Cunningham's

TEXTBOOK of VETERINARY PHYSIOLOGY

SIXTH EDITION

BRADLEY G. KLEIN

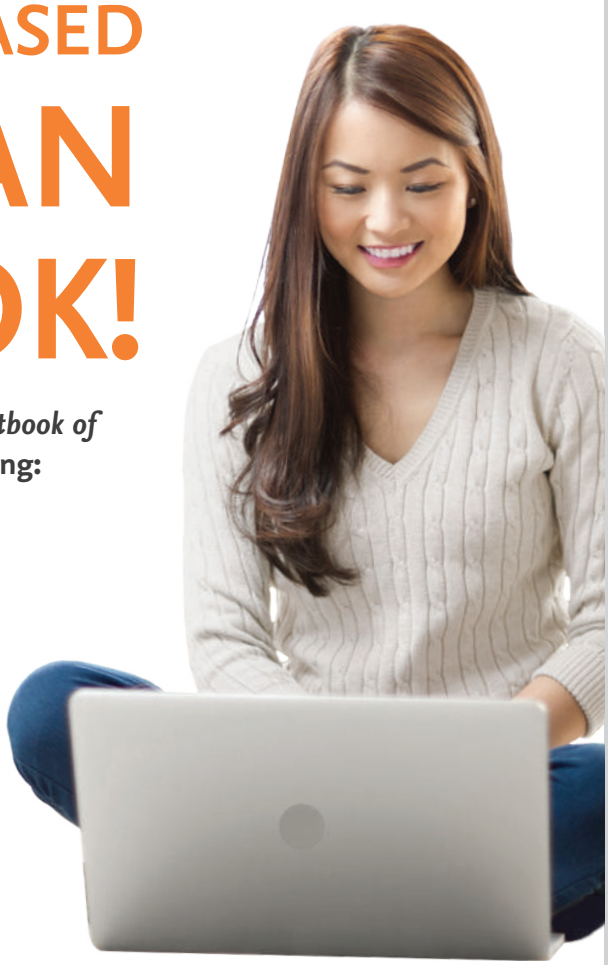
Evolve®

Student Resources on Evolve
Access Code Inside

ELSEVIER

Evolve®

YOU'VE JUST PURCHASED MORE THAN A TEXTBOOK!



Evolve Student Resources for *Klein: Cunningham's Textbook of Veterinary Physiology, Sixth Edition*, include the following:

- Animations
- Clinical Correlations
- Glossary
- Practice Tests

Activate the complete learning experience that comes with each NEW textbook purchase by registering with your scratch-off access code at

http://evolve.elsevier.com/Klein/physiology/

If you rented or purchased a used book and the scratch-off code at right has already been revealed, the code may have been used and cannot be re-used for registration. To purchase a new code to access these valuable study resources, simply follow the link above.

Place Sticker Here

REGISTER TODAY!

ELSEVIER

You can now purchase Elsevier products on Evolve!
Go to evolve.elsevier.com/shop to search and browse for products.

2018v1.0

Cunningham's Textbook of

VETERINARY PHYSIOLOGY

Sixth Edition

Bradley G. Klein, PhD

Associate Professor of Neuroscience
Department of Biomedical Sciences and Pathobiology
Virginia-Maryland College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

ELSEVIER

ELSEVIER

3251 Riverport Lane
St. Louis, Missouri 63043

CUNNINGHAM'S TEXTBOOK OF VETERINARY PHYSIOLOGY, SIXTH EDITION

ISBN: 978-0-323-55227-1

ISBN 978-0-323-55227-1

Senior Content Strategist: Jennifer Flynn-Briggs
Senior Content Development Manager: Ellen Wurm-Cutter
Content Development Specialist: Kathleen Nahm
Publishing Services Manager: Deepthi Unni
Senior Project Manager: Kamatchi Madhavan
Project Manager: Haritha Dharmarajan
Designer: Bridget Hoette

Printed in China

Last digit is the print number 9 8 7 6 5 4 3 2 1

In loving and lasting memory of Ira Matthew Klein (1985-2013).

Contributors

S. Ansar Ahmed, DVM, PhD
Professor and Associate Dean for Research
Department of Biomedical Sciences and Pathobiology
Virginia-Maryland College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Steven P. Brinsko, DVM, MS, PhD, DACT
Professor and Associate Department Head of Academics
Department of Large Animal Clinical Sciences
College of Veterinary Medicine and Biomedical Sciences
Texas A&M University
College Station, Texas

Autumn P. Davidson, DVM, MS, DACVIM (Internal Medicine)
Clinical Professor
Veterinary Medical Teaching Hospital
Department of Medicine and Epidemiology
School of Veterinary Medicine
University of California-Davis
Davis, California

Susan L. Ewart, DVM, PhD, DACVIM
Professor and Special Assistant to the Dean
Department of Large Animal Clinical Sciences
College of Veterinary Medicine
Michigan State University
East Lansing, Michigan

Deborah S. Greco, DVM, PhD, DACVIM-SAIM
Senior Research Scientist
Nestle Purina Petcare
St. Louis, Missouri

Steven R. Heidemann, PhD
Professor Emeritus
Department of Physiology
Michigan State University
East Lansing, Michigan

Thomas H. Herdt, DVM, MS, DACVIM, DACVN
Professor and Chief of Nutrition
Department of Large Animal Clinical Sciences and Veterinary Diagnostic Laboratory
College of Veterinary Medicine
Michigan State University
East Lansing, Michigan

Bradley G. Klein, PhD
Associate Professor of Neuroscience
Department of Biomedical Sciences and Pathobiology
Virginia-Maryland College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Xin M. Luo, MS, PhD
Associate Professor of Immunology
Department of Biomedical Sciences and Pathobiology
Virginia-Maryland College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Brian K. Petroff, DVM, PhD
Associate Professor and Section Chief, Endocrinology
Department of Pathobiology and Diagnostic Investigation
Diagnostic Center for Population and Animal Health
College of Veterinary Medicine
Michigan State University
East Lansing, Michigan

Juan E. Romano, DVM, MS, PhD, DACT
Professor
Department of Large Animal Clinical Sciences
College of Veterinary Medicine and Biomedical Sciences
Texas A&M University
College Station, Texas

John H. Rossmeisl Jr., DVM, MS, DACVIM
Professor
Neurology and Neurosurgery
Department of Small Animal Clinical Sciences
Virginia-Maryland College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Gerhardt G. Schurig, DVM, MS, PhD
Professor and International Programs Strategist
Department of Biomedical Sciences and Pathobiology
Virginia-Maryland College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

George H. Stabenfeldt, DVM, PhD (deceased)
Professor
Department of Reproduction
School of Veterinary Medicine
University of California-Davis
Davis, California
(Contributor of past editions)

Robert B. Stephenson, PhD
Associate Professor
Department of Physiology
Michigan State University
East Lansing, Michigan

Jill W. Verlander, DVM
Scientist
Department of Medicine
Division of Nephrology, Hypertension, and Renal Transplantation
University of Florida College of Medicine
Gainesville, Florida

Sharon G. Witonsky, DVM, PhD, DACVIM
Associate Professor
Equine Field Service
Department of Large Animal Clinical Sciences
Virginia-Maryland College of Veterinary Medicine
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Preface

Physiology is the study of the normal functions of the body—the study of the body's molecules, cells, and organ systems, and the interrelationships among them. Because the study of medicine is the study of abnormal functions of the body, it is essential to understand normal physiology if one is to understand the mechanisms of disease. For this reason, physiology and other important sciences basic to medicine are introduced early in the veterinary curriculum.

Physiology is a vast subject, and the normal time constraints on veterinary students are usually prohibitive for learning all that is known about it. Therefore an effort was made to limit the concepts presented in this book to those germane to the practice of veterinary medicine. Because the scope of physiology encompasses many scientific disciplines and levels of analysis, the contributing authors represent not only the field of physiology but others, among them neuroscience, cell biology, molecular biology, immunology, and various clinical specialties. Most of these authors are also veterinarians or have consulted with veterinary clinicians regarding content. Sections on the immune system and cancer underscore the intimate relationship between the understanding of cell and molecular biology, physiological function, and veterinary medicine.

This book is designed for first-year veterinary students. The goal is to introduce the student to the principles and concepts of physiology pertinent to the practice of veterinary medicine. Other goals are to introduce the reader to physiopathology and clinical problem-solving techniques and to help the reader understand the relationship between physiology and the practice of veterinary medicine.

This book is designed to be as student friendly as possible. New concepts in the text are introduced by a declarative statement designed to summarize the essential point. This format also helps the reader survey the chapter or review for an examination. These declarative statements are also listed at the beginning of the chapter as an outline of Key Points.

Chapters include Clinical Correlations at the end. Concordant with the increasing trend toward early integration of basic science and clinical material in veterinary curricula, these Clinical Correlations are designed to show the reader how knowledge of physiology is applied to the diagnosis and treatment of veterinary patients. They also provide the student with an additional way to think through the principles and concepts presented, and they can serve as a basis for classroom case discussions.

Several Practice Questions are included in each chapter as another method for students to review the book's content. The Bibliography for each chapter is designed to lead the reader to more advanced textbooks, as the aforementioned time constraints can often prevent veterinary students from reading original literature. However, for those who may find the time, some original literature references are also included in several chapters.

Accompanying resources for the text can be found on Elsevier's Evolve website. These include an online repository of the book's Practice Questions and Clinical Correlations, as well as some relevant animations from Elsevier's existing collection. Students will appreciate the ability to access the Practice Questions as Practice Tests. Instructors will appreciate the items in the illustration bank, which can be downloaded for lecture use. There is also a Glossary covering many of the italicized words from the printed text.

Immense gratitude is due to the contributing authors who have worked hard to ensure that the information in this latest edition is accurate and up-to-date. This has resulted in new illustrations and Clinical Correlations, as well as addition of material on new discoveries (e.g., a brain lymphatic system; neutrophil extracellular traps). The expertise of two additional section authors is welcomed, Dr. Susan L. Ewart for Respiratory Function and Homeostasis, and Dr. Brian K. Petroff for Endocrinology. Dr. Xin M. Luo contributed some material to the Immunology section. Notable recognition is in order for Dr. N. Edward Robinson, who is retiring from this project after contributing to all of its previous editions, and who authored the initial material on respiratory physiology and on homeostasis. Thanks also to Dr. Ayman Sayegh for contributing to an earlier edition of the introductory chapter on regulation of gastrointestinal function. Suggestions of ways to improve this text in subsequent editions are always welcome.

Particular gratitude is due to the book's medical illustrator, Jeanne Robertson, who drew the new illustrations for this edition and revised much of the preexisting artwork. Thanks are also in order for the folks at Elsevier who were instrumental in producing the sixth edition, among them Penny Rudolph, Jennifer Flynn-Briggs, Ellen Wurm-Cutter, Anna Miller, Madhavan Kamatchi, and Haritha Dharmarajan. Special recognition is due for Kathleen Nahm, the Elsevier content development specialist who shepherded almost all aspects of this project through various trials and tribulations and who did an admirable job managing my not infrequent tardiness yet otherwise fastidious nature. Drs. Jonathan Abbott, Virginia Buechner-Maxwell, and Shireen Hafez unselfishly provided their valuable opinions on various aspects of the book that resulted in its improvement. Dr. Bonnie Smith was particularly helpful in providing advice on the accuracy of the cover illustration. Again, this book would not exist without the invaluable expertise of the section authors who worked so hard to make this the best veterinary physiology text possible.

The most immense debt is due to Dr. Jim Cunningham, whose vision, guidance, and expertise made the *Textbook of Veterinary Physiology* a reality and a success, and has made "Cunningham's" a household name among veterinary students and faculty of several generations. Placing his name in the book's title is in partial recognition of that debt. He is also recognized here for authoring the initial material on neurophysiology upon which I have built over the years. The instructional style he instituted continues in this edition and will continue in future editions of the text.

I would also like to thank my family whose love and support sustains my existence, and Ira, we miss you so very much. You are always in our hearts and our memories.

Finally, this book is a tribute to the veterinary students throughout the world. It is these students who give pleasure, meaning, and value to our teaching.

Brad Klein

Contents

†Deceased

1

The Molecular and Cellular Bases of Physiological Regulation

STEVEN R. HEIDEMANN

KEY POINTS

1. All physiological change is mediated by proteins.
2. Protein function depends on protein shape and shape changes.
3. A series of enzymatic reactions converts tyrosine into the signaling molecules dopamine, norepinephrine, and epinephrine.
4. Muscle contraction and its initiation and cessation depend on the binding specificity and allosteric properties of proteins.
5. Biological membranes are a mosaic of proteins embedded in a phospholipid bilayer.

Transport

1. Only small, uncharged molecules and oily molecules can penetrate biomembranes without the aid of proteins.
2. Molecules move spontaneously from regions of high free energy to regions of lower free energy.
3. Important transport equations summarize the contributions of the various driving forces.
4. Starling's hypothesis relates fluid flow across the capillaries to hydrostatic pressure and osmotic pressure.
5. Membrane proteins that serve the triple functions of selective transport, catalysis, and coupling can pump ions and molecules to regions of higher free energy.
6. Many membrane proteins selectively facilitate the transport o f ions/molecules from high to low electrochemical potential.
7. Passive transport of K^+ across the plasma membrane creates an electrical potential.
8. Spatial organization of active and passive transport proteins enables material to pass completely through the cell.
9. Membrane fusion allows for a combination of compartmentalization and transport of material.

Information Transmission and Transduction

1. Cell signaling often occurs by a lengthy chain of sequential molecular interactions.
2. Signaling pathways begin with the binding of an extracellular molecule to a receptor.
3. Specific physiological information is inherent in the receptor/ligand complex, not in the hormone/neurotransmitter molecule.
4. G-protein–coupled receptors are the largest family (a *superfamily*) of receptors and help regulate almost all physiological processes.
5. Most G-protein–linked information is sent to the cytoplasm by *second messengers*.
6. Ca^{2+} transport across plasma and intracellular membranes is an important second messenger.
7. Cyclic adenosine monophosphate (AMP) is produced by activation of a membrane-bound enzyme in response to hormone/neurotransmitter binding to receptors.
8. The receptor-mediated hydrolysis of a rare phospholipid of the plasma membrane produces two different second messengers with different actions.
9. Steroid hormones and other lipid signals interact with nuclear receptors, which are transcription factors within the cell.

Physiology is the study of the regulation of change within organisms, in this case higher animals. Our understanding of physiology has changed dramatically in the past 35 years as a result of insight into the molecular basis of biological regulation. This chapter summarizes (and simplifies considerably) our current understanding of the molecular and cellular basis of that regulation. Most of the principles in this chapter apply to all animal cells. The approach taken is one of functional molecular anatomy. That is, the molecular structure of the cell is examined with particular emphasis on the physiological function, in the intact animal, of the molecules and supramolecular structures responsible for the function. Only those aspects of cell function that illuminate the medical physiology of the higher animals are discussed; the reader is referred to the Bibliography for more complete coverage of the cell. Some review of basic concepts and vocabulary is presented. However, the discussion assumes that the reader is familiar with the cell and its constituent molecules as presented in courses in general biology and undergraduate courses in biochemistry.

All Physiological Change Is Mediated by Proteins

All physiological change is mediated by a single class of polymeric macromolecules (large molecules), the proteins. Protein function can be subdivided into a number of categories: catalysis, reaction coupling, transport, structure, and signaling.

Catalysis is the ability to increase greatly the rate of a chemical reaction without altering the equilibrium of the reaction. The majority of biochemical reactions occur at a physiologically useful rate only because of protein catalysts, called *enzymes.* Examples of enzymatic catalysis in the synthesis of a class of physiological regulator molecules, catecholamines, are given later in this chapter.

In *reaction coupling,* two reactions are joined together with the transfer of energy. Energy from a spontaneous reaction (similar to water flowing downhill) is funneled to a nonspontaneous reaction (e.g., sawing wood) so that the sum of the two reactions is spontaneous. That is, the energy liberated by the "downhill" reaction is used to drive the "uphill" reaction. This is the basic function of a motor; the "downhill" burning of gasoline is coupled with the "uphill" movement of the car. The ability of proteins to couple spontaneous and nonspontaneous reactions allows cells to be chemical motors, using chemical energy to do various jobs of work. One such job of work, the contraction of striated muscle, is discussed later with particular emphasis on the proteins involved.

Proteins provide a pathway for the *membrane transport* of most molecules and all ions into and out of the cell. Transport and transport proteins are discussed more fully after a discussion of the lipid bilayer membrane, the major obstacle to transport.

Proteins that form filaments and glue cells to each other and to their environment are responsible for the *structure* and organization of cells and multicellular assemblies (i.e., the tissues and organs of animals). The internal structure of the muscle cell, as well as its ability to do work, is a result of the properties of the muscle proteins discussed later.

At its most basic level, *signaling* requires only a controlled change or difference. Human signaling occurs by way of open and closed electrical circuits (telegraphy), puffs of smoke in the air, and complex black marks on a contrasting background (numbers and letters). As discussed next, a fundamental property of proteins is the ability to change shape. The cell can use changes of protein shape directly to send signals, and the function of some proteins is purely informational. That is, all that some proteins do by changing shape is to transmit and transduce information. *Information* can be defined as "any difference that makes a difference," or more simply, any difference that regulates something. Catalysis, coupling, transport, structural, and signaling functions can be combined on individual protein molecules. As will become apparent, such multifunctional proteins carry out many important physiological functions. Also important is that a change in one or more of these protein functions can be used to carry information, to serve as a signal within the cell. Thus in addition to proteins specialized exclusively to carry information, changes in enzymatic activity or ion transport can also make a difference, transmitting information and triggering an appropriate response.

Protein Function Depends on Protein Shape and Shape Changes

Protein function is founded on two molecular characteristics: (1) proteins can bind to other molecules very specifically; and (2) proteins change shape, which in turn alters their binding properties and their function. The binding specificity of proteins is the result of their complex three-dimensional structure. Grooves or indentations on the surface of protein molecules, called *binding sites,* permit specific interactions with a molecule of a complementary shape, called the *ligand.* This complementary-shape mechanism underlying binding is similar to the shape interaction between a lock and key.

Several aspects of the lock-and-key analogy are worth noting. As with a lock, only a small part of the protein is engaged in binding. The binding is very specific; small changes in the shape of the binding site (keyhole) or the shape of the ligand (key) can cause major changes in protein (lock) behavior. Similar to the lock and key, the complementary-shape interaction serves a recognition function; only those molecules with the right shape affect protein function. This recognition function plays a primary role in information transfer. The protein recognizes a particular signal by binding to it, thus changing the protein's shape and its function. Unlike the majority of locks, however, proteins frequently have multiple binding sites for multiple ligands.

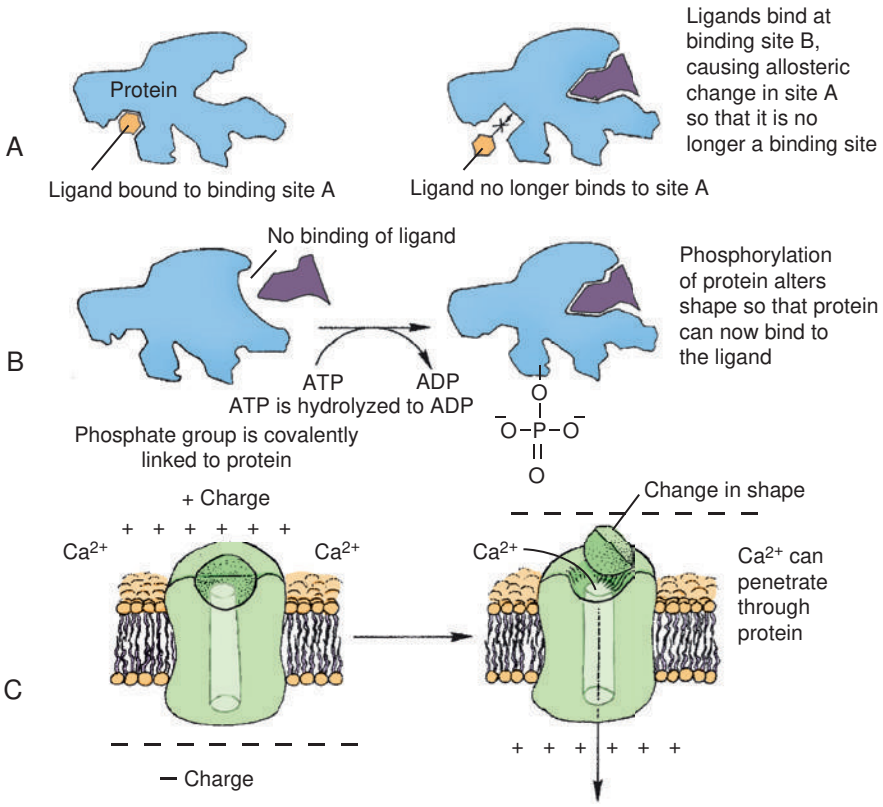
Thus the three-dimensional shape of a protein, its *conformation,* determines protein function. A major force that stabilizes protein conformation is the *hydrophobic interaction.* Oily, *hydrophobic* (water-fearing) amino acids tend to congregate in the middle of a protein away from water, whereas *hydrophilic* (water-loving) amino acids tend to be found on the protein's outer surface interacting with the abundant cellular water. The hydrophobic interaction is also important in stabilizing the interaction of proteins with the lipids of biological membranes, as will be discussed shortly. Protein shape is also stabilized by *hydrogen bonding* between polar amino acid pairs in the polypeptide (protein) chain.

The same weak forces responsible for protein conformation are also used to hold the ligand in the protein-binding site. The position of the ligand in the binding site is stabilized by hydrogen bonds between the polar groups of the ligand and polar amino acid side groups lining the binding site, just as hydrogen bonds within the polypeptide chain stabilize the shape of the polypeptide. Precisely because the same forces are responsible for the shape of the protein and for its binding properties, shape influences binding, and in turn, binding can influence protein shape. The ability of proteins to change shape is called *allostery* (Greek for "other shape").

Allosteric changes in protein conformation arise in four general ways. One way, just mentioned, is that most proteins change shape depending on which ligands are bound at particular binding sites (Fig. 1.1A). The sequence—specific ligand binding → protein shape change → change in protein-binding properties and protein function → this change regulates something—is a common molecular mechanism underlying physiological control. This method involves no alteration in the covalent structure of the protein.

A second method of producing conformational change, however, occurs as a result of the covalent modification of one or more of the amino acid side groups of the protein (see Fig. 1.1B). By far the most common such change is the covalent addition of a phosphate group to the hydroxyl (–OH) group on the side chain of serine, threonine, or tyrosine residues in the protein. This modification is called *phosphorylation.* Because the phosphate group is highly charged, phosphorylation of a protein alters hydrogen bonding and other electrostatic interactions within the protein chain, altering its conformation and functional properties.

In a third method, some physiologically important proteins change shape in response to the electrical field surrounding the protein (see Fig. 1.1C). These respond to a voltage change



• **Fig. 1.1** Three common mechanisms of allosteric shape change in proteins. (A) Ligand binding. Ligand binding to an allosteric site (site B) on a protein changes the protein's conformation such that binding site A is altered; ligand no longer binds at site A because of the binding event at site B. (B) Phosphorylation. Addition of a phosphate group to a serine, threonine, or tyrosine residue of a protein alters the protein's conformation, changing its binding characteristics. In this hypothetical example, phosphorylation activates an otherwise inactive protein. Some proteins inactivate by this mechanism. *ADP,* Adenosine diphosphate; *ATP,* adenosine triphosphate. (C) Voltage-dependent proteins. The conformation of some proteins, particularly ion channels, is altered by the electrical field surrounding the protein. Shown here is the opening (activation) of a voltage-dependent, gated Ca^{2+} channel when the membrane depolarizes.

by altering the position of charged amino acids, thus altering protein shape.

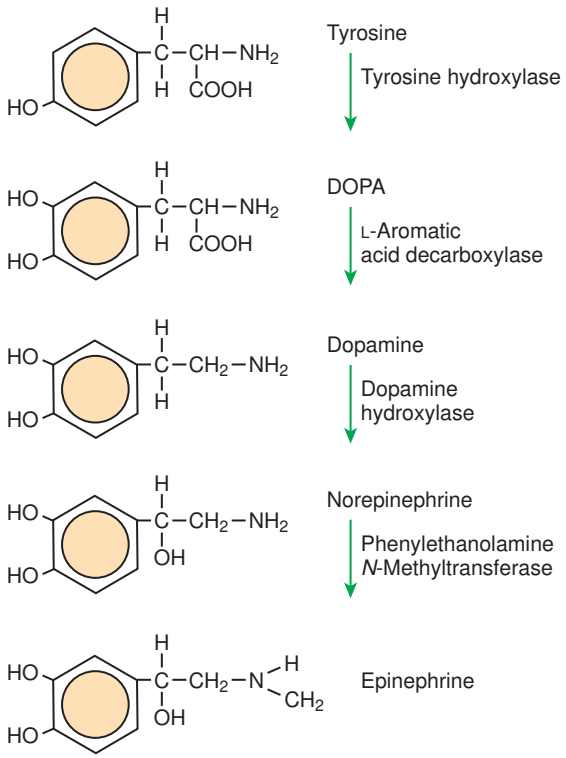
The fourth method of protein shape change is the least well understood (not shown). Some proteins change shape in a controlled manner in response to mechanical forces. Although this is not surprising, because all solids and solid-like substances change shape at least slightly in response to force, we know relatively little about mechanosensitive proteins. The best current example is a protein involved in the very early events of hearing that changes its transport of ions in response to the mechanical stimulation by sound (small changes of air pressure in waves).

The significance of binding specificity and allostery can be better appreciated with two examples of their roles in physiological function. The first example is the role of enzymes in synthesizing three small, structurally similar, nonprotein signaling molecules. This example shows how binding specificity is important in catalytic function and how allostery underlies the regulation of the synthesis. The second example is more complex: the role of proteins in the contraction of muscle. The contraction of muscle shows how proteins can exploit the basic properties of specific binding and allosteric shape change to do more than one job of work at the same time; muscle proteins serve a structural role, serve a catalytic function, and couple the "downhill" hydrolysis of *adenosine triphosphate* (ATP) to do mechanical work, the "uphill" lifting of weight.

A Series of Enzymatic Reactions Converts Tyrosine Into the Signaling Molecules Dopamine, Norepinephrine, and Epinephrine

Fig. 1.2 is a diagram of the series of reactions by which the amino acid tyrosine is converted into three different signaling molecules: (1) *dopamine,* a brain neurotransmitter; (2) *norepinephrine,* a neurotransmitter of the brain and peripheral autonomic nervous system; and (3) *epinephrine,* an autonomic neurotransmitter and hormone. Dopamine, norepinephrine, and epinephrine share a similar structure. All contain a phenyl (benzene) ring with two hydroxyl groups (i.e., catechol) and an amine group (thus the term *catecholamines*). They are among the large number of molecules that function as neurotransmitters. That is, the electrically coded information sent along nerve cells causes the release of a chemical, the neurotransmitter, at the terminal of the neuron, which is next to a target cell, such as another nerve, a muscle, or an endocrine cell. The electrically encoded information of the nerve is transmitted to the target cell by the binding of the neurotransmitter to proteins on the surface of the target cell. Proper neurotransmitter synthesis is crucial to nervous function and physiological regulation.

In the first step of catecholamine biosynthesis, tyrosine binds to the enzyme tyrosine hydroxylase, which catalyzes the addition of



• **Fig. 1.2** Epinephrine biosynthetic pathway. The amino acid tyrosine is metabolized to the neurotransmitters dopamine, norepinephrine, and epinephrine. The diagram shows the names and structural formulas for each compound in the path and the names of the enzymes that catalyze each reaction. *DOPA,* Dihydroxyphenylalanine.

another hydroxyl group to the phenyl group to form dihydroxyphenylalanine, almost always called *DOPA.* This hydroxyl group alters the enzyme-ligand interaction; the key no longer fits the keyhole. DOPA is released from the tyrosine hydroxylase and is then bound by another enzyme, L-aromatic amino acid decarboxylase. As the name implies, this enzyme catalyzes the removal of the carboxyl group, converting DOPA to dopamine. Dopamine is converted into norepinephrine by the activity of dopamine hydroxylase, which adds yet another hydroxyl group, this time to the two-carbon tail of dopamine. Finally, addition of a methyl group to the amino nitrogen by phenylethanolamine *N*-methyltransferase gives rise to epinephrine (also called *adrenalin*). Note the binding specificity of the enzymes; whereas the catecholamine structures are all similar to one another, different enzymes bind each one (e.g., epinephrine does not bind to dopamine hydroxylase).

The allosteric properties of one enzyme in this pathway provide an example of physiological regulation. Certain hormones and neurotransmitters cause the *phosphorylation* of tyrosine hydroxylase, the first enzyme in the pathway, increasing its activity. That is, phosphorylation of the enzyme increases the rate at which it catalyzes the conversion of tyrosine to DOPA. Because this step is the slowest in the pathway, an increase in the activity of this protein increases the net rate of synthesis of all the catecholamines. Regulated decreases in the rate of catecholamine synthesis are achieved by a different allosteric mechanism: binding of end products to the enzyme. Dopamine, norepinephrine, and epinephrine can all bind to tyrosine hydroxylase at a site different than the site for tyrosine. These binding events inhibit the enzymatic activity. The inhibition of the pathway by its own end products makes this a classic case of allosteric control called *end-product inhibition.* Many substances regulate their own synthesis by inhibiting an initial enzyme in the pathway. If the cell has enough end products, these products inhibit further synthesis by allosteric changes in the enzyme. This is an example of the following sequence: specific binding → protein shape change → change in protein-binding properties and protein function → this change regulates something.

Muscle Contraction and Its Initiation and Cessation Depend on the Binding Specificity and Allosteric Properties of Proteins

There are three types of muscle tissue in vertebrates: (1) *skeletal muscle,* responsible for the animal's ability to move; (2) *cardiac muscle,* a muscle type found only in the heart but structurally similar to skeletal muscle; and (3) *smooth muscle,* which surrounds hollow organs such as blood vessels, gut, and uterus. All three produce tensile force by contracting and shortening the length of the muscle. All muscle contraction occurs by the binding and the allosteric properties of two proteins, actin and myosin. Starting and stopping the contraction process depends on two additional proteins in skeletal and cardiac muscle, troponin and tropomyosin. Contraction initiation and cessation in smooth muscle depend on a different system with different proteins and are discussed later in this chapter.

Myosin is a large protein whose shape resembles a two-headed golf club. The elongated tail of the myosin molecule corresponds to the shaft of the golf club, and there are two knobs at one end of the tail that, as with golf clubs, are called *heads.* Myosin tails bind specifically to other myosin tails, forming bipolar aggregates called *thick filaments* (Fig. 1.3). Myosin heads specifically bind ATP and another muscle protein, actin. Actin binds to itself to form long, thin filaments called *thin filaments* in muscle and *F-actin* (filamentous actin) in other cell types. Actin filaments play an important architectural role in all animal cells. Although actin is best understood in muscle cells, all animal cells depend on actin filaments for their shape and for their capacity to migrate in their environment. Actin filaments can be "woven" in various ways to produce different structures, such as ropelike bundles and clothlike networks. These actin bundles and actin networks are used to support the cell in particular shapes, similar to ropes holding up the woven cloth of a tent.

