

# 3

## PROTEIN STRUCTURE AND FUNCTION



Electron density map of the F<sub>1</sub>-ATPase associated with a ring of 10 c-subunits from the F<sub>0</sub> domain of ATP synthase, a molecular machine that carries out the synthesis of ATP in eubacteria, chloroplasts, and mitochondria. [Courtesy of Andrew Leslie, MRC Laboratory of Molecular Biology, Cambridge, UK.]

Proteins, the working molecules of a cell, carry out the program of activities encoded by genes. This program requires the coordinated effort of many different types of proteins, which first evolved as rudimentary molecules that facilitated a limited number of chemical reactions. Gradually, many of these primitive proteins evolved into a wide array of **enzymes** capable of catalyzing an incredible range of intracellular and extracellular chemical reactions, with a speed and specificity that is nearly impossible to attain in a test tube. With the passage of time, other proteins acquired specialized abilities and can be grouped into several broad functional classes: *structural proteins*, which provide structural rigidity to the cell; *transport proteins*, which control the flow of materials across cellular membranes; *regulatory proteins*, which act as sensors and switches to control protein activity and gene function; *signaling proteins*, including cell-surface receptors and other proteins that transmit external signals to the cell interior; and *motor proteins*, which cause motion.

A key to understanding the functional design of proteins is the realization that many have “moving” parts and are capable of transmitting various forces and energy in an orderly fashion. However, several critical and complex cell processes—synthesis of nucleic acids and proteins, signal transduction, and photosynthesis—are carried out by huge

macromolecular assemblies sometimes referred to as *molecular machines*.

A fundamental goal of molecular cell biologists is to understand how cells carry out various processes essential for life. A major contribution toward achieving this goal is the identification of all of an organism’s proteins—that is, a list of the parts that compose the cellular machinery. The compilation of such lists has become feasible in recent years with the sequencing of entire **genomes**—complete sets of genes—of more and more organisms. From a computer analysis of

### OUTLINE

- 3.1 Hierarchical Structure of Proteins
- 3.2 Folding, Modification, and Degradation of Proteins
- 3.3 Enzymes and the Chemical Work of Cells
- 3.4 Molecular Motors and the Mechanical Work of Cells
- 3.5 Common Mechanisms for Regulating Protein Function
- 3.6 Purifying, Detecting, and Characterizing Proteins

genome sequences, researchers can deduce the number and primary structure of the encoded proteins (Chapter 9). The term **proteome** was coined to refer to the entire protein complement of an organism. For example, the proteome of the yeast *Saccharomyces cerevisiae* consists of about 6000 different proteins; the human proteome is only about five times as large, comprising about 32,000 different proteins. By comparing protein sequences and structures, scientists can classify many proteins in an organism's proteome and deduce their functions by homology with proteins of known function. Although the three-dimensional structures of relatively few proteins are known, the function of a protein whose structure has not been determined can often be inferred from its interactions with other proteins, from the effects result-

ing from genetically mutating it, from the biochemistry of the complex to which it belongs, or from all three.

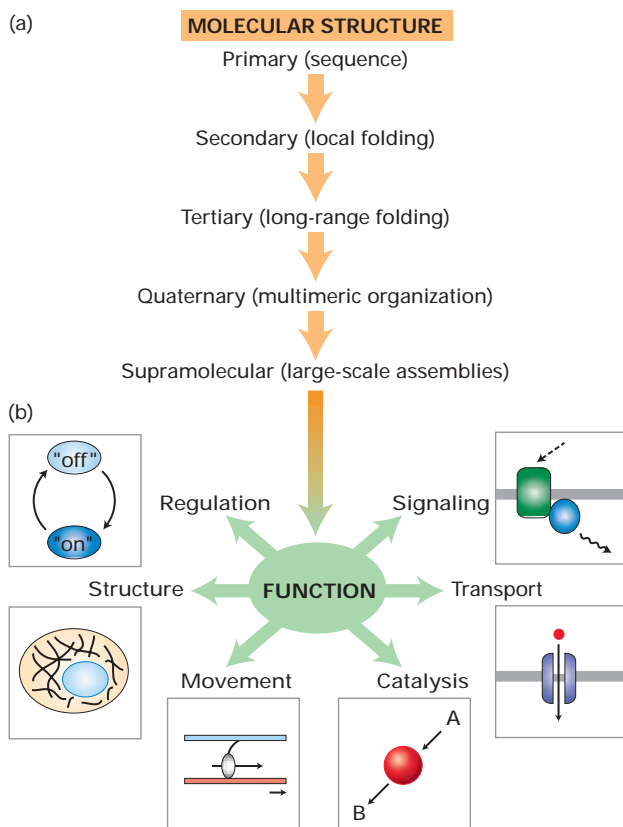
In this chapter, we begin our study of how the structure of a protein gives rise to its function, a theme that recurs throughout this book (Figure 3-1). The first section examines how chains of amino acid building blocks are arranged and the various higher-order folded forms that the chains assume. The next section deals with special proteins that aid in the folding of proteins, modifications that take place after the protein chain has been synthesized, and mechanisms that degrade proteins. The third section focuses on proteins as catalysts and reviews the basic properties exhibited by all enzymes. We then introduce molecular motors, which convert chemical energy into motion. The structure and function of these functional classes of proteins and others are detailed in numerous later chapters. Various mechanisms that cells use to control the activity of proteins are covered next. The chapter concludes with a section on commonly used techniques in the biologist's tool kit for isolating proteins and characterizing their properties.

### 3.1 Hierarchical Structure of Proteins

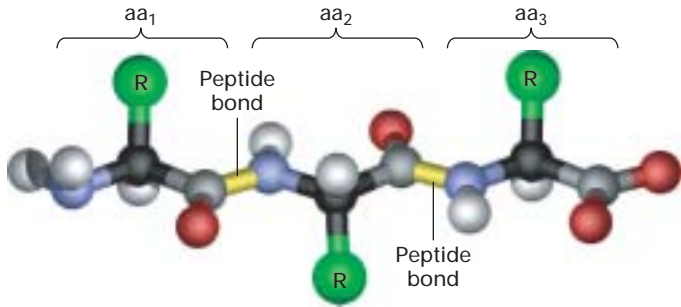
Although constructed by the polymerization of only 20 different **amino acids** into linear chains, proteins carry out an incredible array of diverse tasks. A protein chain folds into a unique shape that is stabilized by noncovalent interactions between regions in the linear sequence of amino acids. This *spatial* organization of a protein—its shape in three dimensions—is a key to understanding its function. Only when a protein is in its correct three-dimensional structure, or **conformation**, is it able to function efficiently. A key concept in understanding how proteins work is that *function is derived from three-dimensional structure, and three-dimensional structure is specified by amino acid sequence*. Here, we consider the structure of proteins at four levels of organization, starting with their monomeric building blocks, the amino acids.

#### The Primary Structure of a Protein Is Its Linear Arrangement of Amino Acids

We reviewed the properties of the amino acids used in synthesizing proteins and their linkage by **peptide bonds** into linear chains in Chapter 2. The repeated amide N,  $\alpha$  carbon ( $C_\alpha$ ), and carbonyl C atoms of each amino acid residue form the **backbone** of a protein molecule from which the various side-chain groups project (Figure 3-2). As a consequence of the peptide linkage, the backbone exhibits directionality because all the amino groups are located on the same side of the  $C_\alpha$  atoms. Thus one end of a protein has a free (unlinked) amino group (the *N-terminus*) and the other end has a free carboxyl group (the *C-terminus*). The sequence of a protein chain is conventionally written with its N-terminal amino acid on the left and its C-terminal amino acid on the right.



▲ **FIGURE 3-1 Overview of protein structure and function.** (a) The linear sequence of amino acids (primary structure) folds into helices or sheets (secondary structure) which pack into a globular or fibrous domain (tertiary structure). Some individual proteins self-associate into complexes (quaternary structure) that can consist of tens to hundreds of subunits (supramolecular assemblies). (b) Proteins display functions that include catalysis of chemical reactions (enzymes), flow of small molecules and ions (transport), sensing and reaction to the environment (signaling), control of protein activity (regulation), organization of the genome, lipid bilayer membrane, and cytoplasm (structure), and generation of force for movement (motor proteins). These functions and others arise from specific binding interactions and conformational changes in the structure of a properly folded protein.



▲ **FIGURE 3-2 Structure of a tripeptide.** Peptide bonds (yellow) link the amide nitrogen atom (blue) of one amino acid (aa) with the carbonyl carbon atom (gray) of an adjacent one in the linear polymers known as peptides or polypeptides, depending on their length. Proteins are polypeptides that have folded into a defined three-dimensional structure (conformation). The side chains, or R groups (green), extending from the  $\alpha$  carbon atoms (black) of the amino acids composing a protein largely determine its properties. At physiological pH values, the terminal amino and carboxyl groups are ionized.

The **primary structure** of a protein is simply the linear arrangement, or *sequence*, of the amino acid residues that compose it. Many terms are used to denote the chains formed by the polymerization of amino acids. A short chain of amino acids linked by peptide bonds and having a defined sequence is called a **peptide**; longer chains are referred to as **polypeptides**. Peptides generally contain fewer than 20–30 amino acid residues, whereas polypeptides contain as many as 4000 residues. We generally reserve the term **protein** for a polypeptide (or for a complex of polypeptides) that has a well-defined three-dimensional structure. It is implied that proteins and peptides are the natural products of a cell.

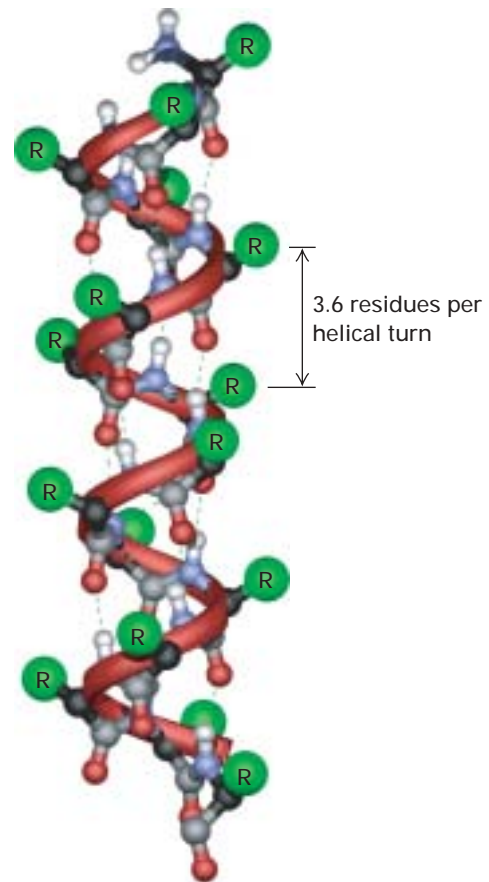
The size of a protein or a polypeptide is reported as its mass in **daltons** (a dalton is 1 atomic mass unit) or as its molecular weight (MW), which is a dimensionless number. For example, a 10,000-MW protein has a mass of 10,000 daltons (Da), or 10 kilodaltons (kDa). In the last section of this chapter, we will consider different methods for measuring the sizes and other physical characteristics of proteins. The known and predicted proteins encoded by the yeast genome have an average molecular weight of 52,728 and contain, on average, 466 amino acid residues. The average molecular weight of amino acids in proteins is 113, taking into account their average relative abundance. This value can be used to estimate the number of residues in a protein from its molecular weight or, conversely, its molecular weight from the number of residues.

