSIOLOGY AND THE ROLE OF HEPCIDIN

Normally, the body-iron content of 3–4 g is maintained such that intestinal mucosal absorption of iron is equal to iron loss. This amount is ~1 mg/d in men and 1.5 mg/d in menstruating women. In hemochromatosis, mucosal absorption is greater than body requirements and amounts to ~4 mg/d. The progressive accumulation of iron increases plasma iron and saturation of transferrin and results in a progressive increase of plasma ferritin (Fig. 414-2). The discovery of a key regulatory hormone that allows the bone marrow and other tissues to communicate their iron requirements has transformed our understanding of the coordination of absorption, mobilization, and storage of iron to meet body iron requirements. It was called hepcidin based upon its antibacterial activity (“HEPatic bacterioCIDal proteIN”). This liver-derived peptide represses basolateral iron export from intestinal enterocytes and iron release from macrophages and other cells by binding to ferroportin. Hepcidin, in turn, responds to signals in the liver mediated by HFE, TFR2, and hemojuvelin (Fig. 414-1). The development of hepcidin agonists represents a promising new therapeutic approach for iron overload disorders caused by low hepcidin levels.

The *HFE* gene encodes a 343-amino-acid protein that is structurally related to MHC class I proteins. The basic defect in *HFE*-associated hemochromatosis is a lack of cell surface expression of HFE (due to the C282Y mutation). The normal (wild-type) HFE protein forms a complex with  $\beta$ -microglobulin and transferrin receptor 1 (TFR1), and the C282Y mutation completely abrogates this interaction. As a result, the mutant HFE protein remains trapped intracellularly. Although the precise function of HFE at the cell surface is not known, mutations in this protein reduce hepcidin production leading to increased dietary iron absorption (Fig. 414-1). In advanced disease, the body may contain 20 g or more of iron, which is deposited mainly in parenchymal cells of the liver, pancreas, and heart. Iron deposition in the pituitary causes hypogonadotropic hypogonadism in both men and women. Tissue injury may result from disruption of iron-laden lysosomes, from lipid peroxidation of subcellular organelles by excess iron, or from stimulation of collagen synthesis by activated stellate cells.

*Secondary iron overload* with iron deposition in parenchymal cells occurs in chronic disorders of erythropoiesis, particularly in those with



**FIGURE 414-2** Sequence of events in genetic hemochromatosis and their correlation with the serum ferritin concentration. Increased iron absorption is present throughout life. Overt, symptomatic disease usually develops between ages 40 and 60, but latent disease can be detected long before this.

defects in hemoglobin synthesis and ineffective erythropoiesis such as sideroblastic anemia and thalassemia (Chap. 98). In these disorders, iron absorption is increased. Moreover, these patients require blood transfusions and are frequently treated inappropriately with iron. PCT, a disorder characterized by a defect in porphyrin biosynthesis (Chap. 416), can also be associated with excessive parenchymal iron deposits. The magnitude of the iron load in PCT is usually insufficient to produce tissue damage. However, some patients with PCT also have mutations in the *HFE* gene, and some have associated hepatitis C virus (HCV) infection. Although the relationship between these disorders remains to be clarified, iron overload accentuates the inherited enzyme deficiency in PCT and should be avoided along with other agents (alcohol, estrogens, haloaromatic compounds) that may exacerbate PCT. Another cause of hepatic parenchymal iron overload is hereditary aceruloplasminemia. In this disorder, impairment of iron mobilization due to deficiency of ceruloplasmin (a ferroxidase) causes iron overload in hepatocytes and a range of other cell types.

*Excessive iron ingestion* over many years rarely results in hemochromatosis. An important exception has been reported in South Africa among groups who brew fermented beverages in vessels made of iron. Hemochromatosis has been described in apparently normal persons who have taken medicinal iron over many years, but such individuals probably had genetic disorders.

The common denominator in all patients with hemochromatosis is *excessive amounts of iron in parenchymal tissues*. Parenteral administration of iron in the form of blood transfusions or iron preparations results predominantly in reticuloendothelial cell iron overload. This appears to lead to less tissue damage than iron loading of parenchymal cells.

In the liver, parenchymal iron is in the form of ferritin and hemosiderin. In the early stages, these deposits are seen in the periportal parenchymal cells, especially within lysosomes in the pericanalicular cytoplasm of the hepatocytes. This stage progresses to perilobular fibrosis and to fibrous septa due to activation of stellate cells. In the advanced stage, a macronodular or mixed macro- and micronodular cirrhosis develops. Hepatic fibrosis and cirrhosis correlate significantly with hepatic iron concentration.

Histologically, iron is increased in many organs, particularly in the liver, heart, and pancreas, and, to a lesser extent, in the endocrine

glands. The epidermis of the skin is thin, and melanin is increased in the cells of the basal layer and dermis. Deposits of iron are present around the synovial lining cells of the joints.

### CLINICAL MANIFESTATIONS

C282Y homozygotes can be characterized by the stage of progression as follows: (1) a genetic predisposition without abnormalities; (2) iron overload without symptoms; (3) iron overload with symptoms (e.g., arthritis and fatigue); and (4) iron overload with organ damage—in particular, cirrhosis. Thus, many subjects with significant iron overload are asymptomatic. For example, in a study of 672 asymptomatic C282Y homozygous subjects (identified by either family screening or routine health examinations) there was hepatic iron overload (grades 2–4) in 56% and 34.5% of male and female subjects, respectively, hepatic fibrosis (stages 2–4) in 18.4% and 5.4%, respectively, and cirrhosis in 5.6% and 1.9%, respectively.

Initial symptoms of hemochromatosis are often nonspecific and include lethargy, arthralgia, skin pigmentation, loss of libido, and features of diabetes mellitus. Hepatomegaly, increased pigmentation, spider angiomas, splenomegaly, arthropathy, ascites, cardiac arrhythmias, congestive heart failure, loss of body hair, testicular atrophy, and jaundice are prominent in advanced disease.

The liver is usually the first organ to be affected, and hepatomegaly is present in >95% of symptomatic patients.

Manifestations of portal hypertension and esophageal varices occur less commonly than in cirrhosis from other causes. Hepatocellular carcinoma develops in ~30% of patients with cirrhosis, and it is the most common cause of death in treated patients—hence the importance of early diagnosis and therapy. The incidence increases with age, it is more common in men, and it occurs almost exclusively in cirrhotic patients.

Excessive skin pigmentation is present in patients with advanced disease. The characteristic metallic or slate-gray hue is sometimes referred to as *bronzing* and results from increased melanin and iron in the dermis. Pigmentation usually is diffuse and generalized.

*Diabetes mellitus* occurs in ~65% of patients with advanced disease and is more likely to develop in those with a family history of diabetes, suggesting that direct damage to the pancreatic islets by iron deposition occurs in combination with other risk factors. The management is similar to that of other forms of diabetes.

*Arthropathy* develops in 25–50% of symptomatic patients. It usually occurs after age 50, but may occur as a first manifestation or long after therapy. The joints of the hands, especially the second and third metacarpophalangeal joints, are usually the first joints involved, a feature that helps to distinguish the chondrocalcinosis associated with hemochromatosis from the idiopathic form (*Chap. 372*). A progressive polyarthritides involving the wrists, hips, ankles, and knees may also ensue. Acute brief attacks of synovitis may be associated with deposition of calcium pyrophosphate (chondrocalcinosis or pseudogout), mainly in the knees. Radiologic manifestations include cystic changes of the subchondral bones, loss of articular cartilage with narrowing of the joint space, diffuse demineralization, hypertrophic bone proliferation, and calcification of the synovium. The arthropathy tends to progress despite removal of iron by phlebotomy. Although the relation of these abnormalities to iron metabolism is not known, the fact that similar changes occur in other forms of iron overload suggests that iron is directly involved.

*Cardiac involvement* is the presenting manifestation in ~15% of symptomatic patients. The most common manifestation is congestive heart failure, which occurs in ~10% of young adults with the disease, especially those with juvenile hemochromatosis. Symptoms of congestive heart failure may develop suddenly, with rapid progression to death if untreated. The heart is diffusely enlarged. This may be misdiagnosed as idiopathic cardiomyopathy if other overt manifestations are absent. Cardiac arrhythmias include premature supraventricular beats, paroxysmal tachyarrhythmias, atrial flutter, atrial fibrillation, and varying degrees of atrioventricular block.

*Hypogonadism* occurs in both sexes and may antedate other clinical features. Manifestations include loss of libido, impotence, amenorrhea, testicular atrophy, gynecomastia, and sparse body hair. These changes are primarily the result of decreased production of gonadotropins due to impairment of hypothalamic-pituitary function by iron deposition.

### DIAGNOSIS

The association of (1) hepatomegaly, (2) skin pigmentation, (3) diabetes mellitus, (4) heart disease, (5) arthritis, and (6) hypogonadism should suggest the diagnosis. However, as stated above, significant iron overload may exist with none or only some of these manifestations. Therefore, a high index of suspicion is needed to make the diagnosis early. Treatment before permanent organ damage occurs can reverse the iron toxicity and restore life expectancy to normal.

The history should be particularly detailed in regard to disease in other family members and should include information on alcohol ingestion; iron intake; and ingestion of large doses of ascorbic acid, which promotes iron absorption (*Chap. 333*). Appropriate tests should be performed to exclude iron deposition due to hematologic disease. The presence of liver, pancreatic, cardiac, and joint disease should be confirmed by physical examination, radiography, and standard function tests of these organs.

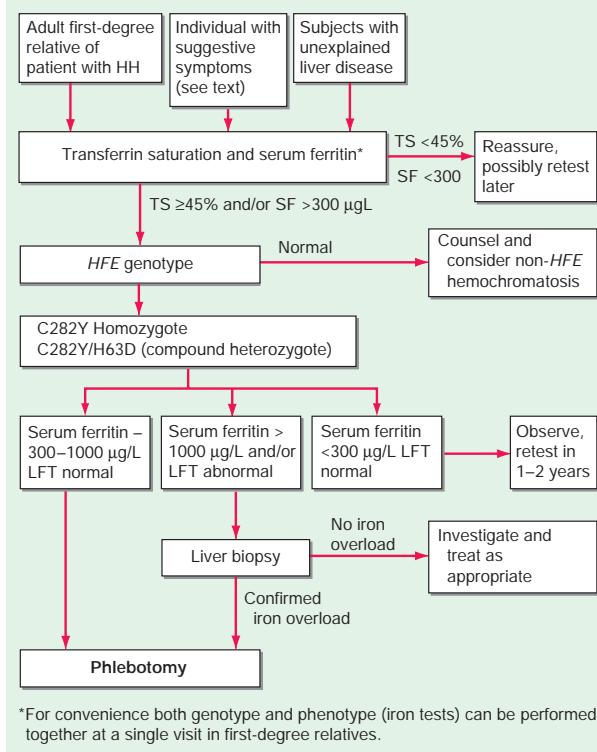
The degree of increase in total body iron stores can be assessed by (1) measurement of serum iron and the percent saturation of transferrin (or the unsaturated iron-binding capacity), (2) measurement of serum ferritin concentration, (3) liver biopsy with measurement of the iron concentration and calculation of the hepatic iron index (**Table 414-2**), and (4) MRI of the liver. In addition, a retrospective assessment of body-iron storage is also provided by performing weekly phlebotomy and calculating the amount of iron removed before iron stores are exhausted (1 mL blood = ~0.5 mg iron).

Each of these methods for assessing iron stores has advantages and limitations. The serum iron level and percent saturation of transferrin are elevated early in the course, but their specificity is reduced by significant false-positive and false-negative rates. For example, serum iron concentration may be increased in patients with alcoholic liver disease without iron overload; in this situation, however, the hepatic iron index is usually not increased as in hemochromatosis (Table 414-2). In otherwise healthy persons, a fasting serum transferrin saturation >45% is abnormal and suggests homozygosity for hemochromatosis.

The serum ferritin concentration is usually a good index of body-iron stores, whether decreased or increased. In fact, an increase of 1 µg/L in serum ferritin level reflects an increase of ~8–10 mg in body stores. In most untreated patients with hemochromatosis, the serum

**TABLE 414-2 Representative Iron Values in Normal Subjects, Patients with Hemochromatosis, and Patients with Alcoholic Liver Disease**

DETERMINATION	NORMAL	SYMPOTOMATIC HEMOCHROMATOSIS	HOMOZYGOTES WITH EARLY, ASYMPTOMATIC HEMOCHROMATOSIS	HETEROZYGOTES	ALCOHOLIC LIVER DISEASE
Plasma iron, µmol/L (µg/dL)	9–27 (50–150)	32–54 (180–300)	Usually elevated	Elevated or normal	Often elevated
Total iron-binding capacity, µmol/L (µg/dL)	45–66 (250–370)	36–54 (200–300)	36–54 (200–300)	Normal	45–66 (250–370)
Transferrin saturation, %	22–45	50–100	50–100	Normal or elevated	27–60
Serum ferritin, µg/L		1000–6000	200–500	Usually <500	10–500
Men	20–250				
Women	15–150				
Liver iron, µg/g dry wt	300–1400	6000–18,000	2000–4000	300–3000	300–2000
Hepatic iron index	<1.0	>2	1.5–2	<2	<2



**FIGURE 414-3** Algorithm for screening for *HFE*-associated hemochromatosis. HH, hereditary hemochromatosis, homozygous subject (*C282Y* *+/+*); LFT, liver function test; SF, serum ferritin concentration; TS, transferrin saturation.

ferritin level is significantly increased (Fig. 414-2 and Table 414-2), and a serum ferritin level  $>1000 \mu\text{g}/\text{L}$  is the strongest predictor of disease expression among individuals homozygous for the *C282Y* mutation. However, in patients with inflammation and hepatocellular necrosis, serum ferritin levels may be elevated out of proportion to body iron stores due to increased release from tissues. Therefore, a repeat determination of serum ferritin should be carried out after acute hepatocellular damage has subsided (e.g., in alcoholic liver disease). Ordinarily, the combined measurements of the percent transferrin saturation and serum ferritin level provide a simple and reliable screening test for hemochromatosis, including the precirrhotic phase of the disease. If either of these tests is abnormal, genetic testing for hemochromatosis should be performed (Fig. 414-3).

The role of liver biopsy in the diagnosis and management of hemochromatosis has been reassessed as a result of the widespread availability of genetic testing for the *C282Y* mutation. The absence of severe fibrosis can be accurately predicted in most patients using clinical and biochemical variables. Thus, there is virtually no risk of severe fibrosis in a *C282Y* homozygous subject with (1) serum ferritin level  $<1000 \mu\text{g}/\text{L}$ , (2) normal serum alanine aminotransferase values, (3) no hepatomegaly, and (4) no excess alcohol intake. However, it should be emphasized that liver biopsy is the only reliable method for establishing or excluding the presence of hepatic cirrhosis, which is the critical factor determining prognosis and the risk of developing hepatocellular carcinoma. Biopsy also permits histochemical estimation of tissue iron and measurement of hepatic iron concentration. Increased density of the liver due to iron deposition can be demonstrated by CT or MRI, and with improved technology, MRI has become more accurate in determining hepatic iron concentration.

### SCREENING FOR HEMOCHROMATOSIS

When the diagnosis of hemochromatosis is established, it is important to counsel and screen other family members (Chap. 467).

Asymptomatic and symptomatic family members with the disease usually have an increased saturation of transferrin and an increased serum ferritin concentration. These changes occur even before iron stores are greatly increased (Fig. 414-2). All adult first-degree relatives of patients with hemochromatosis should be tested for the *C282Y* and *H63D* mutations and counseled appropriately (Fig. 414-3). In affected individuals, it is important to confirm or exclude the presence of cirrhosis and begin therapy as early as possible. For children of an identified proband, testing for *HFE* mutations in the other parent is helpful because if normal, the child is merely an obligate heterozygote and at no risk. Otherwise, for practical purposes, children need not be checked before they are 18 years old.

The role of population screening for hemochromatosis is controversial. Recent studies indicate that it is highly effective for primary care physicians to screen subjects using transferrin saturation and serum ferritin levels. Such screening also detects iron deficiency. Genetic screening of the normal population is feasible but remains controversial in terms of cost-effectiveness.

## TREATMENT

### Hemochromatosis

The therapy of hemochromatosis involves removal of the excess body iron and supportive treatment of damaged organs. Iron removal is best accomplished by weekly or, with gross iron loading, twice-weekly phlebotomy of 500 mL. Although there is an initial modest decline in the volume of packed red blood cells to about 35 mL/dL, the level stabilizes after several weeks. The plasma transferrin saturation remains increased until the available iron stores are depleted. In contrast, the plasma ferritin concentration falls progressively, reflecting the gradual decrease in body-iron stores. One 500-mL unit of blood contains 200–250 mg of iron, and 25 g of iron may have to be removed. Therefore, in patients with advanced disease, weekly phlebotomy may be required for 1–2 years, and it should be continued until the serum ferritin level is  $100 \mu\text{g}/\text{L}$ . Thereafter, phlebotomies are performed at appropriate intervals to maintain ferritin levels at  $100 \mu\text{g}/\text{L}$ . The transferrin saturation fluctuates and may still be elevated but should not dictate further therapy unless it is persistently at 100% when free unbound iron may circulate. Usually one phlebotomy every 3 months will suffice. It is important, however, not to overtreat and render the patient iron deficient.

Chelating agents such as deferoxamine, when given parenterally, remove 10–20 mg of iron per day, which is much less than that mobilized by once-weekly phlebotomy. Phlebotomy is also less expensive, more convenient, and safer for most patients. However, chelating agents are indicated when anemia or hypoproteinemia is severe enough to preclude phlebotomy. Subcutaneous infusion of deferoxamine using a portable pump is the most effective means of its administration.

Effective oral iron chelating agents, deferasirox (Exjade) and deferiprone, are now available. These agents are effective in thalassemia and secondary iron overload, but are expensive and carry the risk of significant side effects.

Alcohol consumption should be severely curtailed or eliminated because it increases the risk of cirrhosis in hereditary hemochromatosis nearly tenfold. Dietary adjustments are unnecessary, although vitamin C and iron supplements should be avoided. The management of hepatic failure, cardiac failure, and diabetes mellitus is similar to conventional therapy for these conditions. Loss of libido and change in secondary sex characteristics are managed with testosterone replacement or gonadotropin therapy (Chap. 391).

End-stage liver disease may be an indication for liver transplantation, although results are improved if the excess iron can be removed beforehand. The available evidence indicates that the fundamental metabolic abnormality in hemochromatosis is reversed by successful liver transplantation.

## PROGNOSIS

The principal causes of death are cardiac failure, hepatocellular failure, or portal hypertension and hepatocellular carcinoma.

Life expectancy is improved by removal of excessive iron stores and maintenance of these stores at near-normal levels. The 5-year survival rate with therapy increases from 33% to 89%. With repeated phlebotomy, the liver decreases in size, liver function improves, pigmentation of skin decreases, and cardiac failure may be reversed. Diabetes improves in ~40% of patients, but removal of excess iron has little effect on hypogonadism or arthropathy. Hepatic fibrosis may decrease, but established cirrhosis is irreversible. Hepatocellular carcinoma occurs as a late sequela in patients who are cirrhotic at presentation. The apparent increase in its incidence in treated patients is probably related to their increased life span. Hepatocellular carcinoma rarely develops if the disease is treated in the precirrhotic stage. Indeed, the life expectancy of homozygotes treated before the development of cirrhosis is normal.

The importance of family screening and early diagnosis and treatment cannot be overemphasized. Asymptomatic individuals detected by family studies should have phlebotomy therapy if iron stores are moderately to severely increased. Assessment of iron stores at appropriate intervals is also important. With this management approach, most manifestations of the disease can be prevented.

## ROLE OF HFE MUTATIONS IN OTHER LIVER DISEASES

 There is considerable interest in the role of *HFE* mutations and hepatic iron in several other liver diseases. Several studies have shown an increased prevalence of *HFE* mutations in PCT patients. Iron accentuates the inherited enzyme deficiency in PCT and clinical manifestations of PCT. The situation in NASH is less clear, but some studies have shown an increased prevalence of *HFE* mutations in NASH patients. The role of phlebotomy therapy, however, is unproven despite an intriguing fall in liver enzyme levels. In chronic HCV infection, *HFE* mutations are not more common, but some subjects have increased hepatic iron. Before initiating antiviral therapy in these patients, it is reasonable to perform phlebotomy therapy to remove excess iron stores, because this reduces liver enzyme levels.

*HFE* mutations are not increased in frequency in alcoholic liver disease. Hemochromatosis in a heavy drinker can be distinguished from alcoholic liver disease by the presence of the C282Y mutation.

End-stage liver disease may also be associated with iron overload of the degree seen in hemochromatosis. The mechanism is uncertain, although studies have shown that alcohol suppresses hepatic hepcidin secretion. Hemolysis also plays a role. *HFE* mutations are uncommon.

Whether subjects homozygous for C282Y are at increased risk of breast and colorectal cancer is controversial.

## GLOBAL CONSIDERATIONS

The *HFE* mutation is of northern European origin (Celtic or Nordic) with a heterozygous carrier rate of ~1 in 10 (1 in 8 in Ireland). Thus, *HFE*-associated hemochromatosis is quite rare in non-European populations, e.g., Asia. However, non-*HFE*-associated hemochromatosis resulting from mutations in other genes involved in iron metabolism (Fig. 414-1) is ubiquitous and should be considered when one encounters iron overload.

African iron overload occurs primarily in sub-Saharan Africa and was previously thought to be due to the consumption of an iron-rich fermented maize beverage. However, recent evidence suggests that it is primarily the result of a non-*HFE*-related genetic trait that is exacerbated by dietary iron loading. A similar form of iron overload has been described in African Americans. Further research is needed to clarify this condition.

## FURTHER READING

- Anderson GJ, Bardou-Jacquet E: Revisiting hemochromatosis: Genetic vs. phenotypic manifestations. *Ann Transl Med* 9:731, 2021.
- Piperno A et al: Inherited iron overload disorders. *Transl Gastroenterol Hepatol* 5:25, 2020.
- Powell LW et al: Haemochromatosis. *Lancet* 388:706, 2016.

415

## Wilson's Disease

Stephen G. Kaler



Wilson's disease is an inherited human disorder of copper transport that primarily impacts the liver and brain. This reflects the critical need for homeostatic mechanisms to properly utilize this trace metal, both systemically and in the central nervous system. Since the initial detailed clinical description in 1912, Wilson's disease has emerged as arguably one of the best-characterized and most effectively managed human inborn errors of metabolism. The condition results from variants in *ATP7B*, a highly evolutionarily conserved P-type ion-motive ATPase that normally mediates copper ion removal from the liver via biliary excretion and prevents brain copper accumulation. Prompt diagnosis in the early symptomatic phase of the illness (or presymptomatic detection) and lifelong treatment are needed to avoid premature mortality in affected individuals.

## HISTORY OF WILSON'S DISEASE

Wilson's disease (hepatolenticular degeneration) was first described in 1912 by neurologist S.A.K. Wilson, who recognized the inherited aspect of the condition. In 1948, the pathologist J.N. Cummings proposed an etiologic connection with copper overload. Several years later, a metal chelator developed to counteract an arsenic-based chemical warfare agent (lewisite) was used to successfully treat advanced Wilson's disease. In 1956, copper chelation by d-penicillamine was introduced and found preferable to anti-lewisite with respect to administration and side effect profile. In the early 1970s, an alternative copper chelator, triethylene tetramine, became the second U.S. Food and Drug Administration (FDA)-approved treatment for Wilson's disease. Also in the early 1970s, the first liver transplants were performed for Wilson's disease, with resultant correction of both hepatic failure and crippling neurologic impairments in patients unresponsive to medical therapies. The treatment potential of zinc salts to reduce gastrointestinal copper absorption in Wilson's disease was recognized in the early 1960s, eventually leading to FDA approval for this indication. Tetrathiomolybdate, which forms a tripartite complex with copper and albumin, and a bacterial peptide, methanobactin, which traverses mitochondrial membranes, are more recently proposed copper chelators with potential for treatment of Wilson's disease.

In 1993, the gene for Wilson's disease was identified and found to encode a copper-transporting ATPase, *ATP7B*, expressed primarily in liver and kidney. In addition to providing a molecular basis for diagnosis and genotype-phenotype correlations, the finding presents current opportunities for viral gene therapy that could impact future management of this illness.

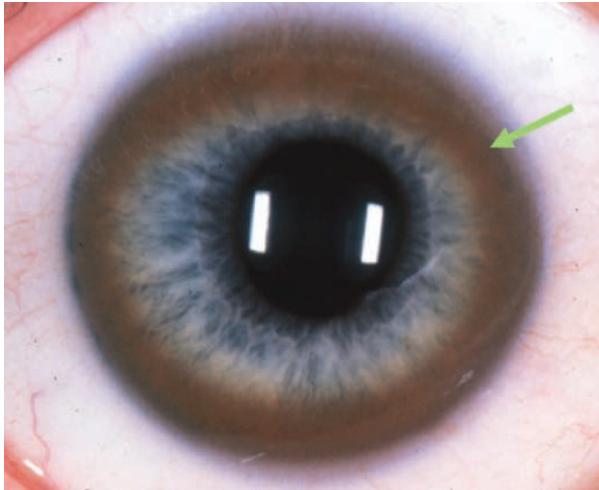
## PHENOTYPES

### CLINICAL

Presenting clinical features of Wilson's disease include nonspecific liver disease, neurologic abnormalities, psychiatric illness, hemolytic anemia, renal tubular Fanconi syndrome, and various skeletal abnormalities. Age influences the specific presentation in Wilson's disease. Nearly all individuals who present with liver disease are <30 years of age, whereas those presenting with neurologic or psychiatric signs may range in age from the first to the fifth decade. This reflects the sequence of events in the pathogenesis of the illness. However, regardless of clinical presentation, some degree of liver disease is invariably present.

**Hepatic Presentation** With hepatic presentations, signs and symptoms include jaundice, hepatomegaly, edema, or ascites. Viral hepatitis and cirrhosis are often initial diagnostic considerations in individuals who, in fact, have Wilson's disease.

**Neurologic Presentation** In patients with neurologic presentations, abnormalities include speech difficulty (dysarthria), dystonia, rigidity, tremor or choreiform movements, abnormal gait, and uncoordinated handwriting. Wilson's disease may be classified as a movement



**FIGURE 415-1** Kayser-Fleischer ring in Wilson's disease, representing copper deposition in Descemet membrane of the cornea. (Image courtesy of Tjaard U. Hoogenraad MD, PhD, Department of Neurology, University Medical Centre Utrecht, Utrecht, The Netherlands.)

disorder. The neurologic signs and symptoms reflect the predilection for basal ganglia (e.g., caudate, putamen) involvement in the brains of affected persons. Parkinson disease or other movement disorders may be mistakenly diagnosed.

**Psychiatric Presentation** In psychiatric presentations, changes in personality (irritability, anger, poor self-control), depression, and anxiety are common symptoms. Typically, patients presenting in this fashion are in their late teens or early twenties, a period during which substance abuse is also a diagnostic consideration. Wilson's disease should be formally excluded in all teenagers and young adults with new-onset psychiatric signs.

**Ocular Manifestations** The eye is a primary site of copper deposition in Wilson's disease, producing a pathognomonic sign, the Kayser-Fleischer ring (Fig. 415-1), a golden to greenish-brown band in the peripheral cornea. This important diagnostic sign first appears as a superior crescent and then develops inferiorly and ultimately becomes circumferential. Slit-lamp or optical coherent tomography examinations are required to detect rings in their early stage of formation. Copper can also accumulate in the lens and produce "sunflower" cataracts. Approximately 95% of Wilson's disease patients with neurologic signs manifest the Kayser-Fleischer ring compared to ~65% of those with hepatic presentations. Copper chelation therapy causes fading and eventual disappearance of corneal copper.

**Other Clinical Manifestations** Secondary endocrine effects of Wilson-associated liver disease may include delayed puberty or amenorrhea. Renal tubular dysfunction in Wilson's disease leads to abnormal losses of amino acids, electrolytes, calcium, phosphorus, and glucose. Presumably, this effect is related to copper toxicity. High copper levels have been noted previously in the kidneys of patients with Wilson's disease. Treatment with copper chelation often improves the renal disturbances. There can also be skeletal effects of Wilson's disease, including osteoporosis and rickets, and these may be attributable to renal losses of calcium and phosphorus. Osteoarthritis primarily affecting the knees and wrists may involve excess copper deposition in the bone and cartilage.

Hemolytic anemia due to the direct toxic effects of copper on red blood cell membranes is usually associated with release of massive quantities of hepatic copper into the circulation, a phenomenon that can be sudden and catastrophic.

## BIOCHEMICAL

Laboratory findings that support the diagnosis of Wilson's disease include low levels of serum copper and serum ceruloplasmin, elevated

hepatic transaminase levels, aminoaciduria, and hemolytic anemia. Incorporation of radiolabeled  $^{64}\text{Cu}$  into serum ceruloplasmin, measured as the appearance of copper in the serum after an oral load, is a highly specific diagnostic test; patients with Wilson's disease incorporate very little  $^{64}\text{Cu}$  into ceruloplasmin.

Increased urinary excretion of copper ( $>100 \mu\text{g}/24 \text{ h}$ ) is an easily performed and important diagnostic test for Wilson's disease. Acid-washed (copper-free) collection containers should be used. The penicillamine challenge is a variation using serial urine copper measurements in which 500 mg of penicillamine are administered orally after collecting a baseline 24-h urine. The penicillamine dose is repeated after 12 h, at the midpoint of the second 24-h urine collection. A severalfold increase in copper excretion in the second collection is suggestive of the diagnosis.

Although invasive, percutaneous needle liver biopsy for measurement of hepatic copper remains a gold standard technique for Wilson's disease diagnosis. Hepatic copper values  $>200 \mu\text{g}$  per gram of dry weight (normal 20–50  $\mu\text{g}$ ) are characteristic of Wilson's disease. Inductively coupled plasma mass spectrometry and atomic absorption spectrometry are preferred quantitative methods; histochemical staining for copper in liver biopsy specimens is unreliable.

## MOLECULAR

Wilson's disease is caused by loss-of-function variants in *ATP7B*. Despite similar genomic structures, large deletions are much less common in *ATP7B* than in *ATP7A*, the closely related X-linked gene responsible for Menkes disease. Several *ATP7B* missense variants are common (H1069Q, M645R, and R778L), with various allelic frequencies reflecting geographic, racial, and/or ethnic differences. Major *ATP7B* databases list  $>650$  pathogenic or likely pathogenic variants. Population-based and genomic-based estimates of prevalence range from 1 in 7000 to 1 in 30,000, with genome-based ascensions supporting the higher prevalence. This disparity may reflect incomplete penetrance, although there is little doubt that some affected individuals unfortunately escape medical attention.

## DIAGNOSIS

The formal diagnosis of Wilson's disease relies on a combination of clinical, biochemical, and molecular features (Table 415-1). A scoring system (Leipzig) that weights and collates various signs and symptoms was produced by an international expert group in 2001 and remains a valuable guide to diagnosis endorsed by the European Association for the Study of the Liver (EASL).

## TREATMENT

### Wilson's Disease

#### COPPER CHELATION

The era of successful treatment of Wilson's disease began with the use of British anti-lewisite (BAL) by a defined regimen of intramuscular injections. An orally administered alternative was d-penicillamine (Cuprimine), a free thiol that binds copper. This chelating drug does not formally correct the basic defect of impaired copper excretion in the bile. However, it greatly enhances urinary excretion of copper and thereby corrects and prevents copper overload and its effects. Pyridoxine (vitamin B<sub>6</sub>) is usually prescribed concomitantly to counter the tendency for deficiency of this vitamin to develop during chronic penicillamine administration.

Certain individuals are intolerant of penicillamine, however, encountering significant side effects that include nephrotoxicity, hematologic abnormalities, and a distinctive rash, *elastosis perforans serpiginosa* (usually involving the neck and axillae). Furthermore, in some Wilson's disease patients with neurologic presentations, penicillamine treatment induces paradoxical worsening of neurologic status. Triethylenetetramine dihydrochloride (trientine hydrochloride [Syprine]) is a suitable alternative chelating agent with a somewhat less extensive side effect profile.

Tetrathiomolybdate (TM) is another molecule in the Wilson's disease therapeutic armamentarium. TM forms stable tripartite

**TABLE 415-1 Main Diagnostic Features of Wilson's Disease**

CLINICAL SIGNS/SYMPOMTS	BIOCHEMICAL/LABORATORY FINDINGS	MOLECULAR FINDINGS
Hepatic:	Low serum copper Low serum ceruloplasmin Increased urinary copper excretion Elevated liver enzymes Hypoalbuminemia	Variants in <i>ATP7B</i> on both chromosomes Variants or polymorphisms in other genes ( <i>CAT</i> , <i>SOD2</i> , <i>MTHFR</i> ) may influence clinical expression of Wilson's disease in some individuals
Neurologic:	Increased liver copper level Fatty liver Cirrhotic liver Hemolytic anemia Renal Fanconi syndrome	
Ocular:	Kayser-Fleischer ring Sunflower cataract (rare)	
Psychiatric:	Decline in school Personality change Mood disorder Schizophrenia	

complexes among albumin, copper, and itself. This drug functions both to decrease copper absorption and to reduce circulating free copper. It is fast acting and can restore normal copper balance within several weeks compared to the several months required with other copper chelators or with zinc. This drug is the subject of recent clinical trials and may one day be an approved treatment for advanced breast cancer as well as Wilson's disease.

**Copper Chelation Treatment During Pregnancy** Spontaneous miscarriage is increased in women with untreated Wilson's disease. From a benefit/risk perspective, it is important to maintain copper chelation treatment during pregnancy to prevent hepatic or neurologic relapse, as well as to lower risk of pregnancy loss. Some academic centers favor copper chelator dose reduction during pregnancy, although if zinc monotherapy (see below) is in place at the time of conception, evidence suggests it is safe to maintain the usual daily dose. Since all anticopper medications enter breast milk, breast-feeding is not recommended for mothers with Wilson's disease.

#### REDUCTION OF COPPER ABSORPTION

Zinc acetate (Galzin) has proven highly effective for treatment of Wilson's disease. The mechanism involves induction of metallothionein synthesis in intestinal epithelial cells; increased metallothionein synthesis results in greater binding of dietary copper and thus decreased absorption. Zinc therapy has particular value in (1) young, presymptomatic patients; (2) patients who are pregnant, given the possible fetal teratogenic effects of other compounds; and (3) as maintenance therapy for patients after their initial "de-coppering" is accomplished. Zinc acetate has minimal side effects. The only drawback to its use is the relatively long time (4–6 months) needed for restoration of proper copper balance when used as monotherapy in the initial stages of treatment.

#### LIVER TRANSPLANTATION

Liver transplantation is a consideration for Wilson's disease in advanced stages and/or when the condition is unresponsive to

medical therapy. This is generally necessary only in cases where delayed diagnosis or poor compliance results in irreversible hepatic damage. A recently proposed alternative for this circumstance is methanobactin, a bacterial peptide that binds copper avidly and dramatically improves mitochondrial copper overload and restores normal mitochondrial morphology in a preclinical (rat) model.

#### GENE THERAPY

In a different preclinical (mouse) model of Wilson, proof of principle that adeno-associated virus-mediated *ATP7B* addition to hepatocytes can be effective was recently demonstrated. Transduction of only 20% of hepatocytes was sufficient to normalize copper homeostasis in the animal model. These results potentially pave the way for clinical trials of gene therapy in Wilson's disease patients.

#### FUTURE OUTLOOK

Wilson's disease is arguably one of the best-characterized human inborn errors of metabolism from combined clinical, biochemical, and molecular perspectives, related to the detailed attention devoted to this condition. As noted, novel copper chelators are still being evaluated, and generic formulations of established drugs are contributing to increased affordability for patients and their families. Viral gene therapy to provide working versions of *ATP7B* to the liver, kidney, and brain or that delivers gene-editing molecules to correct specific mutant alleles is now an emerging prospect. In addition, advances in newborn screening technology may eventually enable wider population-based screening for Wilson's disease, which could help address lingering questions about clinical penetrance. Such future progress in newborn screening would also avert the tragedy that missed diagnoses of this eminently treatable disorder of copper transport represent.