In muscle, the interaction of myosin, ATP, and actin produces contraction and force, as shown in Fig. 1.4:

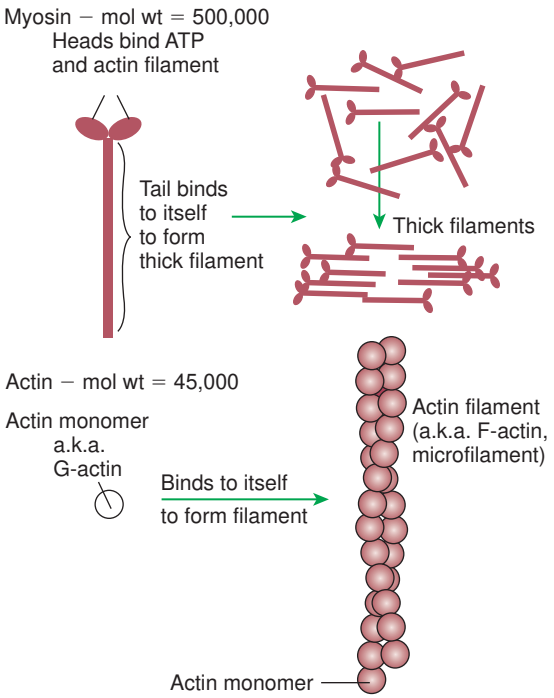
Step A: ATP binds to a myosin head; in this conformation, myosin has little ability to bind to actin.

Step B: Enzymatic activity associated with the myosin head, an *adenosinetriphosphatase* (ATPase), rapidly causes a partial hydrolysis of ATP to *adenosine diphosphate* (ADP) and *inorganic phosphate* (P_i), both of which stay bound to the myosin. With ADP and P_i bound, myosin has a slightly different shape that binds avidly to nearby actin filaments.

Step C: When myosin binds to actin, called *cross-bridging,* the myosin head couples the complete hydrolysis of ATP to a forceful flexing of the myosin head. This allosteric change causes the actin filament to slide past the thick filament. This sliding puts the actin filament under tension, which in turn causes the muscle to contract (shorten) against the load of the muscle (i.e., lifting a weight or pumping out blood). All muscle contraction depends on this *sliding filament mechanism* of actin- and myosin-based filaments. This same

allosteric change of myosin also alters myosin-binding properties so that it releases the ADP and P_i.

Step D: The binding of a new ATP molecule to the myosin head again causes myosin to change shape; the head unflexes and loses its affinity for actin, releasing the cross-bridge, and the cycle can start over. *Rigor mortis* of dead animals is caused by a lack of new ATP to bind to myosin heads. In the absence of ATP, myosin heads remain in Step C (i.e., bound to actin). The muscle is stiff because it is completely cross-bridged together.



• **Fig. 1.3** Assembly of myosin and actin to form filamentous structure. Myosin tails aggregate with one another to form a thick filament, a substructure of striated muscle. Actin monomers (G-actin) are a single polypeptide chain forming a globular protein that can bind to other actin monomers to form actin filament, also called microfilaments. The actin filament is the basic structure of striated muscle thin filaments; thin filaments also have troponin and tropomyosin as part of their structure. *ATP,* Adenosine triphosphate.

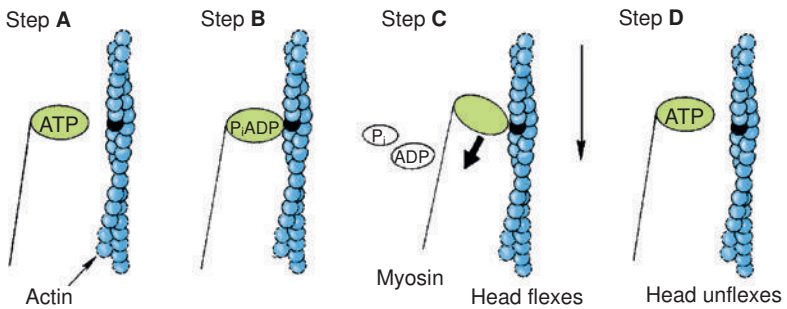
This *actomyosin motor* uses the binding and allosteric properties of proteins to (1) create structural filaments capable of withstanding and transmitting mechanical force, (2) catalyze the hydrolysis of ATP, and (3) couple the "downhill" ATP hydrolysis to the "uphill" contraction to produce force. For just the one protein, myosin, there are a number of examples of the characteristic sequence described earlier: specific binding → protein shape change → change in protein-binding properties and protein function → this change makes a difference.

This system of contractile proteins requires some control so that, for example, the heart beats rhythmically and skeletal muscle contraction is coordinated. At the organismal level, skeletal and cardiac muscle contraction is primarily under control by electrical stimulation from nerves or other electrically active cells (see Chapter 6). The transmission of electrical excitation to the actomyosin system is called *excitation-contraction coupling. Excitation-contraction coupling in all types of muscle depends on changes in intracellular calcium ion (Ca^{2+}) concentration.* In skeletal and cardiac muscle, but not smooth muscle, two additional thin-filament proteins, troponin and tropomyosin, are required for this coupling. (Excitation-contraction coupling for smooth muscle is discussed later in this chapter.) In striated muscles, *troponin* binds to tropomyosin and to Ca^{2+}. *Tropomyosin* is a long, thin protein that binds in the groove of the actin filament in such a way that its positions, high in the groove or snuggled down deep in the groove, allow or prevent the myosin head access to the thin filament (Fig. 1.5). Excitation-contraction coupling of striated muscle works as follows:

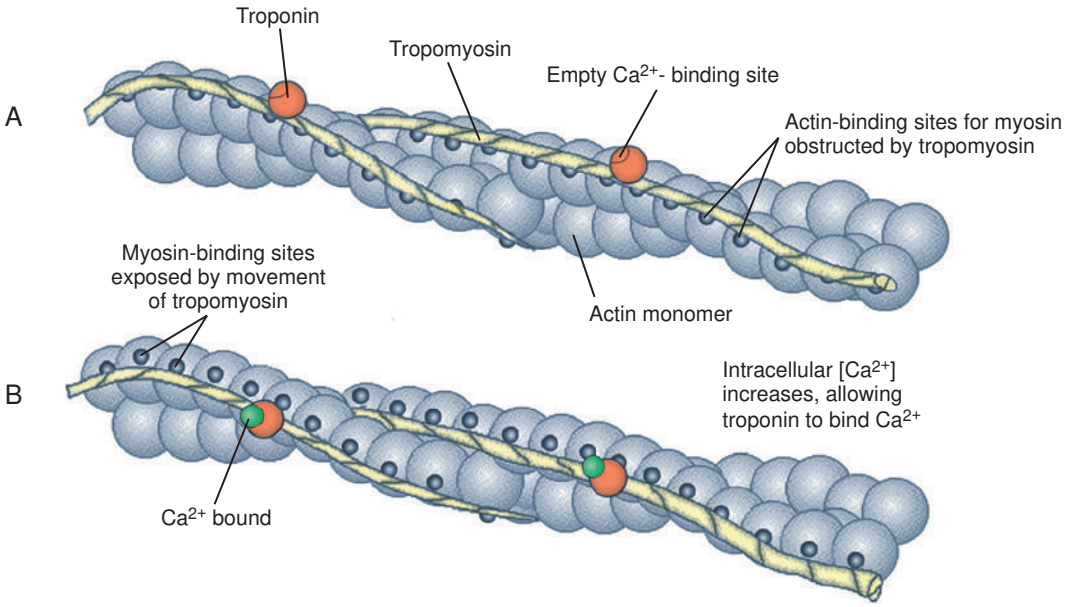
Step A: Electrical excitation of a striated muscle cell causes an increase in the intracellular concentration of Ca^{2+}.

Step B: The additional Ca^{2+} binds to troponin, causing an allosteric change in troponin.

Step C: Because Ca^{2+} is bound to troponin, which in turn is bound to tropomyosin, the Ca^{2+}-induced change in troponin conformation is transmitted to the tropomyosin molecule. When troponin binds Ca^{2+}, tropomyosin changes its binding to actin in such a way that it exposes the actin site for myosin



• **Fig. 1.4** Power stroke of actomyosin. (A) The myosin head has bound to adenosine triphosphate (ATP). In this conformation, myosin has little affinity to bind to actin. (B) ATP is partially hydrolyzed to adenosine diphosphate (ADP) and inorganic phosphate (P_i); the hydrolysis is partial because the products remain bound to the myosin head. The change in what is bound to the myosin (ADP and P_i, not ATP) changes the conformation of myosin so that it binds to actin with high affinity. (C) Hydrolysis is complete; myosin releases ADP and P_i. This change in what is bound at the myosin head causes an allosteric change in the head; it flexes. Because the myosin head is still bound to the thin filament, the flexion causes the thin filament to slide past the thick filament. (D) New ATP molecule binds to the myosin head; as for step A, myosin had little affinity for actin in this state, and the head releases from the thin filament and unflexes.



• **Fig. 1.5** Regulation of the actomyosin adenosinetriphosphatase (ATPase) and striated muscle contraction by Ca^{2+}. (A) In the absence of high concentrations of Ca^{2+}, tropomyosin sits in the groove of the actin filament to obstruct the binding sites on actin for myosin. (B) In the presence of higher Ca^{2+} concentrations, the ion binds to troponin, causing an allosteric change in the interaction of troponin with tropomyosin. This allosteric change in turn changes the interaction of tropomyosin with the actin filament to expose the myosin-binding sites on actin.

cross-bridging. (Tropomyosin snuggles down deeper in its actin groove, revealing actin to the myosin head.) As long as troponin binds Ca^{2+}, the muscle contracts by the actomyosin cycle outlined earlier.

Step D: When the Ca^{2+} concentration drops to normal, however, troponin no longer binds Ca^{2+}. This causes tropomyosin to move up in the thin filament groove so that it again blocks the myosin-binding sites on actin. Myosin heads can no longer cross-bridge, and muscle contraction stops.

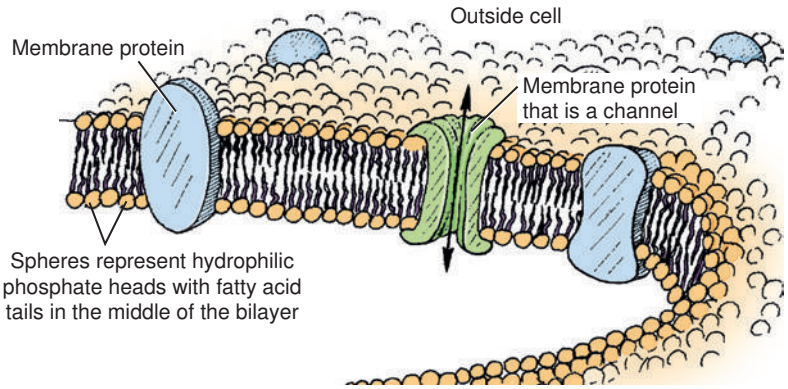
As with the actomyosin force generation itself, its regulation also shows many examples of the specific binding function. The specific binding of Ca^{2+} to troponin is a purely informational use of protein binding and shape change; that is, troponin has no catalytic, transport, or structural function, but transmits the "on" signal to the next protein. The binding of tropomyosin to actin serves not only a regulatory role but also a structural role; the actin filament is stabilized by tropomyosin, making it less likely to disassemble into actin subunits. The change in the binding geometry of tropomyosin that directly regulates myosin access to actin is a good example of the importance of allosteric change and the following sequence: specific binding (troponin to tropomyosin) → protein (tropomyosin) shape change → change in protein-binding properties (tropomyosin to actin) → a difference in the position of tropomyosin, which in turn regulates the actomyosin motor.

Biological Membranes Are a Mosaic of Proteins Embedded in a Phospholipid Bilayer

Before continuing the discussion of the cellular basis of physiological control, an additional basic structure must be introduced. This is the phospholipid bilayer of the biomembranes of cells. *Phospholipids* are molecules that have two long tails of hydrophobic fatty acid and a head containing a charged, hydrophilic phosphate group. Under appropriate aqueous conditions, these molecules spontaneously form an organized membrane structure, similar to the film of a soap bubble. This filmy layer is composed of two layers (a bilayer) of phospholipid molecules. In both layers, the hydrophilic heads point outward to hydrogen bond with water, and the oily, fatty-acid tails point inward, toward one another and away from the water. Proteins embedded in this lipid bilayer, called *intrinsic membrane proteins* or just *membrane proteins,* produce the fluid mosaic structure of biomembranes shown in Fig. 1.6. All biological membranes share this *fluid mosaic* structure, whether the membrane is the outer plasma membrane separating cytoplasm from extracellular fluid or the membrane surrounding intracellular membranous organelles such as endoplasmic reticulum or lysosomes. It is called a fluid mosaic because of the mosaic of proteins among phospholipids, and because the phospholipid layer is fluid; proteins can move around and diffuse within the plane of the bilayer "like icebergs floating in a phospholipid sea" (the apt phrase of S.J. Singer, one of the originators of the model).

Biological membranes are another crucial molecular structure underlying physiological control. The basic fluid mosaic structure serves four broad functions: (1) compartmentation, (2) selective transport, (3) information processing and transmission, and (4) organizing biochemical reactions in space.

Compartmentation is the ability to separate and segregate different regions by composition and function. For example, the lysosome is a membranous organelle within cells that contains hydrolytic (digestive) enzymes that can potentially digest the cell. The lysosomal membrane compartmentalizes these potentially harmful enzymes, segregating them from the bulk cytoplasm. The rigor mortis, mentioned earlier, that begins shortly after death is transitory because, on death, the lysosomes begin to break open, releasing their enzymes, and the actomyosin cross-bridges are eventually digested apart.



• **Fig. 1.6** Fluid mosaic model for biomembranes. Biomembranes consist of a lipid bilayer in which membrane proteins are embedded.

Clearly, the membrane cannot keep a compartment perfectly sealed; material must enter and leave the cell and its internal compartments. *Selective transport* results partly from the properties of the phospholipid bilayer but mostly from transport proteins embedded in the membrane. These proteins are characteristically selective in their transport functions; for example, the protein that is the specialized ion channel underlying neuronal signaling is 15 times more permeable to sodium ions (Na^+) than to potassium ions (K^+). Transport is a major topic of cell physiology and is discussed in more detail later.

If the cells of an organism are to respond to external changes, they must receive information about the state of the outside world. Just as we higher animals have our sensory organs—eyes, ears, nose, and so forth—arrayed on our outside surface, so too do cells have most of their environmental information processing and transmission apparatus on their external surfaces. These are intrinsic membrane proteins of the plasma membrane, called *membrane receptors,* that serve a purely informational function, as discussed earlier.

At first glance it might seem odd that a fluid membrane could provide spatial organization for biochemical reactions. However, returning to the "icebergs in a phospholipid sea" analogy, random collisions are much more likely for material in the two-dimensional membrane surface than for material moving through the three-dimensional volume of the cytoplasm. (If the Titanic had been able to dive or fly, it would have had additional ways to avoid the iceberg!) This much larger collision probability is exploited by the cell in a number of physiological processes. Membranes can also be fenced off into distinct regions, across which there is limited diffusion of membrane proteins. For example, certain cells in the kidney have two membrane regions that are quite distinct with respect to transport proteins, which is important in the regulation of salt and water balance by the animal.

Transport

Only Small, Uncharged Molecules and Oily Molecules Can Penetrate Biomembranes Without the Aid of Proteins

Charged particles (i.e., ions) do not pass through a pure phospholipid bilayer because of the inner, hydrophobic region of bilayer. *Polar molecules* (molecules with no net charge but with electrical imbalances) with a molecular weight greater than about 100 daltons are also unable to pass readily through a pure lipid bilayer, thus excluding all sugar molecules (monosaccharides), amino acids, nucleosides, as well as their polymers (polysaccharide, proteins, nucleic acids). On the other hand, some crucially important polar molecules (e.g., water, urea) are small enough to pass through the lipid bilayer. Small, moderate-sized, and large molecules that are soluble in oily solvents readily pass through a pure lipid bilayer. Physiologically important molecules in this class include O_2, N_2, and the steroid hormones (see Chapters 33 and 34). However, many toxic, synthetic molecules, such as insecticides, are also in this category.

Molecules Move Spontaneously From Regions of High Free Energy to Regions of Lower Free Energy

The majority of biochemicals do not pass readily through a phospholipid bilayer. Transport of this molecular majority requires a protein pathway across the biomembrane. Also needed is a force causing movement along the pathway. Before elaborating on membrane proteins as pathways through the lipid bilayer, the energy factors that drive the transport are considered.

Objects fall spontaneously because of gravity. This is a manifestation of the principle that movement occurs to minimize the potential energy of the object. Indeed, all change in the universe (at scales greater than the subatomic particles) occurs to minimize the potential energy, also called the *free energy,* of the system. The movement of molecules is strongly affected by forces such as concentration, pressure (both part of chemical potential), and voltage (electrical potential). Molecules move spontaneously from a region of higher concentration to lower concentration, from higher to lower pressure, and from higher to lower electrical potential. Each of these factors—concentration, pressure, and electrical potential—is a source of free energy. The transport of a molecule does not depend necessarily on any one factor; rather, the sum of all the free energy contributions is the determinant of transport. The sum of all the free energy contributions on a substance is usually expressed on a per-mole basis as the electrochemical potential. The *electrochemical potential* is the free energy of the substance, from all sources, per mole of the substance.

For spontaneous transport to occur, there must be a difference in the electrochemical potential of the substance between two regions. The two regions are usually two compartments separated by a membrane. This difference in electrochemical potential is called the *driving force.* Typically, students have little difficulty

understanding the direction of spontaneous flow as long as only one factor contributes to the electrochemical potential, pressure, or concentration or the voltage. However, understanding physiological transport, both across cells and across tissues, requires an understanding of the contribution of each factor to the driving force. For example, the flow of fluid from the capillaries of the vascular system depends on the balance between both the hydrostatic pressure difference and the concentration difference of solutes (osmotic pressure) across the capillary. Similarly, movement of Na^+ and K^+ ions across the plasma membrane of nerve cells depends on the driving forces contributed by both voltage differences and ion concentration differences across the membrane.

Material moves spontaneously from regions of high electrochemical potential to low electrochemical potential. Such transport is called *diffusion* or *passive transport.* Net movement of material (i.e., diffusion) stops when the electrochemical difference between regions equals zero. The state at which the free energy or the electrochemical potential difference is zero is called *equilibrium.* Equilibrium means "balance," not equality. Equilibrium is reached when the free energy (electrochemical potential) is balanced; the value on one side is the same as the other. In most cases the source of the free energies on the two sides never becomes equal; the concentrations, the pressure, and the voltages remain different, but their differences "balance out" so that the sum of the free energy differences is zero.

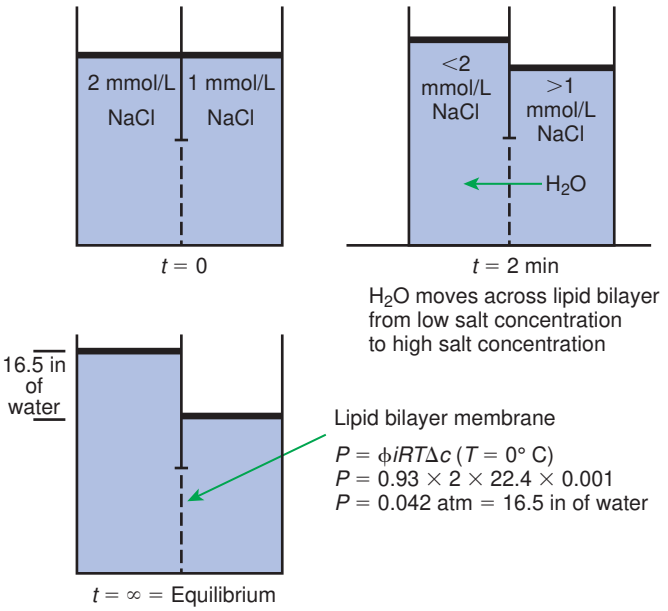
Equilibrium is a particularly important concept because it describes the state toward which change occurs if no work is put into the system. When the system reaches equilibrium, no further net change occurs unless some work is done on the system. The words *net change* are important. Molecules at equilibrium still move and exchange places, but as much goes in one direction as in the other, so there is no net flow of material.

If the cell requires material to move from low to high electrochemical potential (i.e., in the direction away from equilibrium), thus increasing the difference in free energy between two regions, then some driving force, some work, must be provided by some other decrease in free energy. This type of transport is *active transport.* Active transport uses proteins that combine transport and reaction coupling functions; the protein couples the "uphill" movement of material to a "downhill" reaction such as ATP hydrolysis.

Important Transport Equations Summarize the Contributions of the Various Driving Forces

It is worthwhile developing some quantitative aspects of transport, beginning with simple examples and developing equations for the effect of more than one driving force. These equations can be seen as summaries of the physical laws. In most cases the equations describe phenomena with which we have experience by living in a technological society. In these equations, *c* stands for concentration, *V* for volume, *P* for pressure, and so forth; these are common concepts. It is important, however, to think about these equations in real-life terms, not as abstract symbols.

One of these equations relates a hydrostatic (pressure) driving force for water movement that just balances a driving force caused by a chemical potential difference. *Osmosis* is the movement of water across a semipermeable membrane in response to the difference in the electrochemical potential of water on the two sides of the membrane (Fig. 1.7). The chemical potential of water is lower in 1 liter (L) of water (H_2O) in which is dissolved 2 millimoles (mmol) of sodium chloride (NaCl) than in 1 L of H_2O in which is dissolved 1 mmol of NaCl. If these two solutions are separated by a pure lipid bilayer, Na^+ and Cl^- ions cannot move to equilibrate the concentration. Rather, the freely permeable water moves from the side with the higher water potential (low concentration of solute) to the side with the lower water potential (higher concentration of solute). Thus *water follows solute* (a good summary of osmosis), and this water movement dilutes the 2 mmol solution. However, water movement never produces equal concentrations of salt. Rather, another driving force appears as the water moves. The hydrostatic pressure of water increases on the side to which the water moves, increasing the electrochemical potential of the water on that side. Net water movement stops when the increase in water potential from hydrostatic pressure exactly balances the decrease in water potential from the dissolved salt, so that the electrochemical potential becomes equal on both sides of the membrane.



• **Fig. 1.7** Osmosis. At time *(t)* = 0, two compartments are separated by a lipid bilayer membrane (no transport proteins) that contains salt solutions of differing concentrations. At *t* = 2 minutes, the salt ions cannot move across the membrane to equilibrate their concentration, but water can move. Water moves from the region of higher water potential (low salt) to the region of lower water potential (high salt). Water continues to pass the lipid bilayer until at *t* = equilibrium; the difference in the height of water between the two sides creates a difference in pressure equal but opposite to the difference in the water potential between the two sides. That is, the free energy difference resulting from differing salt concentrations is equilibrated by an equal but opposite free energy difference caused by pressure.

The initial potential difference of water shown in Fig. 1.7 is caused by the difference in the concentration of material dissolved in the water. A proper explanation of why the water in a solution has a lower chemical potential than pure water (and why water in a concentrated solution has a lower potential than in a dilute solution) is beyond the scope of this chapter. However, readers familiar with the concept of *entropy* will realize that the disorder of a system increases with the introduction of different particles into a pure substance and with the number of different particles introduced. An analogy would be that a canister with mixed sugar and salt is more disordered, and therefore at higher entropy, than a canister with only pure salt or pure sugar. Also, the disorder of the system increases as more sugar is added to salt (up to 50:50); a pinch of sugar in a canister of salt only increases the disorder

slightly. Because an increase in entropy causes a decrease in free energy, the free energy of a solution is decreased as the mole fraction of solute increases.

Osmosis is important to cells and tissues because, generally, water can move freely across them, whereas much of the dissolved material cannot. Given a concentration difference of some nonpermeable substances, the *van't Hoff equation* relates how much water pressure is required to bring the system to equilibrium, that is, the free energy contributed by a pressure difference across the membrane that exactly balances an opposing free energy contribution caused by a concentration difference:

$$\Pi = iRT\Delta c$$

Π = Osmotic pressure, the driving force for water movement expressed as an equivalent hydrostatic pressure in atmospheres (1 atm = 15.2 lb/in^2 = 760 mm Hg). Osmotic pressure is symbolized by Π to distinguish it from other types of pressure terms.
i = Number of ions formed by dissociating solutes (e.g., 2 for NaCl, 3 for $CaCl_2$).
R = Gas constant = 0.082 L atm/mol degree.
T = Temperature on the Kelvin scale; 0° C = 273° K. (RT is a measure of the free energy of 1 mol of material because of its temperature. At 0° C, RT = 22.4 L atm/mol.)
Δc = Difference in the molar concentration of the *impermeable* substance across the membrane.

This equation summarizes a balance of driving forces; P amount of hydrostatic (osmotic) pressure is the same driving force as a particular concentration difference, Δc. The osmotic pressure depends only on the concentration difference of the substance; no other property of the substance need be taken into account. (Those phenomena that depend only on concentration, such as osmotic pressure, freezing-point depression, and boiling-point elevation, are called *colligative properties.*) The van't Hoff equation is strictly true only for ideal solutions that are approximated in our less-than-ideal world only by very dilute solutions. Real solutions require a "fudge factor," called the *osmotic coefficient,* symbolized by Φ (phi). The osmotic coefficient can be looked up in a table, and then plugged into the equation as follows:

$$\Pi = \Phi iRT\Delta c$$

The term Φic for a given substance represents the osmotically effective concentration of that substance and is often called the *osmolar* or *osmotic concentration,* measured in osmoles per liter (Osm/L). In general, the osmolar concentration of a substance is approximated by the usual concentration times the number of ions formed by the substance; the osmotic coefficient provides a small correction. The osmolarity of a 100 mmol NaCl solution (0.1 mol) is then 0.93 (Φ for NaCl) × 2(NaCl → $Na^+ + Cl^-$) × 0.1 mol = 0.186 Osm = 186 mOsm.

The previous equation summarizes a phenomenon crucial for physiological function. The greater the concentration difference of an impermeable substance across a membrane, the greater is the tendency for water to move to the side of high concentration. (Water follows solute.) Indeed, if you plug some numbers into this equation, you may be surprised at the large pressures required to balance modest concentration differences. For example, an NaCl concentration difference of 0.1 mol (5.8 g/L) is equilibrated by a pressure (4.2 atm) equal to a column of water 141 feet high (divers must be wary of the bends when ascending from below 70 feet of water). The importance of this is that a small concentration difference can produce a strong force for moving water. The body makes effective use of this to transport water in many tissues: ions/molecules are transported into or out of a compartment → and water follows by osmosis.

Starling's Hypothesis Relates Fluid Flow Across the Capillaries to Hydrostatic Pressure and Osmotic Pressure

An excellent practical example of how a balance of driving forces is responsible for the flow of water and permeable substances across a semipermeable membrane is the movement of water and ions across the single layer of cells (endothelial cells) that compose blood capillaries. The single cell layer composes, in effect, a semipermeable membrane with different transport qualities than that of a simple lipid-bilayer membrane. The junctions between cells are sufficiently permeable to allow small molecules and ions to diffuse between compartments. Only large molecules, most importantly proteins, are unable to move through the holes. The difference in protein concentration between the blood and the water solution surrounding tissue cells, called the *extracellular fluid (ECF)* or *interstitial fluid (ISF),* creates an osmotic pressure for the movement of water with all its dissolved small molecules and ions. This osmotic pressure resulting from dissolved proteins has a special name: c*olloid osmotic pressure* or *oncotic pressure.* Protein is more concentrated in the blood than in the ISF, producing an oncotic pressure of about 0.02 to 0.03 atm = 15 to 25 mm Hg, driving water into the capillary. On the basis of this driving force alone, one would expect the capillaries to fill up with water, thus dehydrating the tissue spaces. However, the heart is a pump that exerts a true hydrostatic pressure on the blood, tending to drive the water (and other permeable molecules) out of the capillaries. The net driving force is the algebraic sum of the oncotic pressure difference and hydrostatic pressure difference between the capillaries and the ISF, as follows:

$$\text{Net driving force in capillary} = (P_c - P_i) - (\pi_c - \pi_i)$$

P_c = Hydrostatic pressure in the capillary.
P_i = Hydrostatic pressure in the interstitial space (usually near 0).
π_c = Oncotic pressure of blood plasma in capillary (~28 mm Hg).
π_i = Oncotic pressure of ISF (~5 mm Hg, but depends on the particular tissue).

This equation has enormous relevance to the function of the circulatory system. On the arterial end of capillaries, the hydrostatic pressure (P_c) is high, about 35 mm Hg. Plugging this number into the equation along with the others, the net pressure in the capillary is +12 mm Hg; fluid is being driven out of the capillary on the arterial side *(capillary filtration).* The flow of fluid through the resistance of the capillary causes a decline in pressure so that the hydrostatic pressure on the venous side is low, P_c = 15 mm Hg. The oncotic pressures have not changed, so the net driving force on the venous side is −8 mm Hg; there is a net absorption of fluid into the capillary on the venous side *(capillary reabsorption).* This arrangement achieves a major function of the circulatory system; in this way the fluid of the blood circulates among the cells and is then recycled back into the circulatory system.

Pathological alterations in this system emphasize the physiological importance of balance of driving forces for transport. Chronic liver disease occurs with some frequency in horses and dogs, among other mammals. The liver is compromised in its ability to synthesize and secrete a major blood protein, serum albumin. The decline in

TABLE 1.1 Concentrations of Various Substances in the Intracellular, Extracellular, and Plasma Fluids

	CONCENTRATION (MMOL/L)		
	Intracellular	Extracellular	Blood Plasma
Na^+	15	140	142
K^+	150	5	4
Ca^{2+}	0.0001	1	2.5
Mg^{2+}	12	1.5	1.5
Cl^-	10	110	103
HCO_3^-	10	30	27
Phosphate	40	2	1
Glucose	1	5.6	5.6
Protein	4.0	0.2	2.5

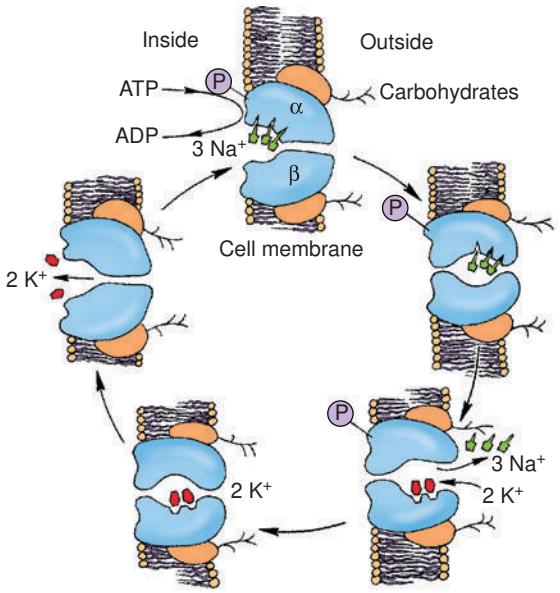
the concentration of serum albumin lowers the oncotic pressure of the blood. As a result, there is more force to drive fluid out of the capillaries on the arterial side and less driving force for net absorption of fluid on the venous side of capillaries. This causes the tissue spaces of the diseased animals to fill with fluid, a painful and visually obvious symptom called *edema.* The Clinical Correlations section at the end of the chapter provides another example of edema in which increased hydrostatic pressure in the veins and capillaries causes increased capillary filtration and less capillary reabsorption.

Membrane Proteins That Serve the Triple Functions of Selective Transport, Catalysis, and Coupling Can Pump Ions and Molecules to Regions of Higher Free Energy

The van't Hoff equation and Starling's hypothesis deal with passive transport (i.e., movement of material in the direction of lower electrochemical potential). However, the cell moves many ions/molecules against their electrochemical potential. That is, this selective transport requires the expenditure of energy by the cell. Transport in a direction requiring an expenditure of energy (i.e., input of work) is called *active transport.* Active transport depends on intrinsic membrane proteins that use specific binding and allostery to achieve the dual functions of selective transport and reaction coupling. Many (but not all) active transport proteins obtain the energy for transport from ATP hydrolysis; these proteins must function also as enzymes (ATPases).

An important example of active transport is the Na^+, K^+ pump (also known as Na^+, K^+-ATPase). This intrinsic membrane protein consists of four polypeptide chains (2 α + 2 β) and has a mass of approximately 300,000 daltons. This molecule catalyzes the hydrolysis of ATP and couples the hydrolysis energy to the movement of Na^+ out of the cell and K^+ into the cell. This ion pump creates and maintains a considerable concentration gradient across the cell membrane for both ions (Table 1.1).