### Secondary Structures Are the Core Elements of Protein Architecture

The second level in the hierarchy of protein structure consists of the various spatial arrangements resulting from the folding of localized parts of a polypeptide chain; these arrangements are referred to as **secondary structures**. A single

polypeptide may exhibit multiple types of secondary structure depending on its sequence. In the absence of stabilizing noncovalent interactions, a polypeptide assumes a *random-coil* structure. However, when stabilizing hydrogen bonds form between certain residues, parts of the backbone fold into one or more well-defined periodic structures: the **alpha** ( $\alpha$ ) **helix**, the **beta** ( $\beta$ ) **sheet**, or a short U-shaped *turn*. In an average protein, 60 percent of the polypeptide chain exist as  $\alpha$  helices and  $\beta$  sheets; the remainder of the molecule is in random coils and turns. Thus,  $\alpha$  helices and  $\beta$  sheets are the major internal supportive elements in proteins. In this section, we explore forces that favor the formation of secondary structures. In later sections, we examine how these structures can pack into larger arrays.

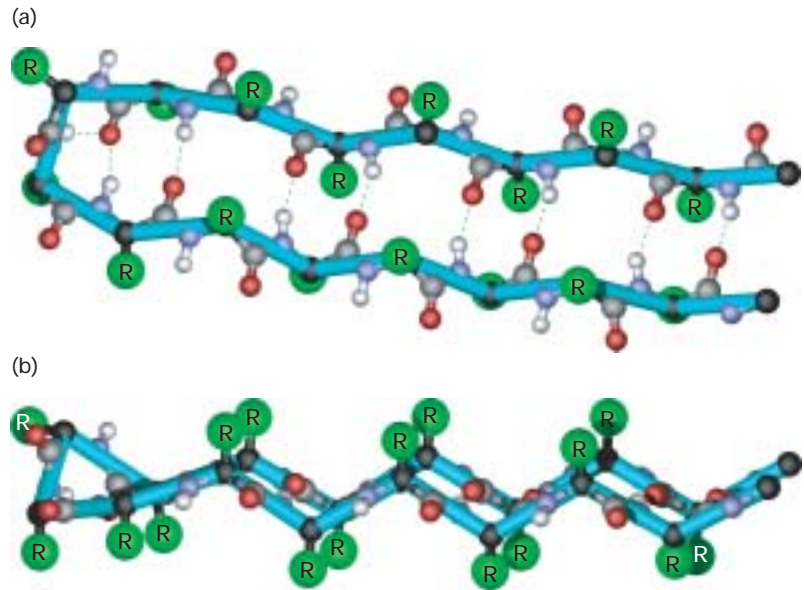
**The  $\alpha$  Helix** In a polypeptide segment folded into an  $\alpha$  helix, the carbonyl oxygen atom of each peptide bond is hydrogen-bonded to the amide hydrogen atom of the amino acid four residues toward the C-terminus. This periodic arrangement of bonds confers a directionality on the helix because all the hydrogen-bond donors have the same orientation (Figure 3-3).



▲ **FIGURE 3-3 The  $\alpha$  helix, a common secondary structure in proteins.** The polypeptide backbone (red) is folded into a spiral that is held in place by hydrogen bonds between backbone oxygen and hydrogen atoms. The outer surface of the helix is covered by the side-chain R groups (green).



► **FIGURE 3-4 The  $\beta$  sheet, another common secondary structure in proteins.** (a) Top view of a simple two-stranded  $\beta$  sheet with antiparallel  $\beta$  strands. The stabilizing hydrogen bonds between the  $\beta$  strands are indicated by green dashed lines. The short turn between the  $\beta$  strands also is stabilized by a hydrogen bond. (b) Side view of a  $\beta$  sheet. The projection of the R groups (green) above and below the plane of the sheet is obvious in this view. The fixed angle of the peptide bond produces a pleated contour.



The stable arrangement of amino acids in the  $\alpha$  helix holds the backbone in a rodlike cylinder from which the side chains point outward. The hydrophobic or hydrophilic quality of the helix is determined entirely by the side chains because the polar groups of the peptide backbone are already engaged in hydrogen bonding in the helix.

**The  $\beta$  Sheet** Another type of secondary structure, the  $\beta$  sheet, consists of laterally packed  $\beta$  strands. Each  $\beta$  strand is a short (5- to 8-residue), nearly fully extended polypeptide segment. Hydrogen bonding between backbone atoms in adjacent  $\beta$  strands, within either the same polypeptide chain or between different polypeptide chains, forms a  $\beta$  sheet (Figure 3-4a). The planarity of the peptide bond forces a  $\beta$  sheet to be pleated; hence this structure is also called a  *$\beta$  pleated sheet*, or simply a *pleated sheet*. Like  $\alpha$  helices,  $\beta$  strands have a directionality defined by the orientation of the peptide bond. Therefore, in a pleated sheet, adjacent  $\beta$  strands can be oriented in the same (parallel) or opposite (antiparallel) directions with respect to each other. In both arrangements, the side chains project from both faces of the sheet (Figure 3-4b). In some proteins,  $\beta$  sheets form the floor of a binding pocket; the hydrophobic core of other proteins contains multiple  $\beta$  sheets.

**Turns** Composed of three or four residues, turns are located on the surface of a protein, forming sharp bends that redirect the polypeptide backbone back toward the interior. These short, U-shaped secondary structures are stabilized by a hydrogen bond between their end residues (see Figure 3-4a). Glycine and proline are commonly present in turns. The lack of a large side chain in glycine and the presence of a built-in bend in proline allow the polypeptide backbone to fold into a tight U shape. Turns allow large proteins to fold into highly compact structures. A polypeptide backbone also may contain longer bends, or *loops*. In contrast with turns, which ex-

hibit just a few well-defined structures, loops can be formed in many different ways.

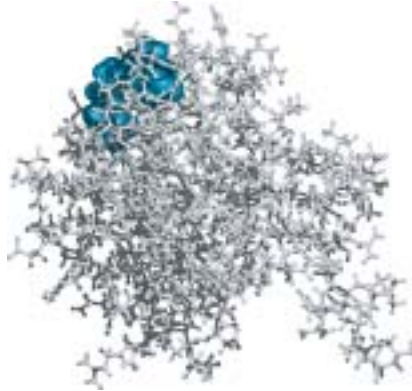
### Overall Folding of a Polypeptide Chain Yields Its Tertiary Structure

**Tertiary structure** refers to the overall conformation of a polypeptide chain—that is, the three-dimensional arrangement of all its amino acid residues. In contrast with secondary structures, which are stabilized by hydrogen bonds, tertiary structure is primarily stabilized by hydrophobic interactions between the nonpolar side chains, hydrogen bonds between polar side chains, and peptide bonds. These stabilizing forces hold elements of secondary structure— $\alpha$  helices,  $\beta$  strands, turns, and random coils—compactly together. Because the stabilizing interactions are weak, however, the tertiary structure of a protein is not rigidly fixed but undergoes continual and minute fluctuation. This variation in structure has important consequences in the function and regulation of proteins.

Different ways of depicting the conformation of proteins convey different types of information. The simplest way to represent three-dimensional structure is to trace the course of the backbone atoms with a solid line (Figure 3-5a); the most complex model shows every atom (Figure 3-5b). The former, a  $C_{\alpha}$  trace, shows the overall organization of the polypeptide chain without consideration of the amino acid side chains; the latter, a ball-and-stick model, details the interactions between side-chain atoms, which stabilize the protein's conformation, as well as the atoms of the backbone. Even though both views are useful, the elements of secondary structure are not easily discerned in them. Another type of representation uses common shorthand symbols for depicting secondary structure—for example, coiled ribbons or solid cylinders for  $\alpha$  helices, flat ribbons or arrows for  $\beta$  strands, and flexible

(a) C<sub>α</sub> backbone trace

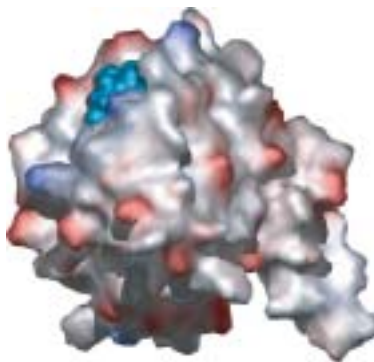
(b) Ball and stick



(c) Ribbons



(d) Solvent-accessible surface



◀ **FIGURE 3-5 Various graphic representations of the structure of Ras, a monomeric guanine nucleotide-binding protein.** The inactive, guanosine diphosphate (GDP)-bound form is shown in all four panels, with GDP always depicted in blue spacefill. (a) The C<sub>α</sub> backbone trace demonstrates how the polypeptide is packed into the smallest possible volume. (b) A ball-and-stick representation reveals the location of all atoms. (c) A ribbon representation emphasizes how β strands (blue) and α helices (red) are organized in the protein. Note the turns and loops connecting pairs of helices and strands. (d) A model of the water-accessible surface reveals the numerous lumps, bumps, and crevices on the protein surface. Regions of positive charge are shaded blue; regions of negative charge are shaded red.

thin strands for turns and loops (Figure 3-5c). This type of representation makes the secondary structures of a protein easy to see.

However, none of these three ways of representing protein structure convey much information about the protein surface, which is of interest because it is where other molecules bind to a protein. Computer analysis can identify the surface atoms that are in contact with the watery environment. On this water-accessible surface, regions having a common chemical character (hydrophobicity or hydrophilicity) and electrical character (basic or acidic) can be mapped. Such models reveal the topography of the protein surface and the distribution of charge, both important features of binding sites, as well as clefts in the surface where small molecules often bind (Figure 3-5d). This view represents a protein as it is “seen” by another molecule.