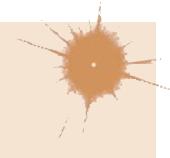
#### FURTHER READING

- Bandmann O et al: Wilson's disease and other neurological copper disorders. *Lancet Neurol* 14:103, 2015.
- European Association for Study of Liver: EASL Clinical Practice Guidelines. Wilson's disease. *J Hepatol* 56:671, 2012.
- Gao J et al: The global prevalence of Wilson disease from next-generation sequencing data. *Genet Med* 21:1155, 2019.
- Kumar M et al: WilsonGen: A comprehensive clinically annotated genomic variant resource for Wilson's disease. *Sci Rep* 10:9037, 2020.
- Murillo O et al: Liver expression of a MiniATP7B gene results in long-term restoration of copper homeostasis in a Wilson disease model in mice. *Hepatology* 70:108, 2019.
- Sandahl TD et al: The prevalence of Wilson's disease: An update. *Hepatology* 71:722, 2020.
- Wallace DF, Dooley JS: ATP7B variant penetrance explains differences between genetic and clinical prevalence estimates for Wilson disease. *Hum Genet* 139:1065, 2020.

416

## The Porphyrias

Robert J. Desnick, Manisha Balwani



#### THE PORPHYRIAS: INTRODUCTION

The porphyrias are metabolic disorders, each resulting from the deficiency or increased activity of a specific enzyme in the heme biosynthetic pathway (Fig. 416-1 and Table 416-1). These enzyme disorders are inherited as autosomal dominant, autosomal recessive, or X-linked traits, with the exception of porphyria cutanea tarda (PCT), which is usually sporadic (Table 416-1). The porphyrias are classified as either *hepatic* or *erythropoietic*, depending on the primary site of

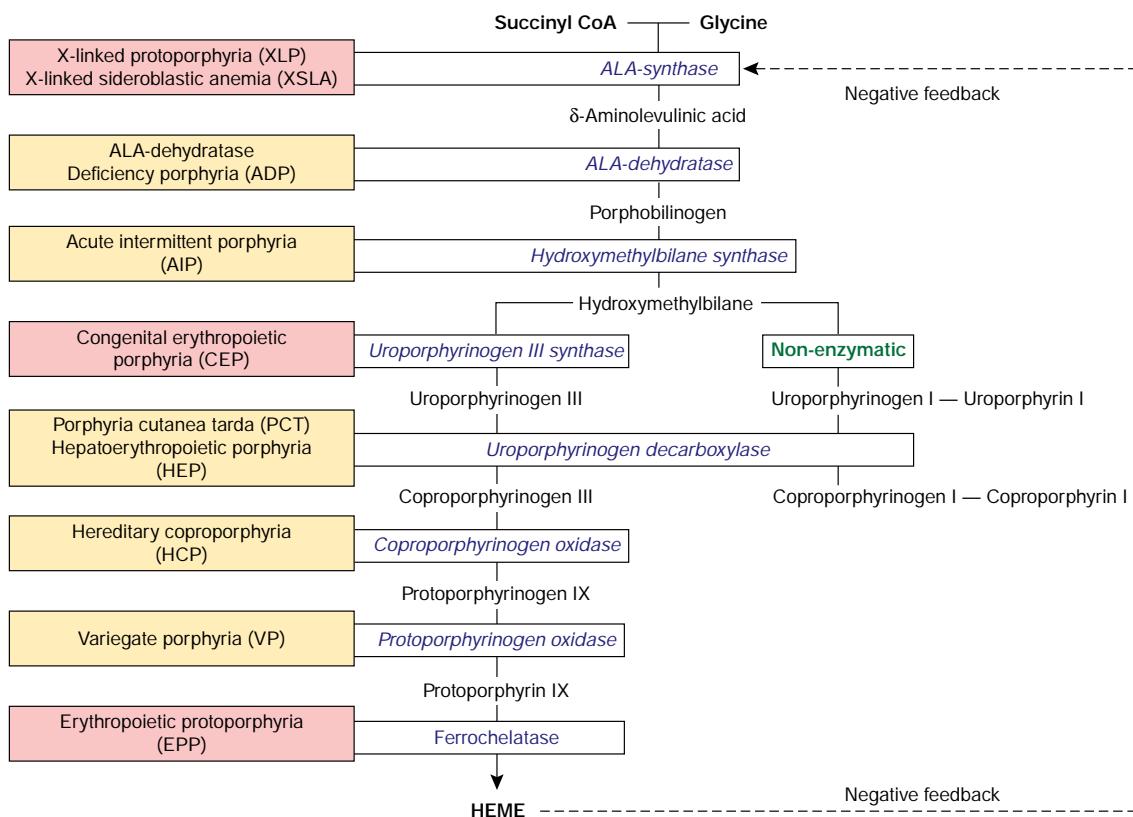


FIGURE 416-1 The human heme biosynthetic pathway indicating in *linked boxes* the enzyme that, when deficient or overexpressed, causes the respective porphyria. Hepatic porphyrias are shown in *yellow boxes* and erythropoietic porphyrias in *pink boxes*.

TABLE 416-1 Human Porphyrias: Major Clinical and Laboratory Features

PORPHYRIA	DEFICIENT ENZYME	INHERITANCE	PRINCIPAL SYMPTOMS: NV OR CP+	ENZYME ACTIVITY % OF NORMAL	INCREASED PORPHYRIN PRECURSORS AND/OR PORPHYRINS		
					ERYTHROCYTES	URINE	STOOL
<b>Hepatic Porphyrias</b>							
5-ALA-dehydratase-deficient porphyria (ADP)	ALA-dehydratase	AR	NV	~5	Zn-protoporphyrin	ALA, coproporphyrin III	—
Acute intermittent porphyria (AIP)	HMB-synthase	AD	NV	~50	—	ALA, PBG, uroporphyrin	—
Porphyria cutanea tarda (PCT)	URO-decarboxylase	AD	CP	~20	—	Uroporphyrin, 7-carboxylate porphyrin	Isocoprotoporphyrin
Hereditary coproporphyria (HCP)	COPRO-oxidase	AD	NV and CP	~50	—	ALA, PBG, coproporphyrin III	Coproporphyrin III
Variegate porphyria (VP)	PROTO-oxidase	AD	NV and CP	~50	—	ALA, PBG, coproporphyrin III	Coproporphyrin III, protoporphyrin
<b>Erythropoietic Porphyrias</b>							
Congenital erythropoietic porphyria (CEP)	URO-synthase	AR	CP	1–5	Uroporphyrin I Coproporphyrin I	Uroporphyrin I <sup>a</sup> Coproporphyrin I <sup>a</sup>	Coproporphyrin I
Erythropoietic protoporphyria (EPP)	Ferrochelatase	AR	CP	~20–30	Protoporphyrin	—	Protoporphyrin
X-linked protoporphyria (XLP)	ALA-synthase 2	XL	CP	>100 <sup>b</sup>	Protoporphyrin	—	Protoporphyrin

<sup>a</sup>Type I isomers. <sup>b</sup>Increased activity due to gain-of-function mutations in *ALAS2* exon 11.

Abbreviations: AD, autosomal dominant; ALA, 5-aminolevulinic acid; AR, autosomal recessive; COPRO, coproporphyrin; CP, cutaneous photosensitivity; NV, neurovisceral; PBG, porphobilinogen; PROTO, protoporphyrin; URO, uroporphyrin; XL, X-linked.

TABLE 416-2 Human HEME Biosynthetic Enzymes and Genes

ENZYME	GENE SYMBOL	CHROMOSOMAL LOCATION	cDNA (bp)	GENE		PROTEIN (aa)	SUBCELLULAR LOCATION	KNOWN MUTATIONS <sup>b</sup>	THREE-DIMENSIONAL STRUCTURE <sup>c</sup>
				SIZE (KB)	EXONS <sup>a</sup>				
ALA-synthase									
Housekeeping	ALAS1	3p21.1	2199	17	11	640	M	—	
Erythroid-specific	ALAS2	Xp11.2	1937	22	11	587	M	>30	—
ALA-dehydratase									
Housekeeping	ALAD	9q32	1149	15.9	12 (1A + 2 – 12)	330	C	12	Y
Erythroid-specific	ALAD	9q32	1154	15.9	12 (1B + 2 – 12)	330	C	—	
HMB-synthase									
Housekeeping	HMBS	11q23.3	1086	11	15 (1 + 3 – 15)	361	C	400	E
Erythroid-specific	HMBS	11q23.3	1035	11	15 (2 – 15)	344	C	10	
URO-synthase									
Housekeeping	UROS	10q26.2	1296	34	10 (1 + 2B – 10)	265	C	45	H
Erythroid-specific	UROS	10q26.2	1216	34	10 (2A + 2B – 10)	265	C	4	
URO-decarboxylase	UROD	1p34.1	1104	3	10	367	C	122	H
COPRO-oxidase	CPOX	3q12.1	1062	14	7	354	M	70	H
PROTO-oxidase	PPOX	1q23.3	1431	5.5	13	477	M	181	—
Ferrochelatase	FECH	18q21.31	1269	45	11	423	M	192	B

<sup>a</sup>Number of exons and those encoding separate housekeeping and erythroid-specific forms indicated in parentheses. <sup>b</sup>Number of known mutations from the Human Gene Mutation Database ([www.hgmd.org](http://www.hgmd.org)). <sup>c</sup>Crystallized from human (H), murine (M), *Escherichia coli* (E), *Bacillus subtilis* (B), or yeast (Y) purified enzyme; references in Protein Data Bank ([www.rcsb.org](http://www.rcsb.org)).

**Abbreviations:** ALA, 5-aminolevulinic acid; C, cytoplasm; COPRO, coproporphyrin; HMB, hydroxymethylbilane; M, mitochondria; PROTO, protoporphyrin; URO, uroporphyrin.

**Source:** Reproduced with permission from KE Anderson et al: Disorders of heme biosynthesis: X-linked sideroblastic anemia and the porphyrias, in Scriven CR: The Metabolic and Molecular Bases of Inherited Diseases. New York, NY: McGraw-Hill; 2001.

overproduction and accumulation of their respective porphyrin precursors or porphyrins (Tables 416-1 and 416-2), although some have overlapping features. For example, PCT, the most common porphyria, is hepatic and presents with blistering cutaneous photosensitivity, which is typically characteristic of the erythropoietic porphyrias.

The major manifestations of the acute hepatic porphyrias are neurologic, including neuropathic abdominal pain, peripheral motor neuropathy, and mental disturbances, with attacks often precipitated by dieting, certain porphyrinogenic drugs, and hormonal changes. While hepatic porphyrias are symptomatic primarily in adults, rare homozygous variants of the autosomal dominant hepatic porphyrias usually manifest clinically prior to puberty. In contrast, the erythropoietic porphyrias usually present at birth or in early childhood with cutaneous photosensitivity or, in the case of congenital erythropoietic porphyria (CEP), even in utero as nonimmune hydrops fetalis. Cutaneous sensitivity to sunlight results from excitation of excess porphyrins in the skin by long-wave ultraviolet light, leading to cell damage, scarring, and disfigurement. Thus, the porphyrias are metabolic disorders in which environmental, physiologic, and genetic factors interact to cause disease.

Because many symptoms of the porphyrias are nonspecific, diagnosis is often delayed. Laboratory measurement of porphyrin precursors (5'-aminolevulinic acid [ALA] and porphobilinogen [PBG]) in the urine or porphyrins in the urine, plasma, erythrocytes, or feces is required to confirm or exclude the various types of porphyria (see below). However, a definite diagnosis requires demonstration of the specific gene defect (Table 416-3). The genes encoding all the heme biosynthetic enzymes have been characterized, permitting identification of the mutations causing each porphyria (Table 416-2). Molecular genetic analyses now make it possible to provide precise heterozygote or homozygote identification and prenatal diagnoses in families with known mutations.

In addition to recent reviews of the porphyrias, informative and up-to-date websites are sponsored by the American Porphyria Foundation ([www.porphyriafoundation.com](http://www.porphyriafoundation.com)) and the European Porphyria Network (<https://porphyria.eu>). An extensive list of unsafe and safe drugs for individuals with acute porphyrias is provided at the Drug Database for Acute Porphyrias ([www.drugs-porphyria.org](http://www.drugs-porphyria.org)).

## GLOBAL CONSIDERATIONS

The porphyrias are panethnic metabolic diseases that affect individuals around the globe. The acute hepatic porphyrias—acute intermittent porphyria (AIP), hereditary coproporphyrin (HCP), and variegate porphyria (VP)—are autosomal dominant disorders. The frequency of symptomatic AIP, the most common acute hepatic porphyria, is ~1 in 20,000 among Caucasian individuals of Western European ancestry, and it is particularly frequent in Scandinavians, with a frequency of ~1 in 10,000 in Sweden. However, recent studies using genomic/exomic databases showed an estimated frequency of pathogenic variants in the HMBS gene as ~1 in 1700. Thus, the penetrance of AIP, and likely the other acute hepatic porphyrias, is low, with ~1% of those with pathogenic mutations experiencing acute attacks (see below).

VP is particularly frequent in South Africa, where its high prevalence (>10,000 affected patients) is in part due to a genetic “founder effect.” The autosomal recessive acute hepatic porphyria, ALA-dehydratase-deficient porphyria (ADP), is very rare, and <20 patients have been reported worldwide.

The erythropoietic porphyrias—CEP, erythropoietic protoporphyrina (EPP), and X-linked protoporphyrin (XLP)—also are panethnic. EPP is the most common porphyria in children, whereas CEP is very rare, with ~200 reported cases worldwide. The frequency of EPP varies globally because most patients have the common low expression ferrochelatase (FECH) mutation that varies in frequency in different populations. The allele rarely occurs in Africans, is present in ~10% of whites, and is frequent (~30%) in the Japanese. The reported prevalence of EPP in the Caucasian population ranges from 1 in ~75,000 to 1 in ~150,000.

The autosomal recessive porphyrias—ADP, CEP, and hepatocerebrovascular porphyria (HEP)—are more frequent in regions with high rates of consanguineous unions. PCT, which is typically sporadic, occurs more frequently in countries in which its predisposing risk factors such as hepatitis C and HIV are more prevalent.

## HEME BIOSYNTHESIS

Heme biosynthesis involves eight enzymatic steps in the conversion of glycine and succinyl-CoA to heme (Fig. 416-2 and Table 416-2). These eight enzymes are encoded by nine genes, as the first enzyme in

TABLE 416-3 Diagnosis of Acute and Cutaneous Porphyrias

SYMPTOMS	FIRST-LINE TEST: ABNORMALITY	POSSIBLE PORPHYRIA	SECOND-LINE TESTING IF FIRST-LINE TESTING IS POSITIVE: TO INCLUDE: URINE (U), PLASMA (P), AND FECAL (F) PORPHYRINS; FOR ACUTE PORPHYRIAS, ADD RED BLOOD CELL (RBC) HMB-SYNTHASE; FOR BLISTERING SKIN LESIONS, ADD P AND RBC PORPHYRINS	CONFIRMATORY TEST: ENZYME ASSAY AND/OR MUTATION ANALYSIS
Neurovisceral	Spot U: ↑↑ALA and normal PBG	ADP	U porphyrins: ↑↑, mostly COPRO III P & F porphyrins: normal or slightly ↑ RBC HMB-synthase: normal	Rule out other causes of elevated ALA; ↓↓RBC ALA-dehydratase activity (<10%); ALA-dehydratase mutation analysis
	Spot U: ↑↑PBG	AIP	U porphyrins: ↑↑, mostly URO and COPRO P & F porphyrins: normal or slightly ↑ RBC HMB-synthase: usually ↓	HMB-synthase mutation analysis
	"	HCP	U porphyrins: ↑↑, mostly COPRO III P porphyrins: normal or slightly ↑ (if skin lesions present) F porphyrins: ↑↑, mostly COPRO III	Measure RBC HMB-synthase: normal activity COPRO-oxidase mutation analysis
	"	VP	U porphyrins: ↑↑, mostly COPRO III P porphyrins: ↑↑ (characteristic fluorescence peak at neutral pH) F porphyrins: ↑↑, mostly COPRO and PROTO	Measure RBC HMB-synthase: normal activity PROTO-oxidase mutation analysis
Blistering skin lesions	P: ↑ porphyrins	PCT and HEP	U porphyrins: ↑↑, mostly URO and heptacarboxylate porphyrin P porphyrins: ↑↑ F porphyrins: ↑↑, including increased isocoproporphyrin RBC porphyrins: ↑↑ zinc PROTO in HEP <sup>a</sup>	RBC URO-decarboxylase activity: half-normal in familial PCT (~20% of all PCT cases); substantially deficient in HEP URO-decarboxylase mutation analysis: mutation(s) present in familial PCT (heterozygous) and HEP (homozygous)
	"	HCP and VP	See HCP and VP above. Also, U ALA and PBG: may be ↑	
	"	CEP	RBC and U porphyrins: ↑↑, mostly URO I and COPRO I F porphyrins: ↑↑; mostly COPRO I	↓↓ RBC URO-synthase activity (<15%) URO-synthase mutation analysis
Nonblistering photosensitivity	P: porphyrins usually ↑	EPP	RBC porphyrins: ↓↓, mostly free PROTO U porphyrins: normal F porphyrins: normal or ↓, mostly PROTO	FECH mutation analysis
	P: porphyrins usually ↑	XLP	RBC porphyrins: ↑↑, approximately equal free and zinc PROTO U porphyrins: normal F porphyrins: normal or ↑, mostly PROTO	ALAS2 mutation analysis

<sup>a</sup>Nonspecific increases in zinc protoporphyrins are common in other porphyrias.

**Abbreviations:** ADP, 5-ALA-dehydratase-deficient porphyria; AIP, acute intermittent porphyria; ALA, 5-aminolevulinic acid; CEP, congenital erythropoietic porphyria; COPRO I, coproporphyrin I; COPRO III, coproporphyrin III; EPP, erythropoietic protoporphyrina; F, fecal; HCP, hereditary coproporphyria; HEP, hepatoerythropoietic porphyria; ISOCOPRO, isocoproporphyrin; P, plasma; PBG, porphobilinogen; PCT, porphyria cutanea tarda; PROTO, protoporphyrin IX; RBC, erythrocytes; U, urine; URO I, uroporphyrin I; URO III, uroporphyrin III; VP, variegate porphyria; XLP, X-linked protoporphyrina.

**Source:** Data from KE Anderson et al: Recommendations for the diagnosis and treatment of the acute porphyrias. Ann Intern Med 142:439, 2005.

the pathway, ALA-synthase, has two genes that encode unique housekeeping (*ALAS1*) and erythroid-specific (*ALAS2*) isoforms. The first and last three enzymes in the pathway are located in the mitochondria, whereas the other four are in the cytosol. Heme is required for a variety of hemoproteins such as hemoglobin, myoglobin, respiratory cytochromes, and the cytochrome P450 (CYP) enzymes. Hemoglobin synthesis in erythroid precursor cells accounts for ~85% of daily heme synthesis in humans. Hepatocytes account for most of the rest, primarily for the synthesis of CYPs, which are especially abundant in the liver endoplasmic reticulum, and turn over more rapidly than many other hemoproteins, such as the mitochondrial respiratory cytochromes. As shown in Fig. 416-2, the pathway intermediates are the porphyrin precursors, ALA and PBG, and porphyrins (mostly in their reduced forms, known as *porphyrinogens*). At least in humans, these intermediates do not accumulate in significant amounts under normal conditions or have important physiologic functions.

The first enzyme, ALA-synthase, catalyzes the condensation of glycine, activated by pyridoxal phosphate and succinyl-coenzyme A, to form ALA. In the liver, this rate-limiting enzyme can be induced by a variety of drugs, steroids, and other chemicals. Distinct nonerythroid (e.g., housekeeping) and erythroid-specific forms of ALA-synthase are encoded by separate genes located on chromosome 3p21.1 (*ALAS1*) and Xp11.2 (*ALAS2*), respectively. Defects in the erythroid gene *ALAS2* that decrease its activity cause an X-linked sideroblastic anemia (XLSA). Gain-of-function mutations in the last exon (11) of

*ALAS2* that increase its activity cause an X-linked form of EPP, known as XLP.

The second enzyme, ALA-dehydratase, catalyzes the condensation of two molecules of ALA to form PBG. Hydroxymethylbilane synthase (HMB-synthase; also known as PBG-deaminase) catalyzes the head-to-tail condensation of four PBG molecules by a series of deaminations to form the linear tetrapyrrole, HMB. Uroporphyrinogen III synthase (URO-synthase) catalyzes the rearrangement and rapid cyclization of HMB to form the asymmetric, physiologic, octacarboxylate porphyrinogen, uroporphyrinogen (URO<sup>gen</sup>) III.

The fifth enzyme in the pathway, uroporphyrinogen decarboxylase (URO-decarboxylase), catalyzes the sequential removal of the four carboxyl groups from the acetic acid side chains of URO<sup>gen</sup> III to form coproporphyrinogen (COPRO<sup>gen</sup>) III, a tetracarboxylate porphyrinogen. This compound then enters the mitochondrion via a specific transporter, where COPRO-oxidase, the sixth enzyme, catalyzes the decarboxylation of two of the four propionic acid groups to form the two vinyl groups of protoporphyrinogen (PROTO<sup>gen</sup>) IX, a decarboxylate porphyrinogen. Next, PROTO-oxidase oxidizes PROTO<sup>gen</sup> to protoporphyrin IX by the removal of six hydrogen atoms. The product of the reaction is a porphyrin (oxidized form), in contrast to the preceding tetrapyrrole intermediates, which are porphyrinogens (reduced forms). Finally, ferrous iron is inserted into protoporphyrin IX to form heme, a reaction catalyzed by the eighth enzyme in the pathway, FECH (also known as heme synthase or protoheme ferrolyase).

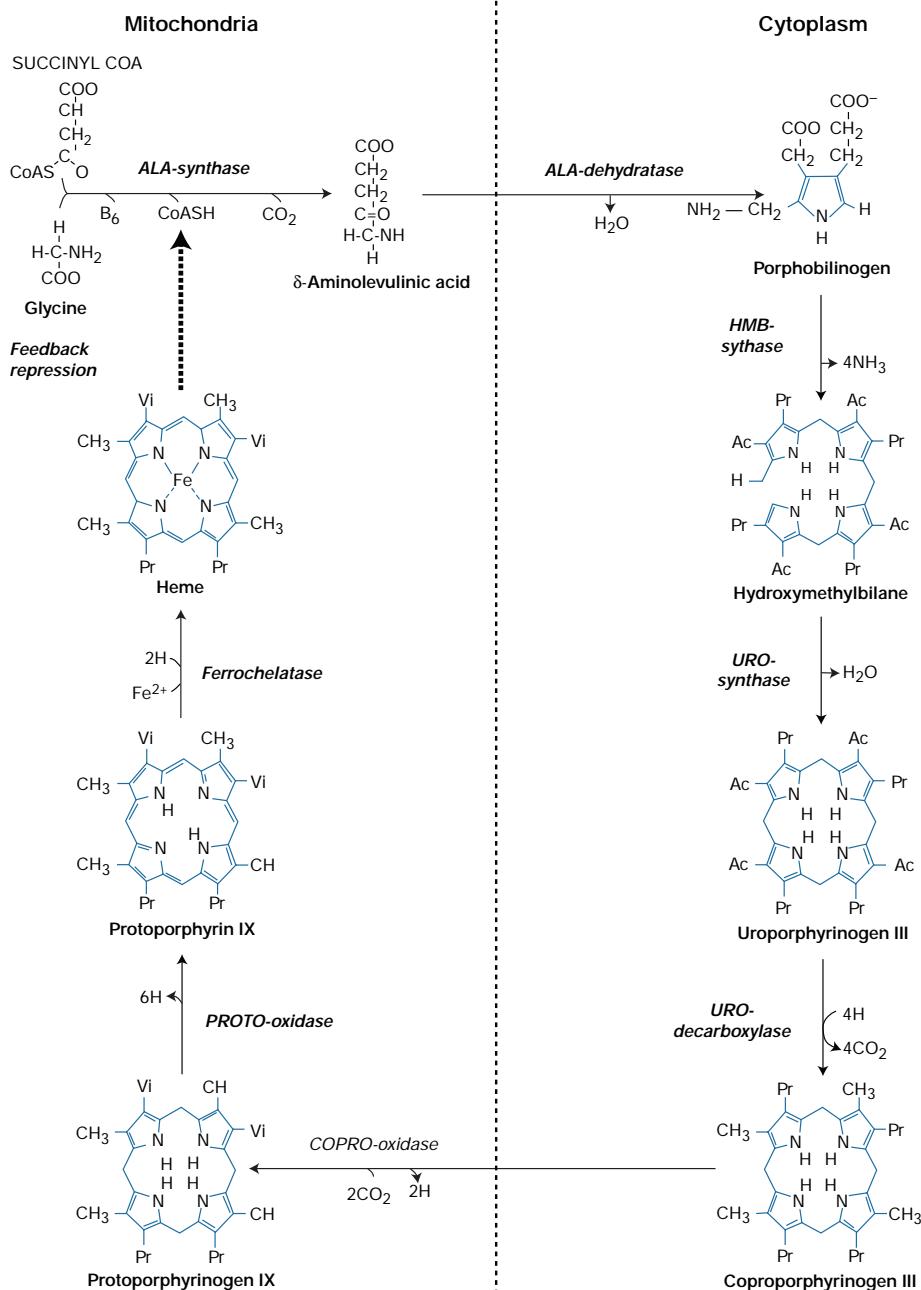


FIGURE 416-2 The heme biosynthetic pathway showing the eight enzymes and their substrates and products. Four of the enzymes are localized in the mitochondria and four in the cytosol.

### REGULATION OF HEME BIOSYNTHESIS

Regulation of heme synthesis differs in the two major heme-forming tissues, the liver and erythron. In the liver, the concentration of "free" heme regulates the synthesis and mitochondrial translocation of the housekeeping form of ALA-synthase 1. Heme represses the synthesis of the ALA-synthase 1 messenger RNA (mRNA) and interferes with the transport of the enzyme from the cytosol into mitochondria. Hepatic ALA-synthase 1 is increased by many of the same chemicals that induce the CYP enzymes in the endoplasmic reticulum of the liver. Because most of the heme in the liver is used for the synthesis of CYP enzymes, hepatic ALA-synthase 1 and the CYPs are regulated in a coordinated fashion, and many drugs that induce hepatic ALA-synthase 1 also induce CYP gene expression. The other hepatic heme

biosynthetic enzymes are presumably expressed at constant levels, although their relative activities and kinetic properties differ. For example, normal individuals have high activities of ALA-dehydratase but low activities of HMB-synthase, the latter being the second rate-limiting step in the pathway.

In the erythron, novel regulatory mechanisms allow for the production of the very large amounts of heme needed for hemoglobin synthesis. The response to stimuli for hemoglobin synthesis occurs during cell differentiation, leading to an increase in cell number. In contrast, the erythroid-specific ALA-synthase 2 is expressed at higher levels than the housekeeping enzyme, and erythroid-specific control mechanisms regulate other pathway enzymes as well as iron transport into erythroid cells. Separate erythroid-specific and nonerythroid or "housekeeping"

**3242** transcripts are known for the first four enzymes in the pathway. As noted above, housekeeping- and erythroid-specific ALA-synthases are encoded by genes on different chromosomes, but for each of the next three genes in the pathway, both erythroid and nonerythroid transcripts are transcribed by alternative promoters from their single respective genes (Table 416-2).

## CLASSIFICATION OF THE PORPHYRIAS

As mentioned above, the porphyrias can be classified as either *hepatic* or *erythropoietic*, depending on whether the heme biosynthetic intermediates that accumulate arise initially from the liver or developing erythrocytes, or as *acute* or *cutaneous*, based on their clinical manifestations. Table 416-1 lists the porphyrias, their principal symptoms, and major biochemical abnormalities. Three of the five hepatic porphyrias—AIP, HCP, and VP—usually present during adult life with acute attacks of neurologic manifestations and elevated levels of one or both of the porphyrin precursors, ALA and PBG, and are thus classified as *acute hepatic porphyrias*. Patients with ADP have presented in infancy and adolescence and typically have elevated ALA with normal or slightly elevated PBG levels. The fifth hepatic disorder, PCT, presents with blistering skin lesions. HCP and VP also may have cutaneous manifestations similar to PCT.

The erythropoietic porphyrias—CEP, EPP, and XLP—are characterized by elevations of porphyrins in bone marrow and erythrocytes and present with cutaneous photosensitivity. The skin lesions in CEP resemble PCT but are usually much more severe, whereas EPP and XLP cause a more immediate, severe, painful, and nonblistering type of photosensitivity. EPP is the most common porphyria to cause symptoms before puberty. About 20% of EPP patients develop minor abnormalities of liver function, with up to ~5% developing hepatic complications that can lead to liver failure requiring liver transplantation. XLP has a clinical presentation similar to EPP causing photosensitivity and liver disease.

## DIAGNOSIS OF PORPHYRIA

A few specific and sensitive first-line laboratory tests should be used whenever symptoms or signs suggest the diagnosis of porphyria (Table 416-3). If a first-line test is significantly abnormal, more comprehensive testing should follow to establish the type of porphyria, including the specific causative gene mutation.

**Acute Hepatic Porphyrias** An acute hepatic porphyria should be suspected in patients with neurovisceral symptoms after puberty. Symptoms include acute abdominal pain, nausea, vomiting, tachycardia, hypertension, and motor neuropathy. As these symptoms are common, other causes should be ruled out. The diagnosis is made by measuring urinary porphyrin precursors (ALA and PBG) in a spot sample of urine (Fig. 416-2). Urinary PBG is always increased during acute attacks of AIP, HCP, and VP and is not substantially increased in any other medical condition. Therefore, this measurement is both sensitive and specific. Results from spot (single-void) urine specimens are highly informative because very substantial increases in PBG are expected during acute attacks of porphyria. A 24-h collection is unnecessary. The same spot urine specimen should be saved for quantitative determination of ALA, PBG, and creatinine, in order to confirm the qualitative PBG result and also to detect patients with ADP. Urinary porphyrins may remain increased longer than porphyrin precursors in HCP and VP. Therefore, it is useful to measure total urinary porphyrins in the same sample, keeping in mind that urinary porphyrin increases are often nonspecific. Measurement of urinary porphyrins alone should be avoided for screening, because these may be increased in disorders other than porphyrias, such as chronic liver disease, and misdiagnoses of porphyria can result from minimal increases in urinary porphyrins that have no diagnostic significance. Measurement of erythrocyte HMB-synthase is not useful as a first-line test. Moreover, the enzyme activity is not decreased in all AIP patients, a borderline low normal value is not diagnostic, and the enzyme is not deficient in other acute porphyrias.

More extensive testing is justified when an initial test is positive. A substantial increase in PBG may be due to AIP, HCP, or VP. These

acute porphyrias can be distinguished by measuring urinary porphyrins (using the same spot urine sample), fecal porphyrins, and plasma porphyrins. Assays for COPRO-oxidase or PROTO-oxidase are not available for clinical testing. More specifically, mutation analysis by sequencing the genes encoding HMB-synthase, COPRO-oxidase, and PROTO-oxidase will detect almost all disease-causing mutations and is diagnostic even when the levels of urinary ALA and PBG have returned to normal or near normal.

**Cutaneous Porphyrias** Blistering skin lesions due to porphyria are virtually always accompanied by increases in total plasma porphyrins. A fluorometric method is preferred, because the plasma porphyrins in VP are mostly covalently linked to plasma proteins and may be less readily detected by high-performance liquid chromatography (HPLC). The normal range for plasma porphyrins is somewhat increased in patients with end-stage renal disease.

Although a total plasma porphyrin determination will usually detect EPP and XLP, an erythrocyte protoporphyrin determination is more sensitive. Increases in erythrocyte protoporphyrin occur in many other conditions. Therefore, the diagnosis of EPP must be confirmed by showing a predominant increase in free protoporphyrin rather than zinc protoporphyrin. In XLP, both free and zinc protoporphyrin are markedly increased. Interpretation of laboratory reports can be difficult, because the term *free erythrocyte protoporphyrin* sometimes actually represents zinc protoporphyrin.

The various porphyrias that cause blistering skin lesions can be differentiated by measuring porphyrins in urine, feces, and plasma. The porphyrias should be confirmed by genetic testing and the demonstration of the causative pathogenic variant. It is often difficult to diagnose or “rule out” porphyria in patients who have had suggestive symptoms months or years in the past and in relatives of patients with acute porphyrias, because porphyrin precursors and porphyrins may be normal. In those situations, detection of the specific gene mutation in the index case can make the diagnosis and facilitate the diagnosis and genetic counseling of at-risk relatives. With the increased access and accuracy of genetic testing, this often precedes secondary biochemical testing in clinical practice. Consultation with a specialist laboratory and physician will assist in selecting the heme biosynthetic gene or genes to be sequenced.

## THE HEPATIC PORPHYRIAS

Markedly elevated plasma and urinary concentrations of the porphyrin precursors, ALA and/or PBG, which originate from the liver, are especially evident during attacks of neurologic manifestations of the four acute porphyrias—ADP, AIP, HCP, and VP. In PCT, excess porphyrins also accumulate initially in the liver and cause chronic blistering of sun-exposed areas of the skin.

### ALA-DEHYDRATASE-DEFICIENT PORPHYRIA

ADP is a rare, autosomal recessive, acute hepatic porphyria caused by a severe deficiency of ALA-dehydratase activity. To date, there are only a few documented cases, some in children or young adults, in which specific gene mutations have been identified. These affected homozygotes had <10% of normal ALA-dehydratase activity in erythrocytes, but their clinically asymptomatic parents and heterozygous relatives had about half-normal levels of activity and did not excrete increased levels of ALA. The frequency of ADP is unknown, but the frequency of heterozygous individuals with <50% normal ALA-dehydratase activity was ~2% in a screening study in Sweden. Because there are multiple causes for deficient ALA-dehydratase activity, it is important to confirm the diagnosis of ADP by mutation analysis.

**Clinical Features** The clinical presentation depends on the amount of residual ALA-dehydratase activity. Four of the documented patients were male adolescents with symptoms resembling those of AIP, including abdominal pain and neuropathy. One patient was an infant with more severe disease, including failure to thrive beginning at birth. The earlier age of onset and more severe manifestations in this patient reflect a more significant deficiency of ALA-dehydratase activity. Another patient developed an acute motor polyneuropathy

at age 63 that was associated with a myeloproliferative disorder. He was heterozygous for an -aminolevulinic acid dehydratase (*ALAD*) mutation that presumably was present in erythroblasts that underwent clonal expansion due to the bone marrow malignancy.

**Diagnosis** All patients had significantly elevated levels of plasma and urinary ALA and urinary coproporphyrin (COPRO) III; ALAD activities in erythrocytes were <10% of normal. Hereditary tyrosinemia type 1 (fumarylacetoacetate deficiency) and lead intoxication should be considered in the differential diagnosis because either succinylacetone (which accumulates in hereditary tyrosinemia and is structurally similar to ALA) or lead can inhibit ALA-dehydratase, increase urinary excretion of ALA and COPRO III, and cause manifestations that resemble those of the acute porphyrias. Heterozygotes are clinically asymptomatic and do not excrete increased levels of ALA but can be detected by demonstration of intermediate levels of erythrocyte ALA-dehydratase activity or a specific mutation in the *ALAD* gene. To date, molecular studies of ADP patients have identified 12 pathogenic mutations, including missense mutations, splice-site mutations, and a two-base deletion in the *ALAD* gene (Human Gene Mutation Database; [www.hgmd.org](http://www.hgmd.org)). The parents in each case were not consanguineous, and the index cases had inherited a different *ALAD* mutation from each parent. Prenatal diagnosis of this disorder is possible by determination of ALA-dehydratase activity and/or gene mutations in cultured chorionic villi or amniocytes.

**Treatment** The treatment of ADP acute attacks is similar to that of AIP (see below). The severely affected infant referred to above was supported by hyperalimentation and periodic blood transfusions but did not respond to intravenous hemin and died after liver transplantation.

### ACUTE INTERMITTENT PORPHYRIA

This hepatic porphyria is an autosomal dominant condition resulting from the half-normal level of HMB-synthase activity. The disease is widespread but is especially common in Scandinavia and Great Britain. Clinical expression is highly variable, and activation of the disease is often related to environmental or hormonal factors, such as drugs, diet, and steroid hormones. Attacks can be prevented by avoiding known precipitating factors. Rare homozygous dominant AIP also has been described in children (see below).

**Clinical Features** Induction and increased expression of the rate-limiting hepatic gene *ALAS1* in heterozygotes who have half-normal HMB-synthase activity is thought to underlie the acute attacks in AIP. The disorder remains latent (or asymptomatic) in the great majority of those who are heterozygous for pathogenic *HMBS* mutations, and this is almost always the case prior to puberty. In patients with no history of acute symptoms, porphyrin precursor excretion is usually normal, suggesting that half-normal hepatic HMB-synthase activity is sufficient and that hepatic ALA-synthase activity is not increased. However, under conditions where heme synthesis is increased in the liver, half-normal HMB-synthase activity may become limiting, and ALA, PBG, and other heme pathway intermediates may accumulate and be excreted in the urine. Common precipitating factors include endogenous and exogenous steroids, porphyrinogenic drugs, alcohol ingestion, and low-calorie diets, usually instituted for weight loss.