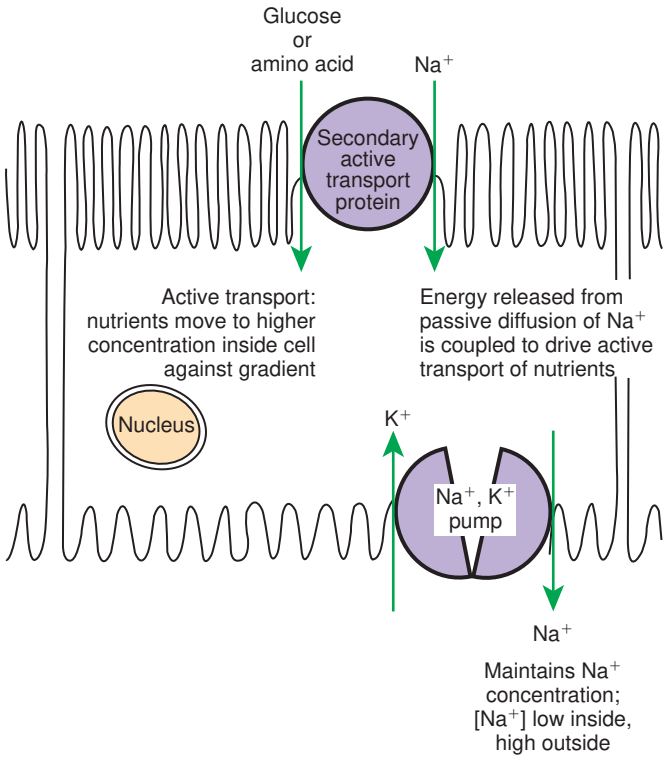
Fig. 1.8 shows our current understanding of this protein's structure and outlines the cycle of binding and conformational changes underlying its transport function. The Na^+, K^+-ATPase



• **Fig. 1.8** Hypothetical transport cycle for Na^+, K^+-ATPase. Changes in the conformation of this transport protein driven by ATP hydrolysis and ion-binding events cause three Na^+ ions to be moved out of the cell against a concentration gradient and two K^+ ions to be moved into the cell, also against a concentration gradient, for each ATP hydrolyzed. *ADP,* Adenosine diphosphate; *ATP,* adenosine triphosphate. (Redrawn from a diagram by Dr. Seth Hootman.)

pumps three Na^+ ions *out* of the cell and two K^+ ions *into* the cell for each ATP molecule hydrolyzed. These directions of ion pumping cause a high Na^+ concentration outside the cell and a low concentration inside, whereas K^+ concentration is high inside and low outside the cell. The different directions of pumping for the two ions depend on differing binding specificity of the pump protein in the two conformational states. The ability of the protein to couple this transport to the enzymatic breakdown of ATP allows the transport to occur against the concentration gradients, from lower to higher electrochemical potentials for both ions. In the particular case of the Na^+, K^+ pump, the number of transported electrical charges is asymmetrical; three positive charges leave for each two positive charges that enter. This asymmetry of electrical charge transport means that the Na^+, K^+ pump is *electrogenic,* making a minor contribution to the electrical potential (voltage) across cell membranes, as discussed later.

Many different intrinsic membrane proteins actively transport a wide variety of ions and molecules against the transported molecules' electrochemical gradient. Many, such as the Na^+, K^+ pump, couple the energy-requiring "uphill" transport with the "downhill" hydrolysis of ATP. However, any potential source of free energy can be coupled to the energy-requiring transport. Indeed, the gradient of Na^+ set up by the Na^+, K^+ pump is itself used frequently as a source of energy. That is, the "downhill" flow of Na^+ from outside the cell to the inside is a spontaneous reaction whose energy can be coupled to some "uphill" reaction (Fig. 1.9). For example, the transport of glucose and many amino acids from the food mass in the small intestine into the cells lining the gut is an active transport process and requires an Na^+ concentration gradient. Transport proteins in the plasma membrane of intestinal epithelial cells couple the spontaneous diffusion of Na^+ into the cell to the inward, energy-requiring transport of the sugar or amino acids. These nutrients are at higher concentration inside the cell



• **Fig. 1.9** Secondary active transport as exemplified by uptake of nutrients by gut epithelia. Nutrients such as glucose and amino acids must be actively transported from relatively low concentration in the gut lumen toward higher concentrations within the cells lining the gut. This active transport process uses the concentration gradient of Na^+ ions set up by Na^+, K^+-ATPase (see Fig. 1.8) as the source of energy for the active transport process. In other words, the energy released by the passive diffusion of Na^+ into the cell along its concentration gradient is coupled to the energy-requiring transport of glucose or amino acids against their concentration gradients. Thus the secondary active transport protein both serves a transport function and couples the "downhill" transport of Na^+ to the "uphill" transport of nutrients. There are many such secondary active transport processes in the body. For example, the same mechanism shown here is used to reabsorb nutrients from blood filtrate in the kidney. *ATPase,* Adenosinetriphosphatase.

than outside, so they must be actively transported into the cell at the expense of the energy stored in the Na^+ electrochemical gradient. That is, the energy from the "downhill" diffusion of Na^+ into the cell is coupled to the "uphill" transport of the nutrient into the cell. Such active transport coupled to Na^+ diffusion across the cell membrane is called *secondary active transport* because of its dependence on the Na^+ concentration gradient established by the primary active transport of the Na^+, K^+ pump.

Examples of transport can be identified in a number of ways. Our examples have been instances in which two ions/molecules must be transported together or not at all, and such transport is called *cotransport.* Cotransport can involve one process of passive transport (diffusion) with an active transport process, as in the two previous examples; it can involve two active transport processes (e.g., Na^+, K^+-ATPase) or two diffusion processes. In the first case, the need for cotransport is energetic; the flow of one ion is needed to drive the other. In the two latter cases, the need for cotransport is a restriction based on the binding properties of the transport protein; it cannot bind one without the other. Cotransport proteins that transport both substances in the same direction are called *symports* or *symporters.* The Na^+/sugar cotransporter in the gut is a symport. Cotransport proteins that transport the two substances in opposite directions (e.g., Na^+, K^+-ATPase) are called *antiports* or *antiporters.* Parenthetically, proteins that transport only a single ion or molecule are called *uniports* or *uniporters.*

Many Membrane Proteins Selectively Facilitate the Transport of Ions/Molecules From High to Low Electrochemical Potential

The movement of ions and of medium and large polar molecules requires a protein molecule to serve as a pathway through the obstruction of the phospholipid bilayer. If the movement of the substance is in the natural direction of its electrochemical gradient (movement from high to low), the transport process is called *facilitated diffusion.* The membrane proteins mediating this transport process through the phospholipid bilayer are *channels* or *carriers* (Fig. 1.10). These are distinguished by the extent to which the protein interacts with the transported substance.

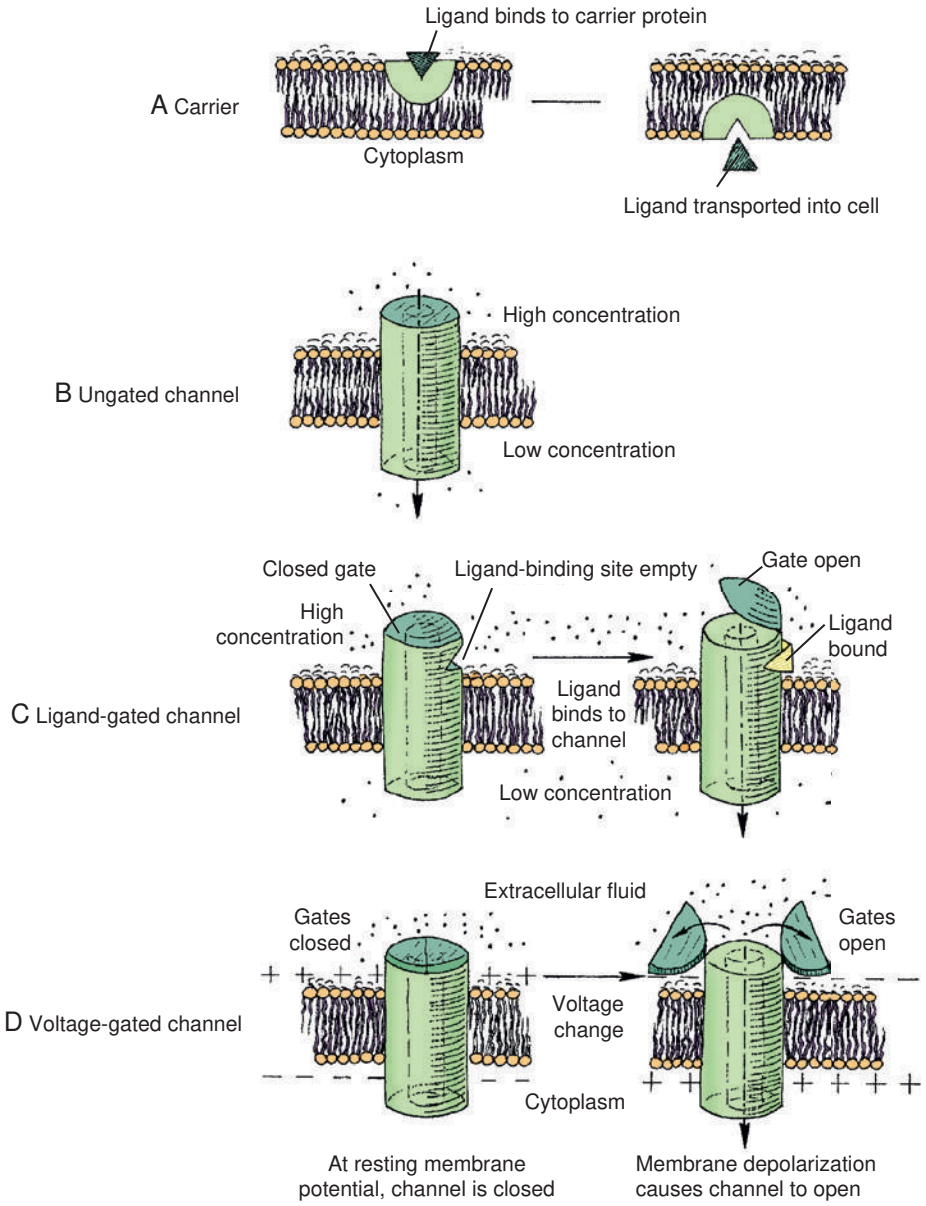
Carriers bind the transported substance in the lock-and-key manner, so there is a site-specific binding of the transported substance to the transport protein (see Fig. 1.10A). Carrier-mediated transport is typically much slower than channel-mediated diffusion because of the relatively slow binding and unbinding processes. The Na^+, K^+ pump and the Na^+/glucose symport are both examples of carriers.

Channels can be thought of as "protein donuts" embedded in the phospholipid bilayer. The hole in the donut is a pore in the membrane through which small ions such as Na^+, K^+, Ca^+, Cl^-, and H^+ are transported. Although most channels transport ions, a class of channels called *aquaporins* comprises channels for water flow. (Although water can flow through a pure lipid bilayer, this transport is too slow for some functions. Kidney cells, for example, are particularly rich in aquaporins, which are required for the water balance function of the kidney.) For all channels, the pore size and the interaction of the transported material with the amino acid side groups lining the pore allow membrane channels to be selective. Only specific molecules or ions can move through a particular channel. Movement of material through channels is almost as rapid as simple diffusion through a water-filled space of the same area as the channel pore.

The plasma membranes of most cells have passive leaks of ions, particularly K^+. These ionic leaks are typically ascribed to *leak channels,* which are open at all times (see Fig. 1.10B). However, most ion channels open or close in response to signals. These latter types are called *gated channels.* The opening and closing of the gates are examples of the allosteric property of proteins. The same signals responsible for allosteric changes in general—ligand binding, phosphorylation, and voltage differences—also control the opening and closing of gated channels, as shown in Fig. 1.10. (Because mechanically gated channels are so poorly understood, these are not discussed here.)

Channels that open in response to ligand binding are called *ligand-gated channels* (see Fig. 1.10C). The nicotinic acetylcholine receptor is a ligand-gated channel found in skeletal muscle membrane directly beneath incoming neurons (nerve cells). This channel is found also in the membrane of neurons in autonomic ganglia and in the brain. As the name implies, the nicotinic acetylcholine receptor binds to the drug nicotine and the neurotransmitter acetylcholine. In both cases, the channel opens in response to ligand binding.

This nicotinic acetylcholine channel plays a key role in transmitting electrical stimulation from neurons to skeletal muscle cells. Briefly, motor neurons release the neurotransmitter acetylcholine



• **Fig. 1.10** Types of transport proteins mediating facilitated diffusion. In all cases the ion moves from a region of high potential (shown here as high concentration) to a region of low potential. (A) Carriers. In a few cases, material is carried by a transport protein that binds tightly to the material, and the complex moves through the lipid bilayer. (B) Leak channels. These channels are thought not to open and close as do gated channels, and thus they support a small but persistent leak of a particular ion through the pore. Although their existence was long postulated, distinct, ungated leak channels have only recently been identified and isolated, as opposed to leaks through normally gated channels. Selectivity of these and other channels is based on the size of the pore and the weak interactions of ions with the atoms lining the pore. (C) Ligand-gated channels. The transport protein again forms a pore through the membrane. In the case of gated channels, access of the ion to the pore is controlled by a gate, a substructure of the transport protein that can open and close the pore. In ligand-gated channels the opening and closing of the gate are controlled by the binding of a ligand to the channel. (D) Voltage-gated channels are similar to ligand-gated channels, except the opening and closing of the gate are controlled by the electrical field around the channel.

in response to the electrical signal coming down the neuron. This acetylcholine binds to and opens the ligand-gated channel on the skeletal muscle. The influx of Na^+ into the muscle cell initiates an electrical response in the muscle, causing the release of Ca^{2+} (through gated channels in the endoplasmic reticulum), in turn causing contraction. (This brief account of neuromuscular transmission, presented only to provide orientation to the function of the acetylcholine channel, is expanded in Chapters 5 and 6). In the case of the nicotinic acetylcholine receptor/channel, the specific binding and allosteric properties of the protein serve the dual functions of selective transport across the membrane and information reception and transmission to the muscle cell.

Channels that open in response to voltage changes across the membrane are called *voltage-gated* or *voltage-dependent channels* (see Fig. 1.10D). This type of channel is largely responsible for the neurons' ability to transmit information along their length and to release neurotransmitter. All voltage-gated channels have a range of membrane potentials that cause them to open; this is the channel's activation range. The minimum membrane potential that causes opening is the channel's *threshold.* The activation range and threshold vary from channel to channel, depending on the conformation of the protein and the electrical properties of the amino acid side groups that form the gate of the channel. In addition to an open and closed configuration, many voltage-dependent channels have a third conformation, called *inactivated.* Like the closed configuration, the inactivated conformation prohibits the diffusion of ions through the channel. Unlike the closed configuration, it does not open immediately in response to changes in membrane potential. Inactivation can be regarded as an enforced rest period for the channel. Voltage-dependent channels that do not inactivate have only open and closed conformations, and they take up one or the other conformation, depending on the membrane potential.

As previously discussed, any of the functions of proteins can be used to transmit information if a difference in the protein function changed the cell. Gated channels, both ligand and voltage gated, are ideal candidates for information transmission because they change their function: opening and closing, permitting or stopping transport. Indeed, the sole physiological function of the nicotinic acetylcholine receptor/channel, as described earlier, is the transmission of information: turning the chemical stimulation by the neuron of the muscle into electrical stimulation (see following discussion) of the muscle membrane, leading to muscle contraction.

Passive Transport of K^+ Across the Plasma Membrane Creates an Electrical Potential

As just discussed, gated ion channels can convert chemical information into electrical information. Electrical signaling in the animal body is the result of electrical imbalances maintained across the plasma membrane of virtually all cells: cells maintain an electrical potential difference across their plasma membrane. That is, the cell membrane is a battery; if one attaches electrodes to the two ends of a battery or to the inside and outside of a cell, one finds a voltage difference between the two ends or sides. If one provides a path for electrical charges to move—a metal wire containing free electrons in the case of a battery or a membrane channel through which ions can move in the case of the cell—an electrical current flows from higher to lower electrical potential. The diversity of battery-powered devices in our society suggests how many ways this electrical potential can be exploited. The physiology of animals also exploits the baseline electrical potential across the plasma membrane, called the *resting membrane potential.* The word "resting" is added to distinguish the baseline potential from the instantaneous values of membrane potential during the passage of membrane currents.

The resting membrane potential is the indirect result of the concentration gradients of ions across the plasma membrane caused by the activity of the Na^+, K^+-ATPase. Partly, this membrane potential is a result of the asymmetry in numbers of ions pumped by the Na^+, K^+-ATPase. However, most of the membrane potential is caused by the passive flow of K^+ through *K^+ leak channels* in response to the concentration gradient of K^+ (high inside, low outside). This concentration gradient sets up an electrical driving force (voltage) that exactly balances the concentration driving force. The concentration of K^+ inside a mammalian cell is about 150 mmol; outside in the ISF, it is about 5 mmol. As a result, K^+ tends to diffuse from the cytoplasm through the leak channel to the ISF. However, when K^+ alone leaves the cytoplasm without an accompanying negative ion, it causes an electrical imbalance. The exit of K^+ ions leaves the inside of the cell with negative charges not neutralized by positive potassium ions, and the ISF now has positive K^+ ions not balanced by negative charges. The cell is building an electrical potential difference across the plasma membrane with the cytoplasm being negative relative to the ISF.

This electrical potential driving force increases until it balances the concentration driving force for K^+. This situation is analogous to osmosis; the concentration-driven flow of water across a semipermeable membrane creates a different driving force, pressure, that eventually balances the concentration-driving force. Similarly, for the resting membrane potential, the concentration-driven flow of K^+ across the semipermeable membrane (semipermeable in the sense that negative ions do not accompany the K^+) creates a different driving force, an electrical voltage, that eventually balances the concentration force. As in the case of osmosis, an equation is used to relate the size of the concentration gradient to the size of the electrical potential that provides an exact balance. This equation is called the *Nernst equation,* as follows:

$$E_X = RT/zF \ln[X_{outside}]/[X_{inside}]$$

E_X = Equilibrium potential for ion X
RT = Gas constant × Absolute temperature
z = Electrical valence for the ion, +1 for Na^+ and K^+, −1 for Cl^-, and so forth
F = Faraday constant = number of coulombs of electrical charge in a mole of ions = 96,500 coulombs/mol
ln = Natural logarithm (i.e., log to base e)
[X] = Concentration of ion X

A simpler form of this equation can be written by taking advantage of the fact that R and F are constants, T is almost constant under physiological conditions, and the natural log of a number is 2.3 times the common log ($\log_{10}$), as follows (mV, millivolts):

$$E_X = -60 \text{ mV}/z \log[X_{inside}]/[X_{outside}]$$

Because the state of balance between the electrical driving force and the concentration driving force is equilibrium the value of the electrical potential is called the *equilibrium potential* of the ion. Given the previous concentrations for K^+ inside (150 mmol) and outside (5 mmol) the cell, the equilibrium potential for K^+ is

$$\begin{aligned} E_{K+} &= -60 \text{ mV}/+1 \times \log 150/5 \\ &= -60 \text{ mV} \log 30 \\ &= -60 \text{ mV} \times 1.47 \\ &= -88.2 \text{ mV} \end{aligned}$$

Indeed, the measured resting membrane potential across a human muscle cell is −90 mV.

Several aspects of this important equation are worth discussing. If the equilibrium potential for a particular ion is the same as the measured membrane potential, the net driving force for the ion is zero. In this case there is no net movement, even in the presence

of wide-open channels, to provide a path through the membrane. However, for any gradient of a specific ion, if the measured membrane potential is *not* the equilibrium potential of that ion, there is a driving force for the transport of that ion. That is, when the membrane potential is anything other than the equilibrium potential, that ion will flow across the membrane if an appropriate channel is open. Thus the equilibrium potential for an ion provides a "baseline" for comparison with the actual membrane potential to determine whether an ion will tend to move across the plasma membrane. If the measured membrane potential has the same sign but is larger in magnitude than the equilibrium potential, the ion flows in the direction of the electrical potential. If the sign is the same but the magnitude lower, the concentration-driving force determines the direction of flow of the ion. If the measured potential has the opposite sign of the equilibrium potential, both electrical and concentration forces are acting on the ion in the same direction. Flow of ions across the plasma membrane (i.e., electrical current) in response to the balance of force between concentration and voltage produces the electrical changes in neurons that underlie the nervous system, as discussed in Chapter 4.

It would be reasonable but incorrect to assume that the transport of ions required to set up the electrical potential measurably alters the concentration gradient. This is untrue because of the large amount of energy required to separate electrical charges. The separation of charge arising from the transport of a few ions balances the energy of quite substantial concentration gradients. Indeed, so few ions move that they cannot be measured by chemical means. Thus electrical, not chemical, measurements are used routinely to assess transport of ions in cells. The measurable voltage changes caused by immeasurably small concentration changes of ions means also that the electrical phenomena at the membrane persist for many hours, even if the Na^+, K^+-ATPase is inactivated by a toxin. That is, an existing concentration gradient of K^+ would require hours to dissipate at the rate of K^+ leakage characteristic of the plasma membrane. Using the membrane battery analogy, the Na^+, K^+-ATPase is a battery recharger. A smartphone does not require the minute-to-minute services of a battery recharger. Enough energy is stored in the battery to operate the phone for an appreciable period, although the battery recharger is needed eventually. Similarly, enough energy is stored in the K^+ concentration gradient to maintain the membrane potential for a period of time. The Na^+, K^+-ATPase is not required on a minute-to-minute basis, although it is needed ultimately to maintain the concentration gradient on which the resting membrane potential depends.

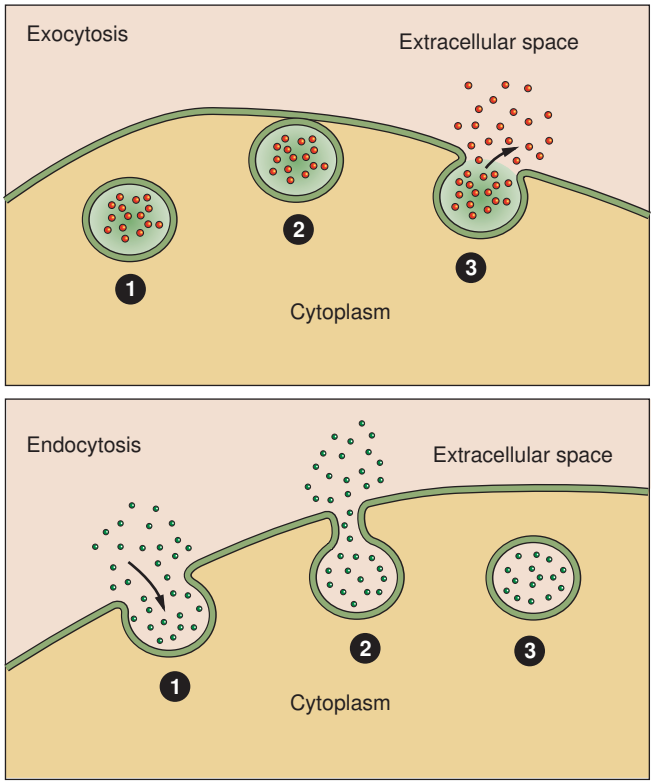
Spatial Organization of Active and Passive Transport Proteins Enables Material to Pass Completely Through the Cell

Although macromolecules and biomembranes clearly underlie physiological function, many phenomena of the intact animal emerge that are not initially apparent as a simple sum of parts. One interesting example is the spatial organization of plasma membrane transport proteins so that ions move across the cell from one ECF compartment to another. This is called *transcellular transport* or, because it typically occurs across a layer of epithelial cells, *epithelial transport.* This epithelial transport is important in the kidney (see Chapter 42). The plasma membrane of the epithelial cells in the proximal tubules of the kidney contains two distinct regions. The apical membrane regions face the lumen of the tubule and the fluid that will become urine, and the basolateral regions are near the capillaries and the blood. The apical surface contains ungated leak channels for Na^+, whereas the basolateral surface contains Na^+, K^+-ATPase molecules. (The membrane proteins in one region are prevented from diffusing into the other by membrane protein "fences" called *tight junctions.*) Na^+ diffuses into the cell on the apical surface from the urinelike fluid driven by both the concentration gradient and the resting membrane potential. When inside the cell, the Na^+ is pumped from the basolateral surface, essentially into the blood, by the Na^+, K^+-ATPase. This allows the kidney to reabsorb and thus conserve Na^+. As long as the Na^+, K^+-ATPase remains restricted to the basolateral surface and the passive channel to the apical membrane, Na^+ can move through the cell from the urinelike fluid in the tubule to the blood in the capillaries. If either protein should lose its spatial restriction, Na^+ would be transported into and out of the cell on the same surface, merely consuming ATP, with no net transport of Na^+ from lumen to capillary.

Membrane Fusion Allows for a Combination of Compartmentalization and Transport of Material

Impermeable molecules can be transported across the cell membrane by means other than membrane proteins. This method involves using membrane itself as a carrier compartment. The lipid bilayer of biological membranes has a structure similar to soap bubbles. As with soap bubbles, small vesicles of biomembrane (essentially "membrane bubbles") can fuse to form larger membrane surfaces. A large membrane surface can also pinch off (requiring fusion of two membrane surfaces) into small vesicles. When these processes occur at the plasma membrane, they are called *exocytosis* and *endocytosis,* respectively (Fig. 1.11). More generally, pinching off of membrane or fusion of two membrane vesicles (e.g., for internal membranes) is referred to as *membrane fusion,* whatever the direction. Membrane fusion underlies much of membrane vesicle traffic around the cell. This traffic creates intracellular vesicles, renews plasma membrane by adding newly synthesized membrane, and transports material within the cell and across the plasma membrane. Because the transport is compartmentalized within a membrane bubble, the transported material can be targeted specifically to one or another region of the cell. Also, changes to the "cargo" can occur within a particular membrane compartment, as seen in cholesterol transport.

Exocytosis and endocytosis are crucial in the transport of cholesterol (Fig. 1.12). *Cholesterol* is an essential lipid component of many animal biomembranes; the plasma membrane lipids of animals are about 15% cholesterol and 60% phospholipids. Cholesterol is also the starting material for the synthesis of the entire group of hormones called *steroids* (see Chapters 33 and 34). Cholesterol can be synthesized by animals and is also absorbed by meat-eating animals from their diet. Because cholesterol is soluble in oil, it passes from food through the plasma membrane without protein mediation into the cells of the gut lining. However, transport of dietary cholesterol through the circulatory system requires that cholesterol molecules form a complex with a protein molecule to create low-density lipoproteins (LDLs). To take up cholesterol from the circulation, cells bind the LDLs to intrinsic membrane proteins that act as LDL receptors, as shown in Fig. 1.12. The receptor/LDL complex then diffuses in the plane of the membrane into specific regions to form coated pits. The coated pit is taken into the cytoplasm by endocytosis. In addition to the transport function, *receptor-mediated endocytosis* functions also to concentrate extracellular material before internalization. The coated pit is not



• **Fig. 1.11** Two membrane fusion processes: exocytosis and endocytosis. (*Top*) In exocytosis, a membrane-bound vesicle from the cytoplasm (*1*) makes contact and fuses with the plasma membrane (*2*). As the vesicle membrane becomes continuous with the plasma membrane, the contents of the vesicle are released to the extracellular space (*3*). (*Bottom*) In endocytosis, some material from the extracellular space is surrounded by plasma membrane (*1*), which continues to invaginate until the edges are able to fuse (*2*), thus pinching off a vesicle from the plasma membrane (*3*). Membrane fusion can occur between any two compartments within cells separated by lipid bilayer membrane, not only between the cytoplasm and extracellular space, as shown here.

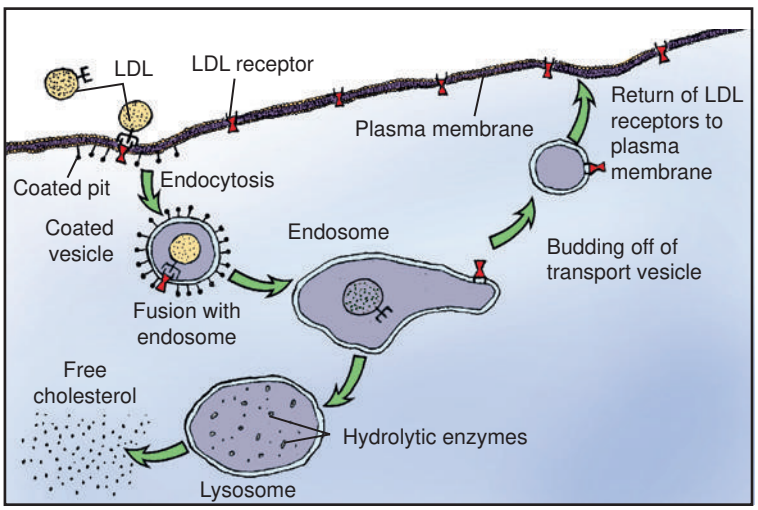
taken into the cell until it has collected the LDLs from a much larger volume of ECF than the cell could "drink." The membrane vesicles formed by this endocytosis fuse subsequently to become an *endosome*. The endosome compartment becomes acidic, which causes dissociation of the LDL and the receptor. Through unknown means, the endosome is then able to further separate and compartmentalize the receptor from the LDL. Membrane vesicles containing the now-vacated LDL receptors return to the plasma membrane and fuse by exocytosis. The LDL receptor is recycled to the plasma membrane to pick up more LDL. Experimental evidence suggests that a single LDL receptor molecule can cycle between the plasma membrane and endosomal vesicles more than 100 times before losing its activity. Meanwhile, the LDL moiety is segregated to another endosomal vesicle, which fuses with the lysosome. The lysosome contains hydrolytic enzymes, thus allowing the internalized LDL to be digested. The cholesterol is now available to the cell for steroid synthesis or incorporation into membrane.

Other molecules are also recycled by endocytosis. As with the LDL receptor, for example, many signal receptors, discussed in the next section, are endocytosed back into the cell that released them, saving the cell the effort of synthesizing new receptors. Not all endocytosed molecules are recycled. Many are broken down after their endosome fuses with a lysosome. Indeed, as described later, this is one method of regulating receptor number on the plasma membrane.

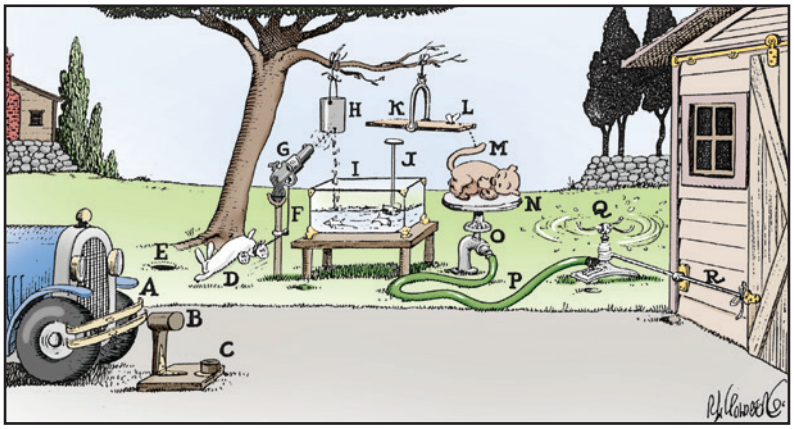
Information Transmission and Transduction

Cell Signaling Usually Occurs by a Lengthy Chain of Sequential Molecular Interactions

One of the areas of most rapid progress in cellular physiology has been our understanding of the mechanisms by which extracellular signals, such as hormones, growth factors, and neurotransmitters, alter cellular function, which in turn alters tissue, organ, and animal function. At the molecular level, almost all chemical signaling



• **Fig. 1.12** Processes of membrane fusion involved in cholesterol uptake by cells. Starting at the left, a low-density lipoprotein (LDL)–containing cholesterol binds to an LDL receptor protein of the plasma membrane and undergoes endocytosis, forming an endosome. The receptor is detached from its LDL ligand in the endosome. The LDL portion of the endosome fuses with a lysosome to digest the LDL and produce free cholesterol, while the receptor-containing portion of the endosome pinches off a vesicle to return to the plasma membrane, thus recycling the receptor. (Redrawn from Alberts B, Bray D, Lewis J, et al. *Molecular Biology of the Cell*, New York: Garland; 1983.)