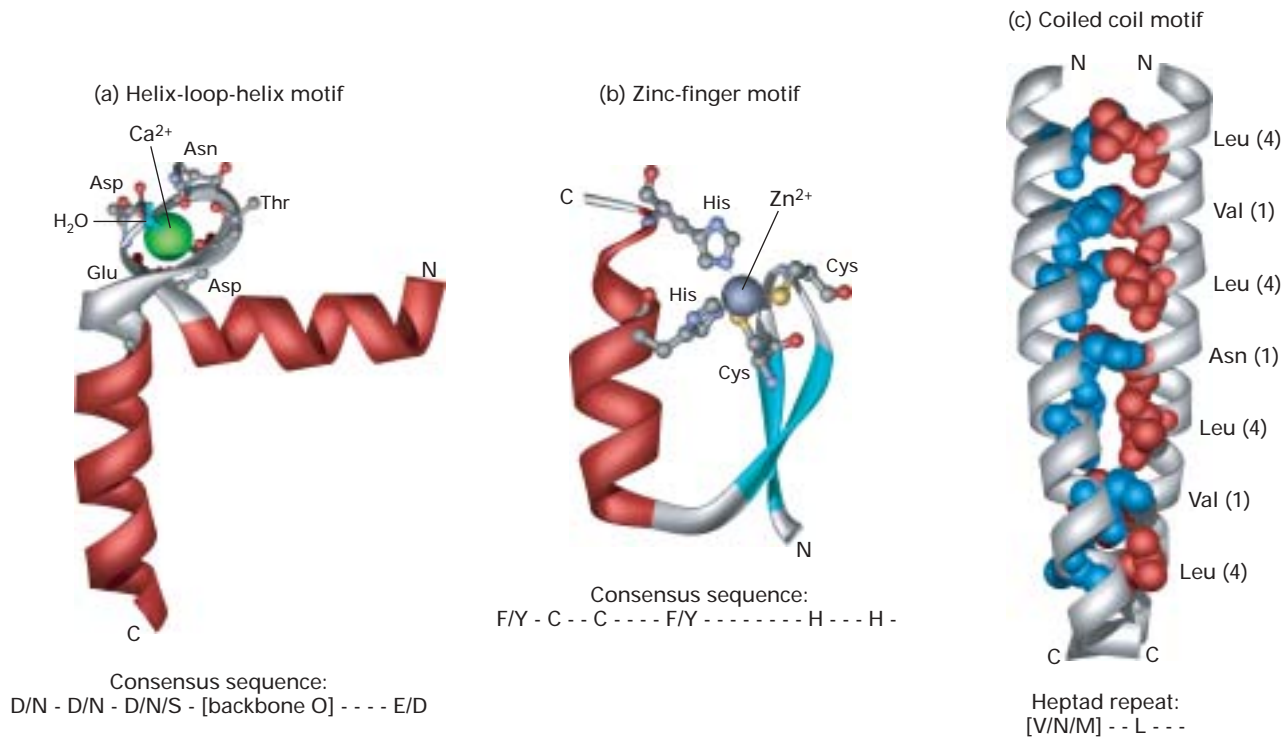
### Motifs Are Regular Combinations of Secondary Structures

Particular combinations of secondary structures, called **motifs** or folds, build up the tertiary structure of a protein. In some cases, motifs are signatures for a specific function. For example, the **helix-loop-helix** is a Ca<sup>2+</sup>-binding motif marked by the presence of certain hydrophilic residues at invariant positions in the loop (Figure 3-6a). Oxygen atoms in

the invariant residues bind a Ca<sup>2+</sup> ion through ionic bonds. This motif, also called the *EF hand*, has been found in more than 100 calcium-binding proteins. In another common motif, the **zinc finger**, three secondary structures—an α helix and two β strands with an antiparallel orientation—form a fingerlike bundle held together by a zinc ion (Figure 3-6b). This motif is most commonly found in proteins that bind RNA or DNA.

Many proteins, especially fibrous proteins, self-associate into oligomers by using a third motif, the **coiled coil**. In these proteins, each polypeptide chain contains α-helical segments in which the hydrophobic residues, although apparently randomly arranged, are in a regular pattern—a repeated heptad sequence. In the heptad, a hydrophobic residue—sometimes valine, alanine, or methionine—is at position 1 and a leucine residue is at position 4. Because hydrophilic side chains extend from one side of the helix and hydrophobic side chains extend from the opposite side, the overall helical structure is **amphipathic**. The amphipathic character of these α helices permits two, three, or four helices to wind around each other, forming a coiled coil; hence the name of this motif (Figure 3-6c).

We will encounter numerous additional motifs in later discussions of other proteins in this chapter and other chapters. The presence of the same motif in different proteins with similar functions clearly indicates that these useful



### ▲ FIGURE 3-6 Motifs of protein secondary structure.

(a) Two helices connected by a short loop in a specific conformation constitute a helix-loop-helix motif. This motif exists in many calcium-binding and DNA-binding regulatory proteins. In calcium-binding proteins such as calmodulin, oxygen atoms from five loop residues and one water molecule form ionic bonds with a  $\text{Ca}^{2+}$  ion. (b) The zinc-finger motif is present in many DNA-binding proteins that help regulate transcription. A  $\text{Zn}^{2+}$  ion is held between a pair of  $\beta$  strands (blue) and a single  $\alpha$  helix (red) by a pair of cysteine residues and a pair of histidine residues. The two invariant cysteine residues are usually at positions 3 and 6 and the two invariant histidine residues are

at positions 20 and 24 in this 25-residue motif. (c) The parallel two-stranded coiled-coil motif found in the transcription factor Gcn4 is characterized by two  $\alpha$  helices wound around one another. Helix packing is stabilized by interactions between hydrophobic side chains (red and blue) present at regular intervals along the surfaces of the intertwined helices. Each  $\alpha$  helix exhibits a characteristic heptad repeat sequence with a hydrophobic residue at positions 1 and 4. [See A. Lewit-Bentley and S. Rety, 2000, EF-hand calcium-binding proteins, *Curr. Opin. Struct. Biol.* **10**:637–643; S. A. Wolfe, L. Nekludova, and C. O. Pabo, 2000, DNA recognition by Cys2His2 zinc finger proteins, *Ann. Rev. Biophys. Biomol. Struct.* **29**:183–212.]

combinations of secondary structures have been conserved in evolution. To date, hundreds of motifs have been cataloged and proteins are now classified according to their motifs.

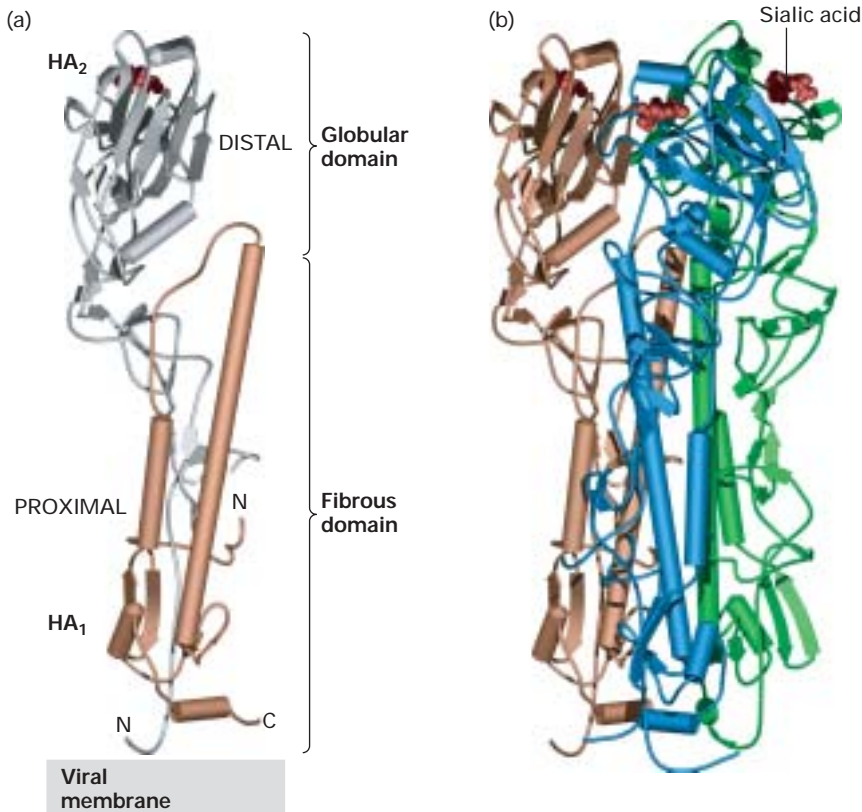
### Structural and Functional Domains Are Modules of Tertiary Structure

The tertiary structure of proteins larger than 15,000 MW is typically subdivided into distinct regions called **domains**. Structurally, a domain is a compactly folded region of polypeptide. For large proteins, domains can be recognized in structures determined by x-ray crystallography or in images captured by electron microscopy. Although these discrete regions are well distinguished or physically separated from one another, they are connected by intervening segments of the polypeptide chain. Each of the subunits in hemagglutinin, for example, contains a globular domain and a fibrous domain (Figure 3-7a).

A structural domain consists of 100–150 residues in various combinations of motifs. Often a domain is characterized by some interesting structural feature: an unusual abundance of a particular amino acid (e.g., a proline-rich domain, an acidic domain), sequences common to (conserved in) many proteins (e.g., SH3, or Src homology region 3), or a particular secondary-structure motif (e.g., zinc-finger motif in the kringle domain).

Domains are sometimes defined in functional terms on the basis of observations that an activity of a protein is localized to a small region along its length. For instance, a particular region or regions of a protein may be responsible for its catalytic activity (e.g., a kinase domain) or binding ability (e.g., a DNA-binding domain, a membrane-binding domain). Functional domains are often identified experimentally by whittling down a protein to its smallest active fragment with the aid of proteases, enzymes that cleave the polypeptide backbone. Alternatively, the DNA encoding a protein can be





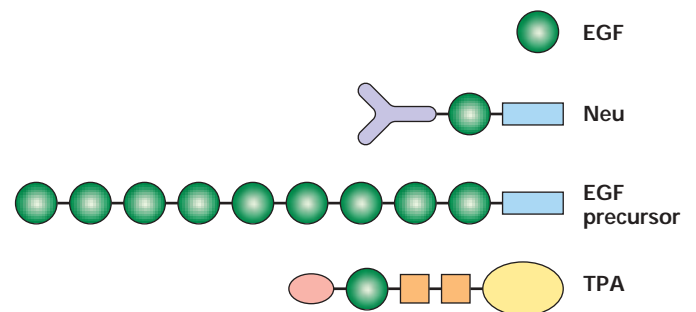
◀ **FIGURE 3-7 Tertiary and quaternary levels of structure in hemagglutinin (HA), a surface protein on influenza virus.** This long multimeric molecule has three identical subunits, each composed of two polypeptide chains, HA<sub>1</sub> and HA<sub>2</sub>. (a) Tertiary structure of each HA subunit constitutes the folding of its helices and strands into a compact structure that is 13.5 nm long and divided into two domains. The membrane-distal domain is folded into a globular conformation. The membrane-proximal domain has a fibrous, stemlike conformation owing to the alignment of two long  $\alpha$  helices (cylinders) of HA<sub>2</sub> with  $\beta$  strands in HA<sub>1</sub>. Short turns and longer loops, which usually lie at the surface of the molecule, connect the helices and strands in a given chain. (b) Quaternary structure of HA is stabilized by lateral interactions between the long helices (cylinders) in the fibrous domains of the three subunits (yellow, blue, and green), forming a triple-stranded coiled-coil stalk. Each of the distal globular domains in HA binds sialic acid (red) on the surface of target cells. Like many membrane proteins, HA contains several covalently linked carbohydrate chains (not shown).

subjected to mutagenesis so that segments of the protein's backbone are removed or changed. The activity of the truncated or altered protein product synthesized from the mutated gene is then monitored and serves as a source of insight about which part of a protein is critical to its function.