The fact that AIP is almost always latent before puberty suggests that adult levels of steroid hormones are important for clinical expression. Symptoms are more common in women, suggesting a role for estrogens or progestins. Premenstrual attacks are probably due to increasing endogenous progesterone during the luteal phase of the menstrual cycle. Acute porphyrias are sometimes exacerbated by exogenous steroids, including oral contraceptive preparations containing progestins. Surprisingly, pregnancy is usually well tolerated, suggesting that beneficial metabolic changes may ameliorate the effects of high levels of progesterone. Extensive lists of unsafe and safe drugs are available on websites sponsored by the American Porphyria Foundation ([www.porphyriafoundation.com](http://www.porphyriafoundation.com)) and the European Porphyria Network (<https://porphyria.eu/>), and at the Drug Database for Acute Porphyrias

website ([www.drugs-porphyria.org](http://www.drugs-porphyria.org)). Reduced intake of calories and carbohydrate, as may occur with illness or attempts to lose weight, can also increase porphyrin precursor excretion and induce attacks of porphyria. Studies in a knockout AIP mouse model indicate that the hepatic *ALAS1* gene is regulated, in part, by the peroxisome proliferator-activated receptor coactivator 1 (PGC-1). Hepatic PGC-1 is induced by fasting, which in turn activates *ALAS1* transcription, resulting in increased heme biosynthesis. This finding suggests an important link between nutritional status and the attacks in acute porphyrias. Attacks also can be provoked by infections, surgery, and ethanol.

Because the neurovisceral symptoms rarely occur before puberty and are often nonspecific, a high index of suspicion is required to make the diagnosis. The disease can be disabling but is rarely fatal. Abdominal pain, the most common symptom, is poorly localized but may be associated with cramping, ileus, abdominal distention, and decreased bowel sounds. However, increased bowel sounds and diarrhea may occur. Abdominal tenderness, fever, and leukocytosis are usually absent or mild because the symptoms are neurologic rather than inflammatory. Nausea; vomiting; constipation; tachycardia; hypertension; mental symptoms; pain in the limbs, head, neck, or chest; muscle weakness; sensory loss; dysuria; and urinary retention are characteristic. Tachycardia, hypertension, restlessness, tremors, and excess sweating are due to sympathetic overactivity.

The peripheral neuropathy is due to axonal degeneration (rather than demyelination) and primarily affects motor neurons. Significant neuropathy does not occur with all acute attacks; abdominal symptoms are usually more prominent. Motor neuropathy affects the proximal muscles initially, more often in the shoulders and arms. The course and degree of involvement are variable and sometimes may be focal and involve cranial nerves. Deep tendon reflexes initially may be normal or hyperactive but become decreased or absent as the neuropathy advances. Sensory changes such as paresthesia and loss of sensation are less prominent. Progression to respiratory and bulbar paralysis and death occurs especially when the diagnosis and treatment are delayed. Sudden death may result from sympathetic overactivity and cardiac arrhythmia.

Mental symptoms such as anxiety, insomnia, depression, disorientation, hallucinations, and paranoia can occur in acute attacks. Seizures can be due to neurologic effects or to hyponatremia. Treatment of seizures is difficult because most antiseizure drugs can exacerbate AIP (clonazepam may be safer than phenytoin or barbiturates). Hyponatremia results from hypothalamic involvement and inappropriate vasoressin secretion or from electrolyte depletion due to vomiting, diarrhea, poor intake, or excess renal sodium loss. When an attack resolves, abdominal pain may disappear within hours, and paresis begins to improve within days and may continue to improve over several years.

Homozygous dominant AIP (HD-AIP) is a rare form of AIP in which patients inherit *HMBS* mutations from each of their heterozygous parents and, therefore, have very low (<2%) enzyme activity. The disease has been described in a Dutch girl, two young British siblings, and a Spanish boy. In these homozygous affected patients, the disease presented in infancy with failure to thrive, developmental delay, bilateral cataracts, and/or hepatosplenomegaly. Urinary ALA and PBG concentrations were markedly elevated. All of these patients' *HMBS* mutations (R167W, R167Q, and R172Q) were in exon 10 within five bases of each other. Studies of the brain magnetic resonance images (MRIs) of children with homozygous AIP have suggested damage primarily in white matter that was myelinated postnatally, while tracks that myelinated prenatally were normal. Most children with homozygous AIP die at an early age. Recently, later-onset HD-AIP was described in an adult with leukoencephalopathy.

**Diagnosis** ALA and PBG levels are substantially increased in plasma and urine, especially during acute attacks. For example, urinary PBG excretion during an attack is usually 50–200 mg/24 h (220–880 µmol/24 h) (normal, 0–4 mg/24 h, [0–18 µmol/24 h]), and urinary ALA excretion is 20–100 mg/24 h (150–760 µmol/24 h) (normal, 1–7 mg/24 h [8–53 µmol/24 h]). Because levels often remain high after symptoms resolve, the diagnosis of an acute attack in a patient with biochemically proven AIP is based primarily on clinical features.

**3244** Excretion of ALA and PBG decreases over a few days after intravenous hemin administration or treatment with givosiran (see below). A normal urinary PBG level before hemin effectively excludes AIP as a cause for current symptoms. Fecal porphyrins are usually normal or minimally increased in AIP, in contrast to HCP and VP. Most AIP heterozygotes with no history of symptoms have normal urinary excretion of ALA and PBG and are classified as latent. Patients can also have high levels of urine PBG and ALA with no clinical symptoms. These patients may have a previous history of an acute attack. These patients are classified as asymptomatic high excretors (ASHE) or chronic high excretors (CHE). Therefore, the detection of the family's *HMBS* mutation will diagnose asymptomatic family members. A urinary ALA and PBG will diagnosis CHE patients who may have a higher risk of an attack if they experience a precipitating factor such as administration of a porphyrinogenic drug.

Patients with *HMBS* mutations in the initiation of translation codon in exon 1 and in the intron 15'-splice donor site have normal enzyme levels in erythrocytes and deficient activity only in nonerythroid tissues. This occurs because the erythroid and housekeeping forms of HMB-synthase are encoded by a single gene, which has two promoters. Thus, the enzyme assay may not be diagnostic, and genetic testing should be used to confirm the diagnosis.

More than 515 *HMBS* mutations have been identified in AIP, including missense, nonsense, and splicing mutations and insertions and deletions, with most mutations found in only one or a few families (Human Gene Mutation Database, [www.hgmd.org](http://www.hgmd.org)). The prenatal diagnosis of a fetus at risk can be made by analysis of the familial mutation in cultured amniotic cells or chorionic villi. However, this is seldom done because the prognosis of individuals with *HMBS* mutations is generally favorable.

## TREATMENT

### Acute Intermittent Porphyria

During acute attacks, narcotic analgesics may be required for abdominal pain, and phenothiazines are useful for nausea, vomiting, anxiety, and restlessness. Chloral hydrate can be given for insomnia, and benzodiazepines are probably safe in low doses if a minor tranquilizer is required. Carbohydrate loading, usually with intravenous glucose (at least 300 g daily), may be effective in milder acute attacks of porphyria (without paresis, hyponatremia, etc.) if hemin is not available. Intravenous hemin is more effective and should be used as first-line therapy for all acute attacks. The standard regimen is 3–4 mg/kg of heme, in the form of lyophilized hematin (Panhematin, Recordati Rare Diseases), heme albumin (hematin reconstituted with human albumin), or heme arginate (Orphan Europe), infused daily for 4 days. Heme arginate and heme albumin are chemically stable and are less likely than hematin to produce phlebitis or an anticoagulant effect. Recovery depends on the degree of neuronal damage and usually is rapid if therapy is started early. Recovery from severe motor neuropathy may require months or years. Identification and avoidance of inciting factors can hasten recovery from an attack and prevent future attacks. Inciting factors are usually multiple, and removal of one or more hastens recovery and helps prevent future attacks. Frequent attacks that occur during the luteal phase of the menstrual cycle may be prevented with a gonadotropin-releasing hormone analogue, which prevents ovulation and progesterone production, or by prophylactic hematin or givosiran administration.

Recently, a hepatic-targeted RNA interference (RNAi) therapy, givosiran (Givlarri, Alnylam Pharmaceuticals), was approved by the U.S. Food and Drug Administration and the European Medicines Agency for the treatment of the acute hepatic porphyrias. Givosiran, a monthly subcutaneous injection of 2.5 mg/kg, is designed to silence the expression of hepatic *ALAS1* mRNA and was initially shown in clinical trials to markedly reduce ALA and PBG levels in CHE patients and in patients with recurrent attacks. In a phase 3 trial in acute hepatic porphyria patients with recurrent attacks, the

RNAi therapy significantly reduced the frequency of acute attacks, decreased hemin utilization, and improved daily pain scores.

The long-term risk of hypertension and chronic renal disease is increased in AIP; a number of patients have undergone successful renal transplantation. Studies have shown that up to 59% of symptomatic AIP patients will develop chronic kidney disease. The PEPT2 receptor polymorphic genotype affects the severity and prognosis of porphyria-associated kidney disease with the high affinity polymorphic PEPT2 \*1 allele and the PEPT2 genotypes \*1\*1 and, to a lesser degree, \*1\*2 associated with decreasing kidney function. Chronic, low-grade abnormalities in liver function tests are common, and the risk of hepatocellular carcinoma is increased. Hepatic imaging is recommended at least every 6 months for early detection of these tumors. Other long-term complications include neuropathy, fatigue, chronic pain, nausea, depression, and/or anxiety.

Orthotopic liver transplantation (OLT) has been successful and is curative in patients with severe, disabling, intractable attacks that are refractory to hemin therapy. Reports from both the United Kingdom and the United States show a marked improvement with no subsequent attacks, an improvement in the neuropathic manifestations, and normalization of the urinary PBG and ALA levels after liver transplantation. OLT is associated with morbidity and mortality and should be considered a treatment of last resort in these patients. In addition, patients who already have advanced neuropathy are considered poor risks for transplantation. Some patients with both recurrent attacks and end-stage renal disease have benefitted from combined liver and kidney transplantation.

Liver-directed gene therapy has proven successful in the prevention of drug-induced biochemical attacks in a murine model of human AIP, and clinical trials of adeno-associated virus vector (AAV)-*HMBS* gene transfer have been initiated. Although the therapy was safe, there was essentially no biochemical evidence of its effectiveness, nor did it prevent recurrent attacks in the treated patients.

### PORPHYRIA CUTANEA TARDÀ

PCT, the most common of the porphyrias, can be either sporadic (type 1) or familial (type 2) and can also develop after exposure to halogenated aromatic hydrocarbons. Hepatic URO-decarboxylase is deficient in all types of PCT, and for clinical symptoms to manifest, this enzyme deficiency must be substantial (-20% of normal activity or less); it is currently attributed to generation of an URO-decarboxylase inhibitor in the liver, which forms a uroporphomethene in the presence of iron and under conditions of oxidative stress. The majority of PCT patients (~80%) have no *UROD* mutations and are said to have sporadic (type 1) disease. PCT patients heterozygous for *UROD* mutations have the familial (type 2) PCT. In these patients, inheritance of a *UROD* mutation from one parent results in half-normal enzyme activity in liver and all other tissues, which is a significant predisposing factor but is insufficient by itself to cause symptomatic PCT. As discussed below, other genetic and environmental factors contribute to susceptibility for both types of PCT. Because penetrance of the genetic trait is low, many patients with familial (type 2) PCT have no family history of the disease. HEP is an autosomal recessive form of porphyria due to the inheritance of two pathogenic *UROD* mutations resulting in the marked systemic deficiency of URO-decarboxylase activity with clinical symptoms in childhood.

**Clinical Features** Blistering skin lesions that appear most commonly on the backs of the hands are the major clinical feature (Fig. 416-3). These rupture and crust over, leaving areas of atrophy and scarring. Lesions may also occur on the forearms, face, legs, and feet. Skin friability and small white papules termed milia are common, especially on the backs of the hands and fingers. Hypertrichosis and hyperpigmentation, especially of the face, are especially troublesome in women. Occasionally, the skin over sun-exposed areas becomes severely thickened, with scarring and calcification that resembles systemic sclerosis. Neurologic features are absent.

A number of susceptibility factors, in addition to inherited *UROD* mutations in type 2 PCT, can be recognized clinically and can affect



**FIGURE 416-3** Typical cutaneous lesions in a patient with porphyria cutanea tarda. Chronic, crusted lesions resulting from blistering due to photosensitivity on the dorsum of the hand of a patient with porphyria cutanea tarda. (Used with permission from Dr. Karl E. Anderson.)

management. These include hepatitis C, HIV, excess alcohol, elevated iron levels, and estrogens. The importance of excess hepatic iron as a precipitating factor is underscored by the finding that the incidence of the common hemochromatosis-causing mutations, hemochromatosis gene (*HFE*) mutations p.C282Y and p.H63D, are increased in patients with types 1 and 2 PCT (Chap. 414). Excess alcohol is a long-recognized contributor, as is estrogen use in women. HIV is probably an independent but less common risk factor that, like hepatitis C, does not cause PCT in isolation. Multiple susceptibility factors that appear to act synergistically can be identified in individual patients. PCT patients characteristically have chronic liver disease and sometimes cirrhosis and are at risk for hepatocellular carcinoma. Various chemicals can also induce PCT; an epidemic of PCT occurred in eastern Turkey in the 1950s as a consequence of wheat contaminated with the fungicide hexachlorobenzene. PCT also occurs after exposure to other chemicals, including di- and trichlorophenols and 2,3,7,8-tetrachlorodibenzo-(*p*)-dioxin (TCDD, dioxin).

**Diagnosis** Porphyrins are increased in the liver, plasma, urine, and stool. The urinary ALA level may be slightly increased, but the PBG level is normal. Urinary porphyrins consist mostly of uroporphyrins and heptacarboxylate porphyrin, with lesser amounts of coproporphyrin and hexa- and pentacarboxylate porphyrins. Plasma porphyrins are also increased, and fluorometric scanning of diluted plasma at neutral pH can rapidly distinguish VP and PCT (Table 416-3). Isocoproporphyrins, which are increased in feces and sometimes in plasma and urine, are diagnostic for hepatic URO-decarboxylase deficiency.

Type 2 PCT and HEP can be distinguished from type 1 by finding decreased URO-decarboxylase in erythrocytes. URO-decarboxylase activity in liver, erythrocytes, and cultured skin fibroblasts in type 2 PCT is ~50% of normal in affected individuals and in asymptomatic heterozygous family members. In HEP, the URO-decarboxylase activity is markedly deficient, with typical levels of 3–10% of normal. Over 145 mutations have been identified in the *UROD* gene (Human Gene Mutation Database; [www.hgmd.org](http://www.hgmd.org)). Of the mutations listed in the database, ~65% are missense or nonsense, and ~8% are splice-site mutations. Many *UROD* mutations have been identified in only one or two families.

## TREATMENT

### Porphyria Cutanea Tarda

Alcohol, estrogens, iron supplements, and, if possible, any drugs that may exacerbate the disease should be discontinued, but this step does not always lead to improvement. A complete response can almost always be achieved by the standard therapy, repeated phlebotomy, to reduce hepatic iron. A unit (450 mL) of blood can

be removed every 1–2 weeks. The aim is to gradually reduce excess hepatic iron until the serum ferritin level reaches the lower limits of normal. Because iron overload is not marked in most cases, remission may occur after only five or six phlebotomies; however, PCT patients with hemochromatosis may require more treatments to bring their iron levels down to the normal range. To document improvement in PCT, it is most convenient to follow the total plasma porphyrin concentration, which becomes normal sometime after the target ferritin level is reached. Hemoglobin levels or hematocrits and serum ferritin should be followed closely to prevent development of iron deficiency and anemia. After remission, continued phlebotomy may not be needed. Plasma porphyrin levels are followed at 6- to 12-month intervals for early detection of recurrences, which are treated by additional phlebotomy.

An alternative when phlebotomy is contraindicated or poorly tolerated is a low-dose regimen of chloroquine or hydroxychloroquine, both of which complex with the excess porphyrins and promote their excretion. Small doses (e.g., 125 mg chloroquine phosphate twice weekly) should be given, because standard doses can induce transient, sometimes marked increases in photosensitivity and hepatocellular damage. Studies indicate that low-dose hydroxychloroquine is as safe and effective as phlebotomy in PCT. Hepatic imaging can diagnose or exclude complicating hepatocellular carcinoma. Treatment of PCT in patients with end-stage renal disease is facilitated by administration of erythropoietin.

Because hepatitis C virus (HCV) is a common precipitating factor causing PCT, the recent development of oral direct-acting antivirals for HCV has proven effective as a first primary treatment in HCV-infected PCT patients.

## HEREDITARY COPROPORPHYRIA

HCP is an autosomal dominant hepatic porphyria that results from the half-normal activity of COPRO-oxidase. The disease presents with acute attacks, as in AIP. Cutaneous photosensitivity also may occur, but much less commonly than in VP. HCP patients may have acute attacks and cutaneous photosensitivity together or separately. HCP is less common than AIP and VP. Homozygous dominant HCP and harderoporphyrinia, a biochemically distinguishable variant of HCP, are present with clinical symptoms in children (see below).

**Clinical Features** HCP is influenced by the same factors that cause attacks in AIP. The disease is latent before puberty, and symptoms, which are virtually identical to those of AIP, are more common in women. HCP is generally less severe than AIP. Blistering skin lesions are identical to PCT and VP and begin in childhood in rare homozygous cases.

**Diagnosis** COPRO III is markedly increased in the urine and feces in symptomatic patients and often persists, especially in feces, when there are no symptoms. Urinary ALA and PBG levels are increased (but less than in AIP) during acute attacks but may revert to normal more quickly than in AIP when symptoms resolve. Plasma porphyrins are usually normal or only slightly increased, but they may be higher in cases with skin lesions. The diagnosis of HCP is readily confirmed by increased fecal porphyrins consisting almost entirely of COPRO III, which distinguishes it from other porphyrias.

Although the diagnosis can be confirmed by measuring COPRO-oxidase activity, the assays for this mitochondrial enzyme are not available and require cells other than erythrocytes. To date, >90 mutations have been identified in the *CPOX* gene, ~70% of which are missense or nonsense (Human Gene Mutation Database; [www.hgmd.org](http://www.hgmd.org)). Detection of a *CPOX* mutation in a symptomatic individual permits the identification of asymptomatic family members.

## TREATMENT

### Hereditary Coproporphyria

Neurologic symptoms are treated as in AIP (see above). Phlebotomy and chloroquine are not effective for the cutaneous lesions.

VP is an autosomal dominant hepatic porphyria that results from the deficient activity of PROTO-oxidase, the seventh enzyme in the heme biosynthetic pathway, and can present with neurologic symptoms, photosensitivity, or both. VP is particularly common in South Africa, where 3 of every 1000 whites have the disorder. Most are descendants of a couple who emigrated from the Netherlands to South Africa in 1688. In other countries, VP is less common than AIP. Rare cases of homozygous dominant VP, presenting in childhood with cutaneous symptoms, also have been reported.

**Clinical Features** VP can present with skin photosensitivity, acute neurovisceral crises, or both. In two large studies of VP patients, ~60% had only skin lesions, 20% had only acute attacks, and ~20% had both. Acute attacks are identical to those in AIP and are precipitated by the same factors as AIP (see above). Blistering skin manifestations are similar to those in PCT but are more difficult to treat and usually are of longer duration. Homozygous VP is associated with photosensitivity, neurologic symptoms, and developmental disturbances, including growth retardation, in infancy or childhood; all cases had increased erythrocyte levels of zinc protoporphyrin, a characteristic finding in all homozygous porphyrias so far described.

**Diagnosis** Urinary ALA and PBG levels are increased during acute attacks but may return to normal more quickly than in AIP. Increases in fecal protoporphyrin and COPRO III and in urinary COPRO III are more persistent. Plasma porphyrin levels also are increased, particularly when there are cutaneous lesions. VP can be distinguished rapidly from all other porphyrias by examining the fluorescence emission spectrum of porphyrins in plasma since VP has a unique fluorescence peak at neutral pH.

Assays of PROTO-oxidase activity in cultured fibroblasts or lymphocytes are not widely available. Over 205 mutations have been identified in the *PPOX* gene from unrelated VP patients (Human Gene Mutation Database; [www.hgmd.org](http://www.hgmd.org)). The missense mutation R59W is the common mutation in most South Africans with VP of Dutch descent. Five missense mutations were common in English and French VP patients; however, most mutations have been found in only one or a few families.

## TREATMENT

### Variegate Porphyria

Acute attacks are treated as in AIP, and hemin should be started early in most cases. Givosiran has proven effective in clinical trials for patients with recurrent attacks. Other than avoiding sun exposure, there are few effective measures for treating the skin lesions.

-Carotene, phlebotomy, and chloroquine are not helpful.

## THE ERYTHROPOIETIC PORPHYRIAS

In the erythropoietic porphyrias, excess porphyrins from bone marrow erythrocyte precursors are transported via the plasma to the skin and lead to cutaneous photosensitivity.

### X-LINKED SIDEROBLASTIC ANEMIA

XLSA results from the deficient activity of the erythroid form of ALA-synthase (ALA-synthase 2) and is associated with ineffective erythropoiesis, weakness, and pallor.

**Clinical Features** Typically, males with XLSA develop refractory hemolytic anemia, pallor, and weakness during infancy. They have secondary hypersplenism, become iron overloaded, and can develop hemosiderosis. The severity depends on the level of residual erythroid ALA-synthase activity and on the responsiveness of the specific mutation to pyridoxal 5'-phosphate supplementation (see below). Peripheral blood smears reveal a hypochromic, microcytic anemia with striking anisocytosis, poikilocytosis, and polychromasia; the leukocytes and platelets appear normal. Hemoglobin content is reduced, and the mean corpuscular volume and mean corpuscular hemoglobin concentration

are decreased. Patients with milder, later-onset disease have been reported recently.

**Diagnosis** Bone marrow examination reveals hypercellularity with a left shift and megaloblastic erythropoiesis with an abnormal maturation. A variety of Prussian blue-staining sideroblasts are observed. Levels of urinary porphyrin precursors and of both urinary and fecal porphyrins are normal. The activity of erythroid ALA-synthase 2 is decreased in bone marrow, but this enzyme is difficult to measure in the presence of the normal ALA-synthase 1 housekeeping enzyme. Definitive diagnosis requires the demonstration of loss-of-function mutations in the erythroid *ALAS2* gene, of which >110 have been identified.

**Treatment** The severe anemia may respond to pyridoxine supplementation. This cofactor is essential for ALA-synthase activity, and mutations in the pyridoxine binding site of the enzyme have been found in several responsive patients. Cofactor supplementation may make it possible to eliminate or reduce the frequency of transfusions. Unresponsive patients may be transfusion dependent and require chelation therapy.

## CONGENITAL ERYTHROPOIETIC PORPHYRIA

CEP, also known as Günther's disease, is an autosomal recessive disorder. It is due to the markedly deficient, but not absent, activity of URO-synthase and the resultant accumulation of URO I and COPRO I isomers. CEP is associated with hemolytic anemia and cutaneous lesions.

**Clinical Features** Severe cutaneous photosensitivity typically begins from birth. The skin over light-exposed areas is friable, and bullae and vesicles are prone to rupture and infection. Skin thickening, focal hypo- and hyperpigmentation, and hypertrichosis of the face and extremities are characteristic. Secondary infection of the cutaneous lesions can lead to disfigurement of the face and hands. Porphyrins are deposited in teeth and in bones. As a result, the teeth are brownish and fluoresce on exposure to long-wave ultraviolet light. Hemolysis is due to the marked increase in erythrocyte porphyrins and leads to splenomegaly. Adults with a milder later-onset form of the disease also have been described.

**Diagnosis** URO and COPRO (mostly type I isomers) accumulate in the bone marrow, erythrocytes, plasma, urine, and feces. The predominant porphyrin in feces is COPRO I. The diagnosis of CEP can be confirmed by demonstration of markedly deficient URO-synthase activity and/or by the identification of specific mutations in the *UROS* gene. The disease can be detected in utero by measuring porphyrins in amniotic fluid and URO-synthase activity in cultured amniotic cells or chorionic villi or by the detection of the family's specific gene mutations. Molecular analyses of the mutant alleles from unrelated patients have revealed the presence of >55 mutations in the *UROS* gene, including six in the erythroid-specific promoter of the *UROS* gene. Genotype/phenotype correlations can predict the severity of the disease. The CEP phenotype may be modulated by sequence variations in the erythroid-specific ALA-synthase 2, the mutation of which typically causes XLP. One mutation (p.ArgR216WTrp) in *GATA1*, encoding the X-linked erythroid-specific transcription factor GATA binding protein 1 (*GATA1*), has been identified in an individual with CEP, thrombocytopenia, and thalassemia.

## TREATMENT

### Congenital Erythropoietic Porphyria

Severe cases often require transfusions for anemia. Chronic transfusions of sufficient fresh packed erythrocytes to suppress erythropoiesis are effective in reducing porphyrin production but result in iron overload. Splenectomy may reduce hemolysis and decrease transfusion requirements. Protection from sunlight and from minor skin trauma is important. Complicating bacterial infections should be treated promptly. Recently, non-transfusion-dependent patients

have been treated by periodic phlebotomies to decrease iron levels, thereby decreasing erythropoiesis and porphyrin accumulation. This approach has not been evaluated in clinical trials to date. Bone marrow and cord blood transplantation has proven curative in transfusion-dependent children, providing the rationale for stem cell gene therapy.

### ERYTHROPOIETIC PROTOPORPHYRIA

EPP is an autosomal recessive disorder resulting from the deficient activity of FECH, the last enzyme in the heme biosynthetic pathway. EPP is the most common erythropoietic porphyria in children and, after PCT, the second most common porphyria in adults. EPP patients have FECH activities as low as 15–25% of normal in lymphocytes and cultured fibroblasts. Protoporphyrin IX accumulates in bone marrow reticulocytes and circulating erythrocytes, is released into the plasma, and then is taken up in the liver where it is excreted in the bile and feces. Plasma protoporphyrin IX taken up by the vascular cells in the skin is photoactivated on exposure to sunlight causing phototoxic cellular damage and excruciatingly painful nonblistering phototoxicity. In most symptomatic patients (>95%) with this disorder, a deleterious mutation in one *FECH* allele was inherited with the relatively common (~10% of Caucasians) intronic 3 (IVS3) alteration (IVS3-48T>C) on the other allele that results in the low expression of the normal enzyme. In ~2% of EPP families, two *FECH* deleterious mutations have been found.

XLP is a less common condition with the same phenotype in affected males, including increased erythrocyte protoporphyrin IX levels resulting from gain-of-function mutations in the last exon of the erythroid-specific form of 5-aminolevulinate-synthase 2 (*ALAS2*). These mutations delete or alter the *ALAS2* C-terminal amino acids resulting in its increased activity and the subsequent accumulation of protoporphyrin IX. Manifestations in female heterozygotes with XLP can range from asymptomatic to as severe as their affected male relatives. The variation in the presence and severity of manifestations in XLP heterozygotes results primarily from random X-chromosomal inactivation. XLP accounts for ~2–10% of cases with the EPP phenotype in Europe and North America. Rare patients with EPP symptoms and elevated erythrocyte protoporphyrin IX levels do not have mutations in *FECH* or *ALAS2* on genetic testing. In an affected family with EPP symptoms and accumulation of protoporphyrin IX, an autosomal dominant mutation was found in human CLPX, a modulator of heme biosynthesis.

**Clinical Features** In EPP and male XLP patients, skin photosensitivity, which differs from that in other cutaneous porphyrias, usually begins in early childhood. The initial symptoms on sun exposure consist of tingling, stinging, itching, or heat/burning sensations on the exposed skin occurring within <10 to 30 min of exposure in >60% of patients; most will have these prodromal symptoms within an hour of sun exposure. The prodromal symptoms are the “warning signal” to get out of the sun, thereby avoiding a severe incapacitating painful attack that can last from 2–5 days. Photosensitivity is associated with substantial elevations in erythrocyte protoporphyrin IX and occurs only in patients with genotypes that result in FECH activities below ~35% of normal. Vesicular lesions are uncommon. Redness and swelling develop after prolonged sun exposure and resemble angioedema (Fig. 416-4). Pain symptoms may seem out of proportion to the visible skin involvement. Chronic skin changes may include lichenification, leathery pseudovesicles, labial grooving, and nail changes. Severe scarring is rare, as are pigment changes, friability, and hirsutism. Unless hepatic or other complications develop, protoporphyrin IX levels and symptoms of photosensitivity tend to remain remarkably stable over many years in most patients. Factors that exacerbate the hepatic porphyrias play no role in EPP or XLP.

The primary source of excess protoporphyrin is the bone marrow erythroid cells. Erythrocyte protoporphyrin IX is free (not complexed with zinc) and is mostly bound to hemoglobin. In plasma, protoporphyrin IX is bound to albumin. Hemolysis and anemia are absent or usually mild.



**FIGURE 416-4** Erythema and edema of the hands due to acute photosensitivity in a 10-year-old boy with erythropoietic protoporphyrinia. (Reproduced with permission from P Poblete-Gutiérrez et al: *The porphyrias: clinical presentation, diagnosis and treatment*. Eur J Dermatol 16:230, 2006.)

Although EPP is an erythropoietic porphyria, up to 27% of EPP patients may have minor abnormalities of liver function, and in ~2–5% of these patients, the accumulation of protoporphyrins causes chronic liver disease that can progress to liver failure requiring transplantation. Protoporphyrin IX is insoluble, and excess amounts form crystalline structures in liver cells (Fig. 416-4) and can decrease hepatic bile flow. Studies in the mouse model of EPP have shown that the bile duct epithelium may be damaged by toxic bile, leading to biliary fibrosis. Thus, rapidly progressive liver disease appears to be related to the cholestatic effects of protoporphyrins and is associated with increasing hepatic protoporphyrin IX levels due to impaired hepatobiliary excretion and increased photosensitivity. The hepatic complications also are often characterized by increasing levels of protoporphyrins in erythrocytes and plasma as well as severe abdominal and back pains, especially in the right upper quadrant. Gallstones composed at least in part of protoporphyrin IX occur in some patients. Hepatic complications appear to be higher in EPP due to two pathogenic *FECH* mutations and in males with XLP.

**Diagnosis** A substantial increase in erythrocyte protoporphyrin IX, which is predominantly free and not complexed with zinc, is the hallmark of EPP. Protoporphyrin levels also are variably increased in bone marrow, plasma, bile, and feces. Erythrocyte protoporphyrin IX concentrations are increased in other conditions such as lead poisoning, iron deficiency, various hemolytic disorders, all homozygous forms of other porphyrias, and sometimes even in acute porphyrias. In all these conditions, however, in contrast to EPP, protoporphyrin IX is complexed with zinc. Therefore, after an increase in erythrocyte protoporphyrin IX is found in a suspected EPP patient, it is important to confirm the diagnosis by an assay that distinguishes free and zinc-complexed protoporphyrin. Erythrocytes in EPP also exhibit red fluorescence under fluorescence microscopy at 620 nm. Urinary levels of porphyrins and porphyrin precursors are normal. FECH activity in cultured lymphocytes or fibroblasts is decreased (<30% of normal mean). DNA diagnosis by mutation analysis is recommended to detect the causative *FECH* mutation(s) and/or the presence of the IVS3-48T>C low expression allele. To date, >220 mutations have been identified in the *FECH* gene, many of which result in an unstable or absent enzyme protein (null alleles) (Human Gene Mutation Database; [www.hgmd.org](http://www.hgmd.org)).

In XLP, the erythrocyte protoporphyrin levels appear to be higher than in EPP, and the proportions of free and zinc protoporphyrin IX may reach 50%. XLP accounts for ~2% of patients with the EPP phenotype in Western Europe. Recent studies show that ~10% of North American patients with the EPP phenotype have XLP.

## Erythropoietic Protoporphyria

Avoiding sunlight exposure and wearing clothing designed to provide protection for conditions with chronic phototoxicity are essential. Various other treatments, including oral -carotene, have proven of little benefit. Afamelanotide, an -melanocyte-stimulating hormone (MSH) analogue that stimulates tanning, has been approved for the treatment of EPP and XLP in the European Union by the European Medicines Agency and in the United States by the U.S. Food and Drug Administration. Dersimelagon, an orally administered, small-molecule, selective melanocortin-1 receptor (MC1R) agonist that increases skin melanin without sun exposure, is currently in phase 3 clinical trials for EPP and XLP.

Treatment of hepatic complications, which may be accompanied by motor neuropathy, is difficult. Cholestyramine and other porphyrin absorbents such as activated charcoal may interrupt the enterohepatic circulation of protoporphyrin and promote its fecal excretion, leading to some improvement. Plasmapheresis and intravenous hemin are sometimes beneficial.

Liver transplantation has been carried out in some EPP and XLP patients with severe liver complications and is often successful in the short term. However, the disease often recurs in the transplanted liver due to continued bone marrow production of excess protoporphyrin. In a retrospective study of 17 liver-transplanted EPP patients, 11 (65%) had recurrent EPP liver disease. Posttransplantation treatment with hematin and plasmapheresis should be considered to prevent the recurrence of liver disease. However, bone marrow transplantation, which has been successful in human EPP and which prevented liver disease in a mouse model, should be considered after liver transplantation, if a suitable donor can be found.

### Acknowledgment

The authors thank Dr. Karl E. Anderson for his review of the manuscript and helpful comments and suggestions. This work is supported in part by the Porphyrias Consortium (U54 DK083909), a part of the National Institutes of Health (NIH) Rare Disease Clinical Research Network (RDCRN), supported through collaboration between the NIH Office of Rare Diseases Research (ORDR) at the National Center for Advancing Translational Science (NCATS) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

### FURTHER READING

- Balwani M et al: Acute hepatic porphyrias: Recommendations for evaluation and long-term management. *Hepatology* 66:1314, 2017.
- Balwani M et al: Clinical, biochemical, and genetic characterization of North American patients with erythropoietic protoporphyrria and X-linked protoporphyrria. *JAMA Dermatol* 153:789, 2017.
- Balwani M et al: Phase 3 trial of RNAi therapeutic givosiran for acute intermittent porphyria. *N Engl J Med* 382:2289, 2020.
- Bissell DM et al: Porphyria. *N Engl J Med* 377:862, 2017.
- Chen B et al: Acute intermittent porphyria: Predicted pathogenicity of HMBS variants indicates extremely low penetrance of the autosomal dominant disease. *Hum Mutat* 37:1215, 2016.
- Kazamel M et al: Porphyric neuropathy: Pathophysiology, diagnosis, and updated management. *Curr Neurol Neurosci Rep* 20:56, 2020.
- Langendonk JG et al: Afamelanotide for erythropoietic protoporphyrria. *N Engl J Med* 373:48, 2015.
- Singal AK et al: Hepatitis C treatment in patients with porphyria cutanea tarda. *Am J Med Sci* 353:523, 2017.
- Tchernitchko D et al: A variant of peptide transporter 2 predicts the severity of porphyria-associated kidney disease. *J Am Soc Nephrol* 28:1924, 2017.
- Yasuda M et al: Liver transplantation for acute intermittent porphyria: Biochemical and pathologic studies of the explanted liver. *Mol Med* 21:487, 2015.

John N. Mecchella, Christopher M. Burns



Purines (adenine and guanine) and pyrimidines (cytosine, thymine, uracil) serve fundamental roles in the replication of genetic material, gene transcription, protein synthesis, and cellular metabolism. Disorders that involve abnormalities of nucleotide metabolism range from relatively common diseases such as hyperuricemia and gout, in which there is increased production or impaired excretion of a metabolic end product of purine metabolism (uric acid), to rare enzyme deficiencies that affect purine and pyrimidine synthesis or degradation. Understanding these biochemical pathways has led, in some instances, to the development of specific forms of treatment, such as the use of allopurinol and febuxostat to reduce uric acid production.