• **Fig. 1.13** Rube Goldberg device (garage door opener, circa 1928) as an analogy for the complex cause-and-effect sequence characteristic of cellular chemical signaling. Automobile (*A*) drives into driveway, causing hammer (*B*) to ignite toy cap (*C*), frightening rabbit (*D*) into its burrow (*E*), and causing pistol (*G*) to fire and so forth, ultimately leading to the opening of the garage door (*R*). As explained in the text, this whimsical "machine" serves as an analogy for chemical signaling within cells because of the multiple control elements, their connection as a cause-and-effect sequence, and the use of household items, similar to the use of evolutionarily conserved proteins of cells in signaling.

shares a common "strategy" of mechanism; signals are sent as a long chain of chemical cause-and-effect interactions transmitted between many sequential chemical steps. Indeed, chemical signaling pathways are structured similar to the whimsical "machines" in the cartoons of Rube Goldberg (1883–1970), a famous American newspaper cartoonist. Fig. 1.13 shows one of his cartoons from 1928 illustrating an outlandish contraption (a *Rube Goldberg device*) to serve as an automatic garage door opener, realistic versions of which had not yet been invented. The automobile (A) drives in, causing a hammer (B) to ignite a toy cap gun (C), which frightens a rabbit (D) with a string (F) tied to its leg, thus firing a pistol (G), and so forth, until a connection to a rotating water sprinkler causes the carriage-house door to slide open (overhead doors also had not yet been invented). Although much of the humor of this parody of a machine is lost on us (our attitudes about machines have changed greatly since Goldberg's heyday), Rube Goldberg devices are a surprisingly useful analogy to the overall mechanism of cellular chemical signaling.

Just as the garage door opener of Fig. 1.13 depends on a series of sequential cause-and-effect interactions, so chemical signaling occurs by a series of cause-and-effect changes in protein shape and binding. Just as the complex events of Goldberg's device are linked to signal and to actuate a garage door opening, so a cascade of changes in protein shape and function are linked to signal and actuate physiological events. Our earlier example of muscle contraction illustrates such a pathway of cause and effect and the analogy to Rube Goldberg devices. Electrical excitation (A) of a muscle cell increases intracellular Ca^{2+} concentration (B), causing Ca^{2+} to bind to troponin (C). This in turn alters the binding of tropomyosin (D) to actin (E) allowing the myosin heads (F) to bind to actin, thus leading to cross-bridging (G) and hydrolysis of ATP and contraction.

As this example indicates, the sequence of cause and effect for both chemical signaling and Rube Goldberg devices is complex. Both involve many different elements, none of which can be identified as *the* controller; all the elements are involved in control. Importantly, this creates multiple sites for regulation and for therapeutic drug action. Just as increasing the caliber of the pistol in the garage door opener would change the response time for opening, so a drug that bound to an element in the middle of a signaling pathway in a cell could increase or decrease the final physiological change in response to a particular hormone, for example. Also related to complexity is that the chain of cause and effect is not obvious; the particular sequence connecting a particular signal (adrenaline binding to a receptor on heart muscle) to a particular outcome (increased cardiac output) must be memorized. However, when the sequence is understood, you can predict from the state of one element in the chain what should happen next. Finally, Rube Goldberg devices were cobbled together from reasonably common household items, such as the bucket, fish tank, sprinkler, and even pistols. Similarly, the elements of chemical-signaling pathways are often highly conserved, and you will see throughout your studies that the same molecules or same basic types of molecules are used in a wide variety of different stimulus-response pathways.

Signaling Pathways Begin With the Binding of an Extracellular Molecule to a Receptor

In addition to the Rube Goldberg–like sequence of signal pathways, another aspect of the overall "strategy" of cellular information transmission is that signaling pathways almost always begin with the environmental signal molecule binding to a protein molecule specialized for information transfer, called a *receptor.* The LDL receptor discussed earlier is involved in the transport of material into cells (see Fig. 1.12). However, most other receptors are proteins whose task is to transmit and transduce information to the cell from the extracellular environment. Receptors distinguish among the large number of external signaling molecules (e.g., various hormones, neurotransmitters, growth factors) through the usual protein mechanism of highly specific binding.

Three broad classes of receptors, called *receptor families,* are particularly important in physiological function and are discussed in this chapter and Chapter 2. Two of these families, the *G-protein–coupled receptors* (GPCRs) and the *receptor tyrosine kinases* (RTKs), are intrinsic membrane proteins of the plasma membrane. These membrane receptors bind the signal molecule in the extracellular environment, and the signal is then communicated intracellularly

through a Rube Goldberg sequence of "differences that make a difference." The third class of receptors is the *nuclear receptors.* These are not membrane proteins but rather intracellular proteins that transduce signals from oily, lipidic molecules that can easily enter the cell. Signaling molecules that bind and activate nuclear receptors include steroid and thyroid hormones, fat molecules in the diet, and derivatives of vitamins A and D. The information transduction pathway of nuclear receptors is simpler than that of the membrane receptors in that nuclear receptors are themselves direct regulators of gene transcription; that is, nuclear receptors are transcription factors. Binding of the signal molecule activates the nuclear receptor so that it is then able to bind directly to specific regions of deoxyribonucleic acid (DNA) and stimulate the binding of ribonucleic acid (RNA) polymerase to, and thus production of, messenger RNA from the particular gene or genes in that region of DNA. An example involving the female-specific production of egg protein in hens is discussed later in the chapter.

Specific Physiological Information Is Inherent in the Receptor/Ligand Complex, Not in the Hormone/Neurotransmitter Molecule

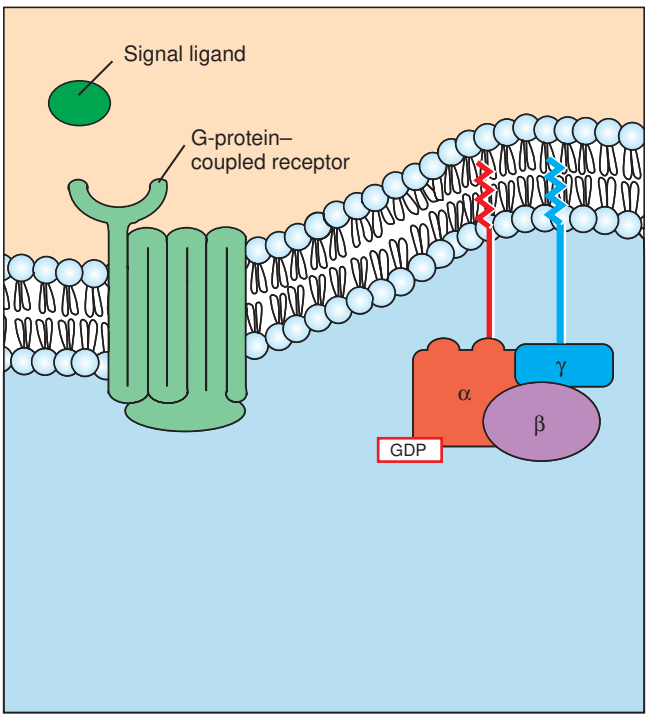
Before discussing specific receptors, it is useful to elaborate on some important points about the nature and regulation of the information transfer between the external signal molecule and receptor. This text provides ample evidence that the same hormone and especially the same neurotransmitter molecule can bind to different receptors. These different receptor-binding events send different information to the cell from the same external signal molecule. For example, acetylcholine is bound by two different receptors, the nicotinic ion channel described earlier and the muscarinic receptor, which is a GPCR, not an ion channel, and sends completely different information to the cell. The hormone/neurotransmitter itself does not contain any specific information; rather, it is a simple signal, such as a phone ringing. One must answer the phone to receive the information. The information content of the hormone/neurotransmitter is really contained in the three-dimensional shape of the receptor molecule. The change in the shape of the receptor on binding the hormone/neurotransmitter is the specific message to the cell.

Cells can make themselves more or less sensitive to the signal of the hormone/neurotransmitter. For example, most cells respond to a prolonged period of exposure to hormone/neurotransmitter by reducing their sensitivity to that molecule. For membrane receptors, one way is to internalize the receptors by endocytosis, fuse the endosome with a lysosome, and digest the receptor. Typically, membrane receptor number is decreased by endocytosis in response to a sustained high concentration of ligand. This is called *downregulation* of the receptor. This process allows the cell to adapt to high ligand concentrations. Receptor-ligand interaction is a true chemical equilibrium, so the proportion of receptor-ligand complexes, which determines physiological response, depends on the concentration of both receptors and ligands. In the presence of a high ligand concentration, a decrease in receptor number returns the binding equilibrium to the normal proportion of bound/unbound receptors. This allows the cell to respond to increases and decreases in ligand, even at high concentrations of ligand. Another way of regulating the response to a hormone/neurotransmitter is to alter the binding function of the receptor, e.g., by phosphorylating it, so that its affinity for the ligand is reduced (*desensitization*) or increased (*hypersensitization*). Nuclear receptors appear to be less subject to short-term regulation of responsiveness, but at least some nuclear receptors require constant turnover by proteolytic breakdown and new synthesis to function.

G-Protein–Coupled Receptors Are the Largest Family (a *Superfamily*) of Receptors and Help Regulate Almost All Physiological Processes

It would be difficult to exaggerate the importance and versatility of information processing that begins with a signal molecule binding to GPCRs. There are approximately 900 GPCRs in humans (Table 1.2). There are an even greater number in animals that depend more on olfaction, with about 1300 in rodents, because smell is mediated by different odorants binding to different GPCRs. An estimated 40% to 50% of all commercial drugs act in a GPCR pathway, exemplifying the importance of GPCRs to medicine. All GPCRs share a similar molecular shape; they are an integral membrane protein composed of a single polypeptide chain that passes in and out of the plasma membrane seven times, resembling a snake (Fig. 1.14). As a result, two other names for GPCRs are *seven-transmembrane receptors* and *serpentine receptors.* However, the name GPCRs is more revealing about their mechanism because all also share the same "next step" in their Rube Goldberg signal sequence; they activate a molecular "on-off switch" known as a *G protein,* so called because it is a *guanosinetriphosphatase* (GTPase).

GPCRs bind to a particular type of G protein (another of the many "families" of informational proteins), which is a membrane-associated trimeric protein composed of α, β, and γ subunits.



• **Fig. 1.14** G-protein–coupled receptor (GPCR) and the heterotrimeric G protein. The hundreds of GPCRs share a similar protein shape, snaking in and out of the membrane seven times. Thus GPCRs are also called *serpentine receptors* and *heptahelical receptors.* These receptors interact with a membrane-associated guanosinetriphosphatase (GTPase) molecule composed of three different polypeptide subunits *(heterotrimeric).* The heterotrimeric G protein is not an intrinsic membrane protein, but rather associates with the membrane through lipid tails inserted into the membrane. *GDP,* Guanosine diphosphate.

TABLE 1.2 Partial List of G-Protein–Coupled Receptors (GPCRs)

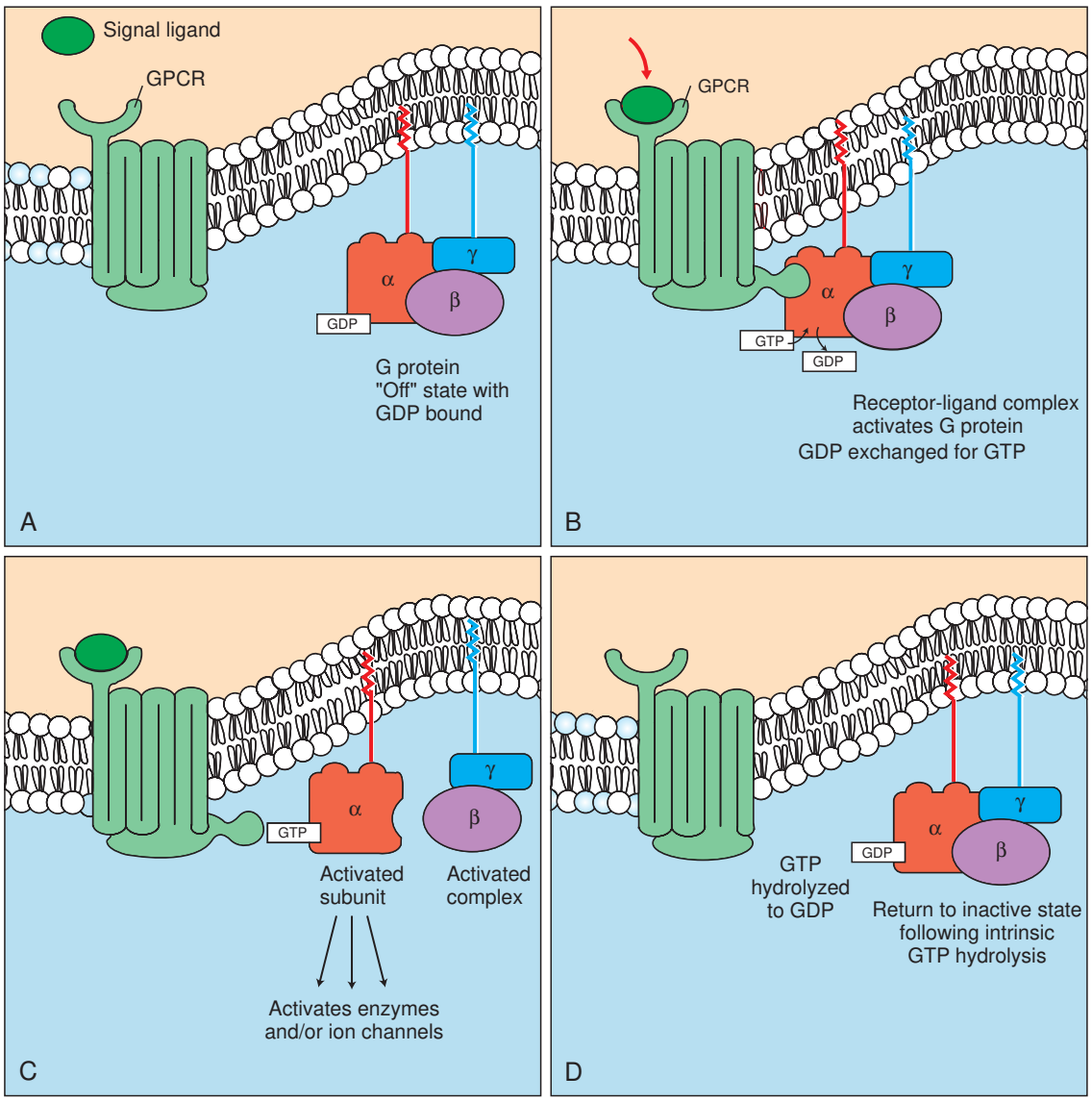
Receptor/Receptor Family[a]	Example of Function	Drug Ligands
α-Adrenergic	Regulates vasculature	Phenylephrine, oxymetazoline
β-Adrenergic	Regulates heart and vasculature	Atenolol, propranolol
Angiotensin	Principal regulator of blood pressure	Losartan
Calcitonin	Regulates bone resorption	[b]
Cannabinoid	Unknown but found widely in brain	Marijuana and derivatives
Dopamine	Movement, cognition, and emotions	Chlorpromazine, bromocriptine
Frizzled	Regulates proliferation and differentiation, particularly in stem cells	[b]
Gastrin	Regulates acid secretion in stomach	Pentagastrin
Glucagon	Regulates "starvation" response	Exendin-4
Histamine	Mediates inflammation and allergy	Diphenhydramine, chlorpheniramine
Muscarinic	Secretion of hormones and neurotransmitters	Atropine, carbachol
Olfactory	Mediates smell	[b]
Opioid	Mediates analgesia	Morphine, codeine, heroin
Opsins	Mediates light transduction in retina	[b]
Prostaglandin	Vasodilation	Sulprostone
Serotonin[c]	Regulates gut motility, behavioral arousal, feeding, circadian rhythms	Sumatriptan, ketanserin
Vasopressin	Regulates water balance of body	Terlipressin, desmopressin

[a]In most cases, receptor is named for its ligand.
[b]None commonly known.
[c]One member of the serotonin receptor family is not G-protein–coupled.

Thus this type of G protein is called the *heterotrimeric G protein* ("three different subunits"). The heterotrimeric G proteins bind directly to the cytoplasmic domain of a GPCR. Although not intrinsic membrane proteins, heterotrimeric G proteins are closely associated with the plasma membrane through lipid molecules that are posttranslationally added to the subunits and insert into the lipid bilayer (see Fig. 1.14).

As noted, G proteins are molecular "on-off switches" that are also GTPases activated by the binding of a signal molecule to its cognate GPCR. That is, in addition to binding to GPCRs, G proteins also bind *guanosine triphosphate* (GTP) and hydrolyze it to *guanosine diphosphate* (GDP). The binding and hydrolysis of the GTP to GDP underlie the biochemical mechanism of the on-off switch. In Fig. 1.15A, the unstimulated GPCR does not bind to the heterotrimeric G protein, which is in its "off" state by virtue of the α subunit having GDP and the β and γ subunits bound to it. In Fig. 1.15B, a signal ligand binds to its GPCR, activating the receptor and the G protein. The activation of the G protein takes the form of dissociation of the β/γ complex from the α subunit, which allows the α subunit to exchange GDP for GTP. The principal "on" activity of the G protein is represented by the G_α subunit with GTP bound to it. GTP-bound G_α stimulates a variety of enzymes and ion channels that send the signal into the cytoplasm (see Fig. 1.15C), as discussed in the next section. However, the $G_{\beta\gamma}$ complex, once thought to be only an inhibitory factor for the G_α subunit, is now known to activate certain K^+ channels itself and inhibit certain voltage-dependent Ca^{2+} channels. After stimulating the next element in the signal pathway, the activated GTP-bound G_α subunit returns to an inactive, "off" state as a result of its intrinsic GTPase activity (see Fig. 1.15D). That is, the bound GTP is hydrolyzed to GDP, and the $G_{\beta\gamma}$ complex rebinds to the G_α subunit, returning it (and the $G_{\beta\gamma}$ complex) to its inactive state, awaiting the next ligand-receptor–binding event.

As noted earlier, one of the aspects of the Rube Goldberg analogy is that the same conserved types of molecules are often used in many different pathways. Among the many protein "differences that make a difference" to transmit information, one of the most widely used is a GTPase that has on-off states based on whether GTP or GDP is bound to it. Thus the heterotrimeric G proteins that couple to GPCRs are only one type of GTPase protein acting as an on-off switch in signaling pathways. Most other members of the G-protein (GTPase) superfamily are simpler and resemble the G_α subunit alone. For example, one such class of these *small G proteins,* called *Rabs,* helps mediate the membrane fusion processes that underlie exocytosis and endocytosis, as discussed earlier. All G proteins share evolutionarily conserved GTP binding and enzymatic hydrolysis sites and a similar on-off mechanism; when GTP is bound, the protein is "on," and when GDP is bound, the protein is "off." Chapter 2 discusses a particular small G protein, Ras, that plays a crucial role in regulating cell division and whose dysfunction plays a major role in cancer. Consequently, G proteins in general are discussed in Chapter 2, and this discussion focuses on signaling mechanisms linked to the heterotrimeric G protein specifically.



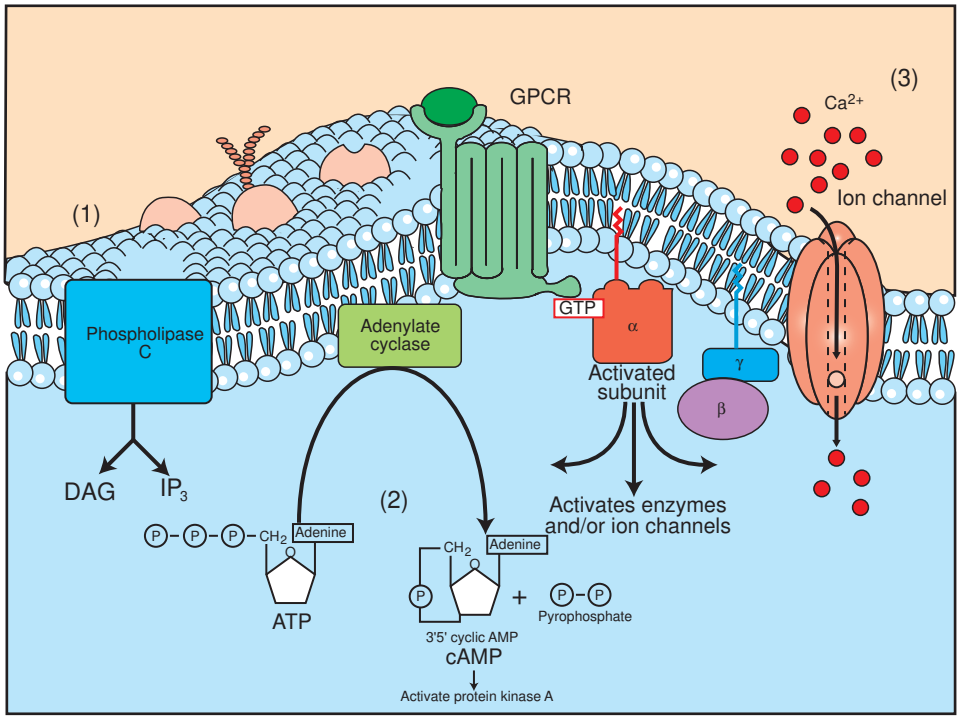
• **Fig. 1.15** Duty cycle of heterotrimeric G-protein, a GTPase that acts as a molecular "on-off switch." See text for further details. (A) Unstimulated GPCR is not bound to the heterotrimeric G protein. (B) Signal ligand binds to its GPCR, activating receptor and G protein. (C) GTP-bound G_α subunit stimulates a variety of enzymes and ion channels that send the signal into the cytoplasm. (D) After stimulating next element in signal pathway, the activated GTP-bound G_α subunit returns to inactive, "off" state because of its intrinsic GTPase activity. *GDP,* Guanosine diphosphate; *GPCR,* G-protein-coupled receptor; *GTP,* guanosine triphosphate.

Most G-Protein–Linked Information Is Sent to the Cytoplasm by Second Messengers

As previously noted, the active (heterotrimeric) G protein stimulates an enzyme or ion channel associated with the plasma membrane. The ensuing change in ion channel or enzyme function can alter the membrane potential or cause certain molecules/ions to change their concentration in the cytoplasm. Those cytoplasmic ions and molecules linked to receptor-ligand binding are called *second messengers.* A second messenger is an ion or molecule that carries the information within the cytoplasm of a cell in response to a signal on the outside surface of a cell (the first message), such as the binding of a hormone or neurotransmitter, or to an electrical event. Most G-protein–linked information is transduced into the cytoplasm in this manner. One of the major advances in our understanding of the molecular basis of physiological signaling is the realization that there are only a few second-messenger systems within animal cells. The most important include the following (Fig. 1.16):

1. Two second messengers, inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG), are produced by G-protein activation of an enzyme phospholipase C (PLC) (see Fig. 1.19 and later discussion).
2. Changes in the concentration of *cyclic adenosine monophosphate* (cAMP).
3. Changes in Ca^{2+} concentration within the cytoplasm.

Clearly, there are many more GPCRs than second messengers. This means that several receptor-mediated events are converted into the same intracellular signal. How does the cell sort out this information? Different cells respond differently to the same



• **Fig. 1.16** Activated α subunit of the G protein (G_α) can activate enzymes and ion channels, leading to second-messenger signaling within the cytoplasm. Three principal second messengers send the G-protein–coupled receptor (GPCR) information to the cytoplasm. These arise from the activation of ion channels and enzymes stimulated by G_α. The second messengers are (*1*) increases in the concentration of inositol 1,4,5-trisphosphate (IP_3) in the cytoplasm and increases in the concentration of diacylglycerol (DAG) in the plasma membrane, both as a result of the breakdown of a rare membrane phospholipid, phosphatidylinositol 4,5-bisphosphate (PIP_2) by phospholipase C, another G_α-stimulated enzyme; (*2*) changes in the concentration of cyclic adenosine monophosphate (cAMP), a special hydrolytic breakdown product of adenosine triphosphate (ATP) created by the enzyme adenylyl cyclase, which can be activated or inhibited by α subunits; and (*3*) changes in Ca^{2+} concentration within the cytoplasm resulting from transport of Ca^{2+} through gated channels stimulated by G_α. *GTP,* Guanosine triphosphate.

second-messenger ion/molecule as a result of the specialized function and makeup of that cell (i.e., the differentiated state it achieved during the development of the animal). For example, smooth muscle cells respond differently to activation of muscarinic acetylcholine receptors (see Table 1.2) than do nerve cells because the two cells have different proteins that are responsible for their specialized tasks. However, this is only part of the answer, and the specificity of response to the same second messenger and to activation of similar or identical receptors remains an important open question in physiology.

Ca^{2+} Transport Across Plasma and Intracellular Membranes Is an Important Second Messenger

The transport of Ca^{2+} ions through gated channels across the plasma membrane and across intracellular membranes (e.g., endoplasmic reticulum) is a major second-messenger system for physiological information transfer. The available evidence suggests that the major role of Ca^{2+} *within* cells is as a physiological signal. In the extracellular compartment, the major physiological function of Ca^{2+} is as the principal mineral of bone. Ca^{2+} is an excellent ion for use as a second messenger because the cytoplasmic concentration of Ca^{2+} is extremely low, about 10^{-7} mol/L in a resting cell. Increases in intracellular Ca^{2+} concentration can be (1) detected easily because the background noise is so low and (2) achieved easily because the Ca^{2+} concentration [Ca^{2+}] in the ECF and in some cellular compartments, such as the endoplasmic reticulum and mitochondria, is 10^4 higher than in the cytoplasm (see Table 1.1). Thus there is an enormous driving force for Ca^{2+} diffusion into the cytoplasm under most conditions.

Although many GPCRs use Ca^{2+} as one part of their intracellular pathway, the interaction is more complex than usual, as discussed shortly. Thus we focus here on a simpler and very important example of Ca^{2+} as a second messenger that has already been discussed: the role of Ca^{2+} in regulating the actomyosin ATPase of muscle.

Increased [Ca^{2+}] in the cytoplasm alters cellular function by binding to any of several Ca^{2+}-binding proteins that serve as control proteins. Troponin is one Ca^{2+}-binding protein already mentioned. Reviewing the example of striated muscle contraction from the Ca^{2+} point of view, Ca^{2+} (second messenger) diffuses through gated channels in the endoplasmic reticulum (sarcoplasmic reticulum) of muscle in response to electrical events (first message) on the plasma membrane of the muscle cell. The diffusion of Ca^{2+} from the concentrated storehouse of the sarcoplasmic reticulum increases [Ca^{2+}] in the cytoplasm of the muscle cell, where it binds to troponin. On binding Ca^{2+}, troponin changes its interaction with tropomyosin, which now moves to allow myosin heads access to the actin of the thin filament. The actomyosin ATPase is activated, and muscle contraction ensues.

Calmodulin is a Ca^{2+}-binding protein that plays an important control function in almost all animal cells. As with troponin, calmodulin binds Ca^{2+} when the cytoplasmic [Ca^{2+}] increases. The

Ca^{2+}/calmodulin complex activates a large number of different cellular processes. In most such cases, but not all, the Ca^{2+}/calmodulin complex binds to and activates an enzyme. One such enzyme, a protein kinase, is involved in the excitation-contraction coupling in smooth muscle (Fig. 1.17), not discussed earlier with the striated muscle types. Protein kinases in general catalyze the hydrolysis of ATP and couple it to the simultaneous phosphorylation of other proteins, as follows:

$$\text{ATP} + \text{Protein} \xrightarrow[\text{Protein kinase}]{Ca^{2+}/\text{calmodulin activated}} \text{Protein phosphate} + \text{ADP}$$

In the case of smooth muscle, the particular protein kinase is *myosin kinase,* which, as its name implies, specifically phosphorylates myosin. This phosphorylation increases the affinity of the myosin heads for actin filaments, thus allowing cross-bridging to actin. On formation of the cross-bridge, myosin strokes past the thin filament, producing filament sliding, contraction, and force production by smooth muscle. Cessation of contraction is achieved by cleavage of the phosphate from the myosin by another enzyme, *myosin phosphatase.*

Thus initiating smooth muscle contraction involves a Rube Goldberg sequence in which environmental stimulation of a smooth muscle cell causes an increase in the intracellular [Ca^{2+}], the second messenger. This in turn leads to a cascade of cause and effect. Increased intracellular [Ca^{2+}] causes calmodulin to bind Ca^{2+}. The Ca^{2+}/calmodulin complex activates the myosin kinase. This enzyme phosphorylates the myosin head, allowing it to cross-bridge to actin. Cross-bridging leads to actomyosin activation, causing filament sliding observed as muscle contraction at the tissue level.

This discussion of Ca^{2+} as a second messenger emphasizes one of its major physiological functions: as the second messenger responsible for mediating contraction for all types of muscle (skeletal, cardiac, and striated), although the details of each pathway differ.

Cyclic AMP Is Produced by Activation of a Membrane-Bound Enzyme in Response to Hormone/Neurotransmitter Binding to Receptors

Changes in the activity of membrane-associated enzyme activities are an important mechanism of transmitting information across the cell membrane and are used by most GPCRs. Binding of a signaling molecule to receptors on the extracellular face of the plasma membrane changes the activity of an enzyme located on the cytoplasmic face. The enzyme catalyzes a breakdown reaction; one or more of the breakdown products released into the cytoplasm are second messengers. One important such second-messenger system, and the first to have been discovered, is the hydrolytic breakdown of ATP to 3′,5′-adenosine monophosphate, or cAMP, by the enzyme *adenylyl cyclase* (previously called *adenyl cyclase* and *adenylate cyclase*). Cyclic AMP is the second messenger, and adenylyl cyclase is turned on or off as a result of the binding of various hormones and neurotransmitters to cell surface receptors.

As summarized in Fig. 1.18, three distinct membrane proteins interact to produce cAMP: (1) any of several receptors, including many GPCRs; (2) the heterotrimeric G protein; and (3) the catalytic protein that actually hydrolyzes ATP to cAMP. Their interaction provides an example of the ability of biomembranes to organize biochemical reactions in space. The likelihood of three proteins colliding and thus being able to interact is much greater in the two-dimensional "phospholipid sea" than in the three-dimensional cytoplasm.