The organization of large proteins into multiple domains illustrates the principle that complex molecules are built from simpler components. Like motifs of secondary structure, domains of tertiary structure are incorporated as modules into different proteins. In Chapter 10 we consider the mechanism by which the gene segments that correspond to domains became shuffled in the course of evolution, resulting in their appearance in many proteins. The modular approach to protein architecture is particularly easy to recognize in large proteins, which tend to be mosaics of different domains and thus can perform different functions simultaneously.

The epidermal growth factor (EGF) domain is one example of a module that is present in several proteins (Figure 3-8). EGF is a small, soluble peptide hormone that binds to cells in the embryo and in skin and connective tissue in adults, causing them to divide. It is generated by proteolytic cleavage between repeated EGF domains in the EGF precursor protein, which is anchored in the cell membrane by a membrane-spanning domain. EGF modules are also present in other proteins and are liberated by proteolysis; these proteins include tissue plasminogen activator (TPA), a protease that is used to dissolve blood clots in heart attack victims;

Neu protein, which takes part in embryonic differentiation; and Notch protein, a receptor protein in the plasma membrane that functions in developmentally important signaling (Chapter 14). Besides the EGF domain, these proteins contain domains found in other proteins. For example, TPA possesses a trypsin domain, a common feature in enzymes that degrade proteins.



▲ **FIGURE 3-8 Schematic diagrams of various proteins illustrating their modular nature.** Epidermal growth factor (EGF) is generated by proteolytic cleavage of a precursor protein containing multiple EGF domains (green) and a membrane-spanning domain (blue). The EGF domain is also present in Neu protein and in tissue plasminogen activator (TPA). These proteins also contain other widely distributed domains indicated by shape and color. [Adapted from I. D. Campbell and P. Bork, 1993, *Curr. Opin. Struct. Biol.* 3:385.]

## Proteins Associate into Multimeric Structures and Macromolecular Assemblies

**Multimeric** proteins consist of two or more polypeptides or subunits. A fourth level of structural organization, **quaternary structure**, describes the number (stoichiometry) and relative positions of the subunits in multimeric proteins. Hemagglutinin, for example, is a trimer of three identical subunits held together by noncovalent bonds (Figure 3-7b). Other multimeric proteins can be composed of any number of identical or different subunits. The multimeric nature of many proteins is critical to mechanisms for regulating their function. In addition, enzymes in the same pathway may be associated as subunits of a large multimeric protein within the cell, thereby increasing the efficiency of pathway operation.

The highest level of protein structure is the association of proteins into macromolecular assemblies. Typically, such structures are very large, exceeding 1 mDa in mass, approaching 30–300 nm in size, and containing tens to hundreds of polypeptide chains, as well as nucleic acids in some

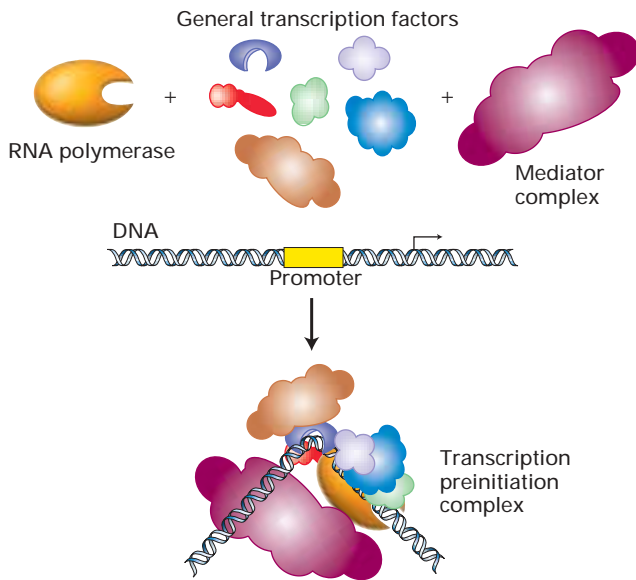
cases. Macromolecular assemblies with a structural function include the **capsid** that encases the viral genome and bundles of cytoskeletal filaments that support and give shape to the plasma membrane. Other macromolecular assemblies act as molecular machines, carrying out the most complex cellular processes by integrating individual functions into one coordinated process. For example, the transcriptional machine that initiates the synthesis of messenger RNA (mRNA) consists of RNA polymerase, itself a multimeric protein, and at least 50 additional components including general transcription factors, promoter-binding proteins, helicase, and other protein complexes (Figure 3-9). The transcription factors and promoter-binding proteins correctly position a polymerase molecule at a **promoter**, the DNA site that determines where transcription of a specific gene begins. After helicase unwinds the double-stranded DNA molecule, polymerase simultaneously translocates along the DNA template strand and synthesizes an mRNA strand. The operational details of this complex machine and of others listed in Table 3-1 are discussed elsewhere.

TABLE 3-1 Selected Molecular Machines

Machine*	Main Components	Cellular Location	Function
Replisome (4)	Helicase, primase, DNA polymerase	Nucleus	DNA replication
Transcription initiation complex (11)	Promoter-binding protein, helicase, general transcription factors (TFs), RNA polymerase, large multisubunit mediator complex	Nucleus	RNA synthesis
Spliceosome (12)	Pre-mRNA, small nuclear RNAs (snRNAs), protein factors	Nucleus	mRNA splicing
Nuclear pore complex (12)	Nucleoporins (50–100)	Nuclear membrane	Nuclear import and export
Ribosome (4)	Ribosomal proteins (>50) and four rRNA molecules (eukaryotes) organized into large and small subunits; associated mRNA and protein factors (IFs, EFs)	Cytoplasm/ER membrane	Protein synthesis
Chaperonin (3)	GroEL, GroES (bacteria)	Cytoplasm, mitochondria, endoplasmic reticulum	Protein folding
Proteasome (3)	Core proteins, regulatory (cap) proteins	Cytoplasm	Protein degradation
Photosystem (8)	Light-harvesting complex (multiple proteins and pigments), reaction center (multisubunit protein with associated pigments and electron carriers)	Thylakoid membrane in plant chloroplasts, plasma membrane of photosynthetic bacteria	Photosynthesis (initial stage)
MAP kinase cascades (14)	Scaffold protein, multiple different protein kinases	Cytoplasm	Signal transduction
Sarcomere (19)	Thick (myosin) filaments, thin (actin) filaments, Z lines, titin/nebulin	Cytoplasm of muscle cells	Contraction

\*Numbers in parentheses indicate chapters in which various machines are discussed.





▲ FIGURE 3-9 The mRNA transcription-initiation machinery.

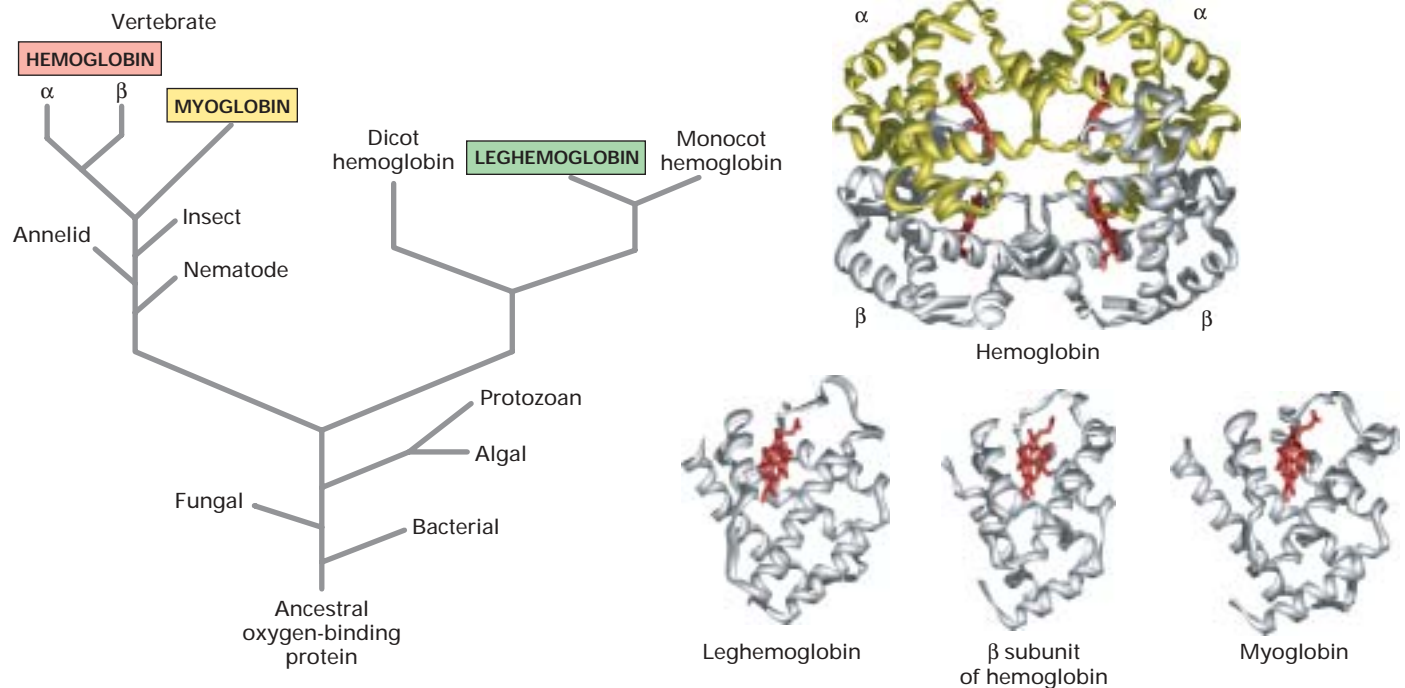
The core RNA polymerase, general transcription factors, a mediator complex containing about 20 subunits, and other protein complexes not depicted here assemble at a promoter in DNA. The polymerase carries out transcription of DNA; the associated proteins are required for initial binding of polymerase to a specific promoter, thereby initiating transcription.

### Members of Protein Families Have a Common Evolutionary Ancestor

Studies on myoglobin and hemoglobin, the oxygen-carrying proteins in muscle and blood, respectively, provided early evidence that function derives from three-dimensional structure, which in turn is specified by amino acid sequence. X-ray crystallographic analysis showed that the three-dimensional structures of myoglobin and the  $\alpha$  and  $\beta$  subunits of hemoglobin are remarkably similar. Subsequent sequencing of myoglobin and the hemoglobin subunits revealed that many identical or chemically similar residues are found in identical positions throughout the primary structures of both proteins.

Similar comparisons between other proteins conclusively confirmed the relation between the amino acid sequence, three-dimensional structure, and function of proteins. This principle is now commonly employed to predict, on the basis of sequence comparisons with proteins of known structure and function, the structure and function of proteins that have not been isolated (Chapter 9). This use of sequence comparisons has expanded substantially in recent years as the genomes of more and more organisms have been sequenced.