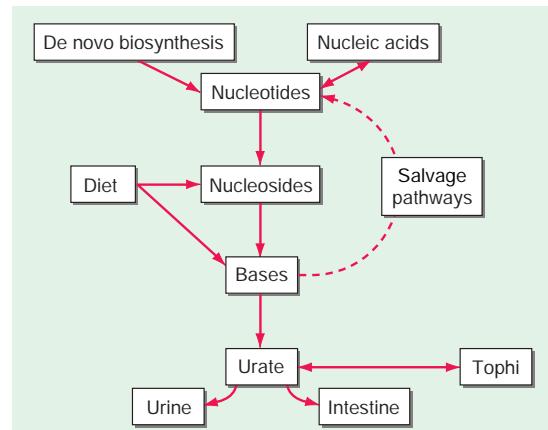
### URIC ACID METABOLISM

Uric acid is the final breakdown product of purine degradation in humans. It is a weak diprotic acid with  $pK_a$  values of 5.75 and 10.3. Urates, the ionized forms of uric acid, predominate in plasma, extracellular fluid, and synovial fluid, with ~98% existing as monosodium urate at pH 7.4.

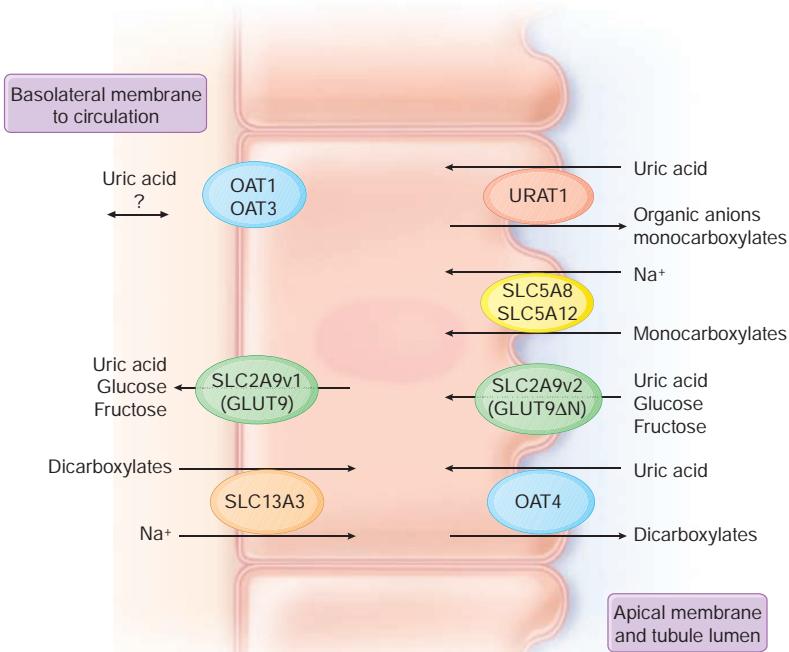
Plasma is saturated with monosodium urate at a concentration of 405  $\mu\text{mol/L}$  (6.8 mg/dL) at 37°C. At higher concentrations, plasma is therefore supersaturated—a situation that creates the potential for urate crystal precipitation. However, plasma urate concentrations can reach 4800  $\mu\text{mol/L}$  (80 mg/dL) without precipitation, perhaps because of the presence of solubilizing substances.

The pH of urine greatly influences the solubility of uric acid. At pH 5.0, urine is saturated with uric acid at concentrations ranging from 360 to 900  $\mu\text{mol/L}$  (6–15 mg/dL). At pH 7.0, saturation is reached at concentrations from 9840 to 12,000  $\mu\text{mol/L}$  (158–200 mg/dL). Ionized forms of uric acid in urine include monosodium, disodium, potassium, ammonium, and calcium urates.

Although purine nucleotides are synthesized and degraded in all tissues, urate is produced only in tissues that contain xanthine oxidase, primarily the liver and small intestine. Urate production varies with the purine content of the diet and with rates of purine biosynthesis, degradation, and salvage (Fig. 417-1). Normally, two-thirds to three-fourths



**FIGURE 417-1** The total-body urate pool is the net result between urate production and excretion. Urate production is influenced by dietary intake of purines and the rates of de novo biosynthesis of purines from nonpurine precursors, nucleic acid turnover, and salvage by phosphoribosyltransferase activities. The formed urate is normally excreted by urinary and intestinal routes. Hyperuricemia can result from increased production, decreased excretion, or a combination of both mechanisms. When hyperuricemia exists, urate can precipitate and deposit in tissues as tophi.



**FIGURE 417-2** Schematic for handling of uric acid by the kidney. A complex interplay of transporters on both the apical and basolateral aspects of the renal tubule epithelial cell is involved in the reabsorption of uric acid. See text for details. Most uricosuric compounds inhibit URAT1 on the apical side, as well as OAT1, OAT3, and GLUT9 on the basolateral side.

of urate is excreted by the kidneys, and most of the remainder is eliminated through the intestines.

The kidneys clear urate from the plasma and maintain physiologic balance by utilizing specific organic anion transporters (OATs), including urate transporter 1 (URAT1, SLC22A12) (Fig. 417-2). In humans, OAT1 (SLC22A6), OAT2 (SLC22A7), and OAT3 (SLC22A8) are located on the basolateral membrane of renal proximal tubule cells. OAT4 (SLC22A11), OAT10 (SLC22A13), and URAT1 are located on the apical brush-border membrane of these cells. The latter transporters carry urate and other organic anions into the tubular cells from the lumen in exchange for intracellular organic anions. Once inside the cell, urate must pass to the basolateral side of the lumen in a process controlled by voltage-dependent carriers, including glucose transporter 9 (GLUT9, SLC2A9). *Uricosuric* compounds (Table 417-1) directly inhibit URAT1 on the apical side of the tubular cell (so-called *cis*-inhibition). In contrast, *antiuricosuric* compounds (those that promote hyperuricemia), such as nicotinate, pyrazinolate, lactate, and other aromatic organic acids, serve as the exchange anion

to the degree of elevation. The prevalence of hyperuricemia is increasing among ambulatory adults and even more markedly among hospitalized patients. The prevalence of gout in the United States more than doubled between the 1960s and the 1990s. Based on NHANES data from 2007 to 2008, these trends continue, with an approximate prevalence of gout among men of 5.9% (6.1 million) and among women of 2.0% (2.2 million). Mean serum urate levels rose to 6.14 mg/dL among men and 4.87 mg/dL among women, with consequent hyperuricemia prevalences of 21.2 and 21.6%, respectively (with *hyperuricemia* defined as a serum urate level of  $>7.0$  mg/dL [415  $\mu$ mol/L] for men and  $>5.7$  mg/dL [340  $\mu$ mol/L] for women). These numbers represent a 1.2% increase in the prevalence of gout, a 0.15-mg/dL increase in the serum urate level, and a 3.2% increase in the prevalence of hyperuricemia over figures reported in NHANES-III (1988–1994). These rises are thought to be driven by increased obesity and hypertension and perhaps also by better medical care and increased longevity.

### CAUSES OF HYPERURICEMIA

Hyperuricemia may be classified as primary or secondary, depending on whether the cause is innate or an acquired disorder. However, it is more useful to classify hyperuricemia in relation to the underlying pathophysiology—i.e., whether it results from increased production, decreased excretion, or a combination of the two (Fig. 417-1, Table 417-2).

**Increased Urate Production** Diet contributes to the serum urate concentration in proportion to its purine content. Strict restriction of purine intake reduces the mean serum urate level by ~60  $\mu$ mol/L (~1 mg/dL) and urinary uric acid excretion by ~1.2 mmol/d (~200 mg/d). Foods high in nucleic acid content include liver, “sweet-breads” (i.e., thymus and pancreas), kidney, and anchovy.

Endogenous sources of purine production also influence the serum urate level (Fig. 417-3). De novo purine biosynthesis is a multistep process that forms inosine monophosphate (IMP). The rates of purine biosynthesis and urate production are predominantly determined

inside the cell, thereby stimulating anion exchange and urate reabsorption (*trans-stimulation*). The activities of URAT1, other OATs, and sodium anion transporters result in excretion of 8–12% of the filtered urate as uric acid.

Most children have serum urate concentrations of 180–240  $\mu$ mol/L (3–4 mg/dL). Levels begin to rise in males during puberty but remain low in females until menopause. The most recent mean serum urate values for men and premenopausal women in the United States are 415 and 360  $\mu$ mol/L (6.14 and 4.87 mg/dL), respectively, according to National Health and Nutrition Evaluation Survey (NHANES) data for 2007–2008. After menopause, values for women increase to approximately those for men. In adulthood, concentrations rise steadily over time and vary with height, body weight, blood pressure, renal function, and alcohol intake.

### HYPERURICEMIA

Hyperuricemia can result from increased production or decreased excretion of uric acid or from a combination of the two processes. Sustained hyperuricemia predisposes some individuals to develop clinical manifestations, including gouty arthritis (Chap. 372), urolithiasis, and renal dysfunction (see below).

In general, hyperuricemia is defined as a plasma (or serum) urate concentration  $>405$   $\mu$ mol/L ( $>6.8$  mg/dL). The risk of developing gouty arthritis or urolithiasis increases with higher urate levels and escalates in proportion

**TABLE 417-1** Medications with Uricosuric Activity

Acetohexamide	Glyceryl guaiacolate
Adrenocorticotrophic hormone	Glycopyrrrolate
Ascorbic acid	Halofenate
Azauridine	Losartan
Benzbromarone	Meclofenamate
Calcitonin	Phenolsulfonphthalein
Chlorprothixene	Phenylbutazone
Citrate	Probenecid
Dicumarol	Radiographic contrast agents
Diflunisal	Salicylates (>2 g/d)
Estrogens	Sulfinpyrazone
Fenofibrate	Tetracycline that is outdated
Glucocorticoids	Zoxazolamine

**TABLE 417-2 Classification of Hyperuricemia by Pathophysiology**

Urate Overproduction		
Primary idiopathic	Myeloproliferative diseases	Rhabdomyolysis
HPRT deficiency	Polycythemia vera	Exercise
PRPP synthetase overactivity	Psoriasis	Alcohol
Hemolytic processes	Paget's disease	Obesity
Lymphoproliferative diseases	Glycogenesis III, V, and VII	Purine-rich diet

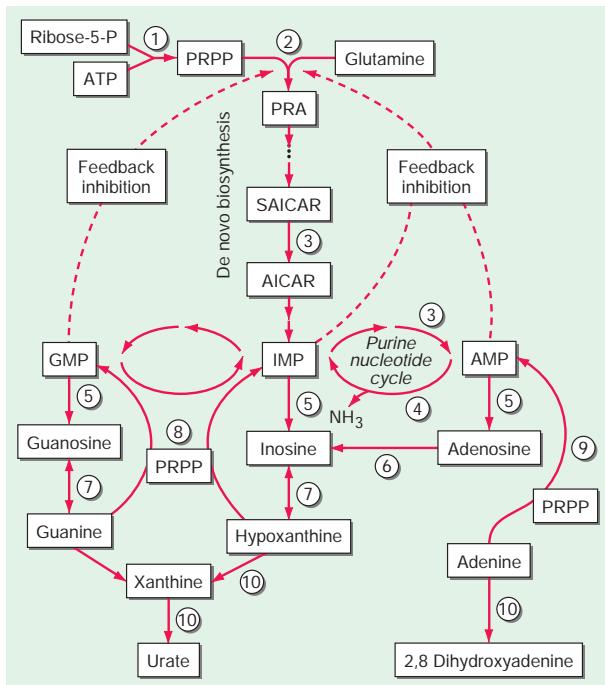
Decreased Uric Acid Excretion		
Primary idiopathic	Starvation ketosis	Drug ingestion
Renal insufficiency	Berylliosis	Salicylates (<2 g/d)
Polycystic kidney disease	Sarcoidosis	Diuretics
Diabetes insipidus	Lead intoxication	Alcohol
Hypertension	Hyperparathyroidism	Levodopa
Acidosis	Hypothyroidism	Ethambutol
Lactic acidosis	Toxemia of pregnancy	Pyrazinamide
Diabetic ketoacidosis	Bartter's syndrome	Nicotinic acid
	Down's syndrome	Cyclosporine

Combined Mechanism		
Glucose-6-phosphatase deficiency	Fructose-1-phosphate aldolase deficiency	Alcohol Shock

Abbreviations: HPRT, hypoxanthine phosphoribosyltransferase; PRPP, phosphoribosylpyrophosphate.

by amidophosphoribosyltransferase (amidoPRT), which combines phosphoribosylpyrophosphate (PRPP) and glutamine. A secondary regulatory pathway is the salvage of purine bases by hypoxanthine phosphoribosyltransferase (HPRT). HPRT catalyzes the combination of the purine bases hypoxanthine and guanine with PRPP to form



**FIGURE 417-3 Abbreviated scheme of purine metabolism.** (1) Phosphoribosylpyrophosphate (PRPP) synthetase, (2) amidophosphoribosyltransferase (amidoPRT), (3) adenylosuccinate lyase, (4) (myo)-Adenylate (AMP) deaminase, (5) 5'-nucleotidase, (6) adenosine deaminase, (7) purine nucleoside phosphorylase, (8) hypoxanthine phosphoribosyltransferase (HPRT), (9) adenine phosphoribosyltransferase (APRT), and (10) xanthine oxidase. AICAR, aminoimidazole carboxamide ribotide; ATP, adenosine triphosphate; GMP, guanylate; IMP, inosine monophosphate; PRA, phosphoribosylamine; SAICAR, succinylaminoimidazole carboxamide ribotide.

the respective ribonucleotides IMP and guanosine monophosphate (GMP).

Serum urate levels are closely coupled to the rates of de novo purine biosynthesis, which is driven in part by the level of PRPP, as evidenced by two X-linked inborn errors of purine metabolism (**Table 417-3**). Both increased PRPP synthetase activity and HPRT deficiency are associated with overproduction of purines, hyperuricemia, and hyperuricaciduria (see below for clinical descriptions).

Accelerated purine nucleotide degradation can also cause hyperuricemia—i.e., with conditions of rapid cell turnover, proliferation, or cell death, as in leukemic blast crises, cytotoxic therapy for malignancy, hemolysis, or rhabdomyolysis. Hyperuricemia can result from excessive degradation of skeletal muscle ATP after strenuous physical exercise or status epilepticus and in glycogen storage disease types III, V, and VII (**Chap. 419**). The hyperuricemia of myocardial infarction, smoke inhalation, and acute respiratory failure may also be related to accelerated breakdown of ATP.

**Decreased Uric Acid Excretion** More than 90% of individuals with sustained hyperuricemia have a defect in the renal handling of uric acid. For any given plasma urate concentration, patients who have gout excrete ~40% less uric acid than those who do not. When plasma urate levels are raised by purine ingestion or infusion, uric acid excretion increases in patients with and without gout; however, in those with gout, plasma urate concentrations must be 60–120  $\mu\text{mol/L}$  (1–2 mg/dL) higher than normal to achieve equivalent uric acid excretion rates.

Diminished uric acid excretion could theoretically result from decreased glomerular filtration, decreased tubular secretion, or enhanced tubular reabsorption. Decreased urate filtration does not appear to cause primary hyperuricemia but does contribute to the hyperuricemia of renal insufficiency. Although hyperuricemia is invariably present in chronic renal disease, the correlation among serum creatinine, urea nitrogen, and urate concentrations is poor. Extrarenal clearance of uric acid increases as renal damage becomes more severe.

Many agents that cause hyperuricemia exert their effects by stimulating reabsorption rather than inhibiting secretion. This stimulation appears to occur through a process of “priming” renal urate reabsorption through the sodium-dependent loading of proximal tubular epithelial cells with anions capable of trans-stimulating urate reabsorption. The sodium-coupled monocarboxyl transporters SMC1 and 2 (SLC5A8, SLC5A12) in the brush border of the proximal tubular cells mediate sodium-dependent loading of these cells with monocarboxylates. A similar transporter, SLC13A3, mediates sodium-dependent influx of dicarboxylates into the epithelial cell from the basolateral membrane. Some of these carboxylates are well known to cause hyperuricemia, including pyrazinamide treatment, niacin (from niacin therapy), and the organic acids lactate, -hydroxybutyrate, and acetoacetate. The mono- and divalent anions then become substrates for URAT1 and OAT4, respectively, and are exchanged for uric acid from the proximal tubule. Increased blood levels of these anions result in their increased glomerular filtration and greater reabsorption by proximal tubular cells. The increased intraepithelial cell concentrations lead to increased uric acid reabsorption by promoting URAT1, OAT4-, and OAT10-dependent anion exchange. Low doses of salicylates also promote hyperuricemia by this mechanism. Sodium loading of proximal tubular cells also provokes urate retention by reducing extracellular fluid volume and increasing angiotensin II, insulin, and parathyroid hormone release. Additional OAT1, OAT2, and OAT3 are involved in the movement of uric acid through the basolateral membrane, although the detailed mechanisms are still being elucidated.

GLUT9 (SLC2A9) is an electrogenic hexose transporter with splicing variants that mediate co-reabsorption of uric acid along with glucose and fructose at the apical membrane (GLUT9 N/SLC2A9v2) as well as through the basolateral membrane (SLC2A9v1) and thus into the circulation. GLUT9 has recently been identified as a high-capacity urate transporter, with rates 45–60 times faster than its glucose/fructose transport activity. GLUT9 may be responsible for the observed

**TABLE 417-3** Inborn Errors of Purine Metabolism

ENZYME	ACTIVITY	INHERITANCE	CLINICAL FEATURES	LABORATORY FEATURES
Hypoxanthine phosphoribosyltransferase	Complete deficiency	X-linked	Self-mutilation, choreoathetosis, gout, and uric acid lithiasis	Hyperuricemia, hyperuricosuria
	Partial deficiency	X-linked	Gout and uric acid lithiasis	Hyperuricemia, hyperuricosuria
Phosphoribosylpyrophosphate synthetase	Overactivity	X-linked	Gout, uric acid lithiasis, and deafness	Hyperuricemia, hyperuricosuria
Adenine phosphoribosyltransferase	Deficiency	Autosomal recessive	2,8-Dihydroxyadenine lithiasis	—
Xanthine oxidase	Deficiency	Autosomal recessive	Xanthinuria and xanthine lithiasis	Hypouricemia, hypouricosuria
Adenylosuccinate lyase	Deficiency	Autosomal recessive	Autism and psychomotor retardation	—
Myoadenylate deaminase	Deficiency	Autosomal recessive	Myopathy with exercise intolerance or asymptomatic	—
Adenosine deaminase	Deficiency	Autosomal recessive	Severe combined immunodeficiency disease and chondro-osseous dysplasia	—
Purine nucleoside phosphorylase	Deficiency	Autosomal recessive	T cell-mediated immunodeficiency	—

association of the consumption of fructose-sweetened soft drinks with an increased risk of hyperuricemia and gout. Genome-wide association studies (GWAS) suggest that polymorphisms in SLC2A9 may play an important role in susceptibility to gout in the Caucasian population. The presence of one predisposing variant allele increases the relative risk of developing gout by 30–70%, most likely by increasing expression of the shorter isoform, SLC2A9v2 (GLUT9 N). GWAS have identified over 30 loci associated with serum urate levels, most encoding transporters in the gut or kidney. Recent meta-analyses suggest that genetic polymorphisms may explain up to 23.9% of the variation in serum urate levels, much higher than previously appreciated. However, the utility of genetic testing for relevant polymorphisms remains investigational with few exceptions. The Q141K variant of ABCG2, which encodes a urate transporter that secretes urate in the small intestine, is associated with early onset and severe gout and resistance to allopurinol. The *HLA-B\*50:10* genotype is associated with allopurinol hypersensitivity in Asian populations. This field is evolving rapidly.

Alcohol promotes hyperuricemia because of increased urate production and decreased uric acid excretion. Excessive alcohol consumption accelerates hepatic breakdown of ATP to increase urate production. Alcohol consumption can also induce hyperlacticacidemia, which blocks uric acid secretion. The higher purine content in some alcoholic beverages may also be a factor. Consumption of beer confers a greater risk of gout than liquor, and moderate wine intake does not increase gout risk. Intake of red meat and fructose increases the risk of gout, whereas intake of low-fat dairy products, purine-rich vegetables, whole grains, nuts and legumes, less sugary fruits, coffee, and vitamin C reduces the risk.

## EVALUATION

Hyperuricemia does not necessarily represent a disease, nor is it a specific indication for therapy. The decision to treat depends on the cause and the potential consequences of hyperuricemia in each individual.

Quantification of uric acid excretion can be used to determine whether hyperuricemia is caused by overproduction or decreased excretion. On a purine-free diet, men with normal renal function excrete <3.6 mmol/d (600 mg/d). Thus, the hyperuricemia of individuals who excrete uric acid above this level while on a purine-free diet is due to purine overproduction; for those who excrete lower amounts on the purine-free diet, it is due to decreased excretion. If the assessment is performed while the patient is on a regular diet, the level of 4.2 mmol/d (800 mg/d) can be used as the discriminating value.

## COMPLICATIONS

The most recognized complication of hyperuricemia is *gouty arthritis*. NHANES 2007–2008 found a prevalence of gout among U.S. adults of 3.9%, with figures of ~6% for men and ~2% for women. The higher the serum urate level, the more likely an individual is to develop gout. In one study, the incidence of gout was 4.9% among individuals with serum urate concentrations >540 µmol/L (>9.0 mg/dL) as opposed

to only 0.5% among those with values between 415 and 535 µmol/L (7.0 and 8.9 mg/dL). The complications of gout correlate with both the duration and the severity of hyperuricemia. **For further discussion of gout, see Chap. 372.**

Hyperuricemia also causes several renal problems: (1) nephrolithiasis; (2) urate nephropathy, a rare cause of renal insufficiency attributed to monosodium urate crystal deposition in the renal interstitium; and (3) uric acid nephropathy, a reversible cause of acute renal failure resulting from deposition of large amounts of uric acid crystals in the renal collecting ducts, pelvis, and ureters.

**Nephrolithiasis** Uric acid nephrolithiasis occurs most commonly, but not exclusively, in individuals with gout. In gout, the prevalence of nephrolithiasis correlates with the serum and urinary uric acid levels, reaching ~50% with serum urate levels of 770 µmol/L (13 mg/dL) or urinary uric acid excretion >6.5 mmol/d (1100 mg/d).

Uric acid stones can develop in individuals with no evidence of arthritis, only 20% of whom are hyperuricemic. Uric acid can also play a role in other types of kidney stones. Some individuals who do not have gout but have calcium oxalate or calcium phosphate stones have hyperuricemia or hyperuricaciduria. Uric acid may act as a nidus on which calcium oxalate can precipitate or lower the formation product for calcium oxalate crystallization.

**Urate Nephropathy** Urate nephropathy, sometimes referred to as *urate nephrosis*, is a late manifestation of severe gout and is characterized histologically by deposits of monosodium urate crystals surrounded by a giant-cell inflammatory reaction in the medullary interstitium and pyramids. The disorder is now rare and cannot be diagnosed in the absence of gouty arthritis. The lesions may be clinically silent or cause proteinuria, hypertension, and renal insufficiency.

**Uric Acid Nephropathy** This reversible cause of acute renal failure is due to precipitation of uric acid in renal tubules and collecting ducts that obstruct urine flow. Uric acid nephropathy develops following sudden urate overproduction and marked hyperuricaciduria. Factors that favor uric acid crystal formation include dehydration and acidosis. This form of acute renal failure occurs most often during an aggressive “blastic” phase of leukemia or lymphoma prior to or coincident with cytolytic therapy but has also been observed in individuals with other neoplasms, following epileptic seizures, and after vigorous exercise with heat stress. Autopsy studies have demonstrated intraluminal precipitates of uric acid, dilated proximal tubules, and normal glomeruli. The initial pathogenic events are believed to include obstruction of collecting ducts with uric acid and obstruction of the distal renal vasculature.

If recognized, uric acid nephropathy is potentially reversible. Appropriate therapy has reduced the mortality rate from ~50% to near zero. Serum levels cannot be relied on for diagnosis because this condition has developed in the presence of urate concentrations varying from 720–4800 µmol/L (12–80 mg/dL). The distinctive feature is the

urinary uric acid concentration. In most forms of acute renal failure with decreased urine output, urinary uric acid content is either normal or reduced, and the ratio of uric acid to creatinine is  $<1$ . In acute uric acid nephropathy, the ratio of uric acid to creatinine in a random urine sample or a 24-h specimen is  $>1$ , and a value that high is essentially diagnostic.

## HYPURICEMIA AND METABOLIC SYNDROME

Metabolic syndrome (Chap. 408) is characterized by abdominal obesity with visceral adiposity, impaired glucose tolerance due to insulin resistance with hyperinsulinemia, hypertriglyceridemia, increased low-density lipoprotein cholesterol, decreased high-density lipoprotein cholesterol, and hyperuricemia. Hyperinsulinemia reduces the renal excretion of uric acid and sodium. Not surprisingly, hyperuricemia resulting from euglycemic hyperinsulinemia may precede the onset of type 2 diabetes, hypertension, coronary artery disease, and gout in individuals with metabolic syndrome.

## TREATMENT

### Hyperuricemia

#### ASYMPTOMATIC HYPURICEMIA

Hyperuricemia is present in ~21% of the population and in at least 25% of hospitalized individuals. The vast majority of hyperuricemic persons are at no clinical risk. In the past, the association of hyperuricemia with cardiovascular disease and renal failure led to the use of urate-lowering agents for patients with asymptomatic hyperuricemia. This practice is no longer recommended except for individuals receiving cytolytic therapy for neoplastic disease, who are treated with urate-lowering agents in an effort to prevent uric acid nephropathy. Because hyperuricemia can be a component of the metabolic syndrome, its presence is an indication to screen for and aggressively treat any accompanying obesity, hyperlipidemia, diabetes mellitus, or hypertension.

Hyperuricemic individuals, especially those with higher serum urate levels, are at risk for the development of gouty arthritis. However, most hyperuricemic persons never develop gout, and prophylactic treatment is not indicated. Furthermore, neither structural kidney damage nor tophi are identifiable before the first attack. Reduced renal function cannot be attributed to asymptomatic hyperuricemia and available evidence does not yet support treatment of asymptomatic hyperuricemia to alter progression of renal dysfunction in patients with renal disease. An increased risk of stone formation in those with asymptomatic hyperuricemia has not been established.

Thus, because treatment with specific antihyperuricemic agents entails inconvenience, cost, and potential toxicity, routine treatment of asymptomatic hyperuricemia cannot be justified other than for prevention of acute uric acid nephropathy. In addition, routine screening for asymptomatic hyperuricemia is not recommended. If hyperuricemia is diagnosed, however, the cause should be determined. Causal factors should be corrected if the condition is secondary, and associated problems such as hypertension, hypercholesterolemia, diabetes mellitus, and obesity should be treated.

#### SYMPOMATIC HYPURICEMIA

**See Chap. 372 for treatment of gout, including urate nephrosis.**

**Nephrolithiasis** Antihyperuricemic therapy is recommended for the individual who has both gouty arthritis and either uric acid- or calcium-containing stones, both of which may occur in association with hyperuricaciduria. Regardless of the nature of the calculi, fluid ingestion should be sufficient to produce a daily urine volume  $>2$  L. Alkalization of the urine with sodium bicarbonate or acetazolamide may be justified to increase the solubility of uric acid. Specific treatment of uric acid calculi requires reducing the urine uric acid concentration with a xanthine oxidase inhibitor, such as allopurinol or febuxostat. These agents decrease the serum urate

concentration and the urinary excretion of uric acid in the first 24 h, with a maximal reduction within 2 weeks. Allopurinol can be given once a day because of the long half-life (18 h) of its active metabolite, oxypurinol. In the febuxostat trials, the generally recommended dose of allopurinol (300 mg/d) was effective at achieving a target serum urate concentration  $<6.0$  mg/dL (357  $\mu$ mol/L) in  $<50\%$  of patients; this result suggested that higher doses should be considered. Allopurinol is effective in patients with renal insufficiency, but the dose should be reduced. Allopurinol is also useful in reducing the recurrence of calcium oxalate stones in patients with gout and in individuals with hyperuricemia or hyperuricaciduria who do not have gout. Febuxostat (40–80 mg/d) is also taken once daily, and doses do not need to be adjusted in the presence of mild to moderate renal dysfunction. Potassium citrate (30–80 mmol/d orally in divided doses) is an alternative therapy for patients with uric acid stones alone or mixed calcium/uric acid stones. A xanthine oxidase inhibitor is also indicated for the treatment of 2,8-dihydroxyadenine kidney stones.

**Uric Acid Nephropathy** Uric acid nephropathy is often preventable, and immediate appropriate therapy has greatly reduced the mortality rate. Vigorous IV hydration and diuresis with furosemide dilute the uric acid in the tubules and promote urine flow to 100 mL/h. The administration of acetazolamide (240–500 mg every 6–8 h) and sodium bicarbonate (89 mmol/L) IV enhances urine alkalinity and thereby solubilizes more uric acid. It is important to ensure that the urine pH remains  $>7.0$  and to watch for circulatory overload. In addition, antihyperuricemic therapy in the form of allopurinol in a single dose of 8 mg/kg is administered to reduce the amount of urate that reaches the kidney. If renal insufficiency persists, subsequent daily doses should be reduced to 100–200 mg because oxypurinol, the active metabolite of allopurinol, accumulates in renal failure. Despite these measures, hemodialysis may be required. Urate oxidase (rasburicase) can also be administered IV to prevent or to treat tumor lysis syndrome.

## HYPOURICEMIA

Hypouricemia, defined as a serum urate concentration  $<120$   $\mu$ mol/L ( $<2.0$  mg/dL), can result from decreased production of urate, increased excretion of uric acid, or a combination of both mechanisms. This condition occurs in  $<0.2\%$  of the general population and  $<0.8\%$  of hospitalized individuals. Hypouricemia causes no symptoms or pathology and therefore requires no therapy.

Most hypouricemia results from increased renal uric acid excretion. The finding of normal amounts of uric acid in a 24-h urine collection from an individual with hypouricemia is evidence for a renal cause. Medications with uricosuric properties (Table 417-1) include aspirin (at doses  $>2.0$  g/d), losartan, fenofibrate, x-ray contrast materials, and glyceryl guaiacolate. Total parenteral hyperalimentation can also cause hypouricemia, possibly a result of the high glycine content of the infusion formula. Other causes of increased urate clearance include conditions such as neoplastic disease, hepatic cirrhosis, diabetes mellitus, and inappropriate secretion of vasopressin; defects in renal tubular transport such as primary Fanconi syndrome and Fanconi syndromes caused by Wilson's disease, cystinosis, multiple myeloma, and heavy metal toxicity; and isolated congenital defects in the bidirectional transport of uric acid. Hypouricemia can be a familial disorder that is generally inherited in an autosomal recessive manner. Most cases are caused by a loss of function mutation in *SLC22A12*, the gene that encodes URAT-1, resulting in increased renal urate clearance. Individuals with normal *SLC22A12* most likely have a defect in other urate transporters. Although hypouricemia is usually asymptomatic, some patients suffer from urate nephrolithiasis or exercise-induced renal failure.

## SELECTED INBORN ERRORS OF PURINE AND PYRIMIDINE METABOLISM

**(See also Table 417-3, Table 417-4, Fig. 417-3, and Fig. 417-4.)** More than 30 defects in human purine and pyrimidine metabolic pathways have been identified thus far. Many are benign, but about half are

**TABLE 417-4** Inborn Errors of Pyrimidine Metabolism

ENZYME	ACTIVITY	INHERITANCE	CLINICAL FEATURES	LABORATORY FEATURES
Uridine-5'-monophosphate synthetase	Deficiency	Autosomal recessive	Orotic acid crystalluria; obstructive uropathy, hypochromic megaloblastic anemia	Orotic aciduria
Pyrimidine 5'-nucleotidase	Deficiency	Autosomal recessive	Hemolytic anemia	Basophilic stippling of erythrocytes; high levels of cytidine and uridine ribonucleotides
Pyrimidine 5'-nucleotidase	Superactivity	Uncertain	Developmental delay, seizures, ataxia, language deficit	Hypouricosuria
Thymidine phosphorylase	Deficiency	Autosomal recessive	Mitochondrial neurogastrointestinal encephalopathy	Hypouricosuria
Dihydropyrimidine dehydrogenase	Deficiency	Autosomal recessive	Seizures, motor and mental retardation	High levels of uracil, thymine, and 5-hydroxymethyluracil and low levels of dihydropyrimidines in urine
Dihydropyrimidinase	Deficiency	Uncertain	Seizures, mental retardation	Dihydropyrimidinuria
Ureidopropionase	Deficiency	Uncertain	Hypotonia, dystonia, developmental delay	High urinary excretion of <i>N</i> -carbamyl-β-alanine and <i>N</i> -carbamyl β-aminoisobutyric acid

associated with clinical manifestations, some causing major morbidity and mortality. Advances in genetics, along with high-performance liquid chromatography and tandem mass spectrometry, have facilitated diagnosis.

## PURINE DISORDERS

**HPRT Deficiency** The HPRT gene is located on the X chromosome. Affected males are hemizygous for the mutant gene; carrier females are asymptomatic. A complete deficiency of HPRT, the Lesch-Nyhan syndrome, is characterized by hyperuricemia, self-mutilative behavior, choreoathetosis, spasticity, and mental retardation. A partial deficiency of HPRT, the Kelley-Seegmiller syndrome, is associated with hyperuricemia but no central nervous system manifestations. In both disorders, the hyperuricemia results from urate overproduction and can cause uric acid crystalluria, nephrolithiasis, obstructive uropathy, and gouty arthritis. Early diagnosis and

appropriate therapy with allopurinol can prevent or eliminate all the problems attributable to hyperuricemia without affecting behavioral or neurologic abnormalities.

**Increased PRPP Synthetase Activity** Like the HPRT deficiency states, PRPP synthetase overactivity is X-linked and results in gouty arthritis and uric acid nephrolithiasis. Neurologic hearing loss occurs in some families.

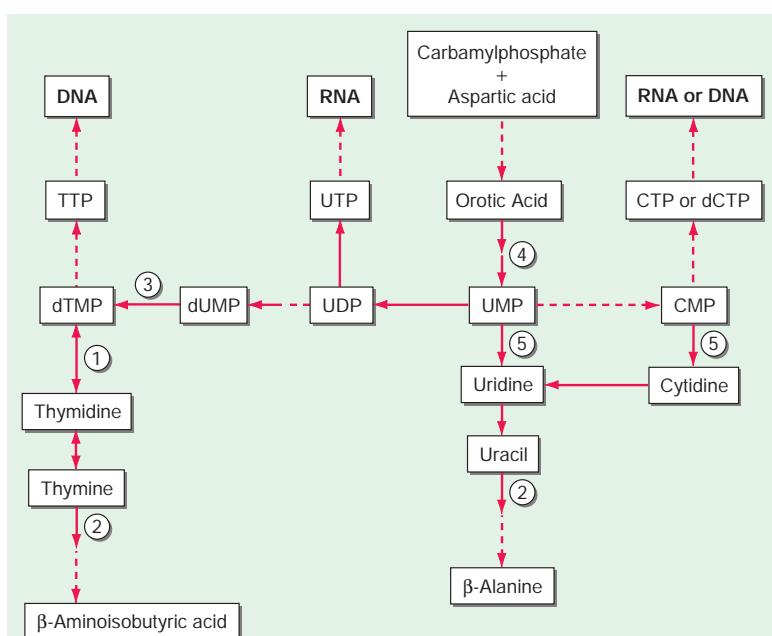
**Adenine Phosphoribosyltransferase (APRT) Deficiency** APRT deficiency is inherited as an autosomal recessive trait. Affected individuals develop kidney stones composed of 2,8-dihydroxyadenine. Caucasians with the disorder have a complete deficiency (type I), whereas Japanese individuals have some measurable enzyme activity (type II). Expression of the defect is similar in the two populations, as is the frequency of the heterozygous state (0.4–1.1 per 100). Allopurinol treatment prevents stone formation.

**Hereditary Xanthinuria** A deficiency of xanthine oxidase causes all purine in the urine to occur in the form of hypoxanthine and xanthine. About two-thirds of deficient individuals are asymptomatic. The remainder develop kidney stones composed of xanthine.

**Myoadenylate Deaminase Deficiency** Primary (inherited) and secondary (acquired) forms of myoadenylate deaminase deficiency have been described. The primary form is inherited as an autosomal recessive trait. Clinically, some persons may have relatively mild myopathic symptoms with exercise or other triggers, but most individuals with this defect are asymptomatic. Therefore, another explanation for the myopathy should be sought in symptomatic patients with this deficiency. The acquired deficiency occurs in association with a wide variety of neuromuscular diseases, including muscular dystrophies, neuropathies, inflammatory myopathies, and collagen vascular diseases.

**Adenylosuccinate Lyase Deficiency** Deficiency of this enzyme is due to an autosomal recessive trait and causes profound psychomotor retardation, seizures, and other movement disorders. All individuals with this deficiency are mentally retarded, and most are autistic.