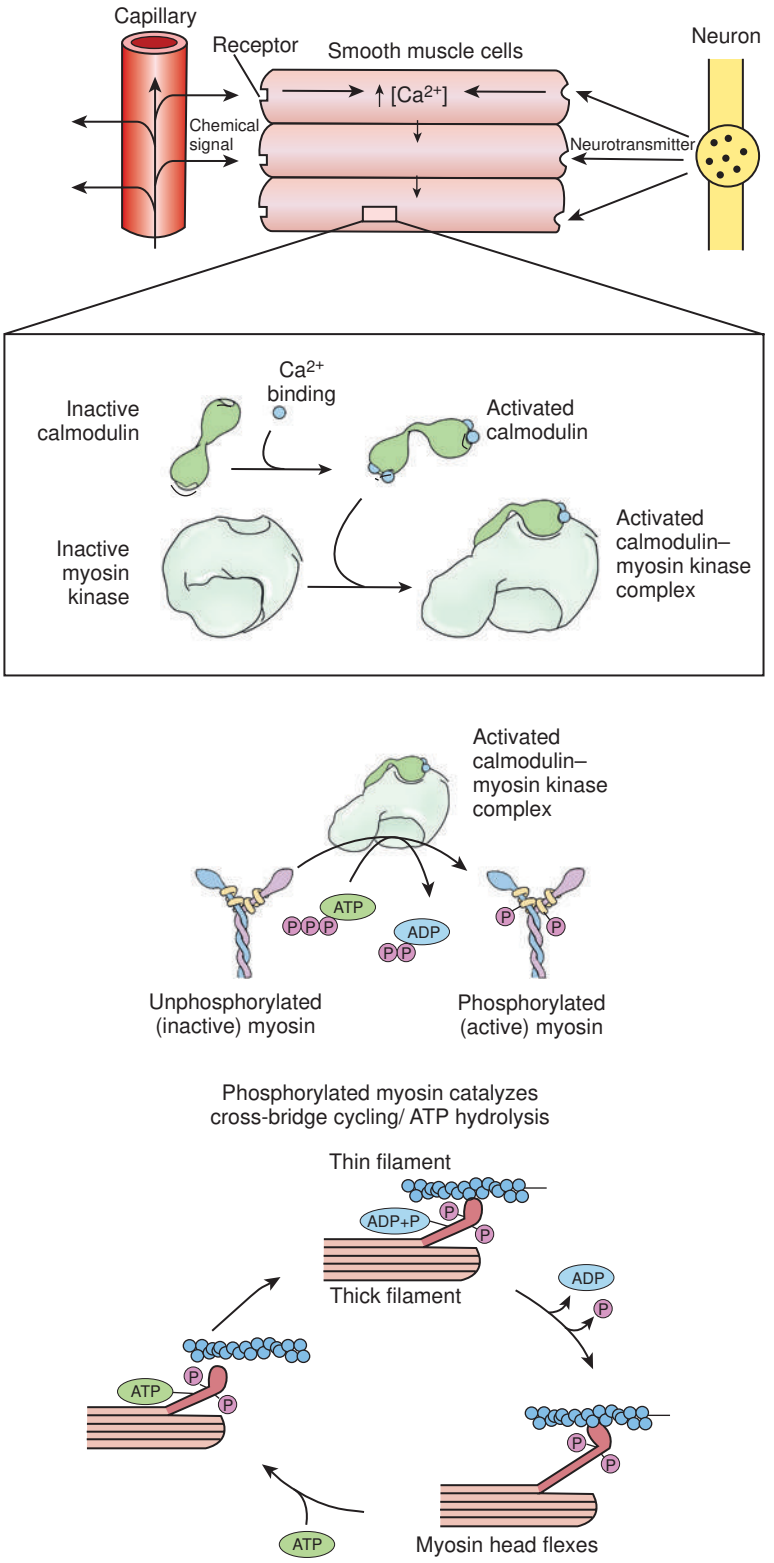
A large number of different hormones/neurotransmitters that bind to different membrane receptors use cAMP to transmit information across the membrane. Among the GPCRs (see Table 1.2) and their hormones/neurotransmitters that use cAMP as their second messenger are *β-adrenergic receptors* that bind epinephrine or norepinephrine, increasing cAMP production and providing important regulation to almost all tissues. The starvation message carried by the binding of glucagon to its receptor (see Chapter 34) is carried to the cytoplasm by an increase in cAMP. *Vasopressin* (also called *antidiuretic hormone,* ADH) binding to its receptors in kidney cells uses cAMP to regulate urine production (see Chapter 33). A number of therapeutic drugs bind to these same receptors and mimic or prevent the physiological action of the hormone/neurotransmitter that normally binds to the receptor.

After ligand binding, the ligand-receptor complex is able to bind to and activate the regulatory G protein (see Fig. 1.15B). The G protein in turn changes shape and binds to the catalytic subunit, altering its shape and regulating its ability to bind ATP, and hydrolyzes the catalytic subunit to cAMP (see Fig. 1.18). There are two types of G proteins in the adenylyl cyclase system, which differ in their α subunit. The G_s (more specifically, $G_{\alpha s}$, s for stimulatory) activates the catalytic subunit; this is the G protein shown in Fig. 1.18. A different G protein, the α subunit of G_i, inhibits adenylyl cyclase when activated. Some diseases are the result of the binding of bacterial toxins to the G proteins. Cholera symptoms result in part from the binding of the toxin of the bacteria *Vibrio cholerae* to the G_s protein, and the irreversible activation of the G_s protein, which in turn irreversibly activates the catalytic subunit. Pertussis (whooping cough) toxin binds irreversibly to and activates G_i, thus inactivating the enzymatic activity.

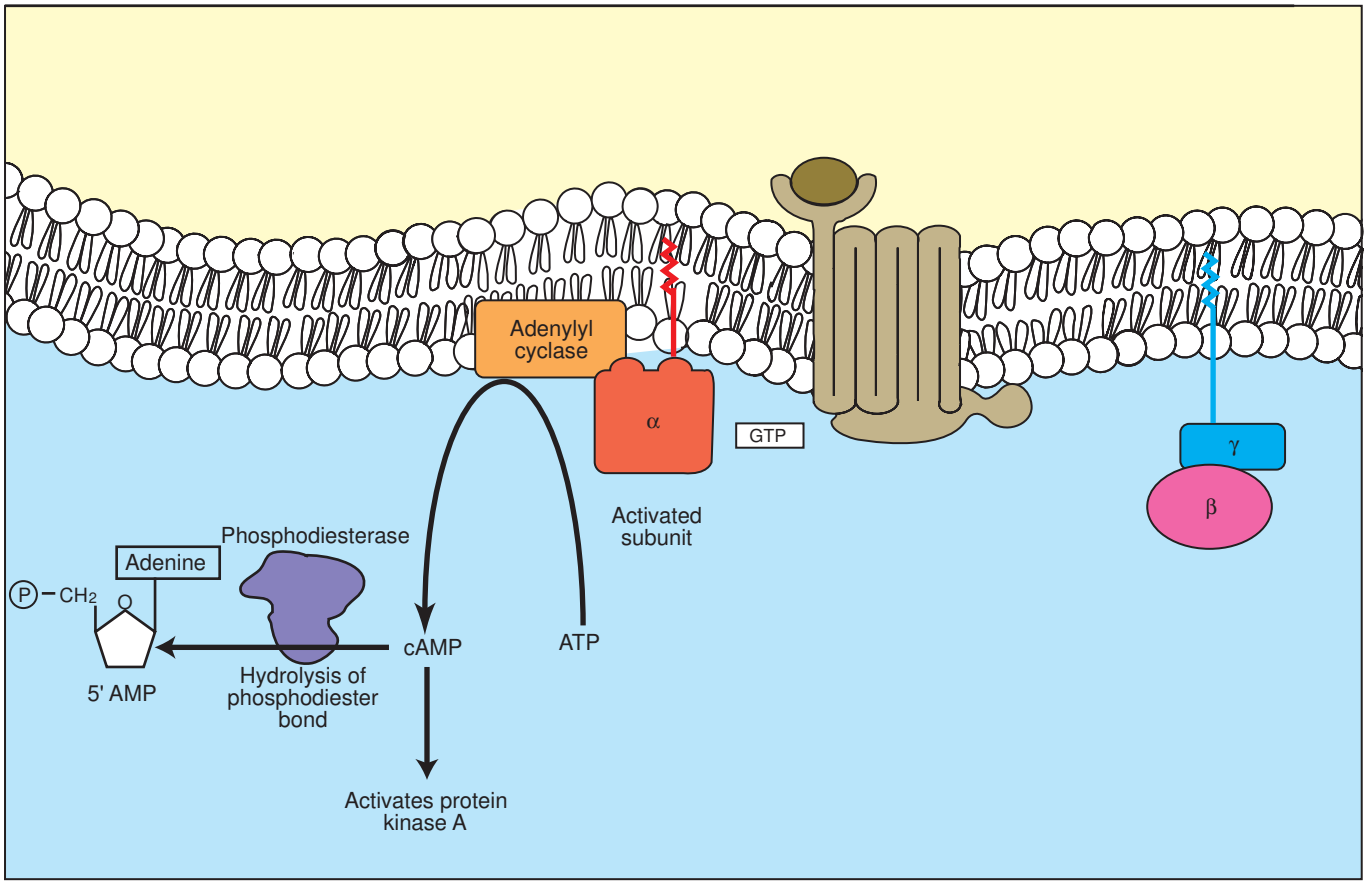
As suggested by the inhibitory G protein (G_i), regulated decreases in cAMP concentrations are an important part of the cAMP second-messenger system. There are two mechanisms for such decreases: decreasing the rate of cAMP production and eliminating cAMP after formation. The former is achieved by G_i inhibiting the catalytic subunit. Certain inhibitory receptors specifically interact with G_i. Opium and drugs derived from it, such as codeine and morphine, are examples of signaling molecules that bind to inhibitory GPCR (opioid) receptors, activate G_i, and inhibit production of cAMP. Other examples are norepinephrine and epinephrine acting through α_2-adrenergic receptors. Recall that these same neurotransmitters activate adenylyl cyclase when bound to β-adrenergic receptors. This is another example of the principle that the receptor/ligand complex contains the information, not the hormone/neurotransmitter itself.

The other control on cAMP levels is elimination of cAMP after formation. This is regulated by the enzyme *cyclic nucleotide phosphodiesterase* (PDE). This enzyme hydrolyzes the 3′ ester bond of the phosphate to the sugar to produce "plain" 5′ AMP (see Fig. 1.18). As with myosin kinase discussed earlier, PDE is a Ca^{2+}/calmodulin-activated enzyme, so in many cells the activities of the Ca^{2+} and cAMP second-messenger systems antagonize one another.

The increase or decrease in cAMP concentrations most often affects cell function through cAMP's interaction with a particular protein kinase. This protein kinase is called *cAMP-dependent protein kinase,* or *protein kinase A* (PKA). This protein kinase is distinct from the Ca^{2+}/calmodulin-dependent protein kinase discussed earlier, although the basic outline of action is similar. PKA is activated by binding cAMP. The higher the concentration of cAMP



• **Fig. 1.17** Role of Ca^{2+} and calmodulin in the regulation of smooth muscle contraction. Smooth muscle regulation is more complex than regulation of striated muscle, and the account here is a simplification. Smooth muscle can be stimulated to contract by a variety of stimuli, including neural signals and soluble chemical signals, as shown here. These external signals all stimulate increased intracellular [Ca^{2+}], which leads to smooth muscle contraction. In the presence of increased intracellular [Ca^{2+}], the Ca^{2+} ions bind to calmodulin, activating it by causing a conformational change. In smooth muscle cytoplasm, the activated Ca^{2+}/calmodulin complex activates myosin kinase, which catalyzes the phosphorylation of myosin. Phosphorylated, activated myosin in turn catalyzes actin-dependent ATP hydrolysis (cross-bridge cycling). Thus smooth muscle contraction is thick filament regulated, because changes in myosin activate cross-bridging, whereas striated muscle contraction is thin filament controlled, because changes in troponin and tropomyosin of the thin filament activate cross-bridging. *ADP,* Adenosine diphosphate; *ATP,* adenosine triphosphate.



• **Fig. 1.18** Activity of cyclic adenosine monophosphate (cAMP) as a second messenger. Cyclic AMP is generated through G-protein–coupled receptor-linked activation of adenylyl cyclase, causing hydrolysis of adenosine triphosphate (ATP) to cAMP. The cAMP thus generated binds to and activates a specific protein kinase, protein kinase A, which in turn can phosphorylate and change the activity of various cellular substrates. When generated, cAMP is broken down by phosphodiesterase, which hydrolyzes cAMP to "normal" AMP (i.e., 5′ AMP). *GTP,* Guanosine triphosphate.

in a cell, the greater is the number of active PKA molecules. The activated kinase binds to proteins and ATP, hydrolyzing the ATP and phosphorylating the protein. As shown in previous examples, this phosphorylation alters the activity of the target protein, altering its particular characteristic function: catalysis, transport, coupling, and so forth.

Mammals respond to a stressful stimulus by increasing the force and rate of heart contraction, among other physiological effects. This increase in force demonstrates the role of cAMP as a second messenger and the role of Ca^{2+} in GPCR signaling, and it is another example of physiological Rube Goldberg devices based on allosteric changes in proteins. The stressful stimulus causes the adrenal medulla to release epinephrine to the blood, and sympathetic nerves release norepinephrine to the heart. Both catecholamines bind to β-adrenergic GPCRs on the cardiac muscle cells. The receptor-ligand interaction stimulates adenylyl cyclase by way of G_s, increasing intracellular [cAMP], thus increasing PKA activity. PKA phosphorylates a number of substrates in the cardiac muscle cells, including voltage-dependent Ca^{2+} channels in the plasma membrane. In the phosphorylated state, these channels remain open somewhat longer in response to membrane potentials above threshold. Consequently, more Ca^{2+} enters the cell for a given electrical stimulation than at lower levels of cAMP. The increase in Ca^{2+} allows more troponin to bind Ca^{2+}; more tropomyosin moves out of the way of myosin heads, causing more cross-bridging and more force production. (Rube Goldberg would have loved modern physiology!)

Another cyclic nucleotide, *cyclic guanosine monophosphate* (cGMP), also serves as a second messenger but is not nearly as widely used as cAMP. Cyclic GMP is the second messenger stimulated by opsins (see Table 1.2) in the rod cells of the retina underlying vision and also causes relaxation of some vascular smooth muscle, including that responsible for penile erection and clitoral engorgement (i.e., blood flow into the corpus cavernosum of both tissues). The role of cGMP in erections is mediated by its activation of cGMP protein kinases, similar to cAMP action via PKA. Activation of cGMP-dependent protein kinase causes relaxation of certain smooth muscles, including those responsible for blood flow to the corpus cavernosum. This has an important clinical correlate; the drug Viagra (sildenafil) inhibits the breakdown of cGMP by a cyclic nucleotide phosphodiesterase, thus increasing blood flow to the penis, and aids erection, but only if neural signals (i.e., sexual stimulation) have stimulated cGMP production initially. This is a good example of how the multistep pathway of cell signaling provides multiple potential sites for appropriate therapeutic intervention; a drug that simply stimulated cGMP production

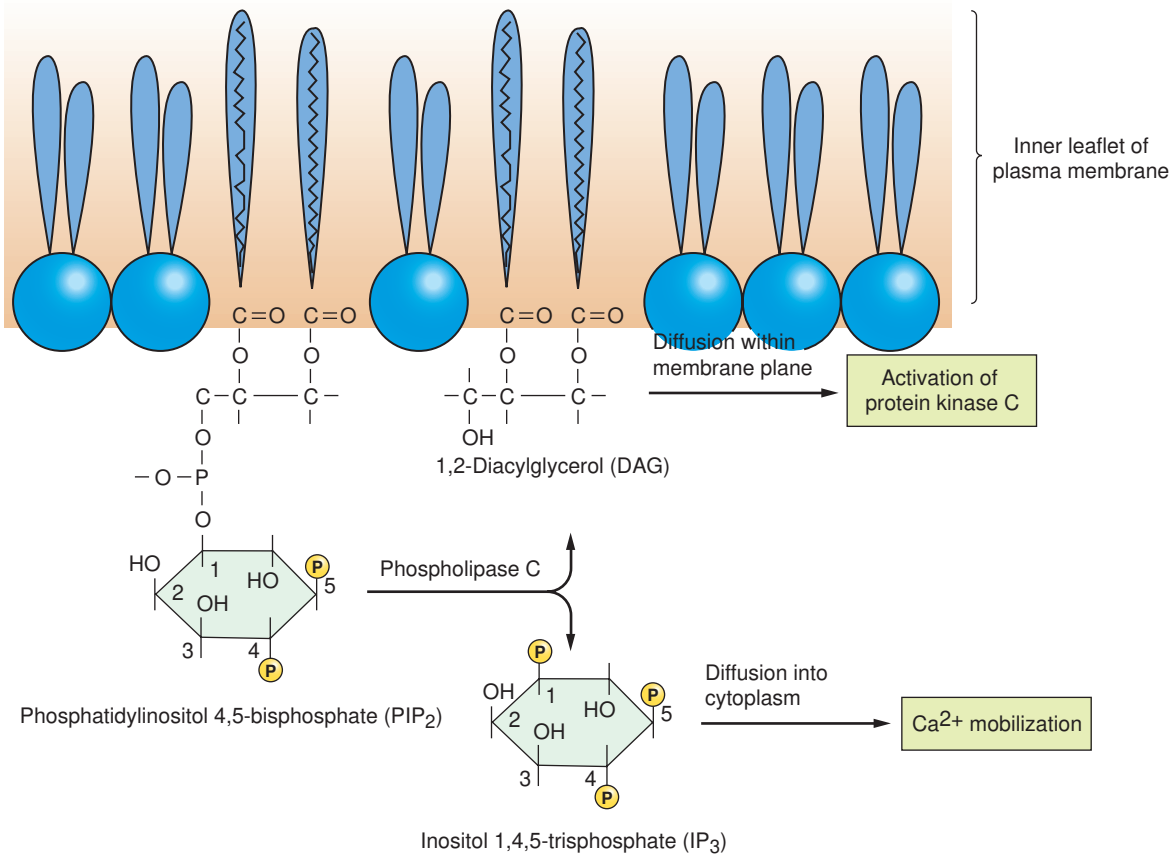
would cause inappropriate erections, whereas inhibiting its breakdown aids timely erections. Although used mostly by men, sildenafil is also occasionally used for stallions to assist them in "covering" a mare.

In addition to activating protein kinases, cAMP and cGMP can also bind directly to and cause opening of a class of ligand-gated ion channels, cyclic nucleotide–gated ion channels. These ion channels are atypical in that their structure resembles voltage-gated K^+ channels, but they open by directly binding a cyclic nucleotide. These channels play an important role in smell, for which cAMP is the relevant second messenger. In vision, as noted earlier, cGMP is the second messenger, and mutations in the cyclic nucleotide–gated ion channels of cones are responsible for most forms of complete color blindness (but which is rare).

The examples of physiological control by second messengers discussed thus far are short time-scale changes (seconds to hours), which historically have been the purview of "physiologists." It has become increasingly clear, however, that most, if not all, major signals have longer-term (days and weeks) effects based on changes of gene transcription, which in turn mediates changes in growth, differentiation, and long-term behavior. For example, cAMP is now known to be an important regulator of gene transcription controlling learning, production of gametes, and cell division. The effect of cAMP on gene expression is the result of PKA phosphorylating a specific transcription factor associated with cAMP signaling ("cyclic AMP response element binding protein," or CREB). Although space does not allow further discussion of the transcriptional roles of "classic" physiological signal pathways, when dealing with signal pathways it is worthwhile to keep in mind the disclaimer in the first paragraph: only a highly simplified account of cell function is presented here!

The Receptor-Mediated Hydrolysis of a Rare Phospholipid of the Plasma Membrane Produces Two Different Second Messengers With Different Actions

Another second-messenger system differs from both Ca^{2+} and cAMP in that *two* distinct second-messenger molecules are produced as a result of an enzymatic activation by a single receptor/ligand complex. *Phosphatidylinositol* (PI) is a membrane phospholipid that can accept additional phosphate groups by reaction with the –OH groups on the inositol (Fig. 1.19). *Phosphatidylinositol 4,*



• **Fig. 1.19** Hydrolysis of a membrane lipid to produce two second messengers. After appropriate receptor and G-protein activation, the rare membrane phospholipid shown to the left, phosphatidylinositol 4, 5-bisphosphate (PIP_2) is hydrolyzed into two separate second messengers by phospholipase C. The phosphate "head" of the PIP_2 molecule is cleaved to produce the soluble messenger inositol 1,4,5-trisphosphate (IP_3), which mobilizes intracellular Ca^{2+}, as well as the lipidic messenger diacylglycerol (DAG), which remains in the membrane and activates protein kinase C.

5-bisphosphate (PIP_2) is the membrane phospholipid that is broken down to produce two important second messengers. PIP_2 is hydrolyzed to DAG and IP_3 by a receptor-mediated enzyme called PLC or phosphoinositidase. Although many distinct processes are controlled through the PIP_2 path, it plays a particularly important role in control of growth and of receptor-mediated secretion. The effect of acetylcholine acting through muscarinic receptors (*not* the nicotinic receptor/ion channel of the nerve-muscle synapse) is often transmitted and transduced through activation of the PIP_2 pathway.

The events involved in the receptor-mediated production of IP_3 and DAG from PIP_2 are similar to those in the production of cAMP. The membrane system appears to consist of three distinct intrinsic membrane proteins: (1) any of several different GPCRs, including the muscarinic acetylcholine receptor and the receptors for some growth factors; (2) a heterotrimeric G protein, similar but not identical to G_s of the cAMP pathway; and (3) the hydrolytic enzyme PLC. A hormone/neurotransmitter or growth factor binds to the receptor, forming a receptor/ligand complex. This complex activates the G protein, which in turn activates the hydrolytic enzyme. At present, only a stimulatory G activity on PLC is known; there is no evidence for an inhibitory G activity in this system.

The activation of the hydrolytic enzyme increases the concentration of IP_3, which is water soluble and thus diffuses through the cytoplasm. IP_3 binds to and opens ligand-gated Ca^{2+} channels in the endoplasmic reticulum. This releases Ca^{2+} from that high [Ca^{2+}] compartment into the cytoplasm. Ca^{2+} thus becomes the "third messenger" in this system (although this term is not in widespread use) and is another example of a role of Ca^{2+} in GPCR signaling. The ensuing increase in cytoplasmic [Ca^{2+}] affects cellular function by the same mechanisms outlined earlier for Ca^{2+} as a second messenger (e.g., binding to calmodulin), with the Ca^{2+}/calmodulin complex in turn activating various enzyme activities. In receptor-mediated secretion, for example, the binding of acetylcholine to muscarinic receptors in the pancreas (the organ that secretes digestive enzymes) causes an increase in PIP_2 breakdown and an increase in cytoplasmic IP_3. The IP_3 opens ligand-gated Ca^{2+} channels in the endoplasmic reticulum, and intracellular [Ca^{2+}] increases. The process then becomes similar to smooth muscle contraction. Calmodulin binds Ca^{2+}, and the complex activates a protein kinase. However, rather than activating myosin, as for smooth muscle, activation of this protein kinase causes exocytosis of secretory vesicles (membrane bubbles full of secretory product) with the plasma membrane, releasing the enzymes into an extracellular space contiguous with the gut.

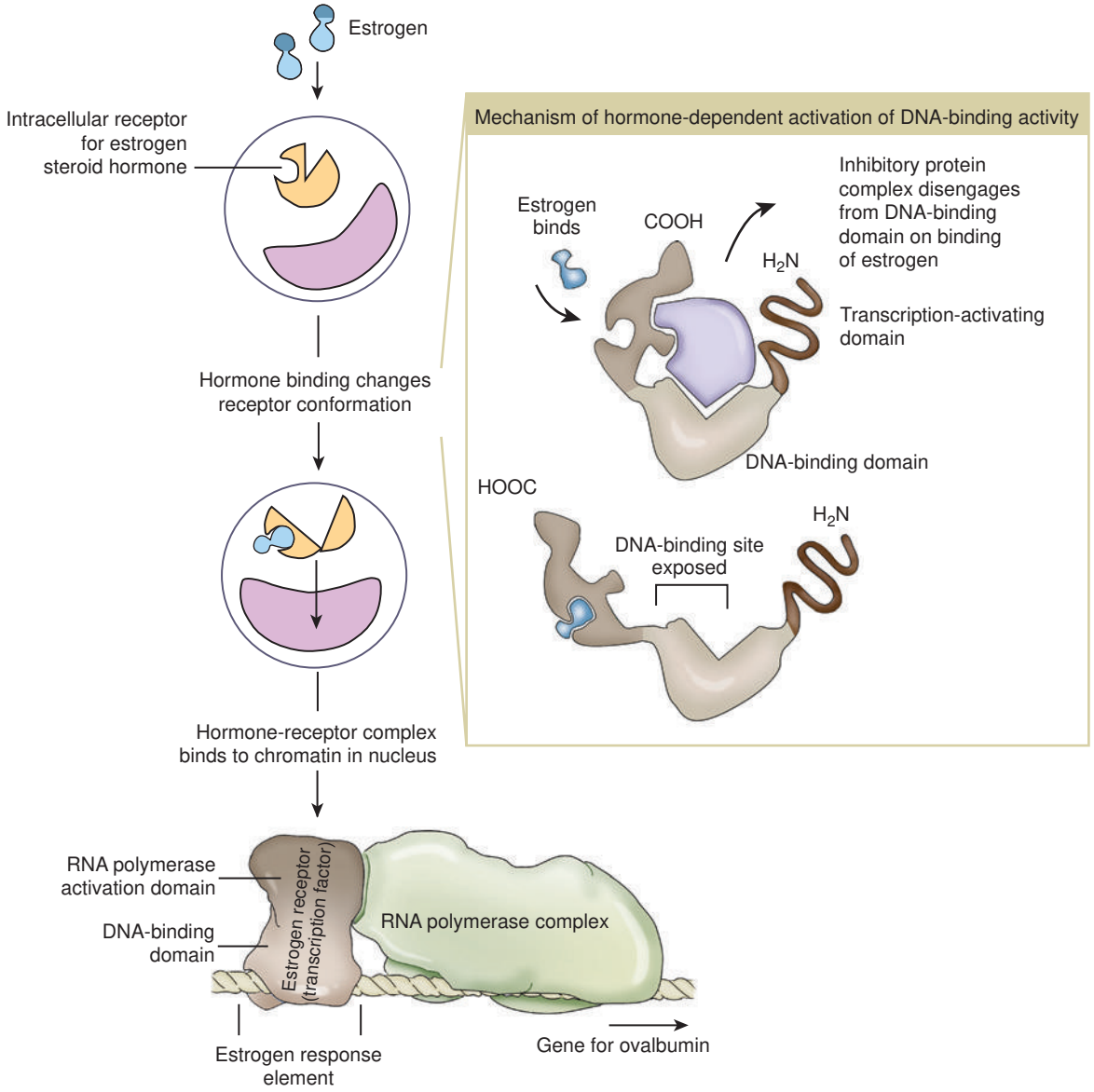
DAG is also produced on activation of PLC, but it is not at all soluble. DAG diffuses in the plasma membrane, binding to and activating membrane-associated protein kinase, *protein kinase C* (PKC). PKC is not an intrinsic membrane protein and can bind reversibly to the cytoplasmic face of the plasma membrane. PKC phosphorylates other proteins and changes their activity. Because of the membrane-bound character of the enzyme, most evidence indicates that PKC phosphorylates membrane proteins such as receptors and ion channels, regulating their function. In the case of the secretory response to some hormone/neurotransmitter stimulus, PKC generally acts separately but additively with IP_3 to produce the response. As with cAMP and PKA, however, much interest focuses on longer-term effects of DAG activation of PKC, particularly its role in growth control and cancer. A class of chemicals long known to promote the onset of cancer, phorbol esters, is a potent substitute for DAG at activating PKC. It is now known that PKC indirectly activates an important transcription factor involved in cell proliferation, nuclear factor kappa B (NF-κB). Thus, as a second messenger and as with cAMP, DAG has both short-term effects and longer-term transcriptional effects.

Steroid Hormones and Other Lipid Signals Interact With Nuclear Receptors, Which Are Transcription Factors Within the Cell

Nuclear receptors are another large class of protein molecules specialized for information transmission and transduction. Nuclear receptors are sufficiently numerous and diverse that they compose a superfamily of evolutionarily conserved and related receptors, as with the GPCRs. All nuclear receptors are transcription factors that respond to the binding of their cognate lipid signal by regulating which genes are expressed within particular cells under particular conditions. Accordingly, one of the conserved features of nuclear receptors is their DNA-binding domain, which can bind directly to specific sequences of DNA (promoter regions) that control the expression of the neighboring gene(s) (Fig. 1.20). As with all other proteins, the DNA-binding function of nuclear receptors is based on their shape. The DNA-binding domain, for example, is a part of the protein shaped into "fingers" by a zinc ion. These *zinc fingers,* also found in many other transcription factors, fit into the grooves of the double helix of DNA at the appropriate base-pair sequence.

Recall that steroid hormones are soluble in oily solvents and are able to diffuse through the lipid bilayer without the mediation of transport proteins. Thyroid hormones are also lipophilic and diffuse through the lipid bilayer. Additionally, several lipid-soluble nutrients are also signaling molecules, including vitamins A and D. Vitamin A is required for vision because it is the covalently bound cofactor for the opsin GPCRs, but it also plays a role in embryonic development. Vitamin D controls Ca^{2+} metabolism. Similarly, saturated and unsaturated fats in the diet are also known to provide signals that control their own breakdown and metabolism and to regulate the differentiation of fat cells (adipose tissue). Consequently, the receptors for these lipid signals are soluble proteins within the target cell. The cellular location of the nuclear receptors varies. Some receptors can be found in the cytoplasm before ligand binding, whereas others are largely restricted to the nucleus (after their initial synthesis in the cytoplasm), but all are functional as transcription factors in the nucleus after activation. The lipid-soluble hormone/nutrient diffuses from the blood into the cell and binds to its receptor, and the hormone/receptor complex is, as in previous examples, the physiologically active entity that ultimately triggers a cellular response. As noted earlier, because the nuclear receptor complex is itself a transcription factor, steroid and thyroid hormones do not require a second messenger; the hormone/receptor complex is itself active within the cell, altering gene expression.

A well-studied example of nuclear receptor action as a regulated transcription factor, with some relevance to veterinary medicine, is the action of estrogen on the reproductive tracts of female chickens (see Fig. 1.20). Estrogen is the principal female sex hormone of birds and mammals, and, of course, hens lay eggs in which the embryo and yolk are surrounded by eggwhite. The principal protein of eggwhite is ovalbumin, which is secreted by the epithelial cells of the avian oviduct as the ovum slides by. Thus one of estrogen's targets in female chickens is oviduct epithelial cells. Estrogen enters the cytoplasm of these cells and binds to its receptor, the *estrogen*



• **Fig. 1.20** Steroid hormone action as illustrated by control of ovalbumin expression by estrogen in hens. The steroid hormone estrogen penetrates the lipid bilayer passively because of the oil solubility of the steroid. Inside the cell, the estrogen binds to a cytoplasmic receptor, the estrogen receptor. The binding of estrogen to its receptor causes the receptor protein to change conformation, which in turn changes the DNA-binding activity of the receptor. The hormone/receptor complex enters the nucleus and binds to regulatory sequences of deoxyribonucleic acid (DNA), the estrogen response element. This binding, in turn, activates ribonucleic acid (RNA) polymerase. This initiates transcription of the ovalbumin gene, an estrogen-responsive gene, to produce messenger RNA (mRNA), which is ultimately translated into the ovalbumin protein for secretion.

receptor. The hormone/receptor complex, but not the ligand-free receptor, is able to mediate estrogen-specific, essentially female-specific, gene transcription. The estrogen receptor complex binds to a sequence of DNA, called an *estrogen response element,* that controls the transcription of a neighboring gene, for ovalbumin in this case. In other cells of the female, binding of the estrogen receptor to the estrogen response elements of other genes would cause these other female-specific genes to be transcribed and ultimately expressed as a protein (e.g., proteins in yolk of egg). Different steroids bind to different receptors (e.g., male sex hormone testosterone binds to testosterone receptor), which bind to different response elements, leading to different genes expressed (e.g., male-specific gene expression).

Differential gene expression and its regulation were initially pursued primarily by molecular biologists. It rapidly gained importance in physiology, however, and will do so soon in veterinary medicine. Humankind will have fewer scruples about controlling gene expression in animals other than their own species (a fact well illustrated by studies of cancer in mice as discussed in the next chapter). Indeed, understanding control of gene expression may prove more important to veterinary students in the near term than for students of human medicine.

CLINICAL CORRELATIONS

Peripheral Edema

History

You examine a 2-year-old cow that has been grazing on poor-quality pasture. The owner states that the cow seems to have a poor appetite, walks slowly, and stands apart from the rest of the herd. The cow has developed swelling beneath the skin of her brisket and ventral thorax.