The molecular revolution in biology during the last decades of the twentieth century also created a new scheme



▲ FIGURE 3-10 Evolution of the globin protein family. (Left)

A primitive monomeric oxygen-binding globin is thought to be the ancestor of modern-day blood hemoglobins, muscle myoglobins, and plant leghemoglobins. Sequence comparisons have revealed that evolution of the globin proteins parallels the evolution of animals and plants. Major junctions occurred with the divergence of plant globins from animal globins and of myoglobin from hemoglobin. Later gene duplication gave rise to the  $\alpha$  and  $\beta$

subunits of hemoglobin. (Right) Hemoglobin is a tetramer of two  $\alpha$  and two  $\beta$  subunits. The structural similarity of these subunits with leghemoglobin and myoglobin, both of which are monomers, is evident. A heme molecule (red) noncovalently associated with each globin polypeptide is the actual oxygen-binding moiety in these proteins. [(Left) Adapted from R. C. Hardison, 1996, *Proc. Natl. Acad. Sci. USA* **93**:5675.]

of biological classification based on similarities and differences in the amino acid sequences of proteins. Proteins that have a common ancestor are referred to as *homologs*. The main evidence for **homology** among proteins, and hence their common ancestry, is similarity in their sequences or structures. We can therefore describe homologous proteins as belonging to a “family” and can trace their lineage from comparisons of their sequences. The folded three-dimensional structures of homologous proteins are similar even if parts of their primary structure show little evidence of homology.

The kinship among homologous proteins is most easily visualized by a tree diagram based on sequence analyses. For example, the amino acid sequences of globins from bacteria, plants, and animals suggest that they evolved from an ancestral monomeric, oxygen-binding protein (Figure 3-10). With the passage of time, the gene for this ancestral protein slowly changed, initially diverging into lineages leading to animal and plant globins. Subsequent changes gave rise to myoglobin, a monomeric oxygen-storing protein in muscle, and to the  $\alpha$  and  $\beta$  subunits of the tetrameric hemoglobin molecule ( $\alpha_2\beta_2$ ) of the circulatory system.

## KEY CONCEPTS OF SECTION 3.1

### Hierarchical Structure of Proteins

- A protein is a linear polymer of amino acids linked together by peptide bonds. Various, mostly noncovalent, interactions between amino acids in the linear sequence stabilize a specific folded three-dimensional structure (conformation) for each protein.
- The  $\alpha$  helix,  $\beta$  strand and sheet, and turn are the most prevalent elements of protein secondary structure, which is stabilized by hydrogen bonds between atoms of the peptide backbone.
- Certain combinations of secondary structures give rise to different motifs, which are found in a variety of proteins and are often associated with specific functions (see Figure 3-6).
- Protein tertiary structure results from hydrophobic interactions between nonpolar side groups and hydrogen bonds between polar side groups that stabilize folding of the secondary structure into a compact overall arrangement, or conformation.
- Large proteins often contain distinct domains, independently folded regions of tertiary structure with characteristic structural or functional properties or both (see Figure 3-7).
- The incorporation of domains as modules in different proteins in the course of evolution has generated diversity in protein structure and function.
- Quaternary structure encompasses the number and organization of subunits in multimeric proteins.

- Cells contain large macromolecular assemblies in which all the necessary participants in complex cellular processes (e.g., DNA, RNA, and protein synthesis; photosynthesis; signal transduction) are integrated to form molecular machines (see Table 3-1).
- The sequence of a protein determines its three-dimensional structure, which determines its function. In short, function derives from structure; structure derives from sequence.
- Homologous proteins, which have similar sequences, structures, and functions, evolved from a common ancestor.

## 3.2 Folding, Modification, and Degradation of Proteins

A polypeptide chain is synthesized by a complex process called **translation** in which the assembly of amino acids in a particular sequence is dictated by **messenger RNA (mRNA)**. The intricacies of translation are considered in Chapter 4. Here, we describe how the cell promotes the proper folding of a nascent polypeptide chain and, in many cases, modifies residues or cleaves the polypeptide backbone to generate the final protein. In addition, the cell has error-checking processes that eliminate incorrectly synthesized or folded proteins. Incorrectly folded proteins usually lack biological activity and, in some cases, may actually be associated with disease. Protein misfolding is suppressed by two distinct mechanisms. First, cells have systems that reduce the chances for misfolded proteins to form. Second, any misfolded proteins that do form, as well as cytosolic proteins no longer needed by a cell, are degraded by a specialized cellular garbage-disposal system.

### The Information for Protein Folding Is Encoded in the Sequence

Any polypeptide chain containing  $n$  residues could, in principle, fold into  $8^n$  conformations. This value is based on the fact that only eight bond angles are stereochemically allowed in the polypeptide backbone. In general, however, all molecules of any protein species adopt a single conformation, called the *native state*; for the vast majority of proteins, the native state is the most stably folded form of the molecule.

What guides proteins to their native folded state? The answer to this question initially came from *in vitro* studies on protein refolding. Thermal energy from heat, extremes of pH that alter the charges on amino acid side chains, and chemicals such as urea or guanidine hydrochloride at concentrations of 6–8 M can disrupt the weak noncovalent interactions that stabilize the native conformation of a protein. The **denaturation** resulting from such treatment causes a protein to lose both its native conformation and its biological activity.

Many proteins that are completely unfolded in 8 M urea and  $\beta$ -mercaptoethanol (which reduces disulfide bonds) spontaneously *renature* (refold) into their native states when the denaturing reagents are removed by dialysis. Because no cofactors

or other proteins are required, *in vitro* protein folding is a self-directed process. In other words, sufficient information must be contained in the protein's primary sequence to direct correct refolding. The observed similarity in the folded, three-dimensional structures of proteins with similar amino acid sequences, noted in Section 3.1, provided other evidence that the primary sequence also determines protein folding *in vivo*.

### Folding of Proteins *In Vivo* Is Promoted by Chaperones

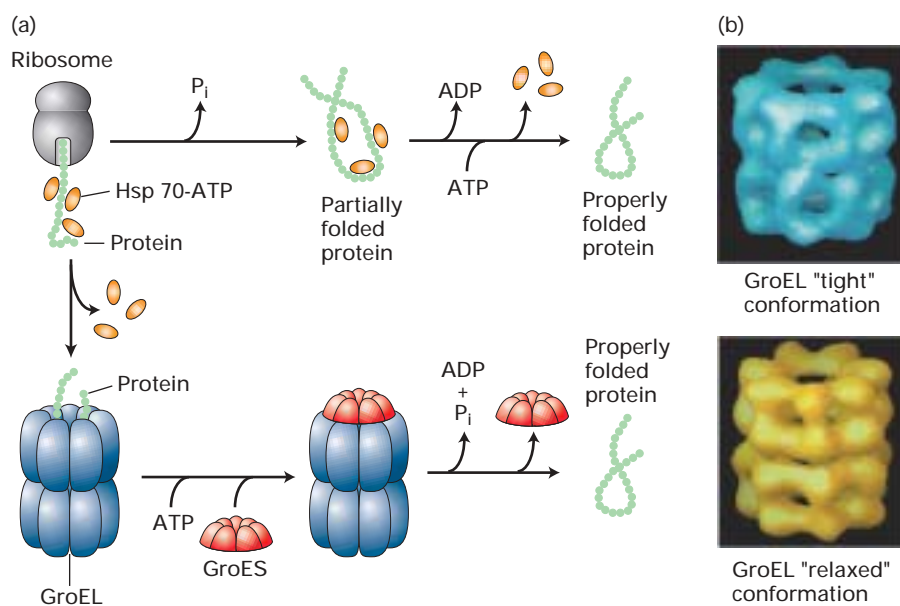
Although protein folding occurs *in vitro*, only a minority of unfolded molecules undergo complete folding into the native conformation within a few minutes. Clearly, cells require a faster, more efficient mechanism for folding proteins into their correct shapes; otherwise, cells would waste much energy in the synthesis of nonfunctional proteins and in the degradation of misfolded or unfolded proteins. Indeed, more than 95 percent of the proteins present within cells have been shown to be in their native conformation, despite high protein concentrations (200–300 mg/ml), which favor the precipitation of proteins *in vitro*.

The explanation for the cell's remarkable efficiency in promoting protein folding probably lies in **chaperones**, a

class of proteins found in all organisms from bacteria to humans. Chaperones are located in every cellular compartment, bind a wide range of proteins, and function in the general protein-folding mechanism of cells. Two general families of chaperones are reconized:

- **Molecular chaperones**, which bind and stabilize unfolded or partly folded proteins, thereby preventing these proteins from aggregating and being degraded
- **Chaperonins**, which directly facilitate the folding of proteins

Molecular chaperones consist of Hsp70 and its homologs: Hsp70 in the cytosol and mitochondrial matrix, BiP in the endoplasmic reticulum, and DnaK in bacteria. First identified by their rapid appearance after a cell has been stressed by heat shock, Hsp70 and its homologs are the major chaperones in all organisms. (Hsc70 is a constitutively expressed homolog of Hsp70.) When bound to ATP, Hsp70-like proteins assume an open form in which an exposed hydrophobic pocket transiently binds to exposed hydrophobic regions of the unfolded target protein. Hydrolysis of the bound ATP causes molecular chaperones to assume a closed form in which a target protein can undergo folding. The exchange of ATP for ADP releases the target protein (Figure 3-11a, *top*). This cycle is



▲ **FIGURE 3-11 Chaperone- and chaperonin-mediated protein folding.** (a) Many proteins fold into their proper three-dimensional structures with the assistance of Hsp70-like proteins (*top*). These molecular chaperones transiently bind to a nascent polypeptide as it emerges from a ribosome. Proper folding of other proteins (*bottom*) depends on chaperonins such as the prokaryotic GroEL, a hollow, barrel-shaped complex of 14 identical 60,000-MW subunits arranged in two stacked rings.

One end of GroEL is transiently blocked by the co-chaperonin GroES, an assembly of 10,000-MW subunits. (b) In the absence of ATP or presence of ADP, GroEL exists in a "tight" conformational state that binds partly folded or misfolded proteins. Binding of ATP shifts GroEL to a more open, "relaxed" state, which releases the folded protein. See text for details. [Part (b) from A. Roseman et al., 1996, *Cell* 87:241; courtesy of H. Saibil.]



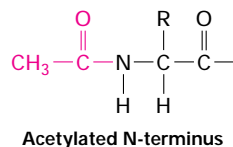
speeded by the co-chaperone Hsp40 in eukaryotes. In bacteria, an additional protein called GrpE also interacts with DnaK, promoting the exchange of ATP for the bacterial co-chaperone DnaJ and possibly its dissociation. Molecular chaperones are thought to bind all nascent polypeptide chains as they are being synthesized on ribosomes. In bacteria, 85 percent of the proteins are released from their chaperones and proceed to fold normally; an even higher percentage of proteins in eukaryotes follow this pathway.