**Adenosine Deaminase Deficiency and Purine Nucleoside Phosphorylase Deficiency** See Chap. 351.



**FIGURE 417-4** Abbreviated scheme of pyrimidine metabolism. (1) Thymidine kinase, (2) dihydropyrimidine dehydrogenase, (3) thymidylate synthase, (4) UMP synthase, (5) 5'-nucleotidase. CMP, cytidine-5'-monophosphate; dTMP, deoxythymidine-5'-monophosphate; dUMP, deoxyuridine-5'-monophosphate; TTP, thymidine triphosphate; UDP, uridine-5'-diphosphate; UMP, uridine-5'-monophosphate; UTP, uridine triphosphate.

The pyrimidine cytidine is found in both DNA and RNA; it is a complementary base pair for guanine. Thymidine is found only in DNA, where it is paired with adenine. Uridine is found only in RNA and can pair with either adenine or guanine in RNA secondary structures. Pyrimidines can be synthesized by a de novo pathway (Fig. 417-4) or reused in a salvage pathway. Although >25 different enzymes are involved in pyrimidine metabolism, disorders of these pathways are rare. Seven disorders of pyrimidine metabolism have been discovered (Table 417-4), three of which are discussed below.

**Orotic Aciduria** Hereditary orotic aciduria is caused by mutations in a bifunctional enzyme, uridine-5'-monophosphate (UMP) synthase, which converts orotic acid to UMP in the de novo synthesis pathway (Fig. 417-4). The disorder is characterized by hypochromic megaloblastic anemia that is unresponsive to vitamin B<sub>12</sub> and folic acid, growth retardation, and neurologic abnormalities. Increased excretion of orotic acid causes crystalluria and obstructive uropathy. Replacement of uridine (100–200 mg/kg per day) corrects anemia, reduces orotic acid excretion, and improves the other sequelae of the disorder.

**Pyrimidine 5'-Nucleotidase Deficiency** Pyrimidine 5'-nucleotidase catalyzes the removal of the phosphate group from pyrimidine ribonucleoside monophosphates (cytidine-5'-monophosphate or UMP) (Fig. 417-4). An inherited deficiency of this enzyme causes hemolytic anemia with prominent basophilic stippling of erythrocytes. The accumulation of pyrimidines or cytidine diphosphate choline is thought to induce hemolysis. There is no specific treatment. Acquired pyrimidine 5'-nucleotidase deficiency has been reported in lead poisoning and in thalassemia.

**Dihydropyrimidine Dehydrogenase Deficiency** Dihydropyrimidine dehydrogenase is the rate-limiting enzyme in the pathway of uracil and thymine degradation (Fig. 417-4). Deficiency of this enzyme causes excessive urinary excretion of uracil and thymine. In addition, this deficiency causes nonspecific cerebral dysfunction with convulsive disorders, motor retardation, and mental retardation. No specific treatment is available.

**Medication Effect on Pyrimidine Metabolism** A variety of medications can influence pyrimidine metabolism. The anticancer agents fluorodeoxyuridine and 5-fluorouracil and the antimicrobial agent fluorocytosine cause cytotoxicity when converted to fluorodeoxyuridylate, a specific suicide inhibitor of thymidylate synthase. Fluorocytosine must be converted to 5-fluorouracil to be effective. This conversion is catalyzed by cytosine deaminase activity. Fluorocytosine's action is selective because cytosine deaminase is present in bacteria and fungi but not in human cells. Dihydropyrimidine dehydrogenase is involved in the degradation of 5-fluorouracil. Consequently, deficiency of this enzyme is associated with 5-fluorouracil neurotoxicity.

Leflunomide, which is used to treat rheumatoid arthritis, inhibits de novo pyrimidine synthesis by inhibiting dihydroorotate dehydrogenase, resulting in an antiproliferative effect on T cells. Allopurinol, which inhibits xanthine oxidase in the purine metabolic pathway, also inhibits the activity of orotidine-5'-phosphate decarboxylase, a step in UMP synthesis. Consequently, allopurinol use is associated with increased excretion of orotidine and orotic acid. There are no known clinical effects of this inhibition.

#### Acknowledgment

The authors are grateful to Robert L. Wortmann for contributions to this chapter in previous editions of the book.

#### FURTHER READING

- Balasubramaniam S et al: Inborn errors of purine metabolism: Clinical update and therapies. *J Inherit Metab Dis* 37:669, 2014.
- Balasubramaniam S et al: Inborn errors of pyrimidine metabolism: Clinical update and therapy. *J Inherit Metab Dis* 37:687, 2014.
- Ben Salem C et al: Drug-induced hyperuricaemia and gout. *Rheumatology (Oxford)* 266:679, 2017.

Bhole V, Krishnan E: Gout and the heart. *Rheum Dis Clin North Am* 40:125, 2014.

Burns CM, Wortman RL: Clinical features and treatment of gout, in *Kelley and Firestein's Textbook of Rheumatology*, 10th ed, GS Firestein et al (eds). Philadelphia, Elsevier, 2017, pp 1620–1644.

Hirano M, Peters GJ: Advances in purine and pyrimidine metabolism in health and diseases. *Nucleosides Nucleotides Nucleic Acids* 35:495, 2016.

Tai V et al: Genetic advances in gout: Potential applications in clinical practice. *Curr Opin Rheumatology* 31:144, 2019.

Terkel taub R (ed): *Gout and Other Crystal Arthropathies*. Philadelphia, Elsevier Health Sciences, 2012.

## 418

## Lysosomal Storage Diseases

Robert J. Hopkin, Gregory A. Grabowski



Lysosomes are heterogeneous subcellular organelles containing specific hydrolases that allow selective processing or degradation of proteins, nucleic acids, carbohydrates, and lipids. There are >50 different lysosomal storage diseases (LSDs), classified according to the nature of the stored material (Table 418-1). Although all are rare diseases, several of the more prevalent disorders are reviewed here: Tay-Sachs disease, Fabry disease, Gaucher disease, Niemann-Pick disease, the mucopolysaccharidoses, Pompe disease, lysosomal acid lipase deficiency (LALD), Krabbe disease, and CLN2-related Batten disease. LSDs should be considered in the differential diagnosis of patients with neurologic, renal, or muscular degeneration and/or unexplained hepatomegaly, splenomegaly, cardiomyopathy, or skeletal dysplasias and deformations. Physical findings are disease specific, and enzyme assays or genetic testing can be used to make a definitive diagnosis. Although the nosology of LSDs segregates the variants into distinct phenotypes, these are heuristic; in the clinic, each disease exhibits—to varying degrees—a spectrum of manifestations, from severe to attenuated variants.

#### PATHOGENESIS

Lysosomal biogenesis involves ongoing synthesis of lysosomal hydrolases, membrane constitutive proteins, and new membranes. Lysosomes originate from the fusion of trans-Golgi network vesicles with late endosomes. Progressive vesicular acidification accompanies the maturation of these vesicles; this gradient facilitates the pH-dependent dissociation of receptors and ligands and also activates lysosomal hydrolases. Lysosomes are components of the lysosome/autophagy/mitophagy system that are regulated by the mTORC1 modulation of the transcription factors TFEB/TFE3. This regulation is disrupted to varying degrees in specific tissues affected by individual LSDs.

Abnormalities at any biosynthetic step can impair enzyme activation and lead to an LSD. After leader sequence clipping, remodeling of complex oligosaccharides (including the lysosomal targeting ligand mannose-6-phosphate as well as high-mannose oligosaccharide chains of many soluble lysosomal hydrolases) occurs during transit through the Golgi. Lysosomal integral or associated membrane proteins are sorted to the membrane or interior of the lysosome by several different peptide signals. Phosphorylation, sulfation, additional proteolytic processing, and macromolecular assembly of heteromers occur concurrently. Such posttranslational modifications are critical to enzyme function, and defects can result in multiple enzyme/protein deficiencies.

The final common pathway for LSDs is the accumulation of specific macromolecules within selected tissues and cells that normally have a high flux of these substrates. The majority of lysosomal enzyme deficiencies result from point mutations or genetic rearrangements

**TABLE 418-1 Selected Lysosomal Storage Diseases**

DISORDER*	ENZYME DEFICIENCY [SPECIFIC THERAPY]	STORED MATERIAL	CLINICAL TYPES (ONSET)	INHERITANCE	CLINICAL FEATURES					
					NEUROLOGIC	LIVER, SPLEEN ENLARGEMENT	SKELETAL DYSPLASIA	OPHTHALMOLOGIC	HEMATOLOGIC	UNIQUE FEATURES
<b>Mucopolysaccharidoses (MPS)</b>										
MPS I H, Hurler	$\alpha$ -L-Iduronidase [ET, HSCT]	Dermatan sulfate Heparan sulfate	Infantile Intermediate	AR	Cognitive degeneration	+++	++++	Corneal clouding	Vacuolated lymphocytes	Coarse facies; cardiovascular involvement; joint stiffness
MPS I H/S, Hurler/Scheie			Childhood/adult		Cognitive degeneration					
MPS I S, Scheie					None					
MPS II, Hunter	Iduronate sulfatase [ET]	Dermatan sulfate Heparan sulfate	Severe infantile Mild juvenile	X-linked	Cognitive degeneration, less in mild form	+++	++++	Retinal degeneration, no corneal clouding	Granulated lymphocytes	Coarse facies; cardiovascular involvement; joint stiffness; distinctive pebbly skin lesions
MPS III A, Sanfilippo A	Heparan-N-sulfatase	Heparan sulfate	Late infantile	AR	Severe cognitive degeneration	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III B, Sanfilippo B	N-Acetyl- $\alpha$ -glucosaminidase	Heparan sulfate	Late infantile	AR	Severe cognitive degeneration	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III C, Sanfilippo C	Acetyl-CoA: $\alpha$ -glucosaminide N-acetyltransferase	Heparan sulfate	Late infantile	AR	Severe cognitive degeneration	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS III D, Sanfilippo D	N-Acetylglucosamine-6-sulfate sulfatase	Heparan sulfate	Late infantile	AR	Severe cognitive degeneration	+	+	None	Granulated lymphocytes	Mild coarse facies
MPS IV A, Morquio A	N-Acetylgalactosamine-6-sulfate sulfatase [ET—trials]	Keratan sulfate Chondroitin-6 sulfate	Childhood	AR	None	+	++++	Corneal clouding	Granulated neutrophils	Distinctive skeletal deformity; odontoid hypoplasia; aortic valve disease
MPS IV B, Morquio	$\beta$ -Galactosidase		Childhood	AR	None	$\pm$	++++			
MPS VI, Maroteaux-Lamy	Arylsulfatase B [ET, BMT]	Dermatan sulfate	Late infantile	AR	None	++	++++	Corneal clouding	Granulated neutrophils and lymphocytes	Coarse facies; valvular heart disease
MPS VII	$\beta$ -Glucuronidase [ET]	Dermatan sulfate Heparan sulfate	Neonatal Infantile Adult	AR	Cognitive degeneration, absent in some adults	+++	+++	Corneal clouding	Granulated neutrophils	Coarse facies; vascular involvement; hydrops fetalis in neonatal form
<b>GM<sub>2</sub> Gangliosidoses</b>										
Tay-Sachs disease	$\beta$ -Hexosaminidase A	GM <sub>2</sub> gangliosides	Infantile Juvenile	AR	Cognitive degeneration; seizures; later juvenile form	None	None	Cherry red spot in infantile form	None	Macrocephaly; hyperacusis in infantile form

(Continued)

**TABLE 418-1 Selected Lysosomal Storage Diseases (Continued)**

DISORDER <sup>a</sup>	ENZYME DEFICIENCY [SPECIFIC THERAPY]	STORED MATERIAL	CLINICAL TYPES (ONSET)	INHERITANCE	CLINICAL FEATURES						
					NEUROLOGIC	LIVER, SPLEEN ENLARGEMENT	SKELETAL DYSPLASIA	OPHTHALMOLOGIC	HEMATOLOGIC	UNIQUE FEATURES	
Sandhoff disease	β-Hexosaminidases A and B	GM <sub>2</sub> gangliosides	Infantile	AR	Cognitive degeneration; seizures	++	±	Cherry red spot	None	Macrocephaly; hyperacusis	
<b>Neutral Glycosphingolipidoses</b>											
Fabry disease	α-Galactosidase A [ET, Chaperone]	Globotriaosylceramide	Childhood	X-linked	Painful acroparesthesias	None	None	Corneal dystrophy, vascular lesions	None	Cutaneous angiokeratomas; hypohydrosis	
Gaucher disease	Acid β-glucosidase [ET, SRT]	Glucosylceramide, glycosylsphingosine	Type 1 Type 2 Type 3	AR	None ++++ +++ -/+++	++++	++++ + ++++	None Eye movements Eye movements	Gaucher cells in bone marrow; cytopenias	Adult form highly variable	
Niemann-Pick disease A and B	Acid sphingomyelinase [ET—trials]	Acid sphingomyelin	Neuronopathic, type A Nonneuronopathic, type B	AR	Cognitive degeneration; seizures	++++	None Osteoporosis	Macular degeneration	Foam cells in bone marrow	Pulmonary infiltrates Lung failure	
<b>Glycoproteinoses</b>											
Fucosidosis	α-Fucosidase	Glycopeptides; oligosaccharides	Infantile Juvenile	AR	Cognitive degeneration	++	++	None	Vacuolated lymphocytes; foam cells	Coarse facies; angiokeratomas in juvenile form	
α-Mannosidosis	α-Mannosidase	Oligosaccharides	Infantile Milder variant	AR	Cognitive degeneration	+++	+++	Cataracts, corneal clouding	Vacuolated lymphocytes, granulated neutrophils	Coarse facies; enlarged tongue	
β-Mannosidosis	β-Mannosidase	Oligosaccharides		AR	Seizures; cognitive degeneration		++	None	Vacuolated lymphocytes, foam cells	Angiokeratomas	
Aspartylglucosaminuria	Aspartylglucosaminidase	Aspartylglucosamine; glycopeptides	Young adult	AR	Cognitive degeneration	±	++	None	Vacuolated lymphocytes, foam cells	Coarse facies	
Sialidosis	Neuraminidase	Sialyloligosaccharides	Type I, congenital Type II, infantile and juvenile	AR	Myoclonus; cognitive degeneration	++, less in type I	++, less in type I	Cherry red spot	Vacuolated lymphocytes	MPS phenotype in type II	
<b>Mucolipidoses (ML)</b>											
ML-II, I-cell disease	UDP-N-Acetylglucosamine-1-phosphotransferase	Glycoprotein; glycolipids	Infantile	AR	Cognitive degeneration	+	++++	Corneal clouding	Vacuolated and granulated neutrophils	Coarse facies; absence of mucopolysacchariduria; gingival hypoplasia	
ML-III, pseudo-Hurler polydystrophy	UDP-N-Acetylglucosamine-1-phosphotransferase	Glycoprotein; glycolipids	Late infantile	AR	Mild cognitive degeneration	None	+++	Corneal clouding, mild retinopathy, hyperopic astigmatism		Coarse facies; stiffness of hands and shoulders	
<b>Leukodystrophies</b>											
Krabbe disease	Galactosylceramidase [BMT/HSCT]	Galactosylceramide Galactosylsphingosine	Infantile	AR	Cognitive degeneration	None	None	None	None	White matter globoid cells	

Metachromatic leukodystrophy	Arylsulfatase A	Cerebroside sulfate	Infantile Juvenile Adult	AR	Cognitive degeneration; dementia; psychosis in adult	None	None	Optic atrophy	None	Gait abnormalities in late infantile form
Multiple sulfatase deficiency	Active site cysteine to C <sub>6</sub> -formylglycine-converting enzyme	Sulfatides; mucopolysaccharides	Late infantile	AR	Cognitive degeneration	+	++	Retinal degeneration	Vacuolated and granulated cells	Absent activity of all known cellular sulfatases
<b>Disorders of Neutral Lipids</b>										
Infantile-onset LALD	Acid lysosomal lipase [ET]	Cholesteryl esters; triglycerides	Infantile	AR	None	+++	None	None	None	Adrenal calcification
Childhood/adult-onset LALD	Acid lysosomal lipase [ET]	Cholesteryl esters	Childhood	AR	None	Hepatomegaly	None	None	None	Fatty liver disease; cirrhosis
Farber disease	Acid ceramidase	Ceramide	Infantile Juvenile	AR	Occasional cognitive degeneration	±	None	Macular degeneration	None	Arthropathy, subcutaneous nodules
<b>Disorders of Glycogen</b>										
Pompe disease	Acid $\alpha$ -glucosidase [ET]	Glycogen	Infantile, late onset	AR	Neuromuscular	±	None	None	None	Myocardiopathy
Late-onset GAA deficiency	Acid $\alpha$ -glucosidase [ET]	Glycogen	Variable: juvenile to adulthood	AR	Neuromuscular	None	None	None	None	Respiratory insufficiency; neuromuscular disease
Danon disease	LAMP-2 (lysosomal associated membrane protein-2)	Glycogen	Variable: childhood to adulthood	X-linked (?Dominant)	Cardiomyopathy Neuromuscular Inconsistent cognitive degeneration	None	None	None	None	Myocardial vacuolar degeneration
<b>Neuronal Ceroid Lipofuscinoses</b>										
CLN2 (a.k.a. NCL2)	TPP1 (tripeptidyl peptidase 1) [ICV ET]	Ceroid lipofuscin	Early childhood	AR	Neurodegenerative Loss of motor skills Myoclonus Loss of vision Cognitive loss Wheelchair bound by adolescence	None	None	Progressive vision loss	None	Symmetric retinal progressive degeneration by 4–6 years

<sup>a</sup>Comprehensive reviews of these lysosomal storage diseases can be found in DL Valle et al: *The Online Metabolic and Molecular Bases of Inherited Disease*, New York, McGraw-Hill, <https://ommbid.mhmedical.com/book.aspx?bookID=2709#225069419>.

*Abbreviations:* AR, autosomal recessive; BMT/HSCT, bone marrow or hematopoietic stem cell transplantation; ET, enzyme therapy; ICV ET, intracerebroventricular enzyme therapy; LALD, lysosomal acid lipase deficiency; SRT, substrate reduction therapy.

at a locus that encodes a single lysosomal hydrolase. However, some mutations cause deficiencies of several different lysosomal hydrolases by alteration of the enzymes/proteins involved in targeting, active site modifications, macromolecular association, or trafficking. Nearly all LSDs are inherited as autosomal recessive disorders except for Hunter (mucopolysaccharidosis type II), Danon, and Fabry diseases, which are X-linked, and two autosomal dominant conditions causing Parry type neuronal ceroid lipofuscinosis (CLN) due to mutations in *DNAJC5* or frontotemporal dementia and *CLN11* due to *GRN* (progranulin) mutations. Substrate accumulation leads to lysosomal distortion/dysfunction, which has significant pathologic consequences. In addition, abnormal amounts of metabolites may also have pharmacologic effects important to disease pathophysiology and propagation, particularly activation of the innate immune responses.

For many LSDs, the accumulated substrates are synthesized within particular tissue sites of pathology. Other diseases have greater exogenous substrate supplies. For example, substrates are delivered by low-density lipoprotein receptor-mediated uptake in Fabry and LALD or by phagocytosis in Gaucher disease type 1. The threshold hypothesis refers to a level of enzyme activity below which disease develops. Small changes in enzyme activity near that threshold can lead to or modify disease. A critical element of this model is that enzymatic activity can be challenged by changes in substrate flux based on genetic background, cell turnover, recycling, or metabolic demands. Thus, a set level of residual enzyme may be adequate for substrate in some tissues or cells but not in others. In addition, several variants of each LSD exist at a clinical level. These disorders therefore represent a spectrum of manifestations that are not easily dissociated into discrete entities. The molecular/genetic bases for such variations have not been elucidated in any detail.

There are European Medicines Agency (EMA) and U.S. Food and Drug Administration (FDA) treatments available for a growing number of LSDs. The first was enzyme replacement therapy (ET) for Gaucher disease; this has been followed by additional ETs, but subsequent developments have included modified enzyme infusion, substrate inhibition, hematopoietic stem cell transplant (HSCT), pharmacologic chaperone therapy (which uses a small molecule to stabilize enzyme produced by the mutated gene and allows it to function), intrathecal enzyme delivery, and gene therapy. The technical ability to intervene for most LSDs now exists but with highly variable impact. Significant additional research is needed to reach the goals of long-term survival with good function and quality of life.

## SELECTED DISORDERS

### TAY-SACHS DISEASE

About 1 in 30 Ashkenazi Jews is a carrier for Tay-Sachs disease (total hexosaminidase A [Hex A] deficiency), resulting from  $\alpha$ -chain gene mutations. The infantile form is a neurodegenerative disease that results in death in infancy. It is characterized by macrocephaly, loss of motor skills, increased startle reaction, and a macular cherry red spot. The juvenile-onset form presents as ataxia and dementia, with death by age 10–15 years. The adult-onset disorder is characterized by clumsiness in childhood; progressive motor weakness in adolescence; and additional spinocerebellar and lower-motor-neuron signs and dysarthria in adulthood; intelligence declines slowly, and psychiatric disorders are common. Screening for Tay-Sachs disease carriers is recommended in the Ashkenazi Jewish population. Sandhoff disease, due to a deficiency in both Hex A and Hex B resulting from defective  $\beta$ -chains, is phenotypically similar to Tay-Sachs disease with the addition of hepatosplenomegaly and bony dysplasias.

### FABRY DISEASE

Fabry disease, an X-linked disorder and likely the most prevalent LSD, results from mutations in *GALA*, which encodes  $\alpha$ -galactosidase A. The estimated prevalence of hemizygous males ranges from 1 in 40,000 to 1 in 3500 in selected populations. Females are expected to have a higher prevalence of mutations, but more variable manifestations. In males, the disease manifests with angiokeratomas (telangiectatic skin

lesions), hypohidrosis, corneal and lenticular opacities, acroparesthesia, and progressive disease of the kidney, heart, and brain vascular systems. Abdominal pain, recurrent diarrhea, and acroparesthesias (debilitating episodic burning pain of the hands, feet, and proximal extremities) may appear in childhood. In females, the overall manifestations vary, except that kidney disease is uncommon. Angiokeratomas often appear in adolescence and are punctate, dark red to blue-black, flat or slightly raised, and usually symmetric; they do not blanch with pressure. They are often small and can be easily overlooked. They usually are most dense between the umbilicus and the knees—the “bathing suit area”—but may occur anywhere, including the mucosal surfaces. Angiokeratomas also occur in several other very rare LSDs. Corneal and lenticular lesions, detectable on slit-lamp examination, may help in establishing a diagnosis of Fabry disease. Acroparesthesia can last from minutes to days and can be precipitated by changes in temperature, exercise, fatigue, or fever. Abdominal pain can resemble that from appendicitis or renal colic. Proteinuria, isosthenuria, and progressive renal dysfunction occur in the second to fourth decades; ~5% of male patients with idiopathic renal failure have *GALA* mutations. Hypertension, left ventricular hypertrophy, anginal chest pain, and congestive heart failure can occur in the third to fourth decades. About 1–3% of patients with idiopathic hypertrophic cardiomyopathy have Fabry disease. Similarly, ~2–5% of patients with idiopathic stroke at 35–50 years of age have *GALA* mutations. Leg lymphedema occurs without hypoalbuminemia. Death is due to cardiovascular, renal, or cerebrovascular disease in untreated patients. Variants with residual  $\alpha$ -galactosidase A activity may have late-onset manifestations that are limited to the cardiovascular system and resemble hypertrophic cardiomyopathy. Cases with predominant cardiac, renal, or central nervous system (CNS) manifestations have been reported. Up to 70% of heterozygous females exhibit clinical manifestations. However, in females, heart disease is the most common life-threatening manifestation, followed in frequency by stroke and renal disease. In males, renal disease followed by cardiovascular disease and stroke are most life-threatening.

Gabapentin and carbamazepine diminish chronic and episodic acroparesthesia. Chronic hemodialysis or kidney transplantation can be lifesaving in patients with renal failure. Intravenous ET clears stored lipids from a variety of cells. More recently a chaperone therapy (migalastat) that stabilizes the residual enzyme made by the patient's body has allowed oral therapy for some patients with amenable mutations. Renal insufficiency, cardiac fibrosis, and stroke are irreversible; therefore, early institution of therapy provides the best opportunity to prevent or slow the progression of life-threatening complications.

### GAUCHER DISEASE

Gaucher disease, a panethnic autosomal recessive disorder, results from defective activity of acid  $\beta$ -glucuronidase; ~600 *GBA1* mutations have been described in such patients. Clinically, disease variants are classified by the absence or presence and progression of primary CNS involvement.

Gaucher disease type 1 is a nonneuronopathic disease (i.e., absence of early-onset or progressive CNS disease) presenting in childhood to adulthood as slowly to rapidly progressive visceral disease. About 55–60% of patients are diagnosed at <20 years of age in white populations and at even younger ages in other groups. This pattern of presentation is distinctly bimodal, with peaks at <10–15 years and at >25 years. Younger patients tend to have greater degrees of hepatosplenomegaly and accompanying blood cytopenias. In contrast, older patients have a greater tendency for chronic bone disease. Hepatosplenomegaly occurs in virtually all clinically identified patients and can be minor or massive. Accompanying anemia and thrombocytopenia are variable and are not directly related to liver or spleen volumes. Severe liver dysfunction is unusual. Splenic infarctions can resemble an acute abdomen. Pulmonary hypertension and alveolar Gaucher cell accumulation are uncommon but life-threatening and can occur at any age. *GBA1* mutations in the heterozygous or homozygous states lead to a significantly increased lifetime risk for developing Parkinson disease. The basic mechanisms for this risk are unknown.

All patients with Gaucher disease have nonuniform infiltration of bone marrow by lipid-laden macrophages termed Gaucher cells. This phenomenon can lead to marrow packing with subsequent infarction, ischemia, necrosis, and cortical bone destruction. Bone marrow involvement spreads from proximal to distal in the limbs and can involve the axial skeleton extensively, causing vertebral collapse. In addition to bone marrow involvement, bone remodeling is defective, with loss of total bone calcium leading to osteopenia, osteonecrosis, avascular infarction, and vertebral compression fractures with spinal cord involvement. Aseptic necrosis of the femoral head is common, as is fracture of the femoral neck. The mechanism by which diseased bone marrow macrophages interact with osteoclasts and/or osteoblasts to cause bone disease is not well understood. Chronic, ill-defined bone pain can be debilitating and poorly correlated with radiographic findings. "Bone crises" are associated with localized excruciating pain and, on occasion, local erythema, fever, and leukocytosis. These crises represent acute infarctions of bone, as evidenced in nuclear scans by localized absent uptake of pyrophosphate agents. Decreased acid -glucosidase activity (0–20% of normal) in nucleated cells establishes the diagnosis. The enzyme is not normally present in bodily fluids. The sensitivity of enzyme testing is poor for heterozygote detection; molecular testing by whole *GBA1* sequencing is the standard. The disease frequency varies from ~1 in 1000 among Ashkenazi Jews to <1 in 100,000 in other populations; ~1 in 12–15 Ashkenazi Jews carries a Gaucher disease allele. Four common mutations account for ~85% of the mutations in that population of affected patients: p.N370S (also known as p.N409S), 84GG (a G insertion at cDNA position 84), p.L444P (also known as p.L483P), and IVS-2<sup>-1</sup> (an intron 2 splice junction mutation).

Genotype/phenotype studies indicate a significant, though not absolute, correlation between disease type and severity and the *GBA1* genotype. The most common mutation in the Ashkenazi Jewish population (p.N370S) shares, either homozygously or heteroallelically, a 100% association with nonneuronopathic or type 1 Gaucher disease. The N370S/N370S and N370S/other mutant allele genotypes are associated with later-onset/less severe disease and with earlier-onset/severe disease, respectively. As many as 40% of individuals with the N370S/N370S genotype do not present clinically. Other alleles include L444P (very low activity), 84GG (null), or IVS-2 (null) and rare/private or uncharacterized alleles. The L444P/L444P patients frequently have life-threatening to very severe/early-onset disease, and many, though not all, develop CNS involvement in the first two decades of life.

Symptom-based treatment of blood cytopenias and joint replacement surgeries continue to have important roles in management. However, regular intravenous ET has been the first-line treatment for significantly affected patients and is highly efficacious and safe in diminishing hepatosplenomegaly and improving hematologic values. An oral substrate reduction therapy (elaglucostat tartrate), which inhibits glycolipid synthesis, is approved as a first-line therapy for adults. Bone disease is decreased and can be prevented, but irreversible damage cannot be reversed, by ET. Adult patients may benefit from adjunctive treatment with bisphosphonates or other interventions to improve bone density. Adults who cannot be treated with enzyme, either because it is not effective or because they have developed an allergy or other hypersensitivities to the enzyme, may receive substrate reduction therapy with either eliglucostat tartrate or miglustat; the latter is approved as a second-line oral therapy.

Gaucher disease type 2 is a rare, severe, progressive CNS disease that leads to death by 2 years of age, depending on supportive care. Gaucher disease type 3 has highly variable manifestations in the CNS and viscera. It can present in early childhood with rapidly progressive, massive visceral disease and slowly progress to static CNS involvement that may not be evident by standard IQ evaluations; in adolescence with dementia; or in early adulthood with rapidly progressive, uncontrollable myoclonic seizures and mild visceral disease. Visceral disease in type 3 is nearly identical to that in type 1 but is generally more severe. Early CNS findings may be limited to defects in lateral gaze tracking, which may remain static for decades. Cognitive degeneration can be slowly progressive or static. Type 3 is much more frequent among individuals of non-Western world descent. Visceral—but not CNS—involvement responds to ET.

## NIEMANN-PICK DISEASES

Niemann-Pick diseases (acid sphingomyelinase deficiency [ASMD]) are autosomal recessive disorders that result from defects in acid sphingomyelinase (ASM). Types A and B are distinguished by the early age of onset and progressive CNS disease in type A. Type A typically has its onset in the first 6 months of life, with rapidly progressive CNS deterioration, spasticity, failure to thrive, and massive hepatosplenomegaly. Type B has a later, more variable onset and is characterized by a progression of hepatosplenomegaly, with eventual development of cirrhosis and hepatic parenchymal and Kupffer cell replacement by foam cells filled with sphingomyelin. Affected patients develop progressive pulmonary disease with dyspnea, hypoxemia, and a reticular infiltrative pattern on chest x-ray. Foam cells are present in alveoli, lymphatic vessels, and pulmonary arteries. Progressive hepatic or lung disease can lead to death in adolescence or early adulthood. The "type B" phenotype includes some patients with slowly progressive CNS involvement.

The diagnosis is established by markedly decreased (1–10% of normal) ASM activity in nucleated cells. There is no approved specific treatment for Niemann-Pick disease, but intravenous ET clinical trials are in phase 3. The efficacies of hepatic transplant (HT) or bone marrow transplantation (BMT) are not established. More complications than expected have occurred with these interventions due to either (1) recurrence of hepatic disease in the transplant following HT by repopulation of bone marrow-derived ASM-deficient myeloid cells or (2) lack of clearance of sphingomyelin in hepatocytes by ASM cross-correction following the BMT of ASM-normal bone marrow stem cells.

Niemann-Pick C diseases are progressive CNS diseases due to mutations in either *NPC1* or *NPC2* mutations in either, lysosomal proteins involved in cholesterol and selected sphingolipid transport out of the lysosome. They can present with liver or splenic disease, but their major manifestations are progressive CNS disease over one to two decades. Treatment with substrate inhibition agents (e.g., miglustat) has shown minor CNS effects, and substrate depletion with cyclodextrin is in clinical trials for NPC1 disease.

## MUCOPOLYSACCHARIDOSES

Mucopolysaccharidosis type I (MPS I) is an autosomal recessive disorder caused by deficiency of -L-iduronidase. The spectrum of involvement traditionally has been divided into three categories: (1) Hurler disease (MPS I H) for severe deficiency with neurodegeneration, (2) Scheie disease (MPS I S) for later-onset disease without neurologic involvement and with relatively less severe disease in other organ systems, and (3) Hurler-Scheie syndrome (MPS I H/S) for patients intermediate between these extremes. MPS I H/S is characterized by severe somatic disease, usually without major overt neurologic deterioration. MPS I often presents in infancy or early childhood as chronic rhinitis, clouding of the corneas, hepatosplenomegaly, and progressive dysmorphia. As the disease progresses, nearly every organ system can be affected. In the more severe forms, cardiac and respiratory diseases become life threatening in childhood. Skeletal disease can be quite severe, resulting in very limited mobility. There are two current treatments for the MPS I diseases. HSCT is the standard treatment for patients presenting at <2 years of age who appear to have or are at risk for neurologic degeneration. Because early diagnosis and intervention are essential, MPS I has been added to the recommended newborn screen (NBS). HSCT results in stabilization of CNS disease and reverses hepatosplenomegaly. It also beneficially affects cardiac and respiratory disease. HSCT does not eliminate corneal disease or result in the resolution of progressive skeletal disease. ET effectively addresses hepatosplenomegaly and alleviates cardiac and respiratory disease. The enzyme does not penetrate the blood-brain barrier and does not directly affect CNS disease. ET and HSCT appear to have similar effects on visceral signs and symptoms. ET poses a lower risk of life-threatening complications and may therefore be advantageous for patients who have attenuated manifestations without CNS disease. A combination of ET and HSCT has been used, with ET initiated prior to transplantation in an attempt to reduce the disease burden. The experience with this approach is not well documented, but it appears to have advantages over HSCT alone. It is clear that HSCT has benefited

patients. However, late cardiac and respiratory complications of MPS I are being reported including obstructive breathing requiring pressure support, cardiomyopathy, and/or valve disease. Regular follow-up for patients with MPS I is required throughout their lives even after successful HSCT.

Hunter disease (MPS II) is an X-linked disorder due to deficiency in iduronate sulfate sulfatase and has manifestations similar to those of MPS I, including some variants with neurologic degeneration. There is no corneal clouding or other eye disease. Like MPS I, MPS II is clinically variable, with CNS and non-CNS variants. HSCT has not been successful in treating CNS disease associated with MPS II. The FDA and EMA have approved ET for the visceral manifestations of MPS II.

MPS IV or Morquio syndrome is a rare autosomal recessive condition (1 in 200,000–300,000) and is different than the other mucopolysaccharidoses in presenting as a spondyloepiphyseal skeletal dysplasia and hyperextensibility of all joints. There are also major heart and respiratory complications. This disorder often presents in childhood, but the age of onset and rate of progression are quite variable. Two variants, type A and type B, are caused by deficiencies in *N*-acetyl-galactosamine-6-sulfatase (*GALNS*) and an acid -galactosidase, respectively. A recombinant human *GALNS* ET (elosulfase alfa) is approved for the treatment of MPS IVA, making it essential to confirm the specific enzyme diagnosis. Treatment has been shown to improve ambulatory mobility and decrease pain. There is no current specific treatment for MPS IVB.

ET for Maroteaux-Lamy disease (MPS VI), arylsulfatase B deficiency, has received FDA approval as well as approval by similar agencies in other countries. This very rare autosomal recessive disorder is characterized by hepatosplenomegaly, bone disease, heart disease, and respiratory compromise. Short stature is also an important manifestation. Visceral signs and symptoms are similar to those in MPS I; however, MPS VI is not associated with neurologic degeneration.

MPS VII, Sly syndrome, is due to mutations in *GUSB*, which encodes -glucuronidase. Severe deficiency in this enzyme may present with fetal hydrops, which can lead to stillbirth or perinatal demise. Other patients with MPS VII may present later with short stature, coarse facial features, and hepatosplenomegaly. There is ET for this disorder (vestronidase alfa-vjbk).

### POMPE DISEASE

Acid maltase (acid -glucosidase deficiency) due to GAA mutation, also called Pompe disease, is the only LSD leading to primary glycogen storage. The classic severe infantile form presents with hypotonia, myopathies, and hepatosplenomegaly. This variant is rapidly progressive and generally results in death in the first year of life. However, as with other LSDs, there are early- and late-onset forms of this disorder.

The late-onset variants may be as common as 1 in 40,000; patients typically present with a slowly progressive myopathy that may resemble limb-girdle muscular dystrophy. Respiratory insufficiency may be the presenting sign or may develop with advancing disease. In late stages of the disease, patients may require mechanical ventilation, report swallowing difficulties, and experience loss of bowel and bladder control. Myopathies are not usually present in late-onset variants of Pompe disease.