Clinical Examination

On clinical examination, you find a listless cow standing in a pasture littered with various metal objects. Examination of the cardiovascular system reveals distended jugular veins and abnormal heart sounds characterized by irregular sloshing sounds throughout the cardiac cycle that drastically muffle the first and second heart sounds. Subcutaneous edema (swelling) can be seen throughout the chest and abdomen but most prominently in the dependent ventral areas of the thorax. Pushing on these swollen areas leaves a dent (pitting edema).

Comment

This is a characteristic history of a cow with hardware disease. The cow, grazing on a pasture littered with metal debris, swallows nails, wire, and so forth. Because these objects are heavier than the feed, they drop into the reticulum, a stomach chamber located just caudal to the diaphragm and heart. With the contractions of the reticulum, a metal object migrates through the reticular wall, diaphragm, and pericardium, leading to an inflammatory response in the pericardium (pericarditis). The resulting process is caused by both inflammation and possible secondary bacterial infections from a contaminated metal object traversing regions of the gastrointestinal tract that contain numerous microorganisms, before the object penetrates the pericardium. An inflammatory exudate fills the pericardial sac; it muffles the heart sounds, and a sloshing sound may be heard on auscultation. As this exudative fluid fills the pericardial sac, it limits the pumping efficiency of the heart by limiting its filling during diastole and by obstructing venous return to the heart (see Chapter 21). The result is left-sided heart failure because the heart cannot circulate (pump) the blood throughout the body. This causes the blood to accumulate initially, leading to an increased hydrostatic pressure in the veins and capillaries. As the capillary hydrostatic pressure rises, capillary filtration is favored over reabsorption, and water leaves the capillary and accumulates in the interstitial space. This accumulated interstitial fluid, primarily as the result of increased capillary filtration, is seen clinically as edema. The other common cause of edema is decreased capillary colloidal osmotic pressure from low serum protein. However, this does not usually play a part in hardware disease.

Treatment

Treatment includes surgical removal of the foreign object or objects, antiinflammatory agents, and antibiotic treatment for the pericarditis. Even though considerable inflammation is present, a secondary bacterial infection often contributes to the response. In such an advanced case, however, treatment often is not completely successful.

Horse With Hyperkalemic Periodic Paralysis

History

A 3-year-old Quarter Horse mare, showing in her third halter class of the evening, began having spontaneous muscle twitches (fasciculations). She collapsed in the ring and was unable to rise. She had a long trailer ride the day before the show, and the rider has been feeding her a slightly different diet because less of her normal hay was available. She has never had these signs before.

Clinical Examination

On examination, her temperature, pulse, and respiratory rate are all increased. She is sweating with muscle fasciculations, and she appears to have muscle cramps. She is not very responsive, and she is unable to rise when she is placed lying on her breastbone (sternal recumbency).

Comment

This mare likely has hyperkalemic periodic paralysis (HYPP). It is the result of an autosomal dominant mutation with a single DNA base-pair substitution. This produces a population of abnormal voltage-gated sodium channels at the muscle cell membrane in which the inactivation gate is thought to have trouble docking into its shut position. It is likely that some combination of multiple shows (increased exercise), the change in diet, and the long trailer ride may have initiated this episode. Different feeds, both natural and commercial, can vary widely in their potassium (K^+) content, and the change in diet may have produced increased serum and extracellular K^+. Increased exercise has been shown to have a similar effect. This excess extracellular K^+ causes a slight depolarization in membrane potential, which then results in an opening of sodium (Na^+) voltage-gated channels, along with a failure of the defective Na^+ voltage-gated channels to inactivate. This leads to a persistent flux of Na^+ into the cell, abnormally prolonging the change in membrane potential. This abnormal voltage change can be associated with prolonged inactivation of normal Na^+ voltage gated channels, resulting in loss of electrical excitability from normal stimulation and thus causing paralysis. Stress, such as that associated with the preceding long trailer ride, has been shown to exacerbate the onset of such episodes in HYPP.

Treatment

Initial treatment for horses experiencing an HYPP episode includes feeding corn syrup to promote insulin-mediated K^+ transport out of the extracellular fluid by the Na^+, K^+ pump. More severe cases can be treated with calcium gluconate; the calcium increases the threshold for muscle stimulation. Also, intravenous dextrose can be used to promote intracellular movement of K^+ by stimulating ATP production, in turn activating the Na^+, K^+ pump. Long-term management includes diet and management modifications to limit stress and K^+ intake. Acetazolamide can also be used to increase renal excretion of K^+, as well as stimulate insulin secretion, which stabilizes blood glucose and K^+. Horses can die of severe episodes. Horses can be tested for this mutation.

Practice Questions

1. Increasing the extracellular K^+ concentration will:
- **a.** Have no effect on the resting membrane potential.
- **b.** Cause the resting membrane potential to decrease (i.e., cause the inside to become less negative with respect to the outside).
- **c.** Cause the resting membrane potential to increase (i.e., cause the inside to become more negative with respect to the outside).
- **d.** Increase the concentration potential for K^+ across the plasma membrane.
- **e.** Require the Na^+, K^+ pump to work harder to pump K^+.

2. G proteins are similar to receptors in that both:
- **a.** Bind extracellular signaling molecules.
- **b.** Interact directly with adenylyl cyclase catalytic subunits.
- **c.** Have activated and inactivated states dependent on ligand binding.
- **d.** Are extracellular protein molecules.
- **e.** Directly activate a protein kinase activity.

3. Which of the following statements concerning intracellular Ca^{2+} is *false*?
- **a.** It is a second messenger for hormones and neurotransmitters.
- **b.** It is responsible for excitation-contraction coupling in smooth muscle.
- **c.** An increase in its concentration in a nerve terminal stimulates the release of a neurotransmitter.
- **d.** It activates protein kinase A.
- **e.** Its concentration is increased in the presence of IP_3.

4. If, in a particular capillary bed, the plasma oncotic pressure were to increase and hydrostatic pressure remained constant:
- **a.** More blood plasma would filter from the capillaries.
- **b.** The transport effect would be similar to decreasing hydrostatic pressure.
- **c.** One would suspect a deficiency in blood protein levels.
- **d.** One would suspect an increase in extracellular fluid protein concentrations.
- **e.** Fluid reabsorption on the venous side of the capillary bed would decline.

5. Substance X is found to be at a much higher concentration on the outside of a cell than in the cytoplasm, but no transport of X from the extracellular fluid to the cytoplasm occurs. Which of the following statements is inconsistent with this situation?
- **a.** Substance X has the same electrochemical potential outside and inside the cell.
- **b.** Substance X is large, is poorly soluble in oil, and has no transport proteins in the membrane.
- **c.** Substance X is an ion, and the measured membrane potential is the equilibrium potential calculated by the Nernst equation.
- **d.** Substance X is a steroid molecule.
- **e.** Substance X is actively transported from the cell to the extracellular fluid.

Bibliography

Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 6th ed. New York: Garland Science; 2014.

Lodish H, Berk A, Kaiser CA, et al. *Molecular Cell Biology*. 8th ed. New York: Macmillan Learning; 2016.

Luttrell LM. Reviews in molecular biology and biotechnology: transmembrane signaling by G protein-coupled receptors. *Mol Biotechnol*. 2008;39(3):239–264.

Novac N, Heinzel T. Nuclear receptors: overview and classification. *Curr Drug Targets Inflamm Allergy*. 2004;3(4):335–346.

Valberg SJ. Diseases of muscles. In: Smith BP, eds. *Large Animal Internal Medicine*. 5th ed. St Louis: Mosby; 2014.

Valberg SJ, Carlson G. Muscle cramping. In: Smith BP, eds. *Large Animal Internal Medicine*. 5th ed. St Louis: Mosby; 2014.

2

Cancer: A Disease of Cellular Proliferation, Life Span, and Death

STEVEN R. HEIDEMANN

KEY POINTS

1. Cancer arises from genetic dysfunction in the regulation of the cell cycle, cell life span, and cell suicide.

Control of the Cell Cycle (Proliferation)

1. Cell division is the result of a clocklike cell cycle.
2. Cyclin-dependent kinases (CDKs) are the "engines" driving the cell cycle.
3. The CDK "engines" are controlled by both throttle (oncogene) and brake (tumor suppressor) controls.

Growth Factor Pathway: Principal Stimulator of Cell Proliferation

1. The cell cycle is stimulated by growth factors that bind to and activate receptor tyrosine kinases.
2. The *ras* oncogene contributes to many cancers and serves as a model for understanding *small G proteins.*
3. The MAP kinase pathway leads to the expression of cyclins and other stimulators of the cell cycle.
4. The MAP kinase pathway also mediates the stimulation of the cell cycle by cell adhesion.

Tumor Suppressors: Inhibitors of Cell Cycle

1. Checkpoints in the cell cycle are manned by tumor suppressors.
2. The retinoblastoma and P53 proteins are the main gatekeepers for the cell cycle.

Mechanisms Regulating Cell Suicide and Cell Life Span

1. Apoptosis is the process of cell suicide.
2. Resistance to apoptosis via the intrinsic pathway is a hallmark of cancer.
3. Many types of cancer cells suppress immune attack and so avoid apoptosis via the extrinsic pathway
4. Cellular life span is determined by DNA sequences at the ends of chromosomes.

Tumor Origin and the Spread of Cancer

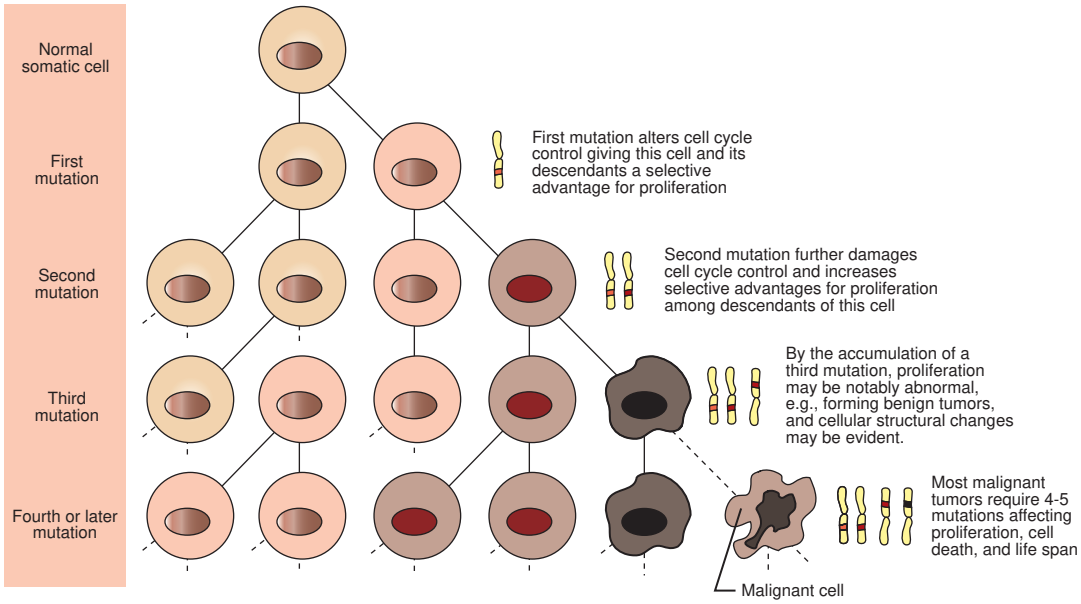
1. Cancer cells may be related to stem cells.
2. Death by cancer is usually the result of its spread, not the original tumor.
3. Growth of solid tumors depends on development of new blood vessels.

Prospective Cancer Therapy

1. Cancer therapy has a hopeful but challenging future.

Traditionally, cancer was (and often still is) first detected in humans and domestic animals by clinicians feeling for an unusual mass of cells, tumor cells. Thus cancer is quite intuitively a disease affecting cellular growth. In the last 30 years, enormous progress has been made in understanding several normal control pathways that regulate cell growth, as well as how these *Rube Goldberg* pathways (see Chapter 1) go wrong in cancer.

The first path to be unraveled, long thought to play a major role in cancer, was the pathway controlling cellular proliferation. By the early twentieth century, cellular proliferation was known to occur by a regular clocklike cycle of chromosomal doubling followed by mitotic division, called the *cell cycle.* However, progress on understanding the molecular control of the cell cycle did not begin until the 1980s from the study of cancer cells but importantly also from the study of the proteins synthesized by fertilized sea urchin eggs; how frogs ovulate; and how yeast cells divide. Cell number depends not only on new cells being formed by cell division but also on cells dying. As a result of studying in detail the history and fate of every cell that arises during embryonic development to form a soil roundworm (a nematode), it was discovered that cells are programmed to commit "suicide." That is, cells can actively kill themselves using metabolic machinery during normal development (e.g., a tadpole losing its tail) and also if the cell has internal damage, such as mutations or oxidative stress. This surprising discovery quickly led to the realization that not only do cancer cells divide inappropriately, but they are also resistant to programmed death and thus continue to divide despite internal damage. The final general process affecting cellular growth is that normal cells, like the organisms they are part of, have a characteristic life span. However, cancer cells were long known to be "immortal," being able to divide indefinitely. How cells age, or become immortal, was not understood until the process of chromosomal duplication was studied in a ciliated protozoan, similar to the familiar *Paramecium* of college biology laboratories.



• **Fig. 2.1** Clonal basis of cancer. Cancer is the result of the accumulation of mutations in a cell lineage of somatic (nongamete) cells of the body. Beginning with a normal cell, mutations occur by chance or by environmental inputs, such as radiation or cancer-causing chemicals, and accumulate to cause cancer.

As these examples illustrate, our understanding of cellular proliferation, cellular life span, and cell suicide came in large part from the study of problems that first seemed distant from the cancer seen in the clinic. As such, the recent progress on cancer is an unusually dramatic example of the importance of understanding basic biology to understand medicine. The vast majority of cancer studies are conducted on humans and in mice, *the* preeminent animal model for cancer, and using cultured cells derived from human and mouse tumors. The much smaller number of studies on domestic animals strongly indicates that the principles derived from humans and mice are generally applicable. However, it is also clear that humans and mice differ in a few aspects of cancer, and thus there are likely to be "special" aspects of cancer for each species. In the case of domestic animals, different species and breeds are known to have differing frequencies of various cancers. For example, the Bibliography at the end of this chapter includes two reviews on the cancer biology of dogs. Veterinary practitioners will need to carefully evaluate the application of knowledge about human and mouse tumors for their patients.

Cancer Arises from Genetic Dysfunction in the Regulation of the Cell Cycle, Cell Life Span, and Cell Suicide

Cancer is a genetic disease (but not usually a hereditary disease) and a uniquely cellular disease. As shown in Fig. 2.1, tumors and other cancers arise from the division of a single mutant cell whose descendants accumulate several additional mutations to become increasingly damaged with respect to control of cellular proliferation, life span, and cell death. That is, cancer is a genetic disease caused by the accumulation of mutations in body cells, such as those of the epithelial cells lining the lungs or the secretory epithelial cells of the mammary glands.

All the cells of a tumor can trace their ancestry back to a single cell that developed an initial deleterious mutation. This first mutation usually occurs in a gene controlling proliferation, such that the cell produces a mutant protein[1] that is a dysfunctional, more permissive regulator of the cell cycle. This greater "permissiveness" provides the mutant cell with more opportunity to proliferate, and it thus has a selective advantage compared with its normal neighbors. Perhaps because of this selective advantage, or because of continued exposure to mutagens (e.g., cigarette smoke, agricultural chemicals), a descendant of this cell accumulates another mutation that also affects some aspect of the cell cycle or cell death. This increases the doubly mutant cell's selective advantage further still, and the downward spiral of increasingly abnormal, dividing cells begins to spin out of control. Scientists agree that this accumulation of mutations in individual genes is necessary for cancer to develop, but some think it is not sufficient. Rather, they argue that cancer only results when the accumulation of mutations eventually leads to large-scale genetic instability, such that whole chromosomes are gained and lost. The majority of spontaneous tumors do have cells with abnormal sets of chromosomes, a phenomenon called *aneuploidy.* Whether aneuploidy is necessary for cancer remains to be seen, but there is no disagreement that cancer cells are in some way badly damaged with respect to genes controlling growth.

The mutations leading to cancer are the same type as those that underlie Mendel's familiar laws of heredity. These include base-pair changes, deletions or additions of nucleotides in the gene, and translocation of one piece of a chromosome to another. However, it is important to understand that the *cells* in which the mutations are occurring are different than those underlying Mendel's laws of inheritance. Mendelian inheritance results from mutations occurring in the *germ line* of the organism. These are

[1]For the many instances when a gene and protein share the same name, this chapter adopts a widely used, but by no means universal, convention for distinguishing genes and their cognate proteins. Gene names are in italics and all lowercase (e.g., *ras*), whereas the protein name is not italicized has one or more capital letters (e.g., Ras). This convention is used throughout in preference to the various species-dependent conventions also used in the literature.

the cells that will produce gametes, either sperm or eggs, and whose deoxyribonucleic acid (DNA) will be passed down to every cell of the offspring following fertilization. The mutations leading to cancer are occurring in nonreproductive cells throughout the body, called *somatic cells.* These are passed down only to a limited number of other somatic cells by cell division, not to offspring through sexual reproduction. Thus although cancer is a genetic disease, only about 10% of the time is it a "hereditary disease," that is, the result of mutation inherited from a parent. In general, cancer appears to be the result of the accumulation of mutations leading to genetic instability in a particular lineage of somatic cells.

Traditionally, cancers are divided into categories based on the cell type involved. Carcinomas are cancers of epithelial cells; sarcomas are derived from connective tissue or muscle; and leukemias are cancers of blood-forming cells. There are many subdivisions based on specific cell types and location of the tumors. However, these names are traditional only; they do not reflect any fundamental differences in the biology of the cancer. Rather, it is now clear that cancers of all types share broadly similar types of dysfunctions controlling cell proliferation, cell suicide, and cell life span.

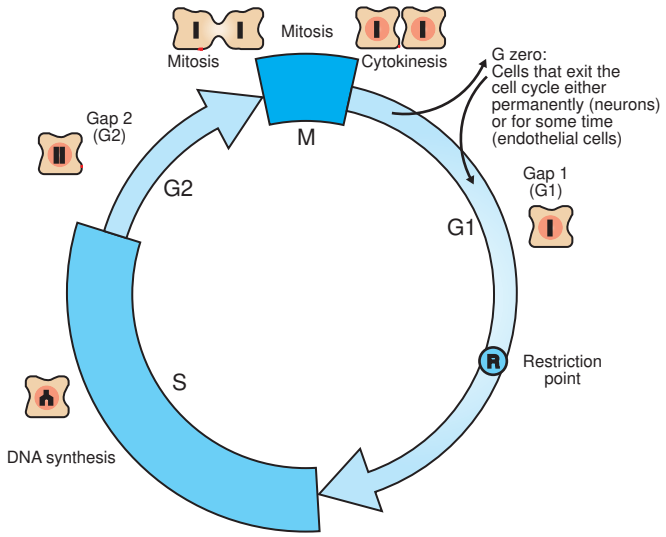
Control of the Cell Cycle (Proliferation)

Cell Division Is the Result of a Clocklike Cell Cycle

The *Rube Goldberg device* that controls cell growth is particularly complex, with many, many more components than the "garage door opener" of Fig. 1.13. To explain these pathways, we begin with the cell cycle that, like the carriage house door, is near the end of the system of control. That is, most of the control elements feed "downstream" to control the cell cycle or intersect with some aspect of cell cycle control.

Fig. 2.2 shows the classic diagram of the cell cycle in which the cell changes its state toward division, progressively going around the diagram, like the hands of a clock. For most mammalian cells in culture, the duration of one cell cycle varies between 18 and 30 hours. Two phases of the cell cycle were identified first and seemed to be where the most important events of the cell cycle occurred. One is *synthesis (S) phase,* during which the DNA is duplicated. The second is *mitosis (M) phase,* during which the duplicated chromosomes are separated to opposite sides of the cell and the cytoplasm divides. In addition to the obvious need for such events if cells are to reproduce, note that both phases must be highly precise. It is crucial for the cell that DNA synthesis produces *exactly* twice the original amount of DNA, no more and no less. Otherwise, there will not be two identical copies of the genetic material to pass on to two identical cells. Similarly, the machinery segregating the duplicated chromosomes during mitosis must partition exactly equal numbers and types of chromosomes to daughter cells, or the cells will be aneuploid. If DNA is not precisely replicated, or if the chromosomes are not properly aligned, the cell cycle is halted by *checkpoints,* as described later.

However, the events during G1 ("gee-one") and G2 phases remained a mystery. The "G" stands for *gap,* because of the decades-long gap in our understanding of what was happening during this time. Although it was suspected that the cell was preparing itself for DNA synthesis during G1 and preparing for mitosis during G2, the nature of these "preparations" proved difficult to determine. In the mid-1980s, work initially conducted on frog oocytes revealed that specialized protein kinases were activated during G1 and G2 to drive the cell into S phase and M phase, respectively.



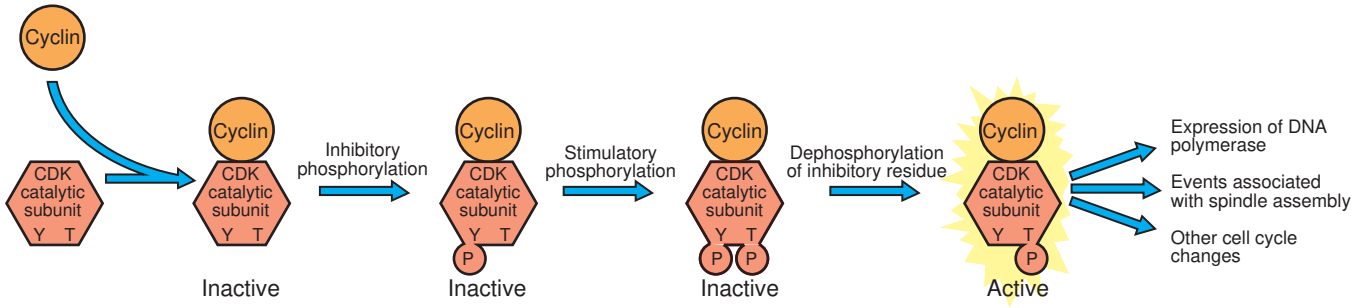
• **Fig. 2.2** The mammalian cell cycle. Cell proliferation occurs by a clock-like progression of phases in which characteristic events occur. The most familiar is M phase (mitosis), during which the cytoplasm and replicated chromosomes are distributed to the daughter cells. Cells then enter G1, during which a "decision" is made whether or not to go forward with the cell cycle; this is the R (restriction) point. The events in G1 then allow S (synthesis) phase to proceed, during which the DNA is replicated to produce exactly two copies. After DNA synthesis, the cell prepares for mitosis during G1, and the cycle is complete. Although cells in culture typically go around the cycle continuously, most cells in the body divide only occasionally. These quiescent cells, as well as cells such as neurons that never divide after differentiation, are in G0, a nondividing phase. Under appropriate stimulation, cells can then exit G0 and are said to reenter the cell cycle.

These special protein kinases are now called *cyclin-dependent kinases* (CDKs).

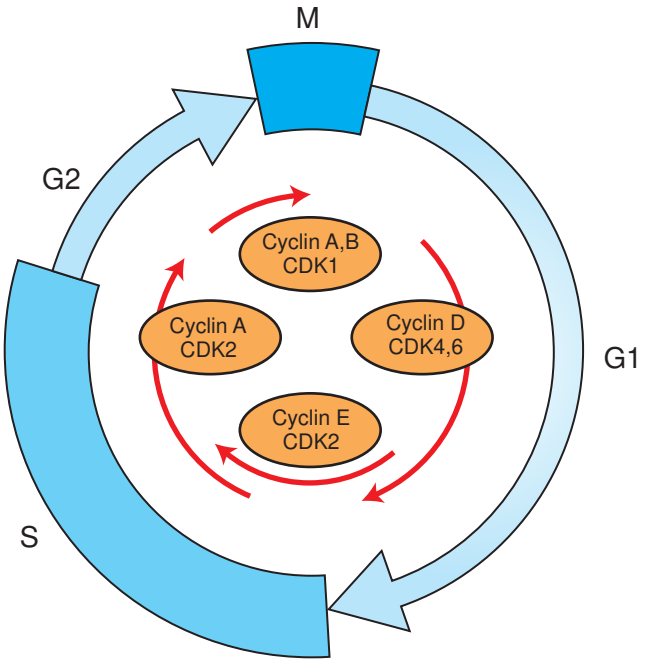
Cyclin-Dependent Kinases Are the "Engines" Driving the Cell Cycle

Recall from Chapter 1 that protein kinases, which are enzymes that phosphorylate other proteins, are important as elements of signaling pathways. For example, the second messenger cyclic adenosine monophosphate (cAMP) acts by activating protein kinase A (see Fig. 1.18), and diacylglycerol as a second messenger activates protein kinase C (see Fig. 1.19). Protein kinases play a major role in many aspects of control of the cell cycle; most importantly, CDKs, when activated, can directly cause a cell to enter either S phase or mitosis, whether the cell is ready or not.

Active CDKs are composed of two different types of protein subunits (Fig. 2.3). The catalytic subunits (numbered CDK1, CDK2, etc.) are the subunits that have enzymatic activity for hydrolyzing adenosine triphosphate (ATP) and transferring the phosphate group to a protein substrate. The other subunit is an activator of the catalytic subunit and is called a *cyclin;* the abundance of this protein increases and decreases during the cell cycle (i.e., the protein concentration cycles up and down during the cell cycle, hence the name). Different cyclins are specific for various CDKs and for the different phases of the cell cycle. The various cyclins are identified by letters, such as cyclin A and cyclin B. Cyclins must reach a threshold concentration to activate the catalytic subunit, and the threshold is achieved as a result of protein accumulation from new synthesis during the G phases.



• **Fig. 2.3** Activation of the cyclin-CDK "engines" of the cell cycle. Activation of CDKs depends on the association of a cyclin with a catalytic subunit and then an appropriate pattern of inhibitory and stimulatory phosphorylations on the catalytic subunit. *CDK,* Cyclin-dependent kinase.



• **Fig. 2.4** Cyclins and CDKs around the cell cycle. Different phases of the cell cycle are associated with and driven by different cyclin-CDK pairs, as shown here. *CDK,* Cyclin-dependent kinase.

When the cyclins have bound to their appropriate catalytic subunit, the cyclin-CDK complex as a whole is activated by achieving a particular state of *phosphorylation.* There are inhibitory sites of phosphorylation around amino acid 15 of the catalytic subunit, and these must be dephosphorylated. There is also a stimulatory phosphorylation site at amino acid 167, and this must be phosphorylated for cyclin-CDK activity. When activated, the CDK phosphorylates various substrates associated with either S phase or M phase. For example, the cyclin-CDK complex responsible for mitosis directly phosphorylates the protein filaments that strengthen the nuclear membrane (lamins). This phosphorylation causes the filaments to disassemble, in turn allowing the nuclear membrane to dissolve, which is an early event of mitosis.

The different phases of the cell cycle are controlled by different cyclin-CDK pairs, as shown in Fig. 2.4. Thus the complex of CDK1 with either cyclin B or cyclin A is the particular CDK pair responsible for driving the cell into mitosis. Cyclins E and A interacting with CDK2 play important roles in initiating and maintaining DNA synthesis in S phase. Cyclin D interacting with either CDK4 or CDK6 functions in late G1 in a "decision" by the cell to commit to DNA synthesis. This decision is called the *restriction (R) point* and is discussed in the later section on tumor suppressors.

Given the importance of cyclins and CDKs in driving the cell cycle, one would expect they would have some connection to cancer. Overexpression of cyclin D is associated with human and mouse breast cancer, and ablation of cyclin D provides some protection against breast cancer in mice. Virtually all multiple myelomas, a type of leukemia, show overexpression of cyclin D. Overexpression of cyclin A is strongly associated with some lung cancers and with testicular cancer of humans, and overexpression of cyclin E is associated with certain human leukemias. Curiously, in contrast to the cyclin subunit, the CDK enzymatic subunit is not known to be mutated in any common cancer.

The CDK "Engines" Are Controlled by Both Throttle (Oncogene) and Brake (Tumor Suppressor) Controls

The CDK-cyclin pairs are controlled by both stimulatory and inhibitory pathways, analogous to an automobile controlled by throttle and brake mechanisms. The throttle mechanisms are largely the result of the cell's environmental inputs. That is, various environmental cues, both soluble signal molecules and insoluble signaling molecules from (or on) neighboring cells, are required for cells to divide. However, the pathways sending inhibitory signals to the cell cycle, the "brakes" for cell division, are largely internal and activated by damage or stress to the cell. In general, these inhibitory signals are like the safety interlocks on an automobile. Just as one cannot start a car in gear, so the cell should not divide if DNA synthesis has not exactly duplicated all the genes and chromosomes, or if something is wrong with the mitotic spindle.

The environmental stimulatory signals for cell division can be as simple and nonspecific as availability of nutrients, to the extent that cells only divide when they have approximately doubled in size through synthetic growth. However, two more specific stimulators of the cell cycle are primarily implicated in cancer. One is the response to soluble growth factors found in the circulation and in the extracellular fluid surrounding cells (see Chapter 1). Growth factors are proteins secreted by a variety of other cell types that are required for the division, and indeed survival, of normal, noncancerous cells. Cancer cells, however, can divide and survive with little or no stimulation from growth factors because of the

acquired ability to synthesize growth factors of their own or the inappropriate activation of downstream elements in the signaling pathway.

The second stimulatory pathway of general importance in cancer is cell attachment. From just a simple mechanical point of view, it is obvious that the cells of multicellular organisms must be tightly attached to one another and to their surrounding matrix (similar to tendon) or we would otherwise be jelly, juice, and bubbles on the floor. Also, however, attachment of cells to their surroundings is a source of specific and complex information to the physiology of the cell. One of the most important such messages is a "permissive" signal to divide. Normal cells must be anchored to some substrate to respond to other signals to divide. That is, most normal animal cells show *anchorage dependence of growth.* For this reason, vertebrate cells in culture are grown on the surface of a dish or flask, not in suspension the way bacteria are cultured. Again, cancer cells have lost this normal restriction on proliferation, and many cancer cells can divide and survive in suspension. The common test for the absence of anchorage dependence is growth in soft agar; cancer cells will, but normal cells will not, divide and form colonies when suspended in soft agar. Thus cancer cells can survive unattached while riding the circulation to relocate in a different tissue than that of the original tumor. In this way, cancer cells "go walkabout" and are able to spread through the body, a process called *metastasis,* which is ultimately the cause of death in most cases of cancer.

If attachment among normal cells had only a stimulatory effect on cell division, one might imagine that it would favor unlimited proliferation, overgrown organs, and tumor formation. It is not surprising then that normal cells with "enough" cell attachment are inhibited in cell division. When normal cells are, indeed, well attached and completely surrounded by their neighbors, proliferation is inhibited. Normal cells detect their surrounding cell density and at a threshold "take their foot off the accelerator," and proliferation ceases. This is called *contact inhibition of growth* and is in part regulated at the level of CDK activation. As a result, normal cells in culture only grow to form a "monolayer," one cell thick. Normal cells will not "pile on" and form multilayered clumps, basically minitumors, on the dish. You will not be surprised to learn that, yet again, cancer cells lose this "good neighbor" regulation of the cell cycle and are thus able to form tumors. Correspondingly in culture, cancer cells "overgrow the monolayer" and, like growth in soft agar, this is a common adhesion-based test to distinguish cancer cells from normal cells.