The proper folding of a large variety of newly synthesized or translocated proteins also requires the assistance of chaperonins. These huge cylindrical macromolecular assemblies are formed from two rings of oligomers. The eukaryotic chaperonin *TriC* consists of eight subunits per ring. In the bacterial, mitochondrial, and chloroplast chaperonin, known as *GroEL*, each ring contains seven identical subunits (Figure 3-11b). The *GroEL* folding mechanism, which is better understood than *TriC*-mediated folding, serves as a general model (Figure 3-11a, *bottom*). In bacteria, a partly folded or misfolded polypeptide is inserted into the cavity of *GroEL*, where it binds to the inner wall and folds into its native conformation. In an ATP-dependent step, *GroEL* undergoes a conformational change and releases the folded protein, a process assisted by a co-chaperonin, *GroES*, which caps the ends of *GroEL*.

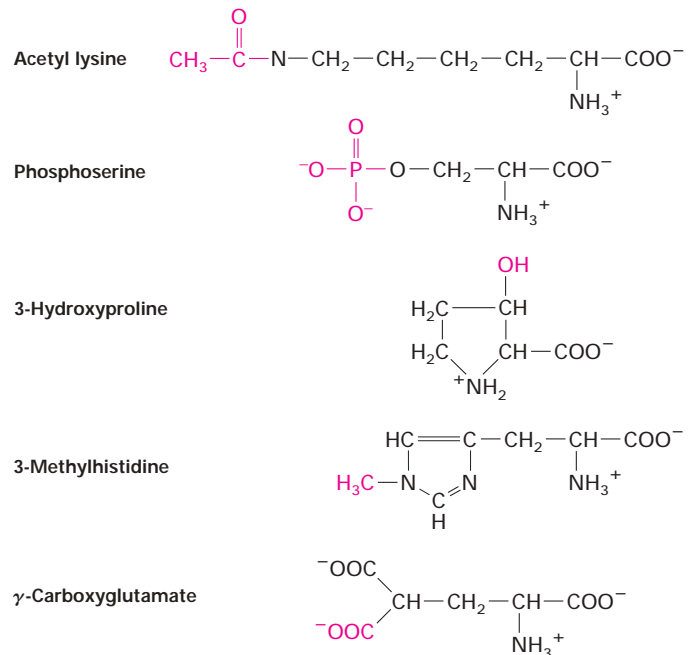
## Many Proteins Undergo Chemical Modification of Amino Acid Residues

Nearly every protein in a cell is chemically modified after its synthesis on a ribosome. Such modifications, which may alter the activity, life span, or cellular location of proteins, entail the linkage of a chemical group to the free  $-\text{NH}_2$  or  $-\text{COOH}$  group at either end of a protein or to a reactive side-chain group in an internal residue. Although cells use the 20 amino acids shown in Figure 2-13 to synthesize proteins, analysis of cellular proteins reveals that they contain upward of 100 different amino acids. Chemical modifications after synthesis account for this difference.

**Acetylation**, the addition of an acetyl group ( $\text{CH}_3\text{CO}$ ) to the amino group of the N-terminal residue, is the most common form of chemical modification, affecting an estimated 80 percent of all proteins:



This modification may play an important role in controlling the life span of proteins within cells because nonacetylated proteins are rapidly degraded by intracellular proteases. Residues at or near the termini of some membrane proteins are chemically modified by the addition of long lipidlike groups. The attachment of these hydrophobic “tails,” which function to anchor proteins to the lipid bilayer, constitutes one way that cells localize certain proteins to membranes (Chapter 5).



▲ FIGURE 3-12 Common modifications of internal amino acid residues found in proteins. These modified residues and numerous others are formed by addition of various chemical groups (red) to the amino acid side chains after synthesis of a polypeptide chain.

Acetyl groups and a variety of other chemical groups can also be added to specific internal residues in proteins (Figure 3-12). An important modification is the **phosphorylation** of serine, threonine, tyrosine, and histidine residues. We will encounter numerous examples of proteins whose activity is regulated by reversible phosphorylation and dephosphorylation. The side chains of asparagine, serine, and threonine are sites for **glycosylation**, the attachment of linear and branched carbohydrate chains. Many secreted proteins and membrane proteins contain glycosylated residues; the synthesis of such proteins is described in Chapters 16 and 17. Other post-translational modifications found in selected proteins include the **hydroxylation** of proline and lysine residues in collagen, the **methylation** of histidine residues in membrane receptors, and the  **$\gamma$ -carboxylation** of glutamate in prothrombin, an essential blood-clotting factor. A special modification, discussed shortly, marks cytosolic proteins for degradation.

## Peptide Segments of Some Proteins Are Removed After Synthesis

After their synthesis, some proteins undergo irreversible changes that do not entail changes in individual amino acid residues. This type of post-translational alteration is sometimes called **processing**. The most common form is enzymatic cleavage of a backbone peptide bond by proteases, resulting in the removal of residues from the C- or N-terminus of a

polypeptide chain. Proteolytic cleavage is a common mechanism for activating enzymes that function in blood coagulation, digestion, and programmed cell death (Chapter 22). Proteolysis also generates active peptide hormones, such as EGF and insulin, from larger precursor polypeptides.

An unusual and rare type of processing, termed *protein self-splicing*, takes place in bacteria and some eukaryotes. This process is analogous to editing film: an internal segment of a polypeptide is removed and the ends of the polypeptide are rejoined. Unlike proteolytic processing, protein self-splicing is an autocatalytic process, which proceeds by itself without the participation of enzymes. The excised peptide appears to eliminate itself from the protein by a mechanism similar to that used in the processing of some RNA molecules (Chapter 12). In vertebrate cells, the processing of some proteins includes self-cleavage, but the subsequent ligation step is absent. One such protein is Hedgehog, a membrane-bound signaling molecule that is critical to a number of developmental processes (Chapter 15).

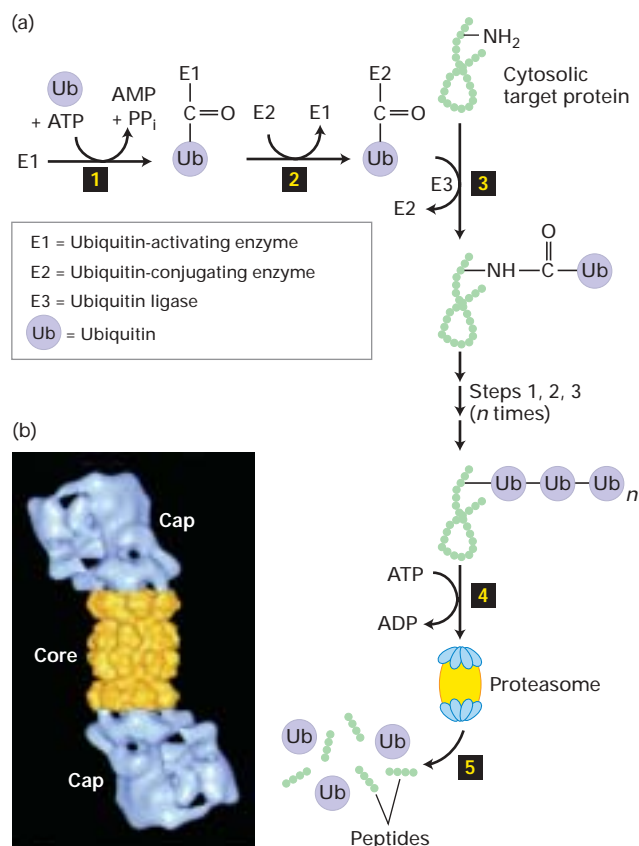
### Ubiquitin Marks Cytosolic Proteins for Degradation in Proteasomes

In addition to chemical modifications and processing, the activity of a cellular protein depends on the amount present, which reflects the balance between its rate of synthesis and rate of degradation in the cell. The numerous ways that cells regulate protein synthesis are discussed in later chapters. In this section, we examine protein degradation, focusing on the major pathways for degrading cytosolic proteins.

The life span of intracellular proteins varies from as short as a few minutes for mitotic cyclins, which help regulate passage through mitosis, to as long as the age of an organism for proteins in the lens of the eye. Eukaryotic cells have several intracellular proteolytic pathways for degrading misfolded or denatured proteins, normal proteins whose concentration must be decreased, and extracellular proteins taken up by the cell. One major intracellular pathway is degradation by enzymes within **lysosomes**, membrane-limited organelles whose acidic interior is filled with hydrolytic enzymes. Lysosomal degradation is directed primarily toward extracellular proteins taken up by the cell and aged or defective organelles of the cell (see Figure 5-20).

Distinct from the lysosomal pathway are cytosolic mechanisms for degrading proteins. Chief among these mechanisms is a pathway that includes the chemical modification of a lysine side chain by the addition of **ubiquitin**, a 76-residue polypeptide, followed by degradation of the ubiquitin-tagged protein by a specialized proteolytic machine. Ubiquitination is a three-step process (Figure 3-13a):

- Activation of *ubiquitin-activating enzyme* (E1) by the addition of a ubiquitin molecule, a reaction that requires ATP
- Transfer of this ubiquitin molecule to a cysteine residue in *ubiquitin-conjugating enzyme* (E2)



▲ **FIGURE 3-13 Ubiquitin-mediated proteolytic pathway.** (a) Enzyme E1 is activated by attachment of a ubiquitin (Ub) molecule (step **1**) and then transfers this Ub molecule to E2 (step **2**). Ubiquitin ligase (E3) transfers the bound Ub molecule on E2 to the side-chain  $-\text{NH}_2$  of a lysine residue in a target protein (step **3**). Additional Ub molecules are added to the target protein by repeating steps **1–3**, forming a polyubiquitin chain that directs the tagged protein to a proteasome (step **4**). Within this large complex, the protein is cleaved into numerous small peptide fragments (step **5**). (b) Computer-generated image reveals that a proteasome has a cylindrical structure with a cap at each end of a core region. Proteolysis of ubiquitin-tagged proteins occurs along the inner wall of the core. [Part (b) from W. Baumeister et al., 1998, *Cell* **92**:357; courtesy of W. Baumeister.]

- Formation of a peptide bond between the ubiquitin molecule bound to E2 and a lysine residue in the target protein, a reaction catalyzed by *ubiquitin ligase* (E3)

This process is repeated many times, with each subsequent ubiquitin molecule being added to the preceding one. The resulting polyubiquitin chain is recognized by a **proteasome**, another of the cell's molecular machines (Figure 3-13b). The numerous proteasomes dispersed throughout the cell cytosol proteolytically cleave ubiquitin-tagged proteins in an ATP-dependent process that yields short (7- to 8-residue) peptides and intact ubiquitin molecules.

Cellular proteins degraded by the ubiquitin-mediated pathway fall into one of two general categories: (1) native cytosolic proteins whose life spans are tightly controlled and (2) proteins that become misfolded in the course of their synthesis in the endoplasmic reticulum (ER). Both contain sequences recognized by the ubiquitinating enzyme complex. The cyclins, for example, are cytosolic proteins whose amounts are tightly controlled throughout the cell cycle. These proteins contain the internal sequence Arg-X-X-Leu-Gly-X-Ile-Gly-Asp/Asn (X can be any amino acid), which is recognized by specific ubiquitinating enzyme complexes. At a specific time in the cell cycle, each cyclin is phosphorylated by a cyclin kinase. This phosphorylation is thought to cause a conformational change that exposes the recognition sequence to the ubiquitinating enzymes, leading to degradation of the tagged cyclin (Chapter 21). Similarly, the misfolding of proteins in the endoplasmic reticulum exposes hydrophobic sequences normally buried within the folded protein. Such proteins are transported to the cytosol, where ubiquitinating enzymes recognize the exposed hydrophobic sequences.