The FDA, EMA, and similar agencies have approved ET for Pompe disease patients of all ages. This treatment clearly prolongs life in the infantile form, consistently resulting in improved cardiac function. Respiratory function is also improved in most treated infants if instituted before age 6 months. Some infants demonstrate marked improvement in motor functions, while others have minor changes in muscle tone or strength. Recently, several states have instituted NBS for Pompe disease. In addition, newer protocols for treatment with methotrexate and rituximab have greatly decreased antidrug antibody formation. The combination of NBS and immunomodulation preceding ET has greatly improved therapeutic response and long-term survival. Prevention of deterioration has been shown with GAA ET in the late-onset forms. Early intervention with acid  $\alpha$ -glucosidase ET in such patients may limit or prevent deterioration, but very advanced disease will have significant irreversible components.

### LYSOSOMAL ACID LIPASE DEFICIENCY

Wolman syndrome (now infantile-onset LALD) and cholesterol ester storage disease (now childhood/adult-onset LALD) are caused by deficiency of lysosomal acid lipase (LAL) due to autosomal recessive mutations in *LIPA*. The diagnosis is established by enzyme or gene analyses of LAL or *LIPA* in serum/plasma or nucleated cells. LAL hydrolyzes cholesterol esters and triglycerides delivered to the lysosome via the LDLR pathway. Accumulation of these in the tissues leads to progressive organ dysfunction including liver disease, intestinal malabsorption, heart dysfunction, and other manifestations. The most severe form presents in early infancy as a medical emergency with severe failure to thrive, vomiting, and hepatosplenomegaly. The infantile-onset LALD patients die without specific treatment by age 1 year (median age of death, 3.7 months). Childhood/adult-onset LALD can have a variable age of initial presentation with nonspecific signs but often involves elevated liver enzymes, nonalcoholic fatty liver disease, cryptogenic cirrhosis, and varying severities of hepatosplenomegaly. Importantly, neither clinical variant manifests primary CNS disease. Disease progresses throughout life and may result in early (adolescence) liver cirrhosis and (early adulthood) atherosclerosis or early death without treatment. Importantly, statins can decrease the hypercholesterolemia but do not alter the basic progressive tissue (e.g., liver) pathology. The majority of the later onset patients are evaluated by hepatology or lipidology physicians. ET for LALD has major effects in reversing disease manifestations and was approved for patients at all ages by the EMA, FDA, and several other country agencies in 2015 and 2016.

### KRABBE DISEASE

Deficiency in galactocerebrosidase (*GALC*) causes Krabbe disease, an autosomal recessive neurodegenerative disorder due to mutations in *GALC*. Krabbe disease is panethnic but quite rare. The early infantile form presents at an average age of 4 months and progresses rapidly, with death at an average age of 18 months. Later onset forms also exist and have onsets and survival that are highly variable. The early-onset form presents with hyperirritability, feeding problems, fever, seizures, and neurodegeneration. Blindness, hypotonia, and loss of voluntary movement develop over time. Later onset forms present with spasticity, ataxia, vision loss, and behavioral problems and progress to dementia and early death. There is no FDA-approved treatment, but early pre-symptomatic HSCT has been used. This results in improved survival, but neurologic problems are still common. More recently studies in mouse and dog models have used gene therapy with dramatic improvement in both neurologic function and survival. Human studies are being implemented.

### NEURONAL CEROID LIPOFUSCINOSIS TYPE 2 (CLN2 OR CLN2)

There are at least 13 genes that have been associated with storage of neuronal ceroid lipofuscin. One of these, CLN2, is due to mutations in *TPP1* and deficiency in tripeptidyl peptidase 1. This autosomal recessive neurodegenerative disorder typically presents between age 2 and 4 years, most commonly with seizures, ataxia, myoclonus, and vision loss. Motor skill losses include sitting, walking, speech, and feeding and lead to severe disability and eventually death at an average age of 12 years. Intellectual disability and behavioral problems also become increasingly severe with age. Most affected children are wheelchair bound in late childhood, and survival beyond adolescence is rare. There are later onset patients, and there is significant clinical overlap between CLN2 and other CLNs; confirmation of the diagnosis by gene sequencing is essential. In 2017, the FDA/EMA approved treatment of CLN2, cerliponase alfa, an ET that is administered by intracerebroventricular injection over several hours every 2 weeks. Administration of cerliponase alfa is facilitated by placement of an intracerebroventricular port to allow reliable access. Currently, this is the only approved ET that is intrathecally administered. CLN2 is the only neuronal ceroid lipofuscinosis that has a specific treatment. Several others are in pre-clinical development.

## FURTHER READING

- Al den hoven M et al: Long-term outcome of Hurler syndrome patients after hematopoietic cell transplantation: An international multicenter study. *Blood* 125:2164, 2015.
- Balwani M et al: Recommendations for the use of eliglustat in the treatment of adults with Gaucher disease type 1 in the United States. *Mol Genet Metab* 117:95, 2016.
- Ortiz A et al: Fabry disease revisited: Management and treatment recommendations for adult patients. *Mol Genet Metab* 123:416, 2018.
- Schoser B et al: Survival and long-term outcomes in late-onset Pompe disease following alglucosidase alfa treatment: A systematic review and meta-analysis. *J Neurol* 264:621, 2017.
- Schulz A et al: Study of intraventricular cerliponase alfa for CLN2 disease. *N Engl J Med* 378:1898, 2018.

continues to expand as is seen in the case of GSDs type II, III, and IX. Advances in the molecular basis of these diseases are now being used to improve diagnosis and management. Some of these disorders are candidates for enzyme replacement therapy, substrate reduction therapy, gene therapy, and other genomic tools, such as small interfering RNA (siRNA) technology and CRISPR genome editing technology.

Historically, the GSDs were categorized numerically in the order in which the enzymatic defects were identified. GSDs are also classified based on the primary organs involved (liver, muscle, and/or heart) and clinical manifestations. In this chapter, the GSDs will be classified based on the organ involvement (**Table 419-1**). The overall frequency of all forms of GSDs is between 1 in 20,000 to 1 in 40,000 live births in the United States and Europe, and up to 1 in 10,000 worldwide for some GSDs. Most are inherited as autosomal recessive traits; however, phosphoglycerate kinase deficiency; two forms of liver and muscle phosphorylase kinase (PhK) deficiency caused by mutations in PHKA2 and PHKA1 genes respectively, and lysosomal-associated membrane protein 2 (LAMP2) deficiency are X-linked disorders. The most common childhood disorders are glucose-6-phosphatase deficiency (GSD type I), lysosomal acid -glucuronidase deficiency (GSD type II), debrancher enzyme deficiency (GSD type III), and liver PhK deficiency (GSD type IX). The most common adult disorder is myophosphorylase deficiency (GSD type V).

## SELECTED LIVER GLYCOGENOSSES

### DISORDERS WITH HEPATOMEGALY AND HYPOGLYCEMIA

**Type I GSD (Glucose-6-Phosphatase or Translocase Deficiency, Von Gierke Disease)** Type I GSD is an autosomal recessive disorder caused by glucose-6-phosphatase or translocase deficiency in liver, kidney, and intestinal mucosa. There are two subtypes of GSD I: type Ia, in which the glucose-6-phosphatase enzyme is defective, and type Ib, in which the translocase that transports glucose-6-phosphate across the microsomal membrane is defective. The defects in both subtypes lead to inadequate conversion of glucose-6-phosphate to glucose in the liver and thus make affected individuals susceptible to fasting hypoglycemia.

**CLINICAL AND LABORATORY FINDINGS** Persons with type I GSD may develop hypoglycemia and lactic acidosis during the neonatal period; however, more commonly, they exhibit hepatomegaly at 3–4 months of age. Hypoglycemia and lactic acidosis can develop after a short fast, typically when infants start sleeping through the night. These children usually have doll-like facies with fat cheeks, relatively thin extremities, short stature, and a protuberant abdomen that is due to massive hepatomegaly. The kidneys are enlarged, but the spleen and heart are of normal size. The hepatocytes are distended by glycogen and fat, with large and prominent lipid vacuoles. Despite hepatomegaly, liver enzyme levels are usually normal or near normal. Easy bruising and epistaxis are associated with prolonged bleeding time as a result of impaired platelet aggregation/adhesion and/or an acquired von Willebrand-like disease. Hyperuricemia is present. Plasma lipids abnormalities includes elevation of triglycerides, total and low-density lipoprotein cholesterol, and phospholipids, compared to the low level of high density cholesterol (HDL). Type Ib patients have additional findings of neutropenia and impaired neutrophil function. Therefore, these patients are prone to recurrent bacterial infections and chronic oral and intestinal mucosal ulceration, which leads to severe diarrhea and malnutrition.

**LONG-TERM COMPLICATIONS** Gout usually becomes symptomatic at puberty as a result of long-term hyperuricemia in untreated patients. Puberty is often delayed. Some women with GSD I have polycystic ovaries and menorrhagia. Several reports of successful pregnancies suggest that fertility is not affected, although symptoms may be exacerbated due to pregnancy-related increases in renal perfusion and maternal blood volume. Secondary to lipid abnormalities, there is an increased risk of pancreatitis. Patients with GSD I may be at increased risk for

419

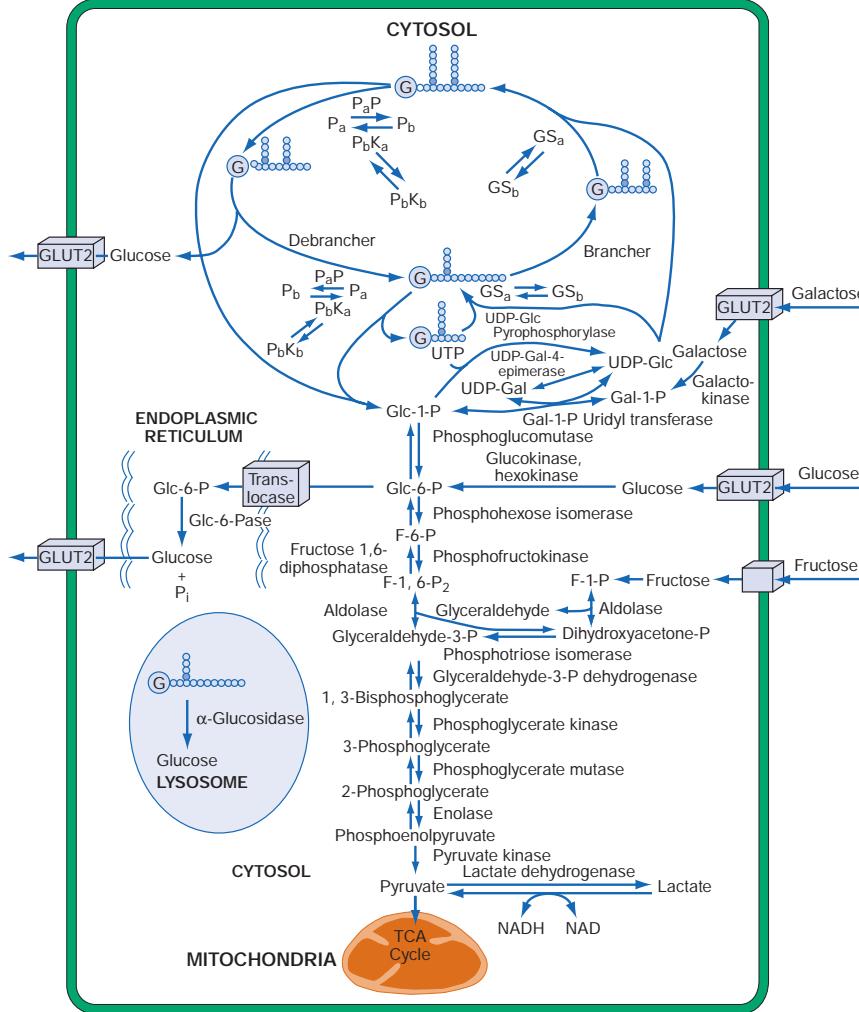
## Glycogen Storage Diseases and Other Inherited Disorders of Carbohydrate Metabolism

Priya S. Kishnani

Carbohydrate metabolism plays a vital role in cellular function by providing the energy required for most metabolic processes. The relevant biochemical pathways involved in the metabolism of these carbohydrates are shown in **Fig. 419-1**. Glucose is the principal substrate of energy metabolism in humans. Metabolism of glucose generates ATP through glycolysis and mitochondrial oxidative phosphorylation. The body obtains glucose through the ingestion of polysaccharides (primarily starch) and disaccharides (e.g., lactose, maltose, and sucrose). Galactose and fructose are two other monosaccharides that serve as sources of fuel for cellular metabolism. However, their role as fuel sources is less significant than that of glucose. Galactose is derived from lactose (galactose + glucose), which is the disaccharide found in milk products, and it is an important component of certain glycolipids, glycoproteins, and glycosaminoglycans. Fructose is found in fruits, vegetables, and honey. Sucrose (fructose + glucose) is another dietary source of fructose and is a commonly used sweetener.

Glycogen, the storage form of glucose in animal cells, is composed of glucose residues joined in straight chains by 1-4 linkages and branched at intervals of 4–10 residues by 1-6 linkages. Glycogen forms a treelike molecule and can have a molecular weight of many millions. Glycogen may aggregate to form structures recognizable by electron microscopy. Defects in glycogen metabolism typically cause an accumulation of glycogen in the tissues—hence the designation *glycogen storage diseases* (GSDs). The accumulated glycogen can be structurally normal or abnormal in the various GSDs. Defects in gluconeogenesis, glycolysis or pathways involving galactose and fructose metabolism usually do not result in glycogen accumulation.

Clinical manifestations of the various disorders of carbohydrate metabolism differ markedly. The symptoms range from minimally harmful to lethal. Unlike disorders of lipid metabolism, mucopolysaccharidoses, or other storage diseases, many disorders of carbohydrate metabolism have been managed with diet therapy. However, diet therapy alone does not prevent long-term complications, and there is a need for definitive therapies. Genes responsible for inherited defects of carbohydrate metabolism have been cloned, and pathogenic variants have been identified. With the use of tools such as DNA sequencing panels, whole exome sequencing, and whole genome sequencing, new GSDs continue to be identified, and the phenotype of known disorders



**FIGURE 419-1** Metabolic pathways related to glycogen storage diseases and galactose and fructose disorders. Nonstandard abbreviations are as follows: GS<sub>a</sub>, active glycogen synthase; GS<sub>b</sub>, inactive glycogen synthase; P<sub>a</sub>P, active phosphorylase; P<sub>b</sub>P, inactive phosphorylase; P<sub>a</sub>K<sub>a</sub>, phosphorylase  $\alpha$  phosphatase; P<sub>b</sub>K<sub>b</sub>, active phosphorylase  $\beta$  kinase; P<sub>b</sub>K<sub>b</sub>, inactive phosphorylase  $\beta$  kinase; G, glycogenin, the primer protein for glycogen synthesis. (Modified with permission from AR Beaudet, in KJ Isselbacher et al: *Harrison's Principles of Internal Medicine*, 13th ed. New York, NY: McGraw Hill; 1994.)

cardiovascular disease such as systemic hypertension. Pulmonary hypertension—although rare—has been reported. In adult patients, frequent fractures can occur, and radiographic evidence of osteopenia/osteoporosis can be found; in prepubertal patients, bone mineral content is significantly reduced. By the second or third decade of life, many patients with type I GSD develop hepatic adenomas that can hemorrhage and, in some cases, become malignant. End-stage renal disease is a serious late complication. Almost all patients aged >20 years have proteinuria, and many have hypertension, kidney stones, nephrocalcinosis, and altered creatinine clearance. In some patients, renal function deteriorates and progresses to end-stage renal disease, requiring dialysis or transplantation.

**DIAGNOSIS** Clinical presentation, hypoglycemia, lactic acidosis, hyperuricemia and abnormal lipids values suggest that a patient may have GSD I, and genetic testing provides a noninvasive means of reaching a definitive diagnosis for most patients with types Ia and Ib disease. Historically, a definitive diagnosis required a liver biopsy to demonstrate the enzyme deficiency.

**TREATMENT** The first line of treatment in GSD I is avoidance of fasting and frequent feedings. A diet high in complex carbohydrates

supplemented by uncooked cornstarch in small, frequent feedings is used for treatment of both GSD Ia and GSD Ib. Modified, extended-release cornstarch products that are longer acting and better tolerated are available, which may help extend the duration of euglycemia and improve metabolic control. Treatment of complications with medications may be necessary, such as citrate supplementation to prevent and/or treat nephrocalcinosis, allopurinol to control hyperuricemia, HMG-CoA reductase inhibitors and fibrate to reduce lipids, as well as angiotensin-converting enzyme inhibitors to treat microalbuminuria. Surgical resection, percutaneous ethanol injections, radiofrequency ablation can be used to treat liver adenoma. Liver transplantation can be lifesaving for those with hepatic adenomatous disease with the risk of malignant transformation, rapid growth in size or number, and/or severe, poor metabolic control. Kidney transplantation may be required in those with renal glomerular dysfunction progressing to renal failure. Individuals with GSD Ib may require further intervention due to the consequences of neutropenia, such as the use of granulocyte colony-stimulating factor. More recently, empagliflozin, renal glucose cotransporter sodium glucose cotransporter 2 (SGLT2), has been effectively used in GSD Ib for the treatment of neutrophil dysfunction and showed improvement of wound healing and symptoms of inflammatory bowel disease.

### Type III GSD (Debrancher Deficiency, Limit Dextrinosis)

Type III GSD is an autosomal recessive disorder caused by a deficiency of glycogen debranching enzyme. Debranching and phosphorylase enzymes are responsible for the complete degradation of glycogen into glucose. When debranching

enzyme is defective, glycogen breakdown is incomplete, resulting in abnormal glycogen accumulation with short outer chains, resembling limit dextrin. GSD III is mainly classified as (1) GSD IIIa, with liver, cardiac, and skeletal muscle involvement (~85% of cases), and (2) GSD IIIb, with primarily liver involvement (~15% of cases).

**CLINICAL AND LABORATORY FINDINGS** The initial presentation of GSD III is similar to that of GSD I with hypoglycemia, hepatomegaly, hyperlipidemia, and short stature, occurring in infancy and early childhood. Hypoglycemia in GSD III can be ketotic or non-ketotic. Patients with GSD III have elevated aminotransferase levels and normal concentrations of blood lactate and uric acid. Patients with GSD IIIa also have a variable skeletal myopathy and cardiomyopathy that can present early. Serum creatine kinase (CK) levels can sometimes be used to identify patients with muscle involvement, but normal levels do not rule out muscle enzyme deficiency. In most patients with GSD III, there is an apparent improvement in hepatomegaly with age; however, many patients present in late adulthood with progressive liver fibrosis, cirrhosis progressing to liver failure, and hepatocellular carcinoma. Hepatic adenomas may occur, although less commonly than in GSD I. Left ventricular hypertrophy, significant scarring of the myocardium, and life-threatening arrhythmias have been reported. Patients with

**TABLE 419-1 Features of Glycogen Storage Diseases and Galactose and Fructose Disorders**

Type/Common Name	Basic Defect	Clinical Features	Comments
<b>Liver Glycogenoses</b>			
<b>Disorders with Hepatomegaly and Hypoglycemia</b>			
Ia/von Gierke	Glucose-6-phosphatase	Growth retardation, enlarged liver and kidney, hypoglycemia, elevated blood lactate, cholesterol, triglycerides, and uric acid	Common, severe hypoglycemia. Complications in adulthood include hepatic adenomas, hepatic carcinoma, osteoporosis, pulmonary hypertension, and renal failure.
Ib	Glucose-6-phosphate translocase	As for Ia, with additional findings of neutropenia and neutrophil dysfunction, increased risk for infections, mucosal ulceration, and periodontal disease, inflammatory bowel disease, hypothyroidism	~20% of type I
IIIA/Cori or Forbes	Liver and muscle debranching enzyme	<i>Childhood:</i> Hepatomegaly, growth failure, muscle weakness, cardiomyopathy, cardiac arrhythmias, hypoglycemia, hyperlipidemia, elevated liver aminotransferases, CK, urinary Glc4	Common, intermediate severity of hypoglycemia, yet severe cases are seen.
		<i>Adulthood:</i> Proximal and distal muscle atrophy and weakness; peripheral neuropathy with preferential median nerve involvement; variable cardiomyopathy, liver fibrosis, cirrhosis, progressive liver failure, risk for HCC in some	Liver fibrosis/cirrhosis, hepatic adenoma and carcinoma can occur. Muscle weakness can progress to need for ambulatory aids such as wheelchair. Risk of life threatening arrhythmia.
IIIB	Liver debranching enzyme (normal muscle debrancher activity)	Liver symptoms same as in type IIIa; no muscle symptoms	~15% of type III
IV/Andersen	Branching enzyme	<i>Hepatic form:</i> Failure to thrive, hypotonia, hepatomegaly, progressive liver cirrhosis and failure (death usually before fifth year); a small subset do not have liver progression with extrahepatic involvement such as myopathy and cardiomyopathy later in life <i>Neuromuscular forms:</i> Perinatal and congenital forms lead to death in the neonatal period. Childhood form presents with myopathy, cardiomyopathy, typical systemic findings. <i>Adult form (APBD):</i> Bilateral lower limb weakness and spasticity, neurogenic bladder, peripheral neuropathy, cognitive impairment	One of the rarer glycogenoses. Other neuromuscular variants exist.
VI/Hers	Liver phosphorylase	Hepatomegaly, variable hypoglycemia, hyperlipidemia, ketosis, growth retardation, liver fibrosis, and hepatocellular carcinoma	Often underdiagnosed, severe cases being recognized
IX/liver PhK deficiency IX α2 (PHKA2) IX β (PHKB) IX γ2 (PHKG2)	Liver PhK Liver and muscle PhK Liver PhK	Hypoglycemia, hyperketosis hepatomegaly, chronic liver disease, hyperlipidemia, elevated liver enzymes, growth retardation Clinical phenotype of IX γ2 is more severe than IX α2; with significant variability among patients, marked hepatomegaly, recurrent hypoglycemia, liver cirrhosis	GSD IXα2 is a common, X-linked, (GSD IX γ2) clinical variability within and between subtypes; severe cases being recognized across different subtypes
00a/liver glycogen synthase deficiency	Glycogen synthase	Fasting hypoglycemia and ketosis, elevated lactic acid, alanine levels and hyperglycemia after glucose load, no hepatomegaly	Decreased liver glycogen stores
GSD XI/Fanconi-Bickel syndrome	Glucose transporter 2 (GLUT2)	Failure to thrive, short stature, hypophosphatemic rickets, metabolic acidosis, hepatomegaly, proximal renal tubular dysfunction, impaired glucose and galactose utilization	Rare, consanguinity in 70%
<b>Muscle Glycogenoses</b>			
<b>Disorders with Muscle-Energy Impairment</b>			
V/McArdle	Muscle phosphorylase	Exercise intolerance, muscle cramps, myoglobinuria on strenuous exercise, increased CK, "second-wind" phenomenon	Common, male predominance
VII/Tarui	Phosphofructokinase—M subunit	As for type V, with additional findings of compensated hemolysis, hyperuricemia, 'out of wind phenomena'	Prevalent in Ashkenazi Jews and Japanese
IX/muscle PhK deficiency IX α1 (PHKA1) IX γ1 (PHKG1)	Muscle PhK	Exercise intolerance, cramps, myalgia, myoglobinuria; no hepatomegaly	X-linked (PHKA1), AR (PHKG1)
X/Phosphoglycerate kinase deficiency	Phosphoglycerate kinase	As for type V, with additional findings of hemolytic anemia and CNS dysfunction	Rare, X-linked
Phosphoglycerate mutase deficiency	Phosphoglycerate mutase—M subunit	As for type V	Rare, most patients African American
Lactate dehydrogenase deficiency	Lactic acid dehydrogenase—M subunit	As for type V, with additional findings of erythematous skin eruption and uterine stiffness resulting in childbirth difficulty in females	Rare
XII/Fructose 1,6-bisphosphate aldolase A deficiency	Fructose 1,6-bisphosphate aldolase A	As for type V, with additional finding of hemolytic anemia, splenomegaly, jaundice	Rare
XIII/β-Enolase deficiency	Muscle β-enolase	Exercise intolerance	Rare

(Continued)

Type/Common Name	Basic Defect	Clinical Features	Comments
<b>Disorders with Progressive Skeletal Muscle Myopathy and/or Cardiomyopathy</b>			
II/Pompe	Lysosomal acid $\alpha$ -glucosidase	<i>Classic infantile</i> : Hypotonia, muscle weakness, cardiac enlargement and failure, fatal early. <i>Nonclassic infantile</i> : Presentation within first year of life with less severe cardiomyopathy and slower progression than the classic form. <i>Late onset (juvenile and adult)</i> : Absence of cardiomyopathy in first year of life. Progressive skeletal muscle weakness and atrophy, proximal muscles and respiratory muscles seriously affected.	Common, undetectable or very low level of enzyme activity in infantile form; variable residual enzyme activity in late-onset form
PRKAG2 deficiency	AMP-activated gamma 2 protein kinase	Severe cardiomyopathy and early heart failure (9–55 years). Congenital fetal form is rapidly fatal with hypertrophic cardiomyopathy and WPW syndrome. Other involvement includes myalgia, myopathy, and seizures.	Autosomal dominant
Danon disease	Lysosomal-associated membrane protein 2 (LAMP2)	Severe cardiomyopathy, WPW pattern, and heart failure (8–15 years); myopathy, retinopathy or maculopathy, learning disability, cognitive and attention deficits may be present.	Very rare, X-linked
XV: Late-onset polyglucosan body myopathy	Glycogenin-1	Adult-onset proximal muscle weakness, severe cardiomyopathy necessitating cardiac transplantation in some cases, nervous system involvement uncommon	Autosomal recessive, rare
<b>Galactose Disorders</b>			
Galactosemia with uridylyltransferase deficiency	Galactose 1-phosphate uridylyltransferase	Vomiting, hepatomegaly, jaundice, cataracts, amino aciduria, failure to thrive	Long-term complications exist despite early diagnosis and treatment.
Galactokinase deficiency	Galactokinase	Cataracts, neonatal bleeding diathesis, encephalopathy and high levels of liver transaminases.	Benign in some cases, more severe phenotype has been reported in others.
Uridine diphosphate galactose 4-epimerase deficiency	Uridine diphosphate galactose 4-epimerase	Similar to transferase deficiency with additional findings of hypotonia and nerve deafness	Benign variant exists.
<b>Fructose Disorders</b>			
Essential fructosuria	Fructokinase	Asymptomatic, positive urine reducing substance	Benign, autosomal recessive
Hereditary fructose intolerance	Fructose 1,6-bisphosphate aldolase B	Vomiting, lethargy, failure to thrive, hepatic failure, aversion to sweets, severity of symptoms depending on age/quantity of sugar ingested	Prognosis good with early diagnosis and fructose restriction, autosomal recessive
Fructose 1,6-diphosphatase deficiency	Fructose 1,6-diphosphatase	Episodic hypoglycemia, hyperlactic acidemia, and ketoacidosis usually following illness, hepatomegaly	Avoid fasting, good prognosis.

Abbreviations: CK, creatine kinase; CNS, central nervous system; HCC, hepatocellular carcinoma; M, muscle; PhK, phosphorylase kinase; WPW, Wolff-Parkinson-White.

GSD IIIa may experience muscle weakness in early childhood that can become severe after the third or fourth decade of life, resulting in use of assistive devices and wheelchair dependence. Patients also experience exercise intolerance. The pattern of muscle weakness is variable, and both proximal and distal muscle weakness are seen. Peripheral neuropathy may become discernible later in life; however there are opposing views concerning the existence of peripheral neuropathy in GSD III. Individuals with GSD IIIa are at an increased risk of osteoporosis. In addition, polycystic ovaries are reported in female patients with GSD III, and some patients develop features of polycystic ovarian syndrome, such as hirsutism and irregular menstrual cycles. Reports of successful pregnancy in women with GSD III suggest that fertility is normal.

**DIAGNOSIS** Hypoglycemia is a presenting symptom in only about half of patients with GSD III, and therefore the diagnosis should be considered in patients with hepatomegaly and typical biochemical parameters. In the past, the diagnosis was confirmed by deficient or absent debrancher enzyme activity in liver, skeletal muscle, or fibroblasts. In patients with GSD IIIb, enzyme activity is low in liver and normal in muscle. With the availability of molecular genetic testing, reliance on invasive tests such as liver and muscle biopsies is declining. DNA-based analyses now provide a noninvasive way of subtyping these disorders in most patients. Liver histology has distended hepatocytes due to glycogen buildup; areas of periportal fibrosis are also noted very early in the disease course along with some fat infiltration.

**TREATMENT** Debrancher enzyme deficiency prevents complete glycogenolysis in GSD III, but gluconeogenesis is intact. Hence, a high-protein diet with complex carbohydrates supplemented with

uncooked cornstarch in small, frequent feedings is effective in preventing hypoglycemia. Individuals with GSD III may benefit from dietary lipid manipulation, such as the implementation of a high-fat diet or a modified ketogenic diet or use of medium-chain triglyceride supplementation, yet careful monitoring of liver function, morphology, lipid profile and growth is necessary given the potential impact on underlying liver disease. Blood ketones and glucose should be evaluated during times of stress. Liver and heart transplantation may be considered in those with severe hepatic or cardiac involvement. Diet therapy is not effective in preventing the progression of hepatic disease, cardiomyopathy, and myopathy. Muscle disease continues to progress and is a significant unmet need for these patients.

#### Type IX GSD (Liver Phosphorylase Kinase Deficiency)

Defects of PhK cause a heterogeneous group of glycogenoses. The PhK enzyme complex consists of four subunits (α, β, γ, and δ). Each subunit is encoded by different genes (X chromosome as well as autosomes) that are differentially expressed in various tissues. PhK deficiency can be divided into several subtypes on the basis of the gene/subunit involved, the tissues primarily affected, and the mode of inheritance.

The most common subtype is GSD IX 2 an X-linked liver PhK deficiency caused by pathogenic variants in the *PHKA2* gene, which is also one of the most common liver glycogenoses. PhK activity may also be deficient in erythrocytes and leukocytes but is normal in muscle. Typically, a child between the ages of 1 and 5 years presents with growth failure and hepatomegaly. Despite delayed onset of puberty and growth continuing well into late teenage years, children typically

attain normal adult stature. Fatty liver and liver fibrosis have been identified in some patients, including children. Cholesterol, triglycerides, and liver enzymes levels are elevated. Fasting ketosis is a feature of the disease, yet is not seen in all patients. Lactic and uric acid levels are usually normal. Hypoglycemia may be mild in some but recurrent in others. Phenotypic variability is being increasingly recognized, with significant disease involvement in some cases of the X-linked form. Liver histology shows distention of hepatocytes due to excess glycogen accumulation; fibrosis is also noted. It is recommended that patients be monitored for hepatic complications with regular CT or MRI scans. Though previously thought to be a mild disease, a broad clinical spectrum of presentations is now being recognized in GSD IX, with more severe cases coming to light, even in the X-linked form. Further research is needed to completely understand the natural history and long-term complications of the X-linked subtype of liver GSD IX.

Treatment of liver GSD IX is symptom-based. Like in GSD III, gluconeogenesis is intact in GSD IX. A high-protein diet with complex carbohydrates in small, frequent feedings is effective in preventing hypoglycemia. Blood ketones and glucose should be evaluated during times of stress. Liver transplantation may be considered in those with severe hepatic involvement.

Other subtypes of type IX liver GSD include GSD IX<sup>-</sup> and GSD IX<sup>2</sup>. Additional subtypes, GSD IX<sup>1</sup> and IX<sup>1</sup>, affect only muscle and are described in a later section. GSD IX<sup>-</sup> (GSD IXb) is an autosomal recessive form of liver and muscle PhK deficiency caused by *PHKB* pathogenic variants. Patients with GSD IX<sup>-</sup> typically present with hepatomegaly. They exhibit a wide clinical spectrum and cannot be distinguished based on clinical findings alone. GSD IX<sup>2</sup> an autosomal recessive form of liver PhK deficiency, is due to *PHKG2* pathogenic variants. This is a severe form of GSD IX that often progresses to liver cirrhosis. GSD IX<sup>2</sup> typically is a more severe phenotype, when compared to GSD IX<sup>-</sup> and GSD IX<sup>1</sup>, with early liver cirrhosis and fibrosis. Previously, infants with severe isolated cardiomyopathy and low PhK activity in the heart and muscle were considered to have a subtype of GSD IX. However, there were no pathogenic variants in the genes encoding for the PhK subunits. This presentation was later considered to be a new syndrome, PRKAG2 syndrome, with a secondary decrease in PhK activity. The condition can be lethal because of massive glycogen deposition in the myocardium. Details about this condition are described under the section about PRKAG2 deficiency.

**Type IV GSD (Branching Enzyme Deficiency, Amylopectinosis, Polyglucosan Disease, or Andersen Disease)** Type IV GSD is caused by deficiency of branching enzyme leading to accumulation of an abnormal glycogen with poor solubility. The disease is clinically heterogeneous, with multisystem organ involvement, yet the primary presentation may be characterized by manifestations in either liver or muscle; thus two main types—hepatic and neuromuscular—are recognized. Individuals with the progressive hepatic form typically present in the first 18 months of life with failure to thrive, hepatosplenomegaly, and progressive liver cirrhosis leading to death in early childhood. Hypoglycemia in GSD IV is secondary to advanced liver disease and considered a late finding. Some patients may develop hepatocellular carcinoma. These patients often have extrahepatic manifestations involving the central and peripheral nervous system as well as cardiac and skeletal muscles. The neuromuscular forms of the disease have four recognized subtypes: perinatal, congenital, childhood, and adult forms. The perinatal and congenital forms are lethal, and death occurs in the neonatal period. The childhood form presents with myopathy or cardiomyopathy, with typical systemic findings. The adult form is known as adult polyglucosan body disease (APBD) and may present with systemic involvement of the central and peripheral nervous system characterized by gait abnormalities due to spastic paraparesis neurogenic bladder, peripheral neuropathy, leukodystrophy, autonomic dysfunction and cognitive impairment in the later stages of the disease. Life expectancy is shortened in APBD patients, yet there is a paucity of systematic long-term natural history studies. Definitive diagnosis of GSD IV requires demonstration of pathogenic variants in the *GBE1*

gene or branching enzyme deficiency in liver, muscle, cultured skin fibroblasts, or leukocytes.

Liver transplantation may be performed for progressive hepatic failure. Extrahepatic manifestations including cardiac and nervous system involvement may occur after transplantation. Treatment for the adult form of GSD IV includes symptomatic support for gait abnormalities and bladder dysfunction, as well as periodic monitoring to uncover any new neurologic deficits.

**Other Liver Glycogenoses with Hepatomegaly and Hypoglycemia** These disorders include hepatic phosphorylase deficiency (Hers disease, type VI) and hepatic glycogenosis with Fanconi-Bickel syndrome. Patients with GSD type VI can have growth retardation, hyperlipidemia, and hyperketosis in addition to hepatomegaly and hypoglycemia. The clinical course can vary from mild to severe. Fanconi-Bickel syndrome is caused by defects in the facilitative glucose transporter 2 (GLUT-2), which transports glucose and galactose in and out of hepatocytes, pancreatic cells, and the basolateral membranes of intestinal and renal epithelial cells. Patients with Fanconi-Bickel syndrome have increased renal clearance of glucose, amino acids, phosphate, and uric acid due to proximal renal tubular dysfunction, impaired glucose and galactose utilization, and accumulation of glycogen in liver and kidney.

## SELECTED MUSCLE GLYCOGENOSSES

### DISORDERS WITH MUSCLE-ENERGY IMPAIRMENT

**Type V GSD (Muscle Phosphorylase Deficiency, McArdle Disease)** Type V GSD is an autosomal recessive disorder caused by deficiency of muscle phosphorylase. McArdle disease is a prototypical muscle-energy disorder, as the enzyme deficiency limits ATP generation by glycogenolysis and results in glycogen accumulation.

**CLINICAL AND LABORATORY FINDINGS** There can be a broad, heterogeneous spectrum of clinical presentations with the neonatal form, which is rapidly fatal at one extreme, and the classical form with myalgia, cramps, and myoglobinuria at the other. Symptom onset as late as the eighth decade has been reported. Patients typically develop muscle stiffness, pain, and weakness induced by exercise. The degree of muscle involvement is variable among the symptomatic patients; however, the exercise intolerance typically worsens over time. Asymptomatic individuals with absent muscle phosphorylase activity have also been identified due to elevated serum CK.