The Rube Goldberg pathways that underlie the proliferative signals of growth factors and adhesion, both stimulatory and inhibitory, are similar and intersect. These "throttle" contraptions begin with a soluble signal binding to a growth factor receptor and a "solid-state" signal about attachment to the surrounding tissue. However, both pathways quickly converge on the same stimulation pathway for conserved cell division. These stimulatory pathways are driven by proteins that were originally identified as being encoded by genes in viruses that caused cancer in animals. Thus these were named *oncogenes,* literally "cancer genes." A major breakthrough came with the discovery that these oncogenes were actually derived from the host genome, not genes normally encoded in the virus. That is, viruses had stolen cell cycle control genes from their animal host cell. Being viruses, they did not take good care of the animal cell cycle genes they stole. The stolen genes mutated into deranged cell cycle regulators. Subsequently, the same mutant genes that were found in viruses were found to explain many spontaneous cancers in humans and in the long-used experimental tumors of mice. The finding that cancer was caused by abnormal host genes helped confirm that cancer was a somatic genetic disease due to mutations in the tumor cells.

Further analysis revealed that these oncogenes often encode normal stimulators of the cell cycle, and the mutations involved had the effect of permanently activating an element in the cell cycle pathway. You can see how this would work based on the Rube Goldberg cartoon of Fig. 1.13. Note that all the elements in the garage door opener are stimulatory; if any one turns "on," a signal is sent "downstream" to cause the garage door to open. If the fish tank of the cartoon were to "mutate" by developing a leak, an "on" signal would be sent downstream of the fish tank, regardless of whether a car had pulled into the driveway. So it is with the oncogene elements controlling the cell cycle. If one of the elements mutates to turn itself "on," that is, acquired a *gain-of-function mutation,* it will stimulate cell division and contribute to cancer. To return to the automobile analogy, oncogenes represent a stuck throttle or accelerator pedal. The normal, well-behaved versions of the oncogene (a watertight fish tank before the bullet, Fig. 1.13) are called *proto-oncogenes.* Thus strictly speaking, oncogenes have their normal equivalent as proto-oncogenes. However, given this awkward usage, increasingly the normal versions are also informally called oncogenes, and it is usually clear from the context whether the mutant or normal version is being discussed. The molecules and molecular events of the oncogene pathway (also called the *growth factor* or *MAP kinase pathway*) are discussed later.

The mechanisms to stop the cell cycle, the "brakes," are called *checkpoints.* Progress through the cell cycle depends on appropriate conditions being reached within the cell before a "decision" is made to go ahead with division. The first such checkpoint occurs before S phase. During G1, the cell checks itself over particularly with respect to DNA damage. The cell has sophisticated pathways to detect and repair DNA damage, such as mismatched bases detected in the double helix. For needed repairs to take place, however, DNA synthesis is delayed; the checkpoint is "engaged." If the DNA is properly repaired, the checkpoint is disengaged, and after the delay, the cell goes ahead into S phase. However, if the DNA damage cannot be repaired, the checkpoint machinery is supposed to signal a more serious consequence. If the checkpoint is not disengaged after about a day, the cell "commits suicide." Thus the checkpoint (or braking machinery) is tied into both the CDK engines and the processes of cell suicide, as described later. Similarly, the second checkpoint is in mitosis and checks for proper mitotic spindle assembly and correct chromosome alignment. Here again, if damage is detected, there are repair mechanisms, and a properly repaired cell will go into M phase after a delay for repair. If no repair can be made, the cell commits suicide.

The molecules and their interactions that underlie both oncogene ("throttle") pathways and checkpoint ("brake") pathways are now covered in greater detail, beginning with the role of growth factors.

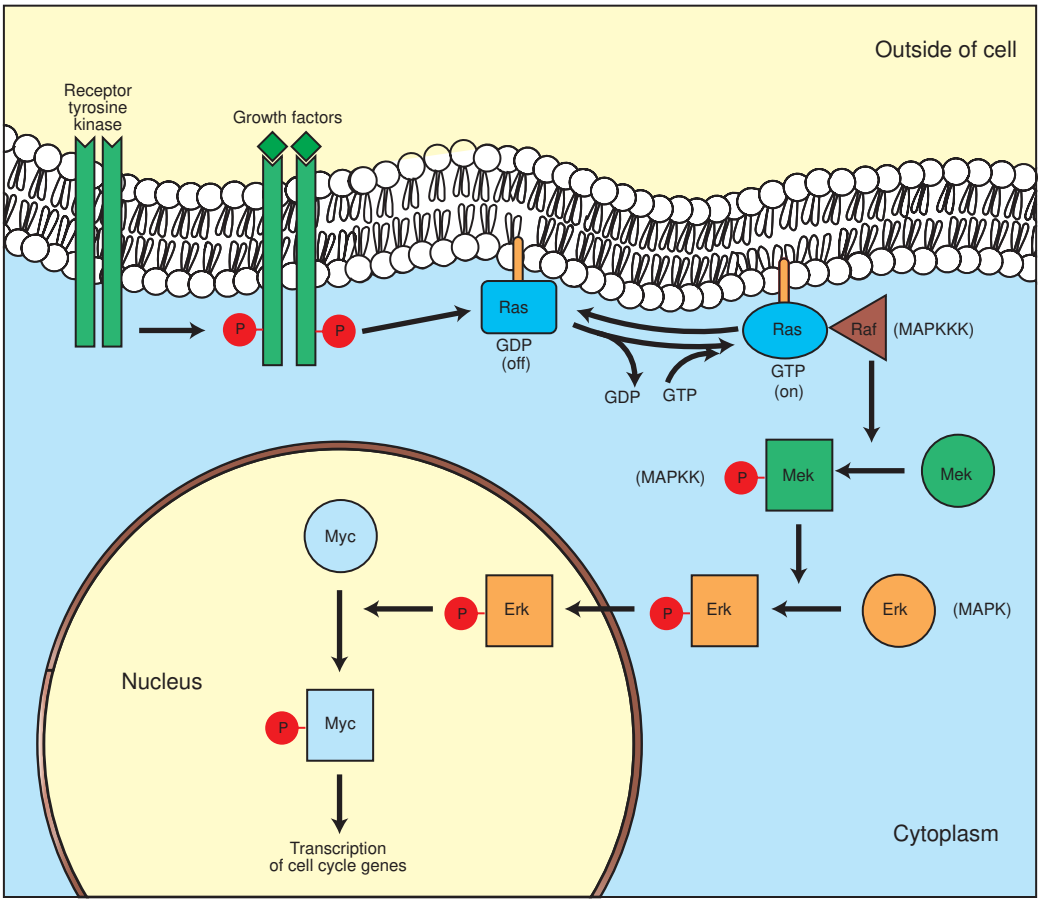
Growth Factor Pathways: Stimulators of Cell Proliferation

The Cell Cycle Is Stimulated by Growth Factors that Bind to and Activate Tyrosine Kinases

The growth factor/oncogene pathway begins with growth factors that function in a familiar way, as discussed in Chapter 1; they bind to and activate an integral membrane protein receptor. Indeed, growth factor receptors belong to the third large receptor

family for environmental signals, the *receptor tyrosine kinase* family. This family of signal transducers has some similarities with the G-protein–coupled receptors (GPCRs), but also some important differences. Receptor tyrosine kinases (RTKs) do not require second messengers, but they do function through protein kinase activity (as many GPCRs do). The structure of RTKs is such that binding of ligand (a growth factor) by the extracellular portion of the receptor directly activates protein kinase activity by the cytoplasmic portion of the protein. The receptor itself is an enzyme (Fig. 2.5). Thus the RTK activates a cytoplasmic signal without the need for a second message. RTKs specifically add a phosphate group to a tyrosine residue of the substrate protein. This differs from the protein kinases discussed in Chapter 1 (PKA and PKC), which add the phosphate to serine or threonine residues. Phosphorylation of tyrosine residues within a protein is largely (but not exclusively) specialized to control cell growth pathways, and therefore tyrosine kinase activity generally is associated with stimulation of proliferation.

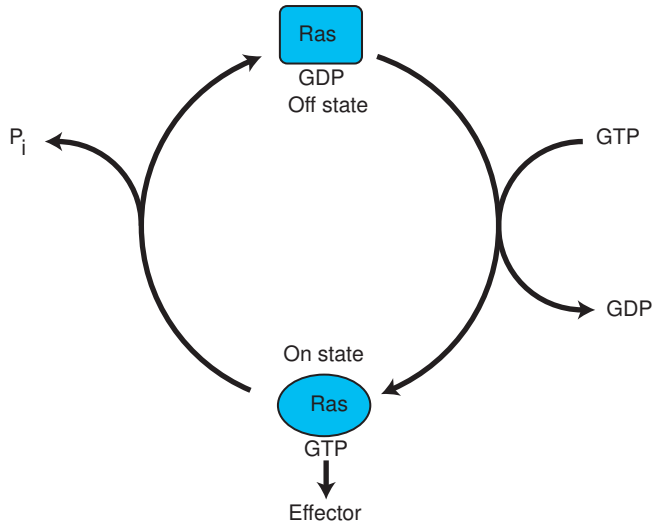
The growth factors that bind to the RTKs are too diverse to be discussed at length in this chapter. Rather, one important similarity for introductory professional students is that these factors are all poorly named, so do not judge the factor by its name. Sometimes growth factors have "growth factor" in their name; some are referred to as *cytokines,* and some are called *colony-stimulating factors* (for growth of colonies in soft agar, as previously mentioned). Further confusion arises because their names always reflect their history but rarely their broader function. Thus "epidermal growth factor" stimulates cell division in many more types of cells than only skin cells, but it was discovered using skin cells. The other, more important similarity among growth factors is that, whatever their name, they share a conserved basic pathway and "strategy" for control, as with the numerous ligands binding GPCRs and nuclear receptors, of their downstream effectors, in this case the CDK engines of the cell cycle. Growth factor activation of RTKs stimulates a pathway involving a G-protein "on-off" molecular switch, the *Ras protein* introduced in Chapter 1, and uses a cascade of protein kinases, both tyrosine and serine-threonine, called the *MAP kinase pathway.* Ultimately, the MAP kinase pathway activates transcription factors, in turn controlling the expression of cyclins, and other direct regulators of CDKs (see Fig. 2.5).



• **Fig. 2.5** Growth factor/oncogene pathway. This diagram shows the normal stimulatory pathway by which growth factors lead to cell division. Growth factors bind to membrane receptors (receptor tyrosine kinases, RTKs) that are themselves protein kinases. As shown here, after activation by binding a growth factor, the first protein to be phosphorylated at tyrosine residues is the receptor protein itself. This in turn causes a small G protein, Ras, to exchange GDP for GTP and thus be "turned on." The activated Ras then activates the first protein kinase in a conserved pathway of three kinases, called the MAP kinase pathway. For more detail on Ras and the MAP kinase pathway, see the text. Finally, this series of activating phosphorylations leads to the activation of transcription factors, such as Myc, in turn leading to the expression of genes directly involved in driving the cell cycle (e.g., expression of cyclin D). In this pathway, gain-of-function mutations of the RTKs, Ras, and Myc are particularly important in human cancers. *GDP,* Guanosine diphosphate; *GTP,* guanosine triphosphate.

The *ras* Oncogene Contributes to Many Cancers and Serves As a Model for Understanding Small G Proteins

After activation of the RTK, the next major step in the growth factor/oncogene pathway in normal cells is activation of the protein product of the *ras* proto-oncogene. Investigations of how it worked revealed that the Ras protein was an important member of the small G-protein family of molecular regulators, all of which have intrinsic guanosinetriphosphatase (GTPase) activity and serve as molecular "on-off switches." These proteins control many basic cellular functions, and the heterotrimeric G protein evolved from Ras-like ancestor proteins (see Chapter 1). Indeed, in yeast it is Ras, not a heterotrimeric G protein, that controls adenyl cyclase and phospholipase C (see Fig. 1.16). Fig. 2.6 illustrates the duty cycle of this on-off switch and its basic similarity to the alpha subunit (G_α) of the heterotrimeric G proteins. Ras, other small G proteins, and G_α all are in the "on" state when they have guanosine triphosphate (GTP) bound to them (because of receptor activation). All are in the "off" state when the G protein hydrolyzes its GTP so that guanosine diphosphate (GDP) is now bound. You can see how this gene could be discovered as an oncogene, that is, a gene in which a gain-of-function mutation contributes to the development of cancer. If the GTPase activity is lost by mutation, this simple, enzymatic on-off switch remains trapped in the "on" position (the accelerator pedal is stuck). It continues to send an activating signal to the downstream cell cycle machinery without the presence of growth factors or the activation of RTKs. In fact, such mutations in Ras underlie its oncogenic function, and it is estimated that 30% of human cancers have activating mutations in their *ras* gene.



• **Fig. 2.6** Duty cycle of the Ras molecular "on-off switch." Ras serves as a model for the activity of small G proteins, of which there are hundreds in the cell. The molecular mechanism of Ras is similar to the alpha subunit of the heterotrimeric G protein, discussed in Chapter 1 and which evolved from Ras-like proteins. As shown here, Ras is in the "off" state when bound to GDP. Activation of receptor tyrosine kinases leads to nucleotide exchange: GDP is lost and GTP is bound. In the GTP-bound form, Ras is in the "on" state and sends a stimulatory signal downstream, in this case to Raf in the MAP kinase pathway (see Fig. 2.4). Normally, Ras rapidly returns to the off state because an intrinsic GTPase activity of the Ras protein hydrolyzes the GTP to GDP. This nucleotide-dependent on-off cycle is characteristic of all normal small G proteins. *GDP,* Guanosine diphosphate; *GTP,* guanosine triphosphate.

Other small G proteins control a myriad of cellular functions, including others involved in cancer. Thus the Rho subfamily of small G proteins is directly involved in the spread of cancer because it helps regulate actin assembly and activity. As described later, the spread of cancer depends on the ability of cells to migrate through tissues. This "crawling" motility in turn depends on a musclelike mechanism based on actin and myosin (see Fig. 1.4). Although the basic on-off activity of Ras and Rho are the same as that shown in Fig. 2.6, Rho is connected to actin, whereas active Ras activates the elements of the MAP kinase pathway.

The MAP Kinase Pathway Leads to the Expression of Cyclins and Other Stimulators of the Cell Cycle

GTP-bound Ras causes the sequential activation of a series of protein kinases, called Raf, Mek, and Erk. Raf phosphorylates and activates Mek, which in turn phosphorylates and activates Erk, as shown in Fig. 2.5. This trio of kinases is called *the mitogen-activated protein kinase,* or MAP kinase, pathway (a *mitogen* is a stimulator of mitosis, e.g., a growth factor). If any of these three protein kinases should experience a gain-of-function mutation irreversibly activating the protein kinase, a stimulatory signal is sent down the remainder of the pathway. Thus, as with *ras,* these three kinase genes act as oncogenes.

One important example of a gain-of-function mutation among the three MAP kinases involves the first of these MAP kinases, Raf. A single–amino-acid mutation in the kinase domain of Raf (a substitution of glutamate for normal valine at amino acid 600) causes permanent activation of Raf in approximately 50% of human melanomas, a very deadly cancer, and is also common in thyroid cancers. As described for mutations in Ras, activation of Raf sends an unregulated stimulatory signal downstream to the other MAP kinases, leading to unregulated proliferation of the cancer cells. Recent clinical progress involving melanoma illustrates the importance of the Raf, Mek, Erk pathway and of understanding which particular mutations are involved in a given patient's cancer. Two recently developed drugs, vemurafenib and dabrafenib, target the mutant Raf and significantly prolongs the life span of those melanoma patients harboring this *raf* mutation but has no effect in cases of melanoma with normal Raf/*raf.* However, melanoma patients typically develop resistance to the two Raf-directed drugs and so these drugs are now used in combination with drugs that specifically inhibit Mek. This combination therapy strengthens the blockade of the Raf, Mek, Erk pathway and leads to better outcomes.

Raf, Mek, and Erk are a specific example of yet another conserved but diverse general module of information transduction. There are MAP kinase trios other than Raf, Mek, and Erk. Although it is not worthwhile to give specific names to all the various pathways, it should be noted that these trios have a systematic set of names for their elements. Raf is a MAP kinase, kinase, kinase (a MAPKKK). Mek is a MAP kinase, kinase (MAPKK), and Erk protein is the MAP kinase (MAPK) itself. This jargon is awkward, but it is widely used and logical, as Fig. 2.5 suggests.

When activated, Erk activates one or more transcription factors that control the transcription and translation of a key regulator of the cyclin-CDK engine. One of these transcription factors, Myc ("mick"), is encoded by another important oncogene/proto-oncogene. As with *ras,* the *myc* gene is mutated in a high frequency of human tumors, giving rise to an oncogenic form able to activate

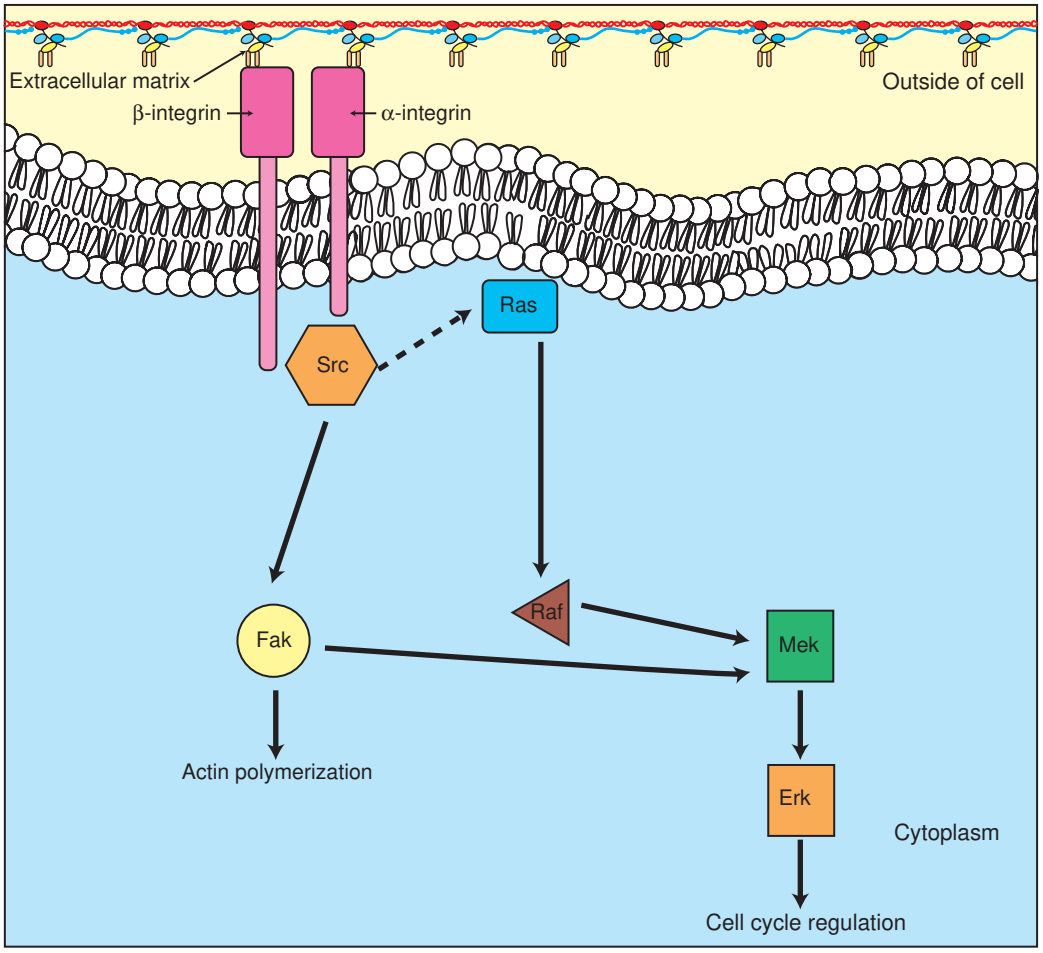
the cell cycle. As shown in Fig. 2.5, Myc protein is involved in the transcription of a variety of cyclins and of the CDK2 catalytic subunit and plays a significant role in allowing the cell to pass from G1 to S phase. Myc is also involved in many other transcription events related to cell growth, differentiation, and cancer.

This completes the growth stimulatory pathway beginning with a growth factor binding to its RTK receptor that, through Ras, a MAP kinase cascade, and a transcription factor, eventually leads to a direct "throttling up" of a cyclin-CDK engine. This same pathway is used similarly to transduce the information about the other major stimulator of cell division, cell attachment.

The MAP Kinase Pathway Also Mediates the Stimulation of the Cell Cycle by Cell Adhesion

As noted earlier, the other major throttle mechanism to regulate the cyclin-CDK engines of the cell cycle is cell adhesion. Cell adhesion, as with growth factor stimulation, ultimately stimulates cyclin-CDK pairs through the MAP kinase pathway. Two types of cell contact are involved in normal growth and proliferation. The most obvious is cell-cell adhesion; most cells are tightly attached to their neighboring cells. The second type is cell adhesion to an extracellular matrix (ECM) of fibrous proteins. Eighty percent of human and mouse cancers arise from epithelial cells (carcinomas), and all epithelial layers are attached to an ECM. The adhesion proteins that bind to other cells or to the ECM are *adhesion receptors.* Adhesion receptors are responsible for the mechanical aspect of attachment but also act similar to other receptors in transducing information across the plasma membrane. In this case adhesion receptors communicate the information that the cell is anchored and can divide.

Both cell-cell and cell-ECM adhesion activate the MAP kinase pathway, similar to growth factors, but the Ras intermediate is less important here. Fig. 2.7 shows the activation of the MAP kinase pathway as a result of cell-ECM adhesion. The adhesion receptors that bind to ECM are called *integrins* and these activate the MAP kinase pathway via two important intermediates that are oncogenes. One is Src ("sark"), a protein tyrosine kinase and the first oncogene (*src*) to be discovered. Unlike the RTKs previously described, Src is not a receptor. However, Src is located on the inside face of the plasma membrane, where it can interact with adhesion receptors. Another important intermediate is also a protein tyrosine kinase, called Fak (focal adhesion kinase). As before, activation of Src and Fak activate the MAP kinase pathway, leading to increased cell



• **Fig. 2.7** Cell adhesion functions through the MAP kinase pathway to stimulate cell division. In addition to the growth factor stimulation of proliferation shown in Fig. 2.5, normal epithelial cells also require stimulation of the MAP kinase pathway through adhesion to the extracellular matrix. The adhesion receptors are integral membrane proteins called integrins, which are activated by binding proteins of the extracellular matrix. Activation of integrins leads to activation of two protein kinases, Src and focal adhesion kinase (Fak), which in turn activate the MAP kinase pathway.

division. Again, mutation or overexpression of *src* and *fak* sends inappropriate stimulation to the cell cycle machinery, which facilitates cancer. As mutant oncogenes, *fak* is associated with aggressive melanomas in humans. The *src* oncogene was named because of its ability to cause sarcomas in chickens.

Several other growth stimulatory pathways work in much the same manner as the growth factor and adhesion pathways. Most stimulatory pathways involve protein kinases and G proteins controlling the transcription of genes encoding proteins that are part of or close to the workings of the cyclin-CDK engines.

Having introduced the fundamentals of stimulatory pathways in the cell cycle, we now change our focus to consider the equally Rube Goldberg–like pathways that provide the brakes to the cell cycle.

Tumor Suppressors: Inhibitors of Cell Cycle

Checkpoints in the Cell Cycle Are Manned by Tumor Suppressors

The cell cycle machinery also has crucial "brake" mechanisms that function as checkpoints, as noted earlier. The existence of brake and checkpoint mechanisms operating in the cell cycle were discovered by fusing a normal cell with a cancer cell of the same type to form a hybrid cell with two nuclei. The resulting hybrid cell invariably showed normal regulation of growth. Apparently, a normal copy of some gene or genes present in the normal cell was able to suppress the altered activity of a mutant gene in the cancer cell. Thus these genes and their encoded proteins were called *tumor suppressors.*

Tumor suppressors play several different functional roles in braking and checking, and they can be divided into two broad types: gatekeepers and caretakers. *Gatekeepers* are genes and proteins that are involved in the actual checkpoint machinery connecting cell damage with a halt in the cell cycle. Thus *P53* (protein of 53-kilodalton mass) is a gatekeeper importantly involved in the pathway that detects DNA damage; it causes a halt in the cell cycle and, if the damage cannot be repaired, signals the cell to undergo programmed death. It is thought that about 50% of human cancers have a mutation in P53. *Caretakers* are usually proteins involved in the repair of damage or the normal maintenance of proteins crucial in the cell cycle. A human example of a caretaker gene and protein is Brca1 (breast cancer 1). This protein is normally involved in the repair of nucleotide mismatches (e.g., G paired with T rather than with C in the complementary DNA strand), and its mutant allele has been found to underlie familial (hereditary) breast cancer in some families.

With these normal functions, one can see how these genes and proteins would suppress tumor activity and cell proliferation. If they are working, DNA is repaired before the cell attempts to divide; this would tend to prevent mutation or other types of genetic instability. However, *loss-of-function mutation* in these genes means the cell now has lost the ability to detect or repair DNA damage. For example, when P53 is nonfunctional, even a badly damaged cell may not receive an adequate signal to commit suicide, and this already-mutant cell can continue to divide. Thus tumor suppressor genes are associated with loss-of-function mutations in cancer, not gain-of-function mutations as for oncogenes. Returning to the automobile analogy of brakes, mutant tumor suppressor genes resemble dysfunctional braking systems or no brakes at all.

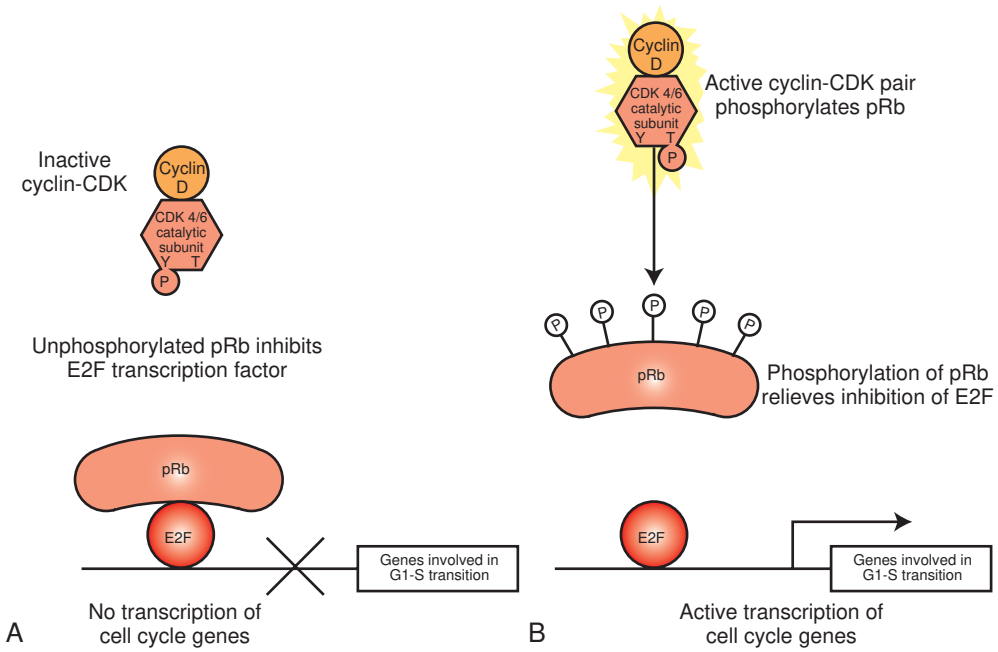
We focus on two gatekeeper-type tumor suppressors because their role and importance in cancer are clear. The role of caretakers such as Brca1 is both more complex and more uncertain (see suggested reading on *brca* in the Bibliography).

The Retinoblastoma and P53 Proteins Are the Main Gatekeepers for the Cell Cycle

Retinoblastoma is a rare, hereditary, childhood cancer of the retina of the eye. Despite its rarity and the fact that it cannot be induced in mice, retinoblastoma has played an important role in the study of cancer. A statistical study of the disease in the early 1970s provided the best evidence then available that human cancer is a genetic disease. Alfred Knudsen showed that children with retinoblastoma typically inherit one mutant copy from a parent (a *germ line mutation*), but then require a second *somatic mutation* in cells giving rise to the retina. Knudsen's *two-hit hypothesis* was a forerunner to the idea that cancer develops by the accumulation of mutations in a cell lineage. (Retinoblastoma tumors do require the accumulation of additional mutations beyond the two retinoblastoma genes being mutant.) Subsequently, the retinoblastoma gene, *rb,* was the first tumor suppressor gene to be cloned. Study of the encoded protein, pRb, showed that it played a central role in controlling the transition from G1 to S phase of the cell cycle.

The *retinoblastoma protein* is a repressor of a transcription factor whose activity is required for the cell to enter S phase from G1 (Fig. 2.8). The transcription factor is E2F, which controls the expression of a wide variety of genes/proteins required for DNA synthesis, including cyclin A, CDK1 (see Fig. 2.4), and subunits of DNA polymerase. The retinoblastoma protein is a potent inhibitor of E2F only when it is bound to E2F directly, which requires pRb to be in an unphosphorylated state. The repression of E2F is released by phosphorylation of pRb by cyclin-CDK pairs operating early in G1 in the cell cycle. As discussed, growth factor stimulation of the MAP kinase pathway leads to expression of cyclin D (see Fig. 2.5), which in turn makes a pair with either CDK4 or CDK6 to make an active CDK. One of the substrates for cyclin D/CDK4, 6 is the retinoblastoma protein. When pRb is phosphorylated by CDK4, 6, it releases from E2F, allowing this transcription factor to promote RNA polymerase activity on genes with E2F promoter regions (see Fig. 2.8). It is this release of inhibition by CDK-mediated phosphorylation of pRb that constitutes the molecular mechanism underlying the R-point "decision" to divide late in G1 mentioned earlier and shown in Fig. 2.2. If both copies of *rb* are mutant, as in retinoblastoma, there will be no active repressor molecules to bind to E2F, and the decision will always be to divide, regardless of other conditions. E2F then promotes uncontrolled expression of S-phase genes whether or not CDK4, 6 has been activated (in part) by growth factors and adhesion, thus making a contribution to unregulated growth and to cancer. Conversely, in its normal, nonmutant form, pRb tends to suppress tumor formation by acting as a gatekeeper, only allowing the cell "to cross the border" between G1 and into S phase if normal growth factor and adhesion signals are received. Thus pRb plays a crucial gatekeeper role in healthy, normal cell cycle control.

The other crucial gatekeeper between G1 and S phase is P53. Unlike pRb, P53 does not participate in healthy cell cycles; P53 is only active in response to cell damage, usually DNA damage, or stress, such as low O_2 concentration or oncogene activation (Fig. 2.9). The role of P53 is to ensure that stressed/damaged cells are either repaired or, if not, commit suicide before being allowed to replicate their DNA. As a gatekeeper, P53's mechanism is also more direct than pRb; P53 is a transcription factor, and



• **Fig. 2.8** Retinoblastoma protein and the G1-to-S transition. (A) In quiescent cells or cells early in G1, retinoblastoma protein (pRb) exists in a nonphosphorylated state that is a direct inhibitor of the E2F transcription factor. The principal CDK pair of G1, cyclin D with CDK4 or CDK6, phosphorylates pRb, releasing its inhibition of E2F. (B) Activated E2F then participates in the expression of a variety of genes required for S phase, including the cyclins and CDKs of S phase and subunits of DNA polymerase.