The immune system also makes use of the ubiquitin-mediated pathway in the response to altered self-cells, particularly virus-infected cells. Viral proteins within the cytosol of infected cells are ubiquitinated and then degraded in proteasomes specially designed for this role. The resulting antigenic peptides are transported to the endoplasmic reticulum, where they bind to class I major histocompatibility complex (MHC) molecules within the ER membrane. Subsequently, the peptide-MHC complexes move to the cell membrane where the antigenic peptides can be recognized by cytotoxic T lymphocytes, which mediate the destruction of the infected cells.

### Digestive Proteases Degrade Dietary Proteins

The major extracellular pathway for protein degradation is the system of digestive proteases that breaks down ingested proteins into peptides and amino acids in the intestinal tract. Three classes of proteases function in digestion. *Endoproteases* attack selected peptide bonds within a polypeptide chain. The principal endoproteases are pepsin, which preferentially cleaves the backbone adjacent to phenylalanine and leucine residues, and trypsin and chymotrypsin, which cleave the backbone adjacent to basic and aromatic residues. *Exopeptidases* sequentially remove residues from the N-terminus (aminopeptidases) or C-terminus (carboxypeptidases) of a protein. *Peptidases* split oligopeptides containing as many as about 20 amino acids into di- and tripeptides and individual amino acids. These small molecules are then transported across the intestinal lining into the bloodstream.

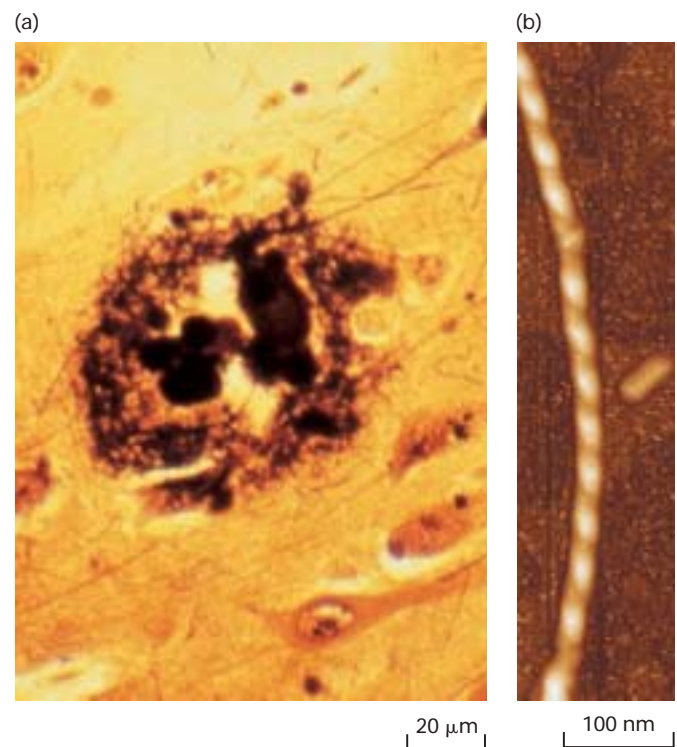
To protect a cell from degrading itself, endoproteases and carboxypeptidases are synthesized and secreted as inactive forms (zymogens): pepsin by chief cells in the lining of the stomach; the others by pancreatic cells. Proteolytic cleavage of the zymogens within the gastric or intestinal lumen yields the active enzymes. Intestinal epithelial cells produce aminopeptidases and the di- and tripeptidases.

### Alternatively Folded Proteins Are Implicated in Slowly Developing Diseases



As noted earlier, each protein species normally folds into a single, energetically favorable conformation that is specified by its amino acid sequence. Recent evidence suggests, however, that a protein may fold into an alternative three-dimensional structure as the result of mutations, inappropriate post-translational modification, or other as-yet-unidentified reasons. Such “misfolding” not only leads to a loss of the normal function of the protein but also marks it for proteolytic degradation. The subsequent accumulation of proteolytic fragments contributes to certain degenerative diseases characterized by the presence of insoluble protein plaques in various organs, including the liver and brain. ■

Some neurodegenerative diseases, including Alzheimer’s disease and Parkinson’s disease in humans and transmissible spongiform encephalopathy (“mad cow” disease) in cows



▲ **EXPERIMENTAL FIGURE 3-14 Alzheimer’s disease is characterized by the formation of insoluble plaques composed of amyloid protein.** (a) At low resolution, an amyloid plaque in the brain of an Alzheimer’s patient appears as a tangle of filaments. (b) The regular structure of filaments from plaques is revealed in the atomic force microscope. Proteolysis of the naturally occurring amyloid precursor protein yields a short fragment, called  $\beta$ -amyloid protein, that for unknown reasons changes from an  $\alpha$ -helical to a  $\beta$ -sheet conformation. This alternative structure aggregates into the highly stable filaments (amyloid) found in plaques. Similar pathologic changes in other proteins cause other degenerative diseases. [Courtesy of K. Kosik.]



and sheep, are marked by the formation of tangled filamentous plaques in a deteriorating brain (Figure 3-14). The *amyloid filaments* composing these structures derive from abundant natural proteins such as amyloid precursor protein, which is embedded in the plasma membrane, Tau, a microtubule-binding protein, and prion protein, an “infectious” protein whose inheritance follows Mendelian genetics. Influenced by unknown causes, these  $\alpha$  helix-containing proteins or their proteolytic fragments fold into alternative  $\beta$  sheet-containing structures that polymerize into very stable filaments. Whether the extracellular deposits of these filaments or the soluble alternatively folded proteins are toxic to the cell is unclear.

### KEY CONCEPTS OF SECTION 3.2

#### Folding, Modification, and Degradation of Proteins

- The amino acid sequence of a protein dictates its folding into a specific three-dimensional conformation, the native state.
- Protein folding in vivo occurs with assistance from molecular chaperones (Hsp70 proteins), which bind to nascent polypeptides emerging from ribosomes and prevent their misfolding (see Figure 3-11). Chaperonins, large complexes of Hsp60-like proteins, shelter some partly folded or misfolded proteins in a barrel-like cavity, providing additional time for proper folding.
- Subsequent to their synthesis, most proteins are modified by the addition of various chemical groups to amino acid residues. These modifications, which alter protein structure and function, include acetylation, hydroxylation, glycosylation, and phosphorylation.
- The life span of intracellular proteins is largely determined by their susceptibility to proteolytic degradation by various pathways.
- Viral proteins produced within infected cells, normal cytosolic proteins, and misfolded proteins are marked for destruction by the covalent addition of a polyubiquitin chain and then degraded within proteasomes, large cylindrical complexes with multiple proteases in their interiors (see Figure 3-13).
- Some neurodegenerative diseases are caused by aggregates of proteins that are stably folded in an alternative conformation.

## 3.3 Enzymes and the Chemical Work of Cells

Proteins are designed to bind every conceivable molecule—from simple ions and small metabolites (sugars, fatty acids) to large complex molecules such as other proteins and nucleic acids. Indeed, the function of nearly all proteins depends on their ability to bind other molecules, or **ligands**, with a high

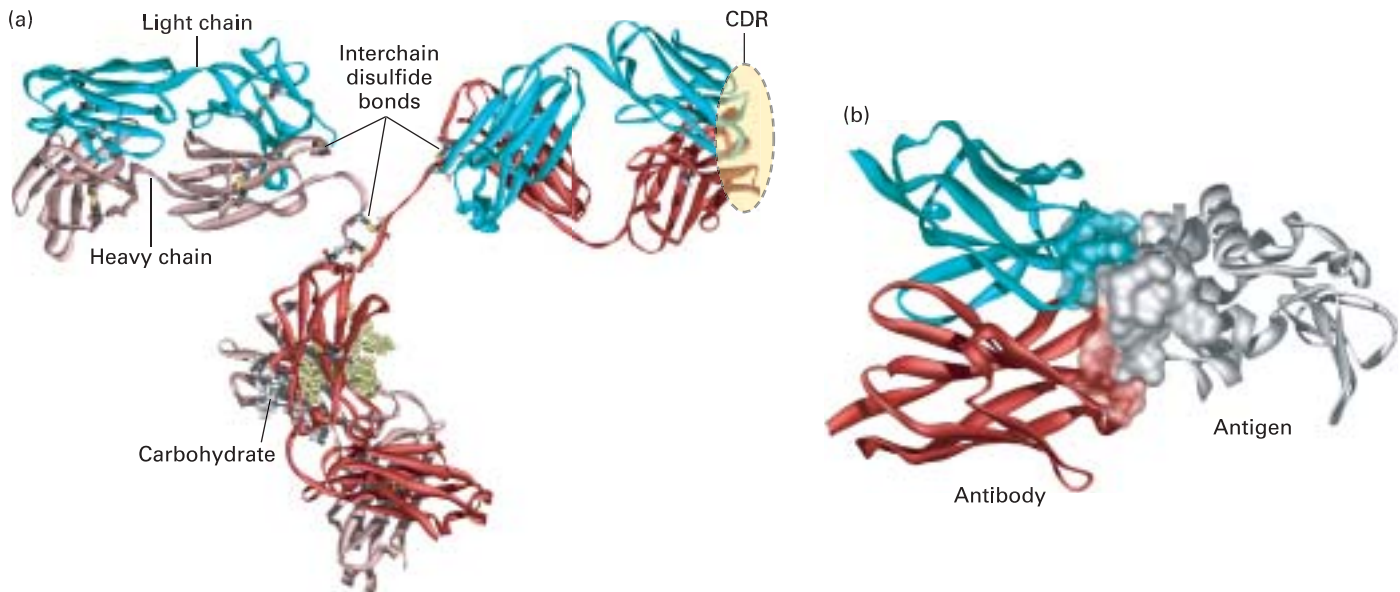
degree of specificity. For instance, an enzyme must first bind specifically to its target molecule, which may be a small molecule (e.g., glucose) or a macromolecule, before it can execute its specific task. Likewise, the many different types of hormone receptors on the surface of cells display a high degree of sensitivity and discrimination for their ligands. And, as we will examine in Chapter 11, the binding of certain regulatory proteins to specific sequences in DNA is a major mechanism for controlling genes. Ligand binding often causes a change in the shape of a protein. Ligand-driven conformational changes are integral to the mechanism of action of many proteins and are important in regulating protein activity. After considering the general properties of protein–ligand binding, we take a closer look at how enzymes are designed to function as the cell’s chemists.