Symptoms can be precipitated by (1) brief, high-intensity activity, such as sprinting or carrying heavy loads; and/or (2) less intense but sustained activity, such as climbing stairs or walking uphill. Most patients can engage in moderate exercise, such as walking on level ground, for long periods. Patients often exhibit the “second-wind” phenomenon, in which, after a short break from the initiation of strenuous physical effort, they are able to continue the activity without pain. This phenomenon is unique to GSD V and is due to the increase of blood glucose supply released from liver glycogen stores and fatty acid oxidation as exercise progresses. Although most patients experience episodic muscle pain and cramping as a result of exercise, 35% report permanent pain that seriously affects sleep and other activities. Burgundy-colored urine is reported after exercise, resulting from myoglobinuria secondary to rhabdomyolysis. Acute renal failure can result from intense myoglobinuria after vigorous exercise.

In rare cases, electromyographic findings may suggest inflammatory myopathy, a diagnosis that may be confused with polymyositis. These patients may be at risk for statin-induced myopathy and rhabdomyolysis.

At rest, the serum CK level is usually elevated; after exercise, the CK level increases even more. Exercise leads to an increase in levels of blood ammonia, inosine, hypoxanthine, and uric acid; these abnormalities reflect residues of accelerated muscle purine nucleotide recycling as a result of insufficient ATP production. NADH is underproduced during physical exertion.

**3266 DIAGNOSIS** Lack of increase in blood lactate and exaggerated blood ammonia elevations after an ischemic exercise test are indicative of a muscle glycogenosis and suggest a defect in the conversion of glycogen or glucose to lactate. This abnormal exercise response, however, can also occur with other defects in glycogenolysis or glycolysis, such as deficiency of muscle phosphofructokinase. A noninvasive, nonischemic forearm exercise test has been developed. Although this test has high sensitivity, is easy to perform, and is cost-effective, the abnormal exercise response does not exclude other muscle glycogenoses and includes some risk. The cycle test detects the hallmark heart rate observed during the second-wind phenomenon. A diagnostic confirmation is established by demonstration of pathogenic variants in the myophosphorylase gene or by enzymatic assay in muscle tissue.

Treatment for muscle phosphorylase deficiency consists of preexercise consumption of simple carbohydrates (e.g., sucrose or sports drinks) to protect muscles and improve exercise tolerance prior to the onset of the second wind. Regular exercise at moderate intensity is recommended to improve exercise capacity. Compared to patients who are physically inactive, those who are physically active are known to have improved cardiorespiratory fitness and a better long-term clinical course. Additionally, poor bone health and significantly lower lean mass have been observed in inactive patients.

#### Type IX GSD (Muscle Phosphorylase Kinase Deficiency)

GSD IX 1 and IX 1 are muscle-specific PhK deficiency caused by pathogenic variants in the *PHKA1* and *PHKG1* genes and are inherited in an X-linked and autosomal recessive manner respectively. Patients with muscle PhK deficiency present from childhood to adulthood with symptoms including exercise intolerance, cramps and myoglobinuria with exercise, fatigue, and progressive muscle weakness and atrophy. Electromyographic and forearm ischemic exercise test findings are typically normal. The heart and liver are not involved. Treatment for muscle PhK deficiency may include physical therapy and nutritional consultation to optimize glucose concentrations based on activity level.

### DISORDERS WITH PROGRESSIVE SKELETAL MUSCLE MYOPATHY AND/OR CARDIOMYOPATHY

**Pompe Disease, Type II GSD (Acid -1,4 Glucosidase Deficiency)** Pompe disease is an autosomal recessive disorder caused by a deficiency of lysosomal acid -glucosidase, an enzyme responsible for the degradation of glycogen in the lysosomes. This disease is characterized by the accumulation of glycogen in the lysosomes as opposed to accumulation in cytoplasm (as in the other glycogenoses).

**CLINICAL AND LABORATORY FINDINGS** The disorder encompasses a range of phenotypes. Each includes myopathy but differs in the age of onset, extent of organ involvement, and clinical severity. The most severe is the classic infantile form, in which infants present with cardiomyopathy at birth and develop a generalized muscle weakness with feeding difficulties, macroglossia, hepatomegaly, and congestive heart failure due to the rapidly progressive hypertrophic cardiomyopathy. Without treatment, patients with the classic infantile form do not survive beyond 2 years of life. A variant form, known as nonclassic infantile Pompe disease, also presents in the first year of life with less severe cardiomyopathy and slower disease progression. All patients with an absence of cardiomyopathy in the first year of life are considered to have the late-onset form. Young children with the late-onset form have delayed motor milestones and difficulty in walking. With disease progression, patients often develop proximal and later a distal muscle weakness, swallowing difficulties, and respiratory insufficiency. With the advent of newborn screening for Pompe disease, delayed motor milestones and other musculoskeletal findings such as scapular winging and pelvic girdle weakness are being recognized as early as the first year of life in some babies with late-onset Pompe disease.

Adults typically present between the second and seventh decades of life with slowly progressive myopathy without overt cardiac involvement. The clinical picture is dominated by slowly progressive, predominantly proximal limb girdle muscle weakness. The pelvic girdle, paraspinal muscles, and diaphragm are most seriously affected.

Respiratory symptoms include sleep apnea, sleep disordered breathing, decreased forced vital capacity, somnolence, morning headache, orthopnea, and exertional dyspnea. Respiratory failure causes significant morbidity and mortality in the late-onset form. In rare instances, patients present with respiratory insufficiency as the initial symptom. Basilar artery aneurysms and dilation of the ascending aorta have been observed in patients with Pompe disease. Ptosis, lingual weakness, hypernasality, speech difficulties, gastrointestinal dysmotility, and incontinence due to poor sphincter tone are now being recognized as part of the clinical spectrum. Small-fiber neuropathy, which presents with painful paresthesia or pins-and-needles sensations, is also seen in some patients with the late-onset form. Individuals with advanced disease often require some form of ventilatory support and are dependent on a walking aid or wheelchair.

Laboratory findings include elevated levels of serum CK, aspartate aminotransferase, alanine aminotransferase, and lactate dehydrogenase. Levels of urine glucose tetrasaccharide ( $\text{Glc}_4$ ), a breakdown product of glycogen, are elevated, especially on the severe end of the disease spectrum, and can be used as a biomarker to monitor disease progression and treatment responsiveness. In the infantile form of the disease, chest x-ray shows massive cardiomegaly, echocardiogram shows severely elevated left ventricular mass index, and electrocardiographic findings include a high-voltage QRS complex and a shortened PR interval. Muscle biopsy shows vacuoles that stain positive for glycogen; the muscle acid phosphatase level is increased, presumably from a compensatory increase of lysosomal enzymes. Electromyography reveals myopathic features, with irritability of muscle fibers and pseudomyotonic discharges, which appears early in the paraspinal muscles. Serum CK is not always elevated in adults, and depending on the muscle biopsied or tested, muscle histology or electromyography may not be abnormal.

**DIAGNOSIS** The confirmatory step for a diagnosis of Pompe disease is enzyme assay demonstrating deficient acid -glucosidase or gene sequencing with two pathogenic variants in the *GAA* gene. Enzyme activity can be measured in muscle, cultured skin fibroblasts, or blood. The latter is increasingly being used and is very reliable when performed in laboratories with experience. Prenatal diagnosis using variant analysis of DNA extracted from fetal cells obtained by amniocentesis or by measuring GAA enzyme activity in chorionic villi or amniocytes is available.

The approval of enzyme replacement therapy (ERT) with alglucosidase alfa in 2006 has changed the natural history and clinical course of Pompe disease. Children with the most severe, classic infantile form respond well to ERT and are living longer. Other adjunctive treatment options include dietary modifications, submaximal aerobic exercise, and respiratory muscle strength training. Early diagnosis with early ERT initiation is the key to treatment efficacy. Gene therapy is under early-phase clinical study as another treatment modality.

Pompe disease is now part of the recommended uniform screening panel (RUSP) for newborns in the United States, and newborn screening (NBS) has been initiated in almost half of all states. In Taiwan, where NBS for Pompe disease is performed routinely for all infants, early disease detection and treatment initiation have led to better treatment outcomes in infantile Pompe patients. Similar evidence is also emerging in the United States.

**Polyglucosan Body Myopathy-2, Type XV GSD** This is an autosomal recessive, slowly progressive skeletal myopathy caused by mutations in the *GYG1* gene blocking glycogenin-1 biosynthesis. *GYG1* pathogenic variants result in a reduced or complete absence of glycogenin-1, which impacts its autoglycosylation and/or its interaction with glycogen synthase, resulting in impaired glycogen synthesis. Affected individuals commonly present with adult-onset proximal muscle weakness prominently affecting the hip and shoulder girdles. The disease course is often progressive, with the most disabling muscle weakness found at older age. Asymmetric muscle involvement has been observed in patients with GSD XV. Individuals with pathogenic variants in the *GYG1* gene may also be identified without musculoskeletal manifestations. In these cases, cardiomyopathy and cardiac failure

necessitating cardiac transplantation may be seen. Liver manifestations have not been identified in patients with this disease.

**GSD Mimicking Hypertrophic Cardiomyopathy** Danon disease is an X-linked glycogen storage disorder caused by pathogenic variants in the *LAMP2* gene. This results in deficiency of lysosomal-associated membrane protein 2 (LAMP2), leading to defective autophagosomes-lysosomal fusion and excessive accumulation of autophagosomes in the heart and skeletal muscle. Patients present primarily with hypertrophic cardiomyopathy but can be distinguished from having the usual causes of hypertrophic cardiomyopathy by their electrophysiologic abnormalities, particularly ventricular preexcitation and conduction defects. In Danon disease, Wolff-Parkinson-White (WPW) syndrome pattern is five times greater in prevalence than in idiopathic and familial hypertrophic cardiomyopathy. Therefore, in a young male with hypertrophic cardiomyopathy, the presence of WPW pattern on electrocardiogram strongly suggests Danon disease. The onset of cardiac symptoms such as chest pain, palpitations, syncope, and cardiac arrest may occur between the ages of 8 and 15 years. Ocular manifestations are often underrecognized and include peripheral pigmentary retinopathy, lens changes, and abnormal electroretinograms. Mild learning disability and cognitive deficits have been noted, as well as speech and language delays, attention deficits, behavioral problems, and dysmetria. The prognosis for *LAMP2* deficiency is poor, with progressive end-stage heart failure early in adulthood. Female carriers can also be symptomatic. Although the disease is less severe in females, cardiomyopathy, skeletal myopathy, retinopathy and cognitive dysfunction have been described. Treatment is mainly symptomatic, and involves management of heart failure, correction of conduction abnormalities, and physical therapy, among others. Cardiac transplantation can be considered for refractory cases of heart failure. Neuropsychological evaluations and special education support may be required for those with intellectual disabilities.

**AMP-ACTIVATED PROTEIN KINASE GAMMA 2 DEFICIENCY (PRKAG2 DEFICIENCY)** AMP-activated protein kinase gamma 2 (PRKAG2) deficiency is caused by pathogenic variants in the *PRKAG2* gene, which is important in many cellular ATP metabolic pathways. Affected individuals present with cardiac abnormalities including hypertrophic cardiomyopathy and conduction system abnormalities, particularly WPW syndrome. The extent of cardiac involvement is variable and includes supraventricular tachycardia, sinus bradycardia, left ventricular dysfunction, or even sudden cardiac death in some cases. In addition to cardiac involvement, there is a broad spectrum of phenotypic presentations including myalgia, myopathy, and seizures. Other manifestations include developmental delays, hypotonia, areflexia, tremors, feeding difficulties, frequent respiratory infections, and failure to thrive. Unlike Danon disease, cardiomyopathy due to *PRKAG2* pathogenic variants is compatible with longer-term survival except for a congenital form that presents in early infancy with a rapid fatal course. *PRKAG2* syndrome should be considered as a differential diagnosis in infants presenting with severe hypertrophic cardiomyopathy. In rare instances, *PRKAG2* patients may be misdiagnosed as having infantile Pompe disease due to phenotypical similarity. Treatment is usually symptomatic and supportive, as in Danon disease. Heart transplantation has been suggested as a preventive measure for noncongenital *PRKAG2* deficiency.

## SELECTED DISORDERS OF GALACTOSE METABOLISM

“Classic” galactosemia is caused by galactose 1-phosphate uridylyltransferase (GALT) deficiency with a GALT enzyme activity that is absent or barely detectable. It is a serious disease with an incidence of 1 in 60,000 and an early onset of symptoms. The newborn infant normally receives up to 40% of caloric intake as lactose (glucose + galactose). Without the transferase, the infant is unable to metabolize galactose 1-phosphate (Fig. 419-1), which consequently accumulates, resulting in injury to parenchymal cells of the kidney, liver, and brain. After the first feeding, infants can present with vomiting, diarrhea, hypotonia, jaundice, and hepatomegaly. There is an increased risk for susceptibility to infection with gram-negative organisms, such as *Escherichia coli* neonatal sepsis

in galactosemic infants, often with the onset of sepsis preceding the diagnosis of galactosemia. Additional findings include hypoglycemia, seizures, poor weight gain, cataracts, bleeding diathesis, renal failure, cerebral edema, and neutropenia.

Widespread NBS for galactosemia has identified these infants early and allowed them to be placed on dietary restriction. Elimination of galactose from the diet reverses growth failure as well as renal and hepatic dysfunction, improving the prognosis. However, long-term developmental outcomes in classic galactosemia are poor, with a majority of patients having speech delays and learning disabilities that increase in severity with age. Impaired motor function and balance (with or without overt ataxia) is frequently seen. Ovarian failure manifesting as primary or secondary amenorrhea is seen in females, with 80–90% or more of women reporting hypergonadotropic hypogonadism. While most female patients are infertile when they reach childbearing age, a few successful pregnancies have been reported. Adults on dairy-free diets have developed cataracts, tremors, and low bone density. The treatment of galactosemia to prevent long-term complications remains a challenge.

Genotype and phenotype relationship is well established in galactosemia with the Q188R mutation in homozygosity causing the above described classical presentation. Several variants appear to be protective, particularly the *Duarte variant* (N314D) and the p.Ser135Leu variant, which is more common in the African-American population. In addition to cataract in neonatal or childhood period, galactokinase deficiency may present with neonatal bleeding diathesis, encephalopathy and high levels of liver transaminases. Intellectual disabilities and developmental delay have been described. Deficiency of *uridine diphosphate galactose 4-epimerase* can be benign when the enzyme deficiency is limited to blood cells but can be as severe as classic galactosemia when the enzyme deficiency is generalized.

## SELECTED DISORDERS OF FRUCTOSE METABOLISM

*Fructokinase* deficiency, or essential fructosemia (Fig. 419-1), causes a benign condition that is incidentally diagnosed from the presence of fructose as a reducing substance in the urine.

Deficiency of *fructose 1,6-bisphosphate aldolase* (aldolase B; hereditary fructose intolerance) is a serious disease in infants. These patients are healthy and symptom-free until fructose or sucrose (table sugar) is ingested (usually from fruit, sweetened cereal, or sucrose-containing formula). Clinical manifestations may include jaundice, hepatomegaly, vomiting, lethargy, irritability, and convulsions. The incidence of celiac disease is higher among patients with hereditary fructose intolerance (>10%) than in the general population (1–3%). Laboratory findings show prolonged clotting time, hypoalbuminemia, elevation of bilirubin and aminotransferase levels, and proximal renal tubular dysfunction. If the disease goes undiagnosed and the deleterious intake of sugar continues, hypoglycemic episodes recur, and eventually death can occur from progressive liver and renal failure. The mainstay of treatment is the elimination of all sources of sucrose, fructose, and sorbitol from the diet. Once dietary control is established, liver and kidney dysfunction improve, and catch-up growth is common; intellectual development is usually not affected. Over time, the patient's symptom intensity improves, even after fructose ingestion. The long-term prognosis is good.

*Fructose 1,6-diphosphatase* deficiency is characterized by childhood life-threatening episodes of hypoglycemia, acidosis, hyperventilation, convulsions, and coma. These episodes are often triggered by foods that contain fructose and include febrile infections and gastroenteritis when oral food intake is low. Hypoglycemic episodes can occur in neonatal period in nearly half of affected patients. Laboratory findings include low blood glucose levels, high lactate, alanine and uric acid levels, and metabolic acidosis. Renal tubular and liver functions are normal, and aversion to sweets is usually not seen, unlike hereditary fructose intolerance. Treatment of acute episodes requires the correction of hypoglycemia and acidosis by IV infusion of dextrose. Further episodes can be prevented by avoidance of fasting and elimination

**3268** of fructose and sucrose from the diet. A complex carbohydrate such as cornstarch, which provides slow and sustained levels of glucose, is useful for the long-term prevention of hypoglycemia. With proper treatment, prognosis is good, and patients who survive childhood develop normally.

## GLOBAL CONSIDERATIONS

The GSDs and other inherited disorders of carbohydrate metabolism, although individually rare, are reported in most ethnic populations. The prevalent genetic variants for each disease may vary in different ethnic populations, but clinical symptoms are remarkably similar, and treatment guidelines apply to all. Symptomatic treatment is available for these disorders, and today, advances in the field have resulted in more definitive diagnosis and treatment approaches. Availability of NBS for Pompe disease has shown that the frequency of Pompe disease is much higher than previously estimated. This has allowed for early treatment initiation and improved outcomes. NBS also mitigates the long diagnostic delays and misdiagnoses often associated with Pompe disease. The lessons learned from Pompe disease have bearing on the other GSDs.

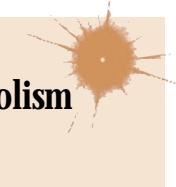
### Acknowledgment

The author is grateful to Dr. Ghada Hijazi, Dr. Aditi Korlimarla, and Cindy Li for their contributions to this chapter.

### FURTHER READING

- Fernandes SA et al: Benign or not benign? Deep phenotyping of liver glycogen storage disease IX. *Mol Genet Metab* 131:299, 2020.
- Grünert SC et al: Improved inflammatory bowel disease, wound healing and normal oxidative burst under treatment with empagliflozin in glycogen storage disease type Ib. *Orphanet J Rare Dis* 15:218, 2020.
- Kishnani PS et al: Glycogen storage disease type III diagnosis and management guidelines. *Genet Med* 12:446, 2010.
- Kishnani PS et al: The new era of Pompe disease: Advances in the detection, understanding of the phenotypic spectrum, pathophysiology, and management. *Am J Med Genet C Semin Med Genet* 160c:1, 2012.
- Kishnani PS et al: Diagnosis and management of glycogen storage disease type I: A practice guideline of the American College of Medical Genetics and Genomics. *Genet Med* 16:e1, 2014.
- Kishnani PS, Goldstein J et al: ACMG Work Group on Diagnosis and Management of Glycogen Storage Diseases Type VI and IX: Diagnosis and management of glycogen storage diseases type VI and IX: A clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet Med* 21:772, 2019.
- Kronn DF et al: Management of confirmed newborn-screened patients with Pompe disease across the disease spectrum. *Pediatrics* 40(Suppl 1):S24, 2017.
- Lévesque S et al: Diagnosis of late-onset Pompe disease and other muscle disorders by next-generation sequencing. *Orphanet J Rare Dis* 11:8, 2016.
- Lopez-Sainz A, Dominguez F et al: European Genetic Cardiomyopathies Initiative Investigators: Clinical features and natural history of PRKAG2 variant cardiac glycogenosis. *J Am Coll Cardiol* 76:186, 2020.
- Porto AG et al: Clinical spectrum of PRKAG2 syndrome. *Circ Arrhythm Electrophysiol* 9:e003121, 2016.
- Quinlivan R et al: Pharmacological and nutritional treatment for McArdle disease (glycogen storage disease type V). *Cochrane Database Syst Rev* 2014:CD003458, 2014.
- Rubio-Gozalbo ME, Derkks B et al: Galactokinase deficiency: lessons from the GalNet registry. *Genet Med* 23:202, 2021.
- Steinmann B et al: Disorders of fructose metabolism. *The Online Metabolic and Molecular Bases of Inherited Disease*. New York, McGraw-Hill, 2013.
- Welling L et al: International clinical guideline for the management of classical galactosemia: Diagnosis, treatment, and follow-up. *J Inher Metab Dis* 40:171, 2017.

Nicola Longo



Amino acids are the building blocks of proteins and serve as neurotransmitters (glycine, glutamate, -aminobutyric acid) or as precursors of hormones, coenzymes, pigments, purines, or pyrimidines. Eight amino acids, referred to as *essential* (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, valine, threonine, and tryptophan), cannot be synthesized by humans and must be obtained from dietary sources. The others can be formed endogenously. Each amino acid has a unique degradative pathway by which its nitrogen and carbon components are used for the synthesis of other amino acids, carbohydrates, and lipids. Disorders of amino acid metabolism and transport (**Chap. 421**) are individually rare—the incidences range from 1 in 10,000 for cystinuria or phenylketonuria to 1 in 200,000 for homocystinuria or alkaptunuria—but collectively, they affect perhaps 1 in 2000 newborns. Almost all are transmitted as autosomal recessive traits.

The features of inherited disorders of amino acid catabolism are summarized in **Table 420-1**. In general, these disorders are named for the compound that accumulates to highest concentration in blood (-emias) or urine (-urias). In the aminoacidopathies, the parent amino acid is found in excess, whereas products in the catabolic pathway accumulate in organic acidemias. Which compound(s) accumulates depends on the site of the enzymatic block, the reversibility of the reactions proximal to the lesion, and the availability of alternative pathways of metabolic “run-off.” Biochemical and genetic heterogeneity are common. Six distinct forms of hyperphenylalaninemia and nine forms of homocystinuria (with or without methylmalonic acidemia) are recognized. Such heterogeneity reflects the complexity of amino acid metabolism requiring multiple enzymes (gene products) for proper functioning.

The manifestations of these conditions differ widely (Table 420-1). Some, such as sarcosinemia, produce no clinical consequences. At the other extreme, complete deficiency of ornithine transcarbamylase is lethal in the untreated neonate. Central nervous system (CNS) dysfunction, in the form of developmental retardation, seizures, alterations in sensorium, or behavioral disturbances, is present in more than half the disorders. Protein-induced vomiting, neurologic dysfunction, and hyperammonemia occur in many disorders of the urea cycle. Metabolic ketoacidosis, often accompanied by hyperammonemia, is frequent in organic acidemias. Some disorders produce focal tissue or organ involvement such as liver disease, renal failure, cutaneous abnormalities, or ocular lesions.

Defects in the synthesis of nonessential amino acids (asparagine, glutamine, serine, proline) involve predominantly the brain with neurologic symptoms, with other organs occasionally affected. Dominant mutations in at least one of these genes cause tremor or spastic paraparesis in adults.

The analysis of plasma amino acids (by ion-exchange chromatography or liquid chromatography/tandem mass spectrometry), urine organic acids (by gas chromatography/mass spectrometry), and plasma acylcarnitine profile (by tandem mass spectrometry) is commonly used to diagnose and monitor most of these disorders. The diagnosis is confirmed by enzyme assay on cells or tissues from the patients or, more commonly, by DNA testing. The clinical manifestations in many of these conditions can be prevented or mitigated if a diagnosis is achieved early and appropriate treatment (e.g., dietary protein or amino acid restriction or vitamin supplementation) is instituted promptly. For this reason, newborn screening programs seek to identify several of these disorders. Infants with a positive screening test need additional metabolic testing (usually suggested by the newborn screening program) to confirm or exclude the diagnosis. Confirmed cases should be referred to a metabolic center for initiation of therapy.

**TABLE 420-1** Inherited Disorders of Amino Acid Metabolism

Amino Acid(s)	Condition	Enzyme Defect	Clinical Findings	Inheritance
Phenylalanine	Phenylketonuria	Phenylalanine hydroxylase	Intellectual disability, microcephaly, hypopigmented skin and hairs, eczema, "mousy" odor	AR
	DHPR deficiency	Dihydropteridine reductase	Intellectual disability, hypotonia, spasticity, myoclonus	AR
	PTPS deficiency	6-Pyruvoyl-tetrahydropterin synthase	Dystonia, neurologic deterioration, seizures, intellectual disability	AR
	GTP cyclohydrolase I deficiency	GTP cyclohydrolase I	Intellectual disability, seizures, dystonia, temperature instability	AR
	Carbinolamine dehydratase deficiency	Pterin-4 $\alpha$ -carbinolamine dehydratase	Transient hyperphenylalaninemia (benign)	AR
	DNAJC12 deficiency	Hydroxylase co-chaperone	Dystonia, parkinsonism, intellectual disability	AR
Tyrosine	Tyrosinemia type I (hepatorenal)	Fumarylacetoacetate hydrolase	Liver failure, cirrhosis, rickets, failure to thrive, peripheral neuropathy, "boiled cabbage" odor	AR
	Tyrosinemia type II (oculocutaneous)	Tyrosine transaminase	Palmoplantar keratosis, painful corneal erosions with photophobia, learning disability	AR
	Tyrosinemia type III	4-Hydroxyphenylpyruvate dioxygenase	Hypertyrosinemia with normal liver function, occasional mental delay	AR
	Hawkinsinuria	4-Hydroxyphenylpyruvate dioxygenase	Transient failure to thrive, metabolic acidosis in infancy	AD
	Alkaptonuria	Homogentisic acid oxidase	Ochronosis, arthritis, cardiac valve involvement, coronary artery calcification	AR
	Maleylacetoacetate isomerase deficiency	Maleylacetoacetate isomerase	No clinical symptoms, elevated succinylacetone in blood and urine.	AR
	Albinism (oculocutaneous)	Tyrosinase	Hypopigmentation of hair, skin, and optic fundus; visual loss; photophobia	AR
	Albinism (ocular)	Different enzymes or transporters	Hypopigmentation of optic fundus, visual loss	AR, XL
	DOPA-responsive dystonia	Tyrosine hydroxylase	Rigidity, truncal hypotonia, tremor, intellectual disability	AR
GABA	4-Hydroxybutyric aciduria	Succinic semialdehyde dehydrogenase	Seizures, intellectual disability, hypotonia	AR
	ABAT deficiency	GABA transaminase	Seizures, intellectual disability, hypotonia	AR
Tryptophan	Hydroxylkynureninuria	Kynureninase	Intellectual disability, spasticity	AR
Histidine	Histidinemia	Histidine-ammonia lyase	Benign	AR
	Urocanic aciduria	Urocanase	Occasional intellectual disability	AR
	Formiminoglutamic aciduria	Formiminotransferase	Occasional intellectual disability	AR
Glycine	Glycine encephalopathy	Glycine cleavage (4 enzymes)	Infantile seizures, lethargy, apnea, profound intellectual disability	AR
	Sarcosinemia	Sarcosine dehydrogenase	Benign	AR
	Hyperoxaluria type I	Alanine:glyoxylate aminotransferase	Calcium oxalate nephrolithiasis, renal failure	AR
	Hyperoxaluria type II	D-Glyceric acid dehydrogenase/glyoxylate reductase	Calcium oxalate nephrolithiasis, renal failure	AR
Serine	3-PGDH deficiency	Phosphoglycerate dehydrogenase	Seizures, microcephaly, intellectual disability	AR
	PSAT1 deficiency	Phosphoserine aminotransferase	Seizures, microcephaly, intellectual disability	AR
	PSP deficiency	Phosphoserine phosphatase	Seizures, microcephaly, intellectual disability	AR
Proline	Hyperprolinemia type I	Proline oxidase	Benign	AR
	Hyperprolinemia type II	$^{1\text{-}}$ Pyrroline-5-carboxylate dehydrogenase	Febrile seizures, intellectual disability	AR
	Hyperhydroxyprolinemia	Hydroxyproline oxidase	Benign	AR
	Polidase deficiency	Polidase	Mild intellectual disability, chronic dermatitis	AR
	PYCR1 deficiency	Pyrroline-5-carboxylate reductase 1	Wrinkly skin, joint laxity, typical facial features, intellectual disability, osteopenia, intrauterine growth retardation, hypotonia	AR
	PYCR2 deficiency	Pyrroline-5-carboxylate reductase 2	Microcephaly, hypomyelination, and reduced cerebral white matter volume, failure to thrive, intellectual disability, movement disorders, seizures	AR
Proline (ornithine, arginine, citrulline)	$^{1\text{-}}$ Pyrroline-5-carboxylate synthase deficiency	$^{1\text{-}}$ Pyrroline-5-carboxylate synthase	Hypotonia, seizures, neurodegeneration, peripheral neuropathy, joint laxity, skin hyperelasticity, subcapsular cataracts, hyperammonemia, adult spastic paraparesis (AD)	AR, AD
Methionine	Hypermethioninemia	Methionine adenosyltransferase	Usually benign	AR
	S-Adenosylhomocysteine hydrolase deficiency	S-Adenosylhomocysteine hydrolase	Hypotonia, intellectual disability, absent tendon reflexes, delayed myelination	AR
	Glycine N-methyltransferase deficiency	Glycine N-methyltransferase	Elevated liver transaminases	AR
	Adenosine kinase deficiency	Adenosine kinase	Intellectual disability, seizures, liver dysfunction	AR

(Continued)

**TABLE 420-1** Inherited Disorders of Amino Acid Metabolism (Continued)

AMINO ACID(S)	CONDITION	ENZYME DEFECT	CLINICAL FINDINGS	INHERITANCE
Homocysteine	Homocystinuria	Cystathione $\beta$ -synthase	Lens dislocation, thrombotic vascular disease, intellectual disability, osteoporosis	AR
	Homocystinuria	5,10-Methylenetetrahydrofolate reductase	Intellectual disability, gait and psychiatric abnormalities, recurrent strokes	AR
	Homocystinuria	Methionine synthase (cblE, G)	Intellectual disability, hypotonia, seizures, megaloblastic anemia	AR
	Homocystinuria and methylmalonic acidemia	Vitamin B <sub>12</sub> lysosomal efflux and metabolism (cblC, -D, -F, -J, -X)	Intellectual disability, lethargy, failure to thrive, hypotonia, seizures, megaloblastic anemia	AR, XL
Cystathione	Cystathioninuria	$\beta$ -Cystathioninase	Benign	AR
Cysteine	Sulfocystinuria	Sulfite oxidase or molybdenum cofactor deficiency	Seizures, intellectual disability, dislocated lenses	AR
Lysine	Hyperlysinemia, saccharopinuria	$\alpha$ -Ketoacidipic semialdehyde synthase	Benign	AR
	Pyridoxine-dependent seizures	L- <sup>1</sup> -Piperideine-6-carboxilate dehydrogenase	Seizures, intellectual disability	AR
Lysine, tryptophan	$\alpha$ -Ketoacidipic acidemia	$\alpha$ -Ketoacidipic acid dehydrogenase DHTKD1	Benign	AR
Lysine, tryptophan	Glutaric acidemia type I	Glutaryl-CoA dehydrogenase	Progressive severe dystonia and athetosis, motor delays	AR
Lysine, tryptophan	Glutaric acidemia type II	Electron transfer flavoproteins (ETF) or ETF:ubiquinone oxidoreductase	Hypoglycemia, metabolic acidosis, "sweaty feet" odor, hypotonia, cardiomyopathy	AR
Ornithine	Gyrate atrophy of the choroid and retina	Ornithine- $\alpha$ -aminotransferase	Myopia, night blindness, loss of peripheral vision, cataracts, chorioretinal degeneration	AR
Urea cycle	Carbamoylphosphate synthase-1 deficiency	Carbamoylphosphate synthase-1	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia	AR
	N-Acetylglutamate synthase deficiency	N-Acetylglutamate synthase	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia	AR
	Ornithine transcarbamylase deficiency	Ornithine transcarbamylase	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia	XL
	Citrullinemia type I	Argininosuccinate synthase	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia, liver failure	AR
	Argininosuccinic acidemia	Argininosuccinate lyase	Lethargy progressing to coma, protein aversion, intellectual disability, hyperammonemia, trichorrhexis nodosa	AR
	Arginase deficiency	Arginase	Spastic tetraparesis, microcephaly, intellectual disability, mild hyperammonemia	AR
	Hyperornithinemia, hyperammonemia, homocitrullinuria	Mitochondrial ornithine carrier ORNT1	Vomiting, lethargy, failure to thrive, intellectual disability, episodic confusion, hyperammonemia, protein intolerance	AR
	Citrullinemia type 2	Mitochondrial aspartate/glutamate carrier CTLN2	Neonatal intrahepatic cholestasis, adult presentation with sudden behavioral changes and stupor, coma, hyperammonemia	AR
Glutamine	Glutamine synthetase deficiency	Glutamine synthase	Brain malformations, pachygryria, seizures, hypotonia, intellectual disability, dysmorphic features, low glutamine	AR
	Glutaminase deficiency	Glutaminase	Epileptic encephalopathy, intellectual disability, ataxia, elevated glutamine	AR
Asparagine	Asparagine synthetase deficiency	Asparagine synthase	Epileptic encephalopathy, seizures, microcephaly, simplified gyration pattern, hypotonia, tetraplegia, intellectual disability	
Valine	Hypervalinemia	Branched chain aminotransferase-2	Headache, memory impairment, failure to thrive, hypotonia, developmental delays	AR
	Isobutyryl-CoA dehydrogenase deficiency	Isobutyryl-CoA dehydrogenase	Benign	AR
Isoleucine, leucine, valine	Maple syrup urine disease	Branched chain ketoacid dehydrogenase (E1 $\alpha$ , E1 $\beta$ , E2, E3 deficiency)	Lethargy, vomiting, encephalopathy, seizures, intellectual disability, "maple syrup" odor, protein intolerance	AR
Leucine	Isovaleric acidemia	Isovaleryl-CoA dehydrogenase	Acidosis, ketosis, vomiting, coma, hyperammonemia, "sweaty feet" odor, protein intolerance	AR
	3-Methylcrotonyl glycineuria	3-Methylcrotonyl-CoA carboxylase	Stress-induced metabolic acidosis, hypotonia, hypoglycemia, "cat's urine" odor	AR
	3-Methylglutaconic aciduria type I	3-Methylglutaconyl-CoA hydratase deficiency	Stress-induced acidosis, leukodystrophy, hypotonia, hepatomegaly	AR
	3-Hydroxy-3-methylglutaric aciduria	3-Hydroxy-3-methylglutaryl-CoA lyase	Stress-induced hypoketotic hypoglycemia and acidosis, encephalopathy, hyperammonemia	AR

(Continued)

**TABLE 420-1** Inherited Disorders of Amino Acid Metabolism (Continued)

Amino Acid(s)	Condition	Enzyme Defect	Clinical Findings	Inheritance
Isoleucine	2-Methylbutyryl-glycinuria	2-Methylbutyryl-CoA dehydrogenase	Benign	AR
	2-Methyl-3-hydroxybutyryl-CoA dehydrogenase deficiency	2-Methyl-3-hydroxybutyryl-CoA dehydrogenase	Developmental regression, seizures, and rigidity sometimes triggered by illnesses	XL
	3-Oxothiolase deficiency	3-Oxothiolase	Fasting-induced acidosis and ketosis, vomiting, lethargy	AR
Isoleucine, methionine, threonine, valine	Propionic acidemia (pccA, -B, -C)	Propionyl-CoA carboxylase	Metabolic ketoacidosis, hyperammonemia, hypotonia, lethargy, coma, protein intolerance, intellectual disability, hyperglycinemia	AR
	Multiple carboxylase/biotinidase deficiency	Holocarboxylase synthase or biotinidase	Metabolic ketoacidosis, diffuse rash, alopecia, seizures, intellectual disability	AR
	Methylmalonic acidemia (mutase, cblA, B, racemase)	Methylmalonyl-CoA mutase/racemase or cobalamin reductase/adenosyltransferase	Metabolic ketoacidosis, hyperammonemia, hypotonia, lethargy, coma, protein intolerance, intellectual disability, hyperglycinemia	AR

Abbreviations: AD, autosomal dominant; AR, autosomal recessive; Cbl, cobalamin; DOPA, dihydroxyphenylalanine; GABA,  $\gamma$ -aminobutyric acid; GTP, guanosine 5'-triphosphate; XL, X-linked.

The parents need to be counseled about the natural history of the disease and its recurrence risk in future pregnancies. In some cases, parents need testing because they might have a disorder themselves (such as glutaric acidemia type 1, methylcrotonyl coenzyme A carboxylase deficiency, primary carnitine deficiency, or fatty acid oxidation defects) since mothers with these conditions can sometimes be identified by abnormal newborn screening results in their offspring. Some metabolic disorders can remain asymptomatic until adult age, presenting only when fasting or severe stress requires full activity of affected metabolic pathways to provide energy.

Selected disorders that illustrate the principles, properties, and problems presented by the disorders of amino acid metabolism are discussed in this chapter.