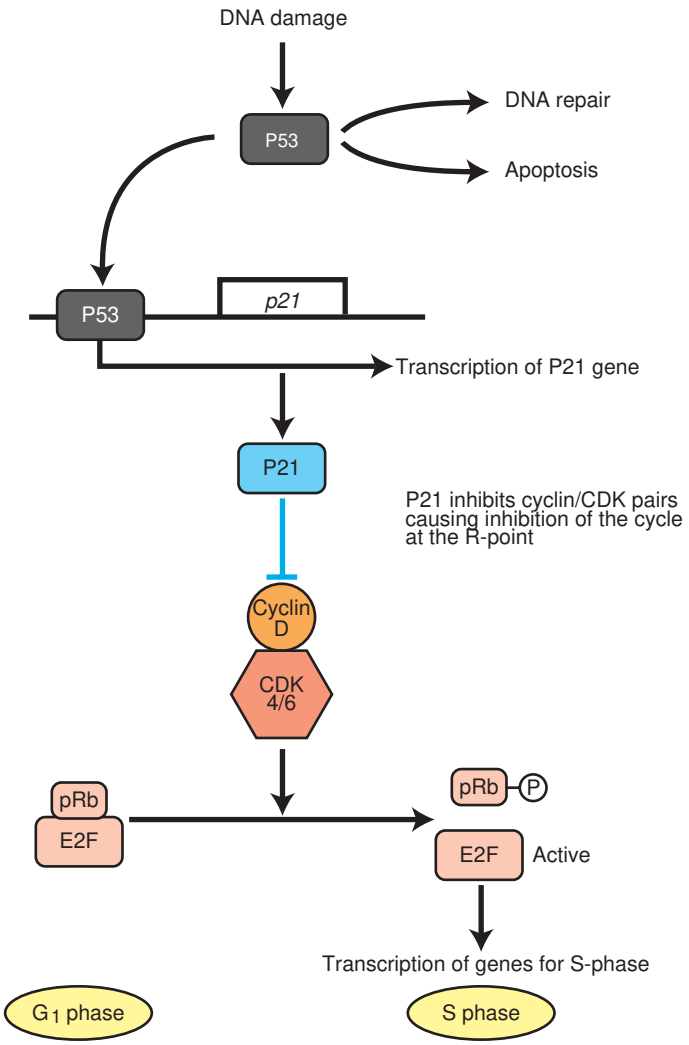
P53 activation stimulates the expression of a protein that is a powerful general inhibitor of all the cyclin/CDK engines. As a transcription factor, P53 also mediates the expression of genes that encode stimulators of cell death, as discussed shortly. Whether the cell responds to P53 by cell cycle arrest to allow repair, or by committing suicide, depends on multiple factors, but presence of an oncogene is among the most important. Normally, the cell cycle arrest activity of P53 is dominant to its death-inducing activity. However, in the presence of oncogenes, including *myc,* suicide is favored. This illustrates clearly the normal tumor suppressor activity of P53; although a cell expressing an oncogene will tend toward increased proliferation, the same oncogene, acting through P53, activates a death pathway to prevent expansion of the mutant cell population.

The activation of P53 occurs in part through mechanisms familiar from previous examples of protein control, including phosphorylation and binding with other proteins. In addition, P53 activity is also regulated simply by an increase in its concentration within the cell. That is, P53 is normally synthesized at a steady but slow rate throughout the cell cycle and is normally degraded at a similar rate. In healthy cells, the half-life for a P53 molecule is about 30 minutes, but this increases threefold to sevenfold in response to DNA damage. Even one double-strand break in DNA has been shown to increase P53 concentration rapidly in some cells. Again, it is clear how P53 serves as both a gatekeeper and a tumor suppressor. Activated P53 prevents a cell with DNA damage from crossing the G1-S boundary (its gatekeeper function), which in turn prevents mutant cells from being allowed to accumulate additional mutations (its tumor suppressor function).

However, if the *p53* gene suffers a loss-of-function mutation and the protein cannot act as a transcription factor, a damaged cell will be able to divide, increasing the probability of accumulating further damage and leading to possible cancer. Thus *p53*/P53 is one of the most important single genes and proteins involved in human cancers; in 1993, the journal *Science* even named it "Molecule of the Year." About 50% of human tumors have a mutation in *p53,* with most of these eliminating DNA binding, disabling its transcription factor activity. When the *p53* gene was "knocked out" in mice, 74% of the animals developed cancers by 6 months of age (young adult). Among experimental mice that had one or two normal copies of the gene, only 1% of animals developed a tumor by 9 months.

In addition to a checkpoint for S phase in which DNA damage provides an important regulatory signal, the other major checkpoint occurs during mitosis. This checkpoint responds to mitotic spindle abnormalities or damage and to abnormalities in the array of chromosomes within the spindle. Again, one can easily see how mutations that disrupted such "safety interlocks" could lead to further damage, by segregating both replicated chromosomes into one daughter cell, for example, with no copy of that chromosome in the other daughter cell. This would lead directly to aneuploidy. Among human cancers, colon cancer is frequently found to have mutations in mitotic checkpoint genes.

However, we leave the topic of mitotic checkpoints at this somewhat intuitive level and do not address the molecular mechanisms. Such an effort would require a lengthy background discussion of the structure, functions, and control of the microtubule-based mitotic spindle, more suitable for a course in cell biology than animal physiology. Instead, we now discuss the controls on cell growth other than proliferation and briefly summarize what is known about programmed cell death and the control of cell life span.



• **Fig. 2.9** P53 and the response to DNA damage. Normally, P53 is maintained at low levels in the cell by continuous synthesis and breakdown. DNA damage inhibits breakdown, allowing P53 to build up to functional levels. P53 is itself a transcription factor, and its targets include *p21,* whose protein is a potent inhibitor of all cyclin-CDK pairs. Thus upregulation of P53 brings the cell cycle to a halt, typically by inhibiting phosphorylation of pRb, as shown here. Subsequently, if the DNA is repaired, P53 returns to low concentration. If the DNA remains damaged, P53 leads to an apoptotic response by mediating expression of proapoptotic proteins, as described in the text. *CDK,* Cyclin-dependent kinase.

Mechanisms Regulating Cell Suicide and Cell Life Span

Apoptosis Is the Process of Cell Suicide

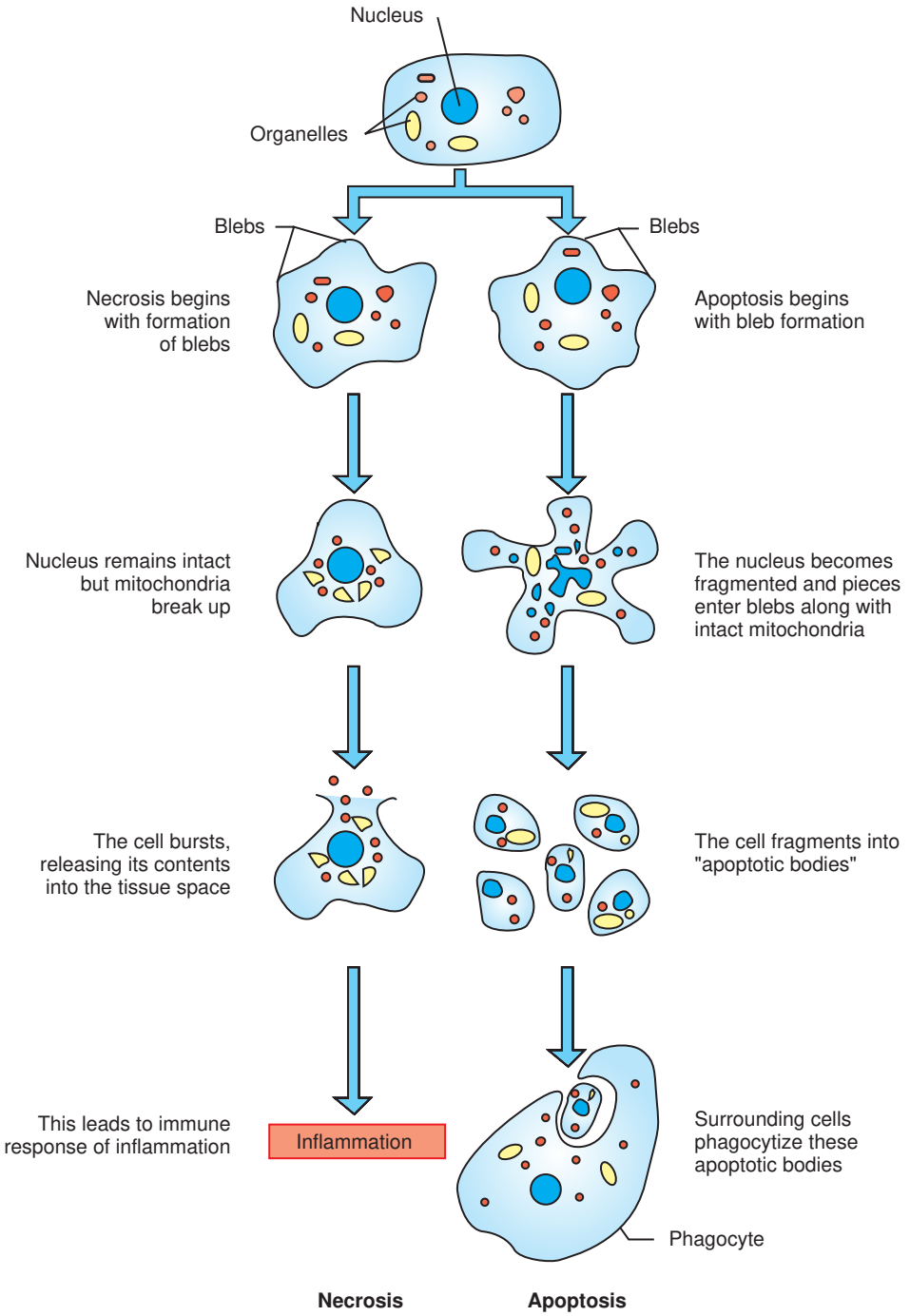
The process of cell death by external damage, involving cellular swelling, bursting, and engagement of the inflammatory response, has been well described for more than 100 years. This form of cell death is called *necrosis* and is familiar from experiences as common as a cut or abrasion. A rather different process of cell death was described in the 1970s in which cells shrink, the DNA fragments in a systematic way, the plasma membrane bubbles and churns, and the cell breaks up into small pieces that are rapidly engulfed by neighboring cells (Fig. 2.10). This neater and cleaner form of cell death was named *apoptosis* (a-pah-toe-sis; Greek, "falling off"). Apoptosis was largely ignored for the next 20 years, until studies of nematode development discovered genes whose only role was to control apoptosis. Further studies revealed the highly conserved mechanisms of apoptosis and its importance in normal development, immune function, and disease. Resistance to apoptosis is clearly a major contributor to cancer. (Conversely, too much apoptosis plays an important role in neurodegenerative diseases and stroke.) Particularly relevant to clinical practice, most cancer drugs and radiation therapy kill the target cells (and unfortunately many bystander cells) by stimulating apoptosis. Also relevant to clinical practice, recently discovered cancer therapies augment the immune system's ability to trigger apoptosis in cancer cells, as discussed shortly.

There are two broad pathways that lead to apoptosis. The *intrinsic pathway* of apoptosis responds to internal damage or stress from within the cell. The *extrinsic pathway* begins with a signal molecule binding to a "death receptor" on the cell surface (Fig. 2.11). However, both pathways converge on the same "executioners." *Caspases* are a family of proteolytic enzymes that have a cysteine amino acid at their active site (the "c" in caspase) and that cleave the substrate proteins at an aspartate amino acid (the "asp" in caspase). Similar to many other proteases, including digestive enzymes and blood-clotting factors, caspases are themselves activated by proteolytic cleavage. That is, as initially translated, the protease contains an inhibitory peptide that must be cleaved away to allow active proteolysis by the enzyme. In the case of the caspases, the activating protease is itself another caspase. Thus caspases are divided into *activating caspases,* which respond directly to one or another element in the intrinsic or extrinsic pathway, and downstream *executioner caspases,* which lead to specific cleavage of cellular structures. Among other tasks, executioner caspases cleave cytoskeletal proteins, leading to cell shrinkage, and activate the DNA-degrading enzymes involved in the systematic fragmentation of DNA.

The basic extrinsic pathway of apoptosis, also called the *death receptor pathway,* is unusually short and straightforward considering the extreme and irreversible outcome. An extracellular signal, which can be either soluble or attached to the surface of another cell, binds to and activates a death receptor on the cell destined to commit suicide. The cytoplasmic domain of the death receptor recruits one or two adapter proteins that directly activate an activating caspase, which in turn activates one or more executioner caspases (see Fig. 2.11). The activating caspase of the extrinsic pathway can also engage in "cross-talk" with the intrinsic pathway, described shortly, to increase the extent of caspase activation. The extrinsic pathway plays a crucial role in regulating the immune system, where the vast majority of immune cells initially generated are eliminated. The extrinsic pathway in cancer is also important in that many cancer cells are able to inhibit the immune system's capacity to kill cancer cells via the extrinsic pathway. Restoration of immune attack is among the most promising new clinical weapons against certain cancers, as discussed shortly.

Resistance to Apoptosis via the Intrinsic Pathway Is a Hallmark of Cancer

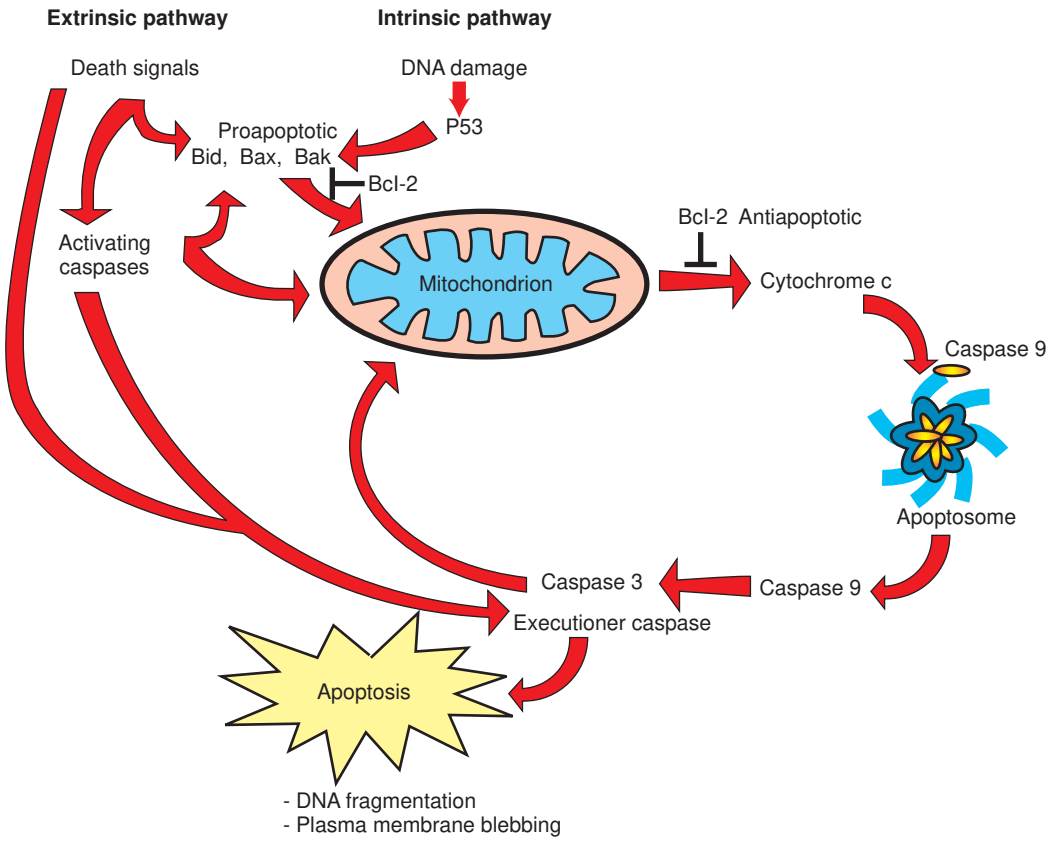
Internal cellular damage or stress, including DNA damage, absence of cell anchorage, too little or too much oxygen metabolism, oncogene activation, and radiation damage, can stimulate the intrinsic pathway of apoptosis in normal cells. Most, and perhaps all, cancer cells are more resistant than normal cells to apoptosis



• **Fig. 2.10** Necrosis versus apoptosis. Necrosis is cell death as a result of external damage to the cell that leads to bursting of the cell and release of cell contents, leading to inflammation. Apoptosis is cell death as a result of intrinsic mechanisms in which the cell is broken down into cell fragments that then undergo phagocytosis by neighboring cells. This produces no inflammatory reaction and is so "tidy" that apoptosis is difficult to observe.

through this pathway. Resistance to apoptosis not only increases the probability that the cell will be able to accumulate further genetic damage, but also reduces the likelihood that cancer cells can be eliminated. This is because the antitumor activity of the immune system, as well as most chemotherapy and radiation treatments, depends on apoptosis. Thus resistance to apoptosis often means resistance to treatment.

The intrinsic pathway is considerably more complex than the extrinsic pathway, and this discussion focuses on three major elements of the pathway involved in activating caspases: P53, the mitochondrion, and the *Bcl family* of proteins (see Fig. 2.11). This family of proteins was originally discovered in a cancer ("Bcl" is from B-cell lymphoma, a type of leukemia in which the first such protein was discovered) and includes both proapoptotic and



• **Fig. 2.11** Extrinsic and intrinsic pathways for apoptosis. See text for details.

antiapoptotic members. The balance between pro- and antiapoptotic members determines whether the cell lives or dies. The resistance of cancer cells to apoptosis arises not only from mutations, such as those already described for *p53,* but also from underexpression of proapoptotic mediators and overexpression of antiapoptotic proteins.

We begin with the *mitochondrion,* familiar as the "powerhouse" of the cell responsible for generating ATP, but also the central control point for the intrinsic pathway of apoptosis. Recall that the mitochondrion has both an inner membrane, responsible for electron transport, and an outer membrane, responsible for compartmentation of this organelle. Proapoptotic signals cause the outer membrane of the mitochondria to become leaky, releasing several proapoptotic proteins not normally found in the cytoplasm. Among the most important is cytochrome c, an electron transport protein that is only loosely attached to the inner membrane. In the cytoplasm, cytochrome c stimulates the assembly of a multiprotein complex (the apoptosome) that directly stimulates the activity of an activating caspase (caspase-9), ultimately leading to the activation of executioner caspases. What then determines the extent of permeability (leakiness) of the mitochondrial outer membrane?

The Bcl family members are major regulators of mitochondrial outer membrane permeability. The proapoptotic members of this family, such as Bax, lead to permeabilization by assembling to form channels in the outer membrane through which cytochrome c can pass. Proapoptotic members of the family can also cause the channel through which ATP normally passes into the cytoplasm to open wider than usual. The antiapoptotic members of the family, such as Bcl-2, seem to function by binding to proapoptotic members, inhibiting their activity. In a healthy cell, antiapoptotic Bcl members are at high enough concentration to neutralize proapoptotic activity. Damage increases the amount of proapoptotic Bcl molecules and leads to membrane permeabilization. Thus the balance between pro- and antiapoptotic members of the family controls the permeability state of mitochondria and the survival of the cell.

With about 20 different members of the Bcl family, the balance between pro- and antiapoptotic Bcl molecules has multiple controls, but P53 activity is certainly a major player. Recall that, when activated (e.g., by DNA damage), P53 acts as a transcription factor, and at least three different proapoptotic Bcl genes are transcriptionally activated by P53. These include Bax, and also the particularly powerful proapoptotic protein, PUMA. Downstream, P53 also activates the transcription of the activating caspase-9 gene, and the gene of a major cytoplasmic component of the apoptosome. In addition to acting as an activating transcription factor, P53 serves as an inhibitory transcription factor for some genes, including that of the antiapoptotic Bcl-2 protein. Finally and independent of transcription, activated P53 can directly activate Bax, which is required for its ability to assemble into channel structures. With these multiple effects on apoptotic genes and proteins, P53 is regarded as a central apoptotic control point, in addition to its role in cell cycle regulation.

As noted earlier in the discussion of P53, the importance of apoptosis to tumorigenesis is that, with normal apoptosis, almost all damaged cells are eliminated. Without apoptosis, damaged cells live to accumulate additional damage, which illustrates why multiple mutations and dysfunctions are required for tumors to reach a clinically significant stage. The resistance of cancer cells to apoptosis arises from many types of mutations and disruptions of normal gene expression. In some cases, mutation of the *p53* gene eliminates its DNA binding and thus transcriptional activity. Related to P53 activity is a protein that regulates P53's normal proteolytic

breakdown (see previous discussion). Overexpression of this protein (MDM2) in various cancers of soft tissues inhibits the accumulation of P53 to active levels and therefore inhibits both cell cycle arrest and apoptosis. The antiapoptotic Bcl-2 protein is overexpressed in a variety of human cancers, including 60% of human follicular lymphomas, but also some lung cancers, melanoma, and prostate cancer. Another common apoptotic lesion seen in cancer cells is overexpression of proteins that bind to and directly inactivate caspases, as well as mutation or loss of expression of the caspases themselves.

Many Types of Cancer Cells Suppress Immune Attack and so Avoid Apoptosis via the Extrinsic Pathway

In contrast to the intrinsic pathway of apoptosis, cancer cells themselves are rarely resistant to the extrinsic pathway of apoptosis. Rather, many types of cancer cells avoid apoptosis via the extrinsic pathway by "jamming" the initiating death signal that would otherwise come from the immune system. The immune system is the "defense establishment" of the body that functions, in large part, by recognizing "self" from foreign, "nonself" cells. These foreign cells are then destroyed, often by apoptosis. The immune system is exceedingly complex; a primer on the immune system can be found in Chapters 54 and 55 of this text. Despite the topic being somewhat out of order, recent highly encouraging advances in cancer therapy mediated through the immune system make it imperative to say just a few words about these new therapies. Indeed, the Nobel Prize in Medicine or Physiology for 2018 was awarded for work, including the discovery of the PD-1 receptor discussed below, that led to this new type of therapy. It is likely that you will hear more about this in your immunology course.

Suffice it to say that cancer cells, although derived from the body's own cells, are sufficiently damaged that the immune system should, and often does, attack them as foreign, nonself cells. This is the same basis on which the immune system attacks virally infected cells (see Fig. 55.3). And the available evidence suggests that spontaneous remission of cancer is the result of the patient's immune system successfully destroying the cancer. One form of immune attack involves immune cells signaling the cancer cell to commit apoptosis via the extrinsic pathway. As with a military defense establishment, it is crucial that only nonself cells (bacteria, fungi, infected and damaged cells) are attacked and not the normal "self" cells, i.e., the latter must avoid "friendly fire" in current military jargon. Consequently, the attacking immune cells have molecular receptors for being activated, or "armed," as well as for being inhibited, or "standing down."

Among the latter receptors is a protein called PD-1 ("programmed death 1"); as the name implies, this receptor was discovered as a "death receptor" whose activation led to apoptosis via the extrinsic pathway. As with the limited accuracy of names of growth factors mentioned earlier, subsequent study revealed that PD-1 only rarely causes apoptosis in some normal immune cells. Rather, this receptor primarily serves as a downregulator of immune cell activation. For example, melanoma cells express high levels of a ligand that activates PD-1 and thus downregulates immune cells that would otherwise attack the melanoma. The therapeutic advance has been the development of drugs (antibodies in this case) that bind to and inactivate the PD-1 receptor, i.e., inhibiting the inhibitor. This enables the immune cells to attack the melanoma and cause remission. This therapy was used on former US President Jimmy Carter to successfully treat his advanced-stage melanoma. The same therapy is used to treat small-cell lung carcinoma and is showing promise for a variety of other solid tissue tumors including breast and renal tumors. More information about this new strategy of immune-directed cancer therapy is provided in a paper by Messerschmidt et al. in the Bibliography for this chapter.

Cellular Life Span Is Determined by DNA Sequences at the Ends of Chromosomes

The final major dysfunction of growth control found within cancer cells is the most recently discovered but also seems to be the most common single molecular lesion in cancers: the expression of a reverse transcriptase called *telomerase.* (A reverse transcriptase is any enzyme that synthesizes DNA from an RNA template.) Telomerase is responsible for elongating *telomeres,* the specialized, noncoding regions of DNA found at the end of chromosomes. However, telomerase is normally expressed only in embryonic cells and in adult *stem cells.* (Stem cells are specialized normal cells that do have limitless replicative potential, such as gamete-generating cells and the blood-forming cells of the bone marrow, as discussed later.) The vast majority of normal somatic cells do not express telomerase, but it is expressed in 85% to 90% of all cancers and is the major determinant of the "immortality" of cancer cells.

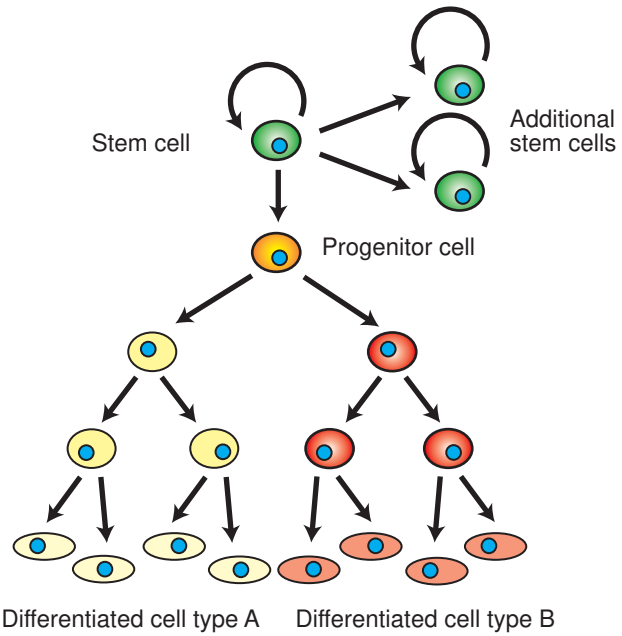
Telomeres are segments of highly repetitive DNA, representing hundreds of repeats of the simple nucleotide sequence TTAGGG (in vertebrates), found at the ends of chromosomes. Telomeres serve as caps at chromosomal ends, protecting them against end-to-end joining of chromosomes. Telomeres also prevent the ends of chromosomes from being recognized as sites of DNA damage (double-strand DNA breaks). Crucially, telomeres also protect against the loss of coding DNA from each chromosomal end with every round of DNA replication; this is needed because normal DNA polymerases have a serious limitation; they cannot fully replicate the end of a double-strand DNA molecule. As a result, the ends of chromosomes become shorter with each round of DNA replication. (Bacteria solve this problem by having circular DNA chromosomes.)

Telomeres are expendable DNA, at the ends of chromosomes, whose progressive shortening does not compromise the coding function of the genome. Although no coding sequence is lost, the shortening of telomeres nevertheless plays an important role in the cell. The shortening of telomeres serves as a kind of clock, measuring the number of times a cell has replicated its DNA, and the length of the telomere reflects the age of the cell. Through poorly understood mechanisms, cells can detect the length of their telomeres, and when they reach a critically short length, the cell ceases to divide and is said to undergo *senescence* (Latin; "growing old"). As noted earlier, normal cells have a finite life span, such that a cell taken from a middle-aged human will divide 20 to 40 times in culture before senescence. When placed into culture, the number of subsequent cell divisions before senescence reflects the original length of the telomeres. Further, various degenerative diseases, including cirrhosis of the liver, have been shown to accelerate telomere shortening. In principle, senescence is a powerful block to cancer because the original damaged cell (see Fig. 2.1) would be unable to divide for a sufficient number of generations to accumulate the necessary multiple mutations required to produce a tumor. Telomerase expression (and other, less common means of elongating telomeres) effectively eliminates this block to cancer development by causing the cells to become immortal.

Telomerase has both protein and RNA components. The protein provides the catalytic reverse transcriptase, allowing the enzyme

to elongate the telomere sequence based on the RNA template it carries. That is, the RNA component of telomerase is complementary to the telomere DNA sequence and is used as the template for telomere DNA elongation. Telomerase is not expressed in normal adult somatic cells except for stem cells, as mentioned earlier. However, immortal tissue culture cells do express telomerase, as do cancer cells. Experimental expression of telomerase in human cells dramatically increases the replicative life span of the cells. Thus the observed expression of telomerase in the vast majority of human cancers permits these cells to divide indefinitely, providing yet another selective advantage for these cells to accumulate additional damage over time.

In the last sections of this chapter, we turn our attention to the cancer cell in the context of a *tumor,* which is a population of cancer cells interacting with one another and with surrounding normal tissue. We end our discussion of the intrinsic growth controls of normal and cancer cells with an experimental result that seems to confirm the importance of the types of damage discussed thus far. This experiment showed that four genetic changes were sufficient to transform normal human kidney cells into cancer cells able to form tumors when transplanted into a mouse host (with no immune system). The four genetic changes were to "engineer" into the cells an activating mutation for the *ras* oncogene, inactivation of both the retinoblastoma and P53 proteins, and expression of the catalytic subunit of telomerase. Thus damage to the genes or expression of these molecules, emphasized here, reflects the minimum requirements for a normal cell to grow as a cancer.



• **Fig. 2.12** Stem cells. Stem cells are self-renewing cells of high, sometimes unlimited, replicative potential. (The word "potential" is crucial; typically stem cells replicate only rarely. It is the progenitor cells that are the rapidly dividing population.) The proliferation forms both additional stem cells and progenitor cells. These progenitor cells divide, often rapidly as noted, and eventually differentiate to become one or more types of differentiated somatic cells specialized for certain tasks (e.g., erythrocytes and monocytes of blood).

Tumor Origin and the Spread of Cancer

Cancer Cells May Be Related to Stem Cells

As noted in the previous section, some normal adult cells do have unlimited replicative potential. These are *stem cells,* a cell type that has been in the news for decades and remains so (see K. Servick, *The Stem Cell Skeptic* in the chapter Bibliography). A *stem cell* is a self-renewing cell of high proliferative potential that can also give rise to differentiated cells. Typically, stem cell division produces one cell that remains a stem cell whereas the other daughter cell differentiates into a specialized cell with the usual limited life span (Fig. 2.12). The cell that continues being a stem cell does not lose any developmental capacity and can divide indefinitely, continuing to produce additional stem cells and additional differentiated cells.

Much of the recent attention in the news centers on *embryonic* stem cells. These are embryonic cells that can either continue to form stem cells or differentiate, in principle, to any and every cell type within the body. Even in the adult, however, the maintenance of many normal tissues is critically dependent on stem cells. *Adult* stem cells, however, can only differentiate into a limited array of different cell types, not every cell type in the body. Best understood is that all the various cells of the blood arise from the division of hematopoietic stem cells in the bone marrow; one daughter cell remains a stem cell in the bone marrow, whereas the other differentiates to become one of the several types of blood cells (but the blood stem cell can only form blood cells, not nonblood cells). The cells lining the gut and those giving rise to hair also arise from a stable population of adult stem cells, some of whose descendants differentiate into specialized gut or hair-producing cells. For this reason, chemotherapy that is intended to cause apoptosis in cancer cells typically also affects these same populations of normal stem cells; common side effects of chemotherapy include anemia, hair loss, and digestive dysfunction.

Cancer cells resemble stem cells in their immortality, but the relationship of cancer cells to stem cells may go further. Based on the presentation thus far, you may have the mental image of a tumor composed of a uniform population of badly damaged cells, any of which would be capable of forming a new tumor if transplanted. In fact, real tumors are not a homogeneous population of cells but rather are composed of a variety of cells that differ significantly in their phenotype, despite all being clonal descendants of a single somatic cell, as shown in Fig. 2.1. (Keep in mind that all somatic cells of the body are clonal descendants of the fertilized egg, so phenotypic differences arising within clonal lines is not surprising by itself.) Further, experiments with a variety of cancers show that only 1% or less of tumor cells are capable of forming another tumor, even in the same patient (or mouse). Thus tumors may contain a small subpopulation of *cancer stem cells* that are responsible for producing the heterogeneous cells in the tumor and are uniquely able to continue cancer growth. This would also give tumors the capacity to adapt to their surroundings; because stem cells can differentiate in various ways, differentiated cells that allowed continued growth and survival would be selected.

This hypothesis has been persuasively supported only in leukemias, but it may apply to other cancers as well. For leukemias, the cancer stem cells express some marker proteins characteristic of normal hematopoietic stem cells. Further, only those leukemia stem cells expressing certain normal markers are capable of forming new cancers when transplanted. Finally, a possible relationship between cancer and stem cells is that perhaps the genetic changes summarized in this chapter must occur in a normal adult stem cell to produce cancer cells. Here again, the best evidence in favor of such a mechanism comes from leukemias. But the blood is unusual in ways other than just being a fluid rather than a solid