## THE HYPERPHENYLALANINEMIAS

The hyperphenylalaninemias (Table 420-1) result from impaired conversion of phenylalanine to tyrosine. The most common and clinically important is *phenylketonuria* (frequency 1:16,500), which is an autosomal recessive disorder characterized by an increased concentration of phenylalanine and its by-products in body fluids and by severe intellectual disability if untreated in infancy. It results from reduced activity of phenylalanine hydroxylase. The accumulation of phenylalanine inhibits the transport of other amino acids required for protein or neurotransmitter synthesis, reduces synthesis and increases degradation of myelin, and leads to inadequate formation of norepinephrine and serotonin. Phenylalanine is a competitive inhibitor of tyrosinase, a key enzyme in the pathway of melanin synthesis, resulting in hypopigmentation of hair and skin. Untreated children with classic phenylketonuria are normal at birth but fail to attain early developmental milestones, develop microcephaly, and demonstrate progressive impairment of cerebral function. Hyperactivity, seizures, and severe intellectual disability are major clinical problems later in life. Electrocerebral abnormalities; “mousy” odor of skin, hair, and urine (due to phenylacetate accumulation); and a tendency to develop hypopigmentation (compared to the family background) and eczema complete the devastating clinical picture. In contrast, affected children who are detected and treated at birth show none of these abnormalities.

## TREATMENT

### Phenylketonuria

To prevent intellectual disability, diagnosis and initiation of dietary treatment of classic phenylketonuria must occur before the child is 2 weeks of age. For this reason, newborns in North America, Australia, and Europe are screened by determinations of blood phenylalanine levels. Abnormal values are confirmed using quantitative analysis of plasma amino acids. Dietary phenylalanine restriction

is usually instituted if blood phenylalanine levels are >360  $\mu$ mol/L. Treatment consists of a special diet low in phenylalanine and supplemented with tyrosine since tyrosine becomes an essential amino acid in phenylalanine hydroxylase deficiency. With therapy, plasma phenylalanine concentrations should be maintained between 120 and 360  $\mu$ mol/L. Dietary restriction should be continued and monitored indefinitely. Compliance with the strict diet is often difficult as patients become older; increased levels of phenylalanine in adults can cause deficits in executive function or psychiatric symptoms. Oral tetrahydrobiopterin (5–20 mg/kg per d), an essential cofactor of phenylalanine hydroxylase, can reduce phenylalanine levels in some patients with phenylketonuria in conjunction with a low-protein diet. Pegvaliase is a pegylated form of phenylalanine ammonia lyase, a bacterial enzyme that converts phenylalanine to trans-cinnamic acid and ammonia. This injectable drug can substantially reduce phenylalanine levels, allowing a normal diet. The bacterial origin of pegvaliase can cause immune reactions that limit its use in some patients with phenylketonuria.

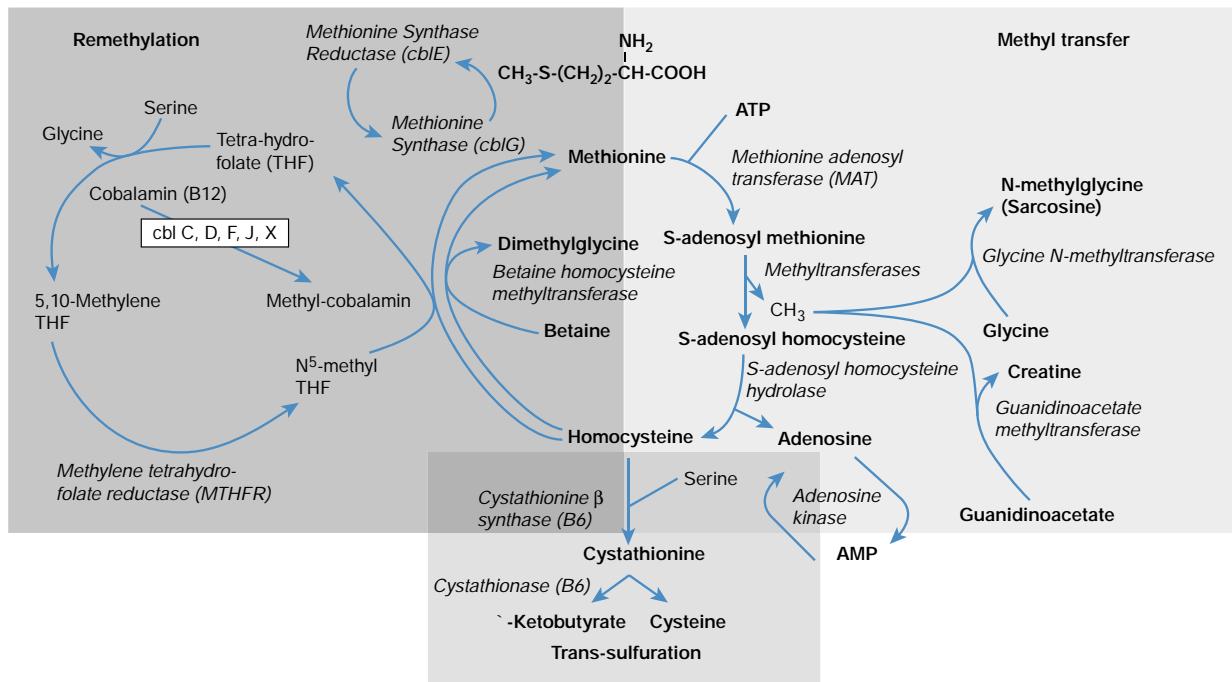
Women with phenylketonuria can become pregnant. If maternal phenylalanine levels are not strictly controlled before and during pregnancy, their offspring are at increased risk for congenital defects and microcephaly (*maternal phenylketonuria*). After birth, these children have severe intellectual disability and growth retardation. Pregnancy risks can be minimized by continuing lifelong phenylalanine-restricted diets and assuring strict phenylalanine restriction 2 months prior to conception and throughout gestation.

## THE HOMOCYSTINURIAS (HYPERHOMOCYSTEINEMIAS)

The homocystinurias are nine biochemically and clinically distinct disorders (Table 420-1) characterized by increased concentration of the sulfur-containing amino acid homocysteine in blood and urine.

Classic homocystinuria, the most common (frequency 1:450,000), results from reduced activity of cystathione  $\beta$ -synthase (Fig. 420-1), the pyridoxal phosphate-dependent enzyme that condenses homocysteine with serine to form cystathione. Most patients present between 3 and 5 years of age with dislocated optic lenses and intellectual disability (in about half of cases). Some patients develop a marfanoid habitus and radiologic evidence of osteoporosis.

Life-threatening vascular complications (affecting coronary, renal, and cerebral arteries) can occur during the first decade of life and are the major cause of morbidity and mortality. Classic homocystinuria can be diagnosed with analysis of plasma amino acids, showing elevated methionine and presence of free homocystine. Total plasma homocysteine is also extremely elevated (usually >100  $\mu$ M). Elevated levels of methionine can be also detected by neonatal screening, but milder variants can be missed by this approach. Treatment consists of a



**FIGURE 420-1** Pathways, enzymes, and coenzymes involved in the homocystinurias. Methionine transfers a methyl group during its conversion to homocysteine. Defects in methyl transfer or in the subsequent metabolism of homocysteine by the pyridoxal phosphate (vitamin B<sub>6</sub>)-dependent cystathione  $\beta$ -synthase increase plasma methionine levels. Homocysteine is transformed into methionine via remethylation. This occurs through methionine synthase, a reaction requiring methylcobalamin and folic acid. Deficiencies in these enzymes or lack of cofactors is associated with decreased or normal methionine levels. In an alternative pathway, homocysteine can be remethylated by betaine:homocysteine methyl transferase.

special diet restricted in protein and methionine. In approximately half of patients, oral pyridoxine (25–500 mg/d) produces a fall in plasma methionine and homocysteine concentration in body fluids. Folate and vitamin B<sub>12</sub> deficiency should be prevented by adequate supplementation. Betaine is also effective in reducing homocysteine levels by favoring its remethylation to methionine.

The other forms of homocystinuria are the result of impaired remethylation of homocysteine to methionine. This can be caused by defective methionine synthase or reduced availability of two essential cofactors, 5-methyltetrahydrofolate and methylcobalamin (methyl-vitamin B<sub>12</sub>). In contrast to cystathione  $\beta$ -synthase, elevated levels of free homocystine are associated with low levels of methionine in the plasma amino acid profile in remethylation defects. Therapy in these cases requires administration of methylfolate, hydroxycobalamin (an activated form of vitamin B<sub>12</sub>), and betaine.

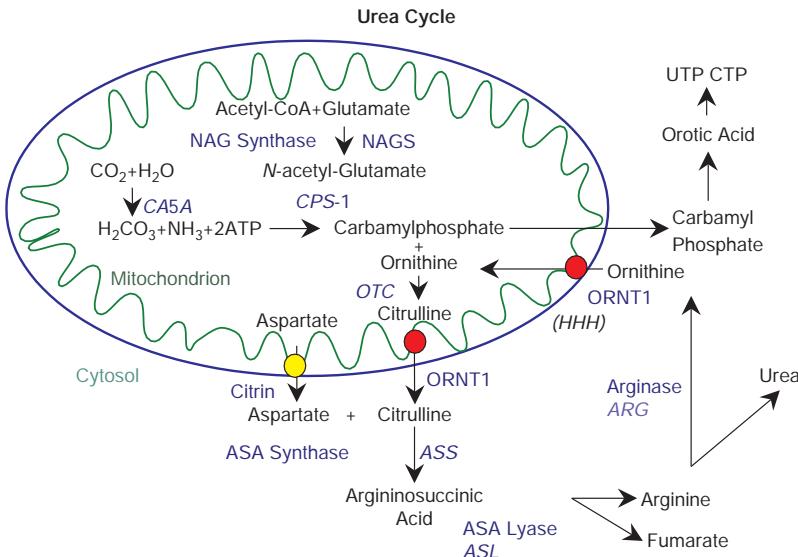
*Hyperhomocysteinemia* refers to increased total plasma concentration of homocysteine with or without an increase in free homocysteine (disulfide form). Hyperhomocysteinemia, in the absence of significant homocystinuria, is found in some heterozygotes for the genetic defects noted above or in homozygotes for milder variants. Changes of homocysteine levels are also observed with increasing age; with smoking; in postmenopausal women; in patients with renal failure, hypothyroidism, leukemias, inflammatory bowel disease, or psoriasis; and during therapy with drugs such as methotrexate, nitrous oxide, isoniazid, and some antiepileptic agents. Homocysteine can act as an atherogenic and thrombophilic agent, and increased total plasma homocysteine has been associated with an increased risk for coronary, cerebrovascular, and peripheral arterial disease as well as for deep-vein thrombosis. In addition, hyperhomocysteinemia and folate and vitamin B<sub>12</sub> deficiencies have been associated with an increased risk of neural tube defects in pregnant women and dementia (Alzheimer's type) in the general population. Vitamin supplements are effective in reducing plasma homocysteine levels in these cases, although there are limited effects on cardiovascular disease.

## ALKAPTONURIA

Alkaptonuria is a rare (frequency 1:200,000) disorder of tyrosine catabolism in which deficiency of homogentisate 1,2-dioxygenase (also known as *homogentisic acid oxidase*) leads to excretion of large amounts of homogentisic acid in urine and accumulation of oxidized homogentisic acid pigment in connective tissues (*ochronosis*). Alkaptonuria may go unrecognized until middle life, when degenerative joint disease develops. Prior to this time, about half of patients might be diagnosed for the presence of urine that becomes dark with standing or addition of alkali. Foci of gray-brown scleral pigment and generalized darkening of the concha, anthelix, and, finally, helix of the ear usually develop after age 30. Low back pain usually starts between 30 and 40 years of age. *Ochronotic arthritis* is heralded by pain, stiffness, and some limitation of motion of the hips, knees, and shoulders. Acute arthritis may resemble rheumatoid arthritis, but small joints are usually spared. Pigmentation of heart valves, larynx, tympanic membranes, and skin occurs, and occasional patients develop pigmented renal or prostatic calculi. Pigment deposition in the heart and blood vessels leads to aortic stenosis necessitating valve replacement, especially after 60 years of age. The diagnosis should be suspected in a patient whose urine darkens to blackness. Homogentisic acid in urine is identified by urine organic acid analysis. Ochronotic arthritis is treated symptomatically with pain medications, spinal surgery, and arthroplasty (Chap. 371). Nitisinone (2-[2-nitro-4-trifluoromethylbenzoyl]-1,3-cyclohexanedione), a drug used in tyrosinemia type I, at low dose (10 mg/d) reduces urinary excretion of homogentisic acid and delays progression and improves clinical signs of alkaptonuria.

## UREA CYCLE DEFECTS

Excess ammonia generated from protein nitrogen is removed by the urea cycle, a process mediated by several enzymes and transporters (Fig. 420-2, Table 420-1). Complete absence of any of these enzymes usually causes severe hyperammonemia in newborns, while milder



**FIGURE 420-2 The urea cycle.** This cycle, which is fully expressed only in the liver, forms urea starting from ammonia ( $\text{NH}_3$ ) derived from the nitrogen group of all amino acids. It requires many enzymes and mitochondrial transporters, any of which can be defective and may impair the function of the urea cycle. Ammonia escaping the urea cycle in peripheral hepatocytes is conjugated with glutamate by glutamine synthase in perivenous hepatocytes to generate glutamine. ARG, arginase; ASA, argininosuccinic acid; ASL, argininosuccinate lyase; ASS, argininosuccinate synthase; CA5A, carbonic anhydrase 5a; citrin (*SLC25A13*), aspartate/glutamate exchanger; CP, carbamylphosphate; CPS-1, carbamylphosphate synthase 1; CTP, cytidine triphosphate; HHH, hyperammonemia, hyperornithinemia, homocitrullinuria syndrome; NAG, *N*-acetylglutamate; NAGS, *N*-acetylglutamate synthase; ORNT1 (*SLC25A15*), ornithine/citrulline mitochondrial transporter; OTC, ornithine transcarbamylase; UTP, uridine triphosphate.

variants can be seen in adults. The accumulation of ammonia and glutamine leads to direct neuronal toxicity and brain edema. Deficiencies in urea cycle enzymes are individually rare, but as a group, they affect ~1:35,000 individuals. They are all transmitted as autosomal recessive traits, with the exception of ornithine transcarbamylase deficiency, which is X-linked and the most frequent urea cycle defect. Hepatocytes of females with ornithine transcarbamylase deficiency express either the normal or the mutant allele due to random X-inactivation and may be unable to remove excess ammonia if mutant cells are predominant.

Infants with classic urea cycle defects present at 1-4 days of life with refusal to eat and lethargy progressing to coma and death. Milder enzyme deficiencies present with protein avoidance, recurrent vomiting, migraine, mood swings, chronic fatigue, irritability, and disorientation that can progress to coma. Some cases have presented with acute or chronic hepatic dysfunction. Females with ornithine transcarbamylase deficiency can present at time of childbirth due to the combination of involuntary fasting and stress that favors catabolism. Administration of systemic corticosteroids or chemotherapy can precipitate hyperammonemia and can be fatal in previously asymptomatic individuals of any age. These patients may be misdiagnosed as having gastrointestinal disorders, food allergies, behavioral problems, or nonspecific hepatitis. The diagnosis requires measurement of plasma ammonia, plasma amino acids, and urine orotic acid, useful for differentiating ornithine transcarbamylase deficiency from carbamyl phosphate synthase-1 and *N*-acetylglutamate synthase deficiency. Increased plasma glutamine is seen with all urea cycle defects since ammonia not removed by the urea cycle in periportal hepatocytes is conjugated to glutamate by glutamine synthase in perivenous hepatocytes. Citrulline is low or undetectable in proximal defects of the urea cycle (*N*-acetylglutamate synthase, carbamylphosphate synthase 1, and ornithine transcarbamylase deficiency), with urine orotic acid being increased only in ornithine transcarbamylase deficiency. Plasma citrulline is markedly increased in argininosuccinic acid synthase deficiency (citrullinemia type 1), with a milder elevation in argininosuccinic acid lyase deficiency in the presence of argininosuccinic acid (argininosuccinic aciduria). Arginine levels are usually normal to low in these conditions and become markedly elevated only in patients with arginase deficiency. In addition to urea cycle defects, hyperammonemia can also be caused by liver disease from any cause and several organic acidemias and fatty

acid oxidation defects (the latter two excluded by the analysis of urine organic acids and plasma acylcarnitine profile).

## TREATMENT

## Urea Cycle Defects

Therapy is aimed at stopping catabolism and ammonia production by providing adequate calories (as IV glucose and lipids in the comatose patient) and, if needed, insulin. Excess nitrogen is removed by IV phenylacetate and benzoate (0.25 g/kg for the priming dose and subsequently as an infusion over 24 h) that conjugate with glutamine and glycine, respectively, to form phenylacetylglutamine and hippuric acid, water-soluble molecules efficiently excreted in urine. Arginine (200 mg/kg per d) becomes an essential amino acid (except in arginase deficiency) and should be provided intravenously to resume protein synthesis. If these measures fail to reduce ammonia, hemodialysis should be initiated promptly. Chronic therapy consists of a protein-restricted diet, phenylbutyrate, glycerol phenylbutyrate (a liquid drug better tolerated by most patients), arginine, or citrulline supplements, depending on the specific diagnosis. Oral carbamylc acid can restore a functional urea cycle in patients with *N*-acetylglutamate synthase deficiency and renders other therapies unnecessary. Liver transplantation should be considered in patients with severe urea cycle defects that are difficult to control medically.

Hyperammonemia due to a functional deficiency of glutamine synthase can occur in patients receiving chemotherapy for different malignancies or undergoing solid organ transplants. It can also be seen with hepatic cirrhosis. Several of these patients have been successfully rescued from hyperammonemia using the protocol described above for urea cycle defects.

## FURTHER READING

- Morris AA et al: Guidelines for the diagnosis and management of cystathione beta-synthase deficiency. *J Inherit Metab Dis* 40:49, 2017.

Ranganath LR et al: Efficacy and safety of once-daily nitisinone for patients with alkaptonuria (SONIA 2): An international, multicentre,

open-label, randomised controlled trial. *Lancet Diabetes Endocrinol* 8:762, 2020.

Smith AD, Refsum H: Homocysteine, B vitamins, and cognitive impairment. *Annu Rev Nutr* 36:211, 2016.

Zori R et al: Long-term comparative effectiveness of pegvaliase versus standard of care comparators in adults with phenylketonuria. *Mol Genet Metab* 128:92, 2019.

## 421

# Inherited Defects of Membrane Transport

Nicola Longo



Specific membrane transporters mediate the passage of amino acids, oligopeptides, sugars, cations, anions, vitamins, water, and many other molecules across cellular membranes. They are encoded by members of the solute-carrier gene (SLC) superfamily. These transporters are located on the plasma membrane or intracellular organelles, and their cellular and tissue distribution in addition to the presence (or absence) of redundant transporters explains organ involvement and possible metabolic disturbances. The first transport disorders identified affected the gut or the kidney, but transport processes are essential for the normal function of every organ, but especially the brain and sensory organs (**Table 421-1**). Inherited defects impairing the transport of selected amino acids that can present in adults are discussed here as examples of the abnormalities encountered; others are considered elsewhere in this text.

### CYSTINURIA

Cystinuria (worldwide frequency of 1 in 7000) is an autosomal recessive disorder caused by defective transporters in the apical brush border of proximal renal tubule and small intestinal cells. It is characterized by impaired reabsorption and excessive urinary excretion of the dibasic amino acids lysine, arginine, ornithine, and cystine. Because cystine is poorly soluble, its excess excretion predisposes to the formation of renal, ureteral, and bladder stones. Such stones are responsible for the signs and symptoms of the disorder.

There are two variants of cystinuria. Homozygotes for both variants have high urinary excretion of cystine, lysine, arginine, and ornithine. Type A heterozygotes usually have normal urinary amino acid excretion, whereas most type B heterozygotes have moderately increased urinary excretion of cystine that under some circumstances can result in the formation of kidney stones. The gene for type A cystinuria (*SLC3A1*, chromosome 2p16.3) encodes a membrane glycoprotein. Type B cystinuria is caused by mutations in *SLC7A9* (chromosome 19q13) that encodes the  $b^{0+}$  amino acid transporter. The glycoprotein encoded by *SLC3A1* favors the correct processing of the  $b^{0+}$  membrane transporter and explains why mutations in two different genes cause a similar disease.

Cystine stones account for 1–2% of all urinary tract calculi and for ~4–5% of stones in children. Cystinuria homozygotes regularly excrete 2400–7200  $\mu\text{mol}$  (600–1800 mg) of cystine daily. Since the maximum solubility of cystine in the physiologic urinary pH range of 4.5–7.0 is ~1200  $\mu\text{mol/L}$  (300 mg/L), cystine needs to be diluted to 2.5–7 L of water to prevent crystalluria. Stone formation usually manifests in the second or third decade but may occur in the first year of life. Symptoms and signs are those typical of urolithiasis: hematuria, flank pain, renal colic, obstructive uropathy, and infection (**Chap. 318**). Recurrent urolithiasis may lead to progressive renal insufficiency.

Cystinuria is suspected after observing typical hexagonal crystals in the sediment of acidified, concentrated, chilled urine or after performing a urinary nitroprusside test. Quantitative urine amino acid analysis

confirms the diagnosis of cystinuria by showing selective overexcretion of cystine, lysine, arginine, and ornithine. Quantitative measurements are important for differentiating heterozygotes from homozygotes and for following free cystine excretion during therapy.

Management is aimed at preventing cystine crystal formation by increasing urinary volume and by maintaining an alkaline urine pH. Fluid ingestion in excess of 4 L/d is essential, and 5–7 L/d is optimal. Urinary cystine concentration should be <1000  $\mu\text{mol/L}$  (250 mg/L). The daily fluid ingestion necessary to maintain this dilution of excreted cystine should be spaced over 24 h, with one-third of the total volume ingested between bedtime and 3 a.m. Cystine solubility rises sharply above pH 7.5, and urinary alkalinization (with potassium citrate) can be therapeutic. Penicillamine (1–3 g/d) and tiopronin (-mercaptopropionylglycine, 800–1200 mg/d in four divided doses) undergo sulfhydryl-disulfide exchange with cystine to form mixed disulfides. Because these disulfides are much more soluble than cystine, pharmacologic therapy can prevent and promote dissolution of calculi. Penicillamine can have significant side effects and should be reserved for patients who fail to respond to hydration alone or who are in a high-risk category (e.g., one remaining kidney, renal insufficiency). When medical management fails, shock wave lithotripsy, ureteroscopy, and percutaneous nephrolithotomy are effective for most stones. Open urologic surgery is considered only for complex staghorn stones or when the patient has concomitant renal or ureteral abnormalities. Occasional patients progress to renal failure and require kidney transplantation.

### LYSINURIC PROTEIN INTOLERANCE

This disorder is characterized by a defect in renal tubular reabsorption and intestinal transport of the three dibasic amino acids lysine, arginine, and ornithine but *not* cystine (*lysinuric protein intolerance*). Lysinuric protein intolerance is most common in Finland (1 in 60,000), southern Italy, and Japan, but is rare elsewhere. The transport defect affects basolateral rather than luminal membrane transport and causes secondary impairment of the urea cycle. The defective gene (*SLC7A7*, chromosome 14q11.2) encodes the  $y^+$ LAT membrane transporter, which associates with the cell-surface glycoprotein 4F2 heavy chain to form the complete sodium-independent transporter  $y^+$ L.

Manifestations are related to impairment of the urea cycle and to immune dysfunction potentially attributable to nitric oxide overproduction secondary to arginine intracellular trapping within macrophages. Affected patients present in childhood with hepatosplenomegaly, protein intolerance, and episodic ammonia intoxication. Older patients may present with severe osteoporosis, impaired renal function, pulmonary alveolar proteinosis, various autoimmune disorders, and an incompletely characterized immune deficiency. Plasma concentrations of lysine, arginine, and ornithine are reduced, whereas urinary excretion of lysine and orotic acid is increased. Hyperammonemia may develop after the ingestion of protein loads or with infections, probably because of insufficient amounts of arginine and ornithine to maintain proper function of the urea cycle. Therapy consists of dietary protein restriction and supplementation of citrulline (2–8 g/d), a neutral amino acid that fuels the urea cycle when metabolized to arginine and ornithine. Pulmonary disease responds to glucocorticoids or recombinant human granulocyte-macrophage colony-stimulating factor in some patients. Women with lysinuric protein intolerance who become pregnant have an increased risk of anemia, toxemia, and bleeding complications during delivery. These can be minimized by aggressive nutritional therapy and control of blood pressure. Their infants can have intrauterine growth restriction but have normal neurologic function.

### CITRULLINEMIA TYPE 2 (CITRIN DEFICIENCY)

Citrullinemia type 2 is a recessive condition caused by deficiency of the mitochondrial aspartate-glutamate carrier AGC2 (citrin). A defect in this transporter reduces the availability of cytoplasmic aspartate to combine with citrulline to form argininosuccinate (**see Fig. 420-1**), impairing the urea cycle and decreasing the transfer of reducing equivalents from the cytosol to the mitochondria through the malate-aspartate NADH shuttle. Mutations in the *SLC25A13* gene on chromosome

TABLE 421-1 Genetic Disorders of Amino Acid Transport

DISORDER	SUBSTRATES	TISSUES MANIFESTING TRANSPORT DEFECT	MOLECULAR DEFECT	MAJOR CLINICAL MANIFESTATIONS	INHERITANCE
Cystinuria	Cystine, lysine, arginine, ornithine	Proximal renal tubule, jejunal mucosa	Shared dibasic-cystine transporter <i>SLC3A1, SLC7A9</i>	Cystine nephrolithiasis	AR
Lysinuric protein intolerance	Lysine, arginine, ornithine	Proximal renal tubule, jejunal mucosa	Dibasic transporter <i>SLC7A7</i>	Protein intolerance, hyperammonemia, intellectual disability	AR
Hartnup disease	Neutral amino acids	Proximal renal tubule, jejunal mucosa	Neutral amino acid transporter <i>SLC6A19</i>	Constant neutral aminoaciduria, intermittent symptoms of pellagra	AR
Histidinuria	Histidine	Proximal renal tubule, jejunal mucosa	Histidine transporter	Intellectual disability	AR
Iminoglycinuria	Glycine, proline, hydroxyproline	Proximal renal tubule, jejunal mucosa	Shared glycine-imino acid transporter <i>SLC6A20, SLC6A18, SLC36A2</i>	None	AR
Dicarboxylic aminoaciduria	Glutamic acid, aspartic acid	Proximal renal tubule, jejunal mucosa	Shared dicarboxylic amino acid transporter <i>SLC1A1</i>	None	AR
Hyperargininemia	Arginine, lysine, ornithine	Ubiquitous	CAT2 cationic amino acid transporter <i>SLC7A2</i>	Hyperargininemia, Hyperammonemia (?)	AR
Brain branched chain amino acid deficiency	Leucine, isoleucine, valine	Plasma membrane of blood-brain barrier	Branched chain amino acid transporter <i>SLC7A5</i>	Microcephaly, intellectual disability, seizures, autism	AR
Citrullinemia type 2	Aspartate, glutamate, malate	Inner mitochondrial membrane	Mitochondrial aspartate/glutamate carrier 2 <i>SLC25A13</i>	Sudden behavioral changes with stupor, coma, hyperammonemia	AR
Hyperornithinemia, hyperammonemia, homocitrullinuria	Ornithine, citrulline	Inner mitochondrial membrane	Mitochondrial ornithine carrier <i>SLC25A15</i>	Lethargy, failure to thrive, intellectual disability, episodic confusion, hyperammonemia, protein intolerance	AR
Epileptic encephalopathy	Aspartate, glutamate, malate	Inner mitochondrial membrane	Mitochondrial aspartate/glutamate carrier 1 <i>SLC25A12</i>	Intellectual disability, epilepsy, hypotonia, cerebral atrophy, and hypomyelination	AR
Epileptic encephalopathy	Glutamate	Inner mitochondrial membrane	Mitochondrial glutamate carrier <i>SLC25A22</i>	Intellectual disability, epilepsy	AR
Epileptic encephalopathy	Glutamic acid, aspartic acid	Presynaptic glutamatergic nerve endings	EEAT2 neuronal dicarboxylic amino acid transporter <i>SLC1A2</i>	Developmental and epileptic encephalopathy	AD
Episodic ataxia	Glutamic acid, aspartic acid	Presynaptic glutamatergic nerve endings	EEAT1 neuronal dicarboxylic amino acid transporter <i>SLC1A3</i>	Episodic ataxia	AD
Brain serine deficiency	Alanine, serine, cysteine, threonine	Neuronal cells	ASCT neutral amino acid transporter <i>SLC1A4</i>	Progressive microcephaly, intellectual disability, spasticity	AR
Glycine encephalopathy with normal serum glycine	Glycine	Astrocytes and neuronal cells	GLYT1 astrocyte glycine transporter <i>SLC6A9</i>	Arthrogryposis, apnea, axial hypotonia, spasticity, intellectual disability	AR
Hyperekplexia-3	Glycine	Neuronal cells	GLYT2 presynaptic glycine transporter <i>SLC6A5</i>	Exaggerated startle response, hypertonia, apnea	AR
Intellectual disability	Proline, glycine, leucine, and alanine, glutamine	Neuronal cells synaptic vesicles	NTT4 synaptic vesicle neutral amino acid transporter <i>SLC6A17</i>	Intellectual disability, tremor	AR
Deafness	Glutamic acid	Neuronal cortical synaptic vesicles	VGLUT3 vesicular glutamate transporter <i>SLC17A8</i>	Deafness	AD
Foveal hypoplasia	Glutamine	Retinal photoreceptors	<i>SLC38A8</i>	Foveal hypoplasia, optic nerve decussation defects, anterior segment dysgenesis	AR
Retinitis pigmentosa	Arginine, lysine, ornithine	Retinal photoreceptors	Cationic amino acid transporter <i>SLC7A14</i>	Retinitis pigmentosa, blindness	AR
Early retinal degeneration	Taurine	Retinal cells	TAUT taurine transporter <i>SLC6A6</i>	Nystagmus, vision loss, retinal degeneration	AR
Cystinosis	Cystine	Lysosomal membranes	Lysosomal cystine transporter	Renal failure, hypothyroidism, blindness	AR

Abbreviations: AD, autosomal dominant; AR, autosomal recessive.

7q21.3 that encodes for this transporter are rare in Caucasians but affect ~1:20,000 people with ancestry from Japan, China, and Southeast Asia with variable penetrance.

The disease can present in children with neonatal intrahepatic cholestasis, failure to thrive, and dyslipidemia but usually presents with sudden onset between 20 and 50 years of age with recurring episodes of hyperammonemia with associated neuropsychiatric symptoms such as altered mental status, irritability, seizures, or coma resembling hepatic encephalopathy. Some patients might come to

medical attention for hypertriglyceridemia, pancreatitis, hepatoma, or fatty liver histologically similar to nonalcoholic steatohepatitis. Without therapy, most patients die with cerebral edema within a few years of onset. Episodes are usually triggered by medications (such as acetaminophen), surgery, alcohol consumption, or high sugar intake, with the latter conditions causing NADH production in the cytoplasm. NADH is not generated by the metabolism of proteins or fats, and many individuals with citrullinemia type 2 spontaneously prefer foods such as meat, eggs, and fish and avoid carbohydrates.

Laboratory studies during an acute attack include elevated ammonia, citrulline, and arginine with low or normal levels of glutamine (the latter is usually increased in classic urea cycle defects). Levels of galactose-1-phosphate in red blood cells are also increased, reflecting defective transfer of reducing equivalents from the cytosol to mitochondria. The diagnosis is confirmed by demonstrating mutations in the *SLC25A13* gene. Liver transplantation prevents progression of the disease and normalizes biochemical parameters. A diet high in fats and proteins and low in carbohydrates with supplements of medium-chain triglycerides, arginine, and pyruvate is also effective in preventing further episodes, at least in the short term.

## HARTNUP DISEASE

Hartnup disease (frequency 1 in 24,000) is an autosomal recessive disorder characterized by pellagra-like skin lesions, variable neurologic manifestations, and neutral and aromatic aminoaciduria. Alanine, serine, threonine, valine, leucine, isoleucine, phenylalanine, tyrosine, tryptophan, glutamine, asparagine, and histidine are excreted in urine in quantities 5–10 times greater than normal, and intestinal transport of these same amino acids is defective. The defective neutral amino acid transporter, B<sup>0</sup>AT1 encoded by the *SLC6A19* gene on chromosome 5p15, requires either collectrin or angiotensin-converting enzyme 2 for surface expression in the kidney and intestine, respectively.

The clinical manifestations result from nutritional deficiency of the essential amino acid tryptophan, caused by its intestinal and renal malabsorption, and of niacin, which derives in part from tryptophan metabolism. Only a small fraction of patients with the chemical findings of this disorder develop a pellagra-like syndrome, implying that manifestations depend on other factors in addition to the transport defect. The diagnosis of Hartnup disease should be suspected in any patient with clinical features of pellagra who does not have a history of dietary niacin deficiency (*Chap. 333*). The neurologic and psychiatric manifestations range from attacks of cerebellar ataxia to mild emotional lability to frank delirium, and they are usually accompanied by exacerbations of the erythematous, eczematoid skin rash. Fever, sunlight, stress, and sulfonamide therapy provoke clinical relapses. Diagnosis is made by detection of the neutral aminoaciduria, which does not occur in dietary niacin deficiency. Treatment is directed at niacin repletion and includes a high-protein diet and daily nicotinamide supplementation (50–250 mg).

## CYSTINOSIS

Cystinosis (frequency 1 in 100,000–200,000) is an autosomal recessive disorder caused by mutations in the *CTNS* gene encoding the lysosomal cystine/proton transporter (cystinosin). In this condition, cystine derived from protein degradation accumulates inside lysosomes and forms crystals due to its poor solubility. Depending on the degree of impairment of transporter function, three clinical forms are recognized.

The most severe form, classic nephropathic cystinosis, causes renal Fanconi syndrome during the first year of life and, without treatment, evolves to renal failure usually by 10 years of age. Intermediate nephropathic cystinosis leads to kidney failure between 15 and 25 years of age, whereas photophobia, caused by deposition of cystine crystals in the cornea, is the only manifestation of ocular nonnephropathic cystinosis. Cystinosis is suspected by the identification of cystine crystals in the cornea by slit lamp examination and diagnosed by measuring cystine content in white blood cells. DNA testing (including deletion analysis) of the *CTNS* gene can further confirm the diagnosis. Therapy consists in the administration of cysteamine that enters lysosomes, forms a mixed disulfide with cysteine, and is exported from the lysosome using a cationic amino acid transporter. Oral cysteamine therapy (60–90 mg/kg per d up to 2 g/d in adults, 0.2–0.3 g/m<sup>2</sup> per d divided into two doses given every 12 h for the extended-release formulation) can delay renal failure and is more effective if started early in the course of the disease. Therapy with cysteamine reduces intracellular cystine accumulation in white blood cells, but compliance with therapy is difficult due to the unpleasant odor of the drug and the need for frequent administration. Cysteamine eye drops can relieve photophobia. Renal replacement therapy with salts, alkali, and activated vitamin D is necessary for renal Fanconi syndrome. Cystine accumulation occurs in virtually all organs and tissues, causing additional complications such as hypothyroidism, hypohydrosis, diabetes, and delayed puberty in both males and females with primary hypogonadism in males. Growth hormone replacement, 1-thyroxine for hypothyroidism, insulin for diabetes mellitus, and testosterone for hypogonadism in males may be necessary. Despite therapy, many patients with cystinosis progress to end-stage renal failure and require kidney transplantation. Late-onset complications include hepatomegaly and splenomegaly that occur in approximately one-third of subjects and a vacuolar myopathy causing weakness (initially involving the distal extremities), swallowing difficulties, gastrointestinal dysmotility, and pulmonary insufficiency. Before the availability of cystine-depleting therapy and renal transplantation, the life span in nephropathic cystinosis was <10 years. With current therapies, affected individuals can survive into the late forties with satisfactory quality of life.

## FURTHER READING

- Servais A et al: Cystinuria: Clinical practice recommendation. *Kidney Int* 99:48, 2021.
- Tanner LM et al: Inhaled sargramostim induces resolution of pulmonary alveolar proteinosis in lysinuric protein intolerance. *JIMD Rep* 34:97, 2017.
- Tărlungeanu DC et al: Impaired amino acid transport at the blood brain barrier is a cause of autism spectrum disorder. *Cell* 167:1481, 2016.
- Yahyaoui R, Pérez-Frías J: Amino acid transport defects in human inherited metabolic disorders. *Int J Mol Sci* 21:119, 2019.