### Specificity and Affinity of Protein–Ligand Binding Depend on Molecular Complementarity

Two properties of a protein characterize its interaction with ligands. *Specificity* refers to the ability of a protein to bind one molecule in preference to other molecules. *Affinity* refers to the strength of binding. The  $K_d$  for a protein–ligand complex, which is the inverse of the equilibrium constant  $K_{eq}$  for the binding reaction, is the most common quantitative measure of affinity (Chapter 2). The stronger the interaction between a protein and ligand, the lower the value of  $K_d$ . Both the specificity and the affinity of a protein for a ligand depend on the structure of the ligand-binding site, which is designed to fit its partner like a mold. For high-affinity and highly specific interactions to take place, the shape and chemical surface of the binding site must be complementary to the ligand molecule, a property termed **molecular complementarity**.

The ability of proteins to distinguish different molecules is perhaps most highly developed in the blood proteins called **antibodies**, which animals produce in response to **antigens**, such as infectious agents (e.g., a bacterium or a virus), and certain foreign substances (e.g., proteins or polysaccharides in pollens). The presence of an antigen causes an organism to make a large quantity of different antibody proteins, each of which may bind to a slightly different region, or **epitope**, of the antigen. Antibodies act as specific sensors for antigens, forming antibody–antigen complexes that initiate a cascade of protective reactions in cells of the immune system.

All antibodies are Y-shaped molecules formed from two identical heavy chains and two identical light chains (Figure 3-15a). Each arm of an antibody molecule contains a single light chain linked to a heavy chain by a disulfide bond. Near the end of each arm are six highly variable loops, called *complementarity-determining regions (CDRs)*, which form the antigen-binding sites. The sequences of the six loops are highly variable among antibodies, making them specific for different antigens. The interaction between an antibody and an epitope in an antigen is complementary in all cases; that is, the surface of the antibody’s antigen-binding site physically matches the corresponding epitope like a glove



▲ **FIGURE 3-15 Antibody structure and antibody-antigen interaction.** (a) Ribbon model of an antibody. Every antibody molecule consists of two identical heavy chains (red) and two identical light chains (blue) covalently linked by disulfide bonds. (b) The hand-in-glove fit between an antibody and an epitope on its antigen—in this case, chicken egg-white lysozyme. Regions

where the two molecules make contact are shown as surfaces. The antibody contacts the antigen with residues from all its complementarity-determining regions (CDRs). In this view, the complementarity of the antigen and antibody is especially apparent where “fingers” extending from the antigen surface are opposed to “clefts” in the antibody surface.

(Figure 3-15b). The intimate contact between these two surfaces, stabilized by numerous noncovalent bonds, is responsible for the exquisite binding specificity exhibited by an antibody.

The specificity of antibodies is so precise that they can distinguish between the cells of individual members of a species and in some cases can distinguish between proteins that differ by only a single amino acid. Because of their specificity and the ease with which they can be produced, antibodies are highly useful reagents in many of the experiments discussed in subsequent chapters.

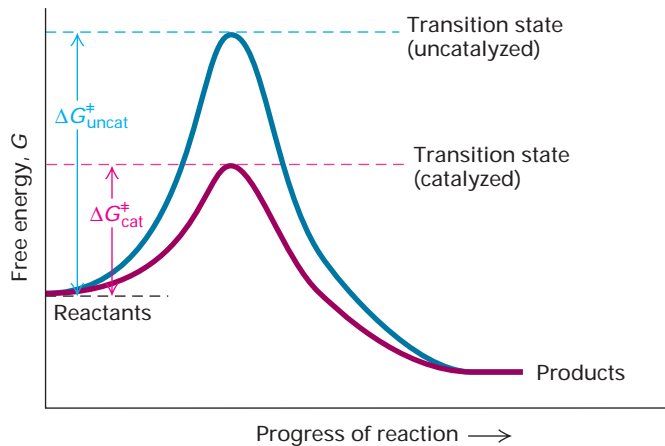
### Enzymes Are Highly Efficient and Specific Catalysts

In contrast with antibodies, which bind and simply present their ligands to other components of the immune system, enzymes promote the chemical alteration of their ligands, called **substrates**. Almost every chemical reaction in the cell is catalyzed by a specific enzyme. Like all catalysts, enzymes do not affect the extent of a reaction, which is determined by the change in free energy  $\Delta G$  between reactants and products (Chapter 2). For reactions that are energetically favorable ( $-\Delta G$ ), enzymes increase the reaction rate by lowering the **activation energy** (Figure 3-16). In the test tube, catalysts such as charcoal and platinum facilitate reactions but usually only at high temperatures or pressures, at extremes of high

or low pH, or in organic solvents. As the cell’s protein catalysts, however, enzymes must function effectively in aqueous environment at 37°C, 1 atmosphere pressure, and pH 6.5–7.5.

Two striking properties of enzymes enable them to function as catalysts under the mild conditions present in cells: their enormous *catalytic power* and their high degree of *specificity*. The immense catalytic power of enzymes causes the rates of enzymatically catalyzed reactions to be  $10^6$ – $10^{12}$  times that of the corresponding uncatalyzed reactions under otherwise similar conditions. The exquisite specificity of enzymes—their ability to act selectively on one substrate or a small number of chemically similar substrates—is exemplified by the enzymes that act on amino acids. As noted in Chapter 2, amino acids can exist as two stereoisomers, designated L and D, although only L isomers are normally found in biological systems. Not surprisingly, enzyme-catalyzed reactions of L-amino acids take place much more rapidly than do those of D-amino acids, even though both stereoisomers of a given amino acid are the same size and possess the same R groups (see Figure 2-12).

Approximately 3700 different types of enzymes, each of which catalyzes a single chemical reaction or set of closely related reactions, have been classified in the enzyme database. Certain enzymes are found in the majority of cells because they catalyze the synthesis of common cellular products (e.g., proteins, nucleic acids, and phospholipids) or take part in the



**▲ FIGURE 3-16 Effect of a catalyst on the activation energy of a chemical reaction.** This hypothetical reaction pathway depicts the changes in free energy  $G$  as a reaction proceeds. A reaction will take place spontaneously only if the total  $G$  of the products is less than that of the reactants ( $-\Delta G$ ). However, all chemical reactions proceed through one or more high-energy transition states, and the rate of a reaction is inversely proportional to the activation energy ( $\Delta G^\ddagger$ ), which is the difference in free energy between the reactants and the highest point along the pathway. Enzymes and other catalysts accelerate the rate of a reaction by reducing the free energy of the transition state and thus  $\Delta G^\ddagger$ .

production of energy by the conversion of glucose and oxygen into carbon dioxide and water. Other enzymes are present only in a particular type of cell because they catalyze chemical reactions unique to that cell type (e.g., the enzymes that convert tyrosine into dopamine, a neurotransmitter, in nerve cells). Although most enzymes are located within cells, some are secreted and function in extracellular sites such as the blood, the lumen of the digestive tract, or even outside the organism.

The catalytic activity of some enzymes is critical to cellular processes other than the synthesis or degradation of molecules. For instance, many regulatory proteins and intracellular signaling proteins catalyze the phosphorylation of proteins, and some transport proteins catalyze the hydrolysis of ATP coupled to the movement of molecules across membranes.

### An Enzyme's Active Site Binds Substrates and Carries Out Catalysis

Certain amino acid side chains of an enzyme are important in determining its specificity and catalytic power. In the native conformation of an enzyme, these side chains are brought into proximity, forming the **active site**. Active sites thus consist of two functionally important regions: one that recognizes and binds the substrate (or substrates) and another that catalyzes the reaction after the substrate has been

bound. In some enzymes, the catalytic region is part of the substrate-binding region; in others, the two regions are structurally as well as functionally distinct.

To illustrate how the active site binds a specific substrate and then promotes a chemical change in the bound substrate, we examine the action of cyclic AMP-dependent protein kinase, now generally referred to as **protein kinase A (PKA)**. This enzyme and other protein kinases, which add a phosphate group to serine, threonine, or tyrosine residues in proteins, are critical for regulating the activity of many cellular proteins, often in response to external signals. Because the eukaryotic protein kinases belong to a common superfamily, the structure of the active site and mechanism of phosphorylation are very similar in all of them. Thus protein kinase A can serve as a general model for this important class of enzymes.

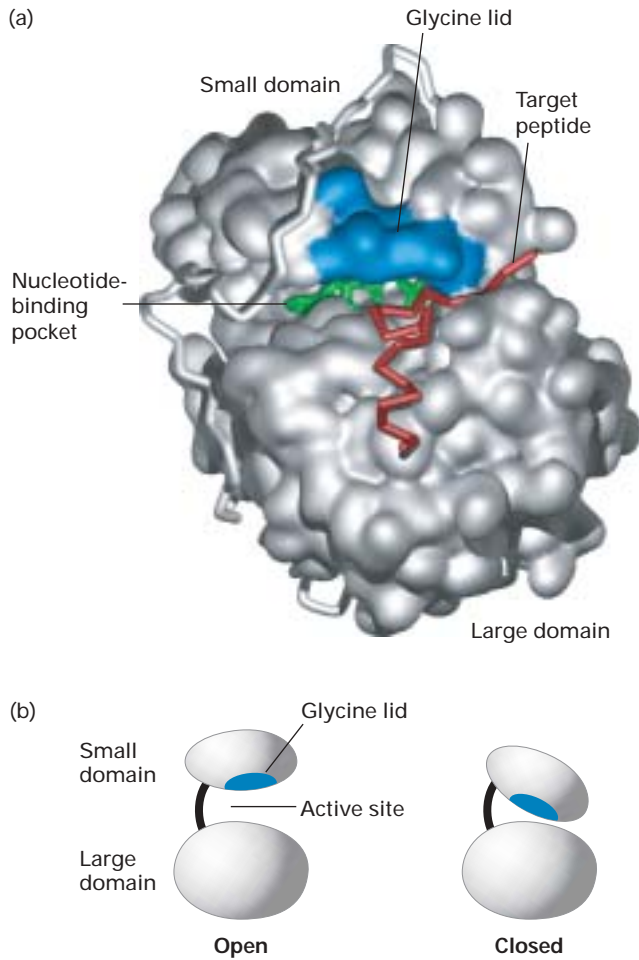
The active site of protein kinase A is located in the 240-residue “kinase core” of the catalytic subunit. The kinase core, which is largely conserved in all protein kinases, is responsible for the binding of substrates (ATP and a target peptide sequence) and the subsequent transfer of a phosphate group from ATP to a serine, threonine, or tyrosine residue in the target sequence. The kinase core consists of a large domain and small one, with an intervening deep cleft; the active site comprises residues located in both domains.

**Substrate Binding by Protein Kinases** The structure of the ATP-binding site in the catalytic kinase core complements the structure of the nucleotide substrate. The adenine ring of ATP sits snugly at the base of the cleft between the large and the small domains. A highly conserved sequence, Gly-X-Gly-X-X-Gly-X-Val (X can be any amino acid), dubbed the “glycine lid,” closes over the adenine ring and holds it in position (Figure 3-17a). Other conserved residues in the binding pocket stabilize the highly charged phosphate groups.

Although ATP is a common substrate for all protein kinases, the sequence of the target peptide varies among different kinases. The peptide sequence recognized by protein kinase A is Arg-Arg-X-Ser-Y, where X is any amino acid and Y is a hydrophobic amino acid. The part of the polypeptide chain containing the target serine or threonine residue is bound to a shallow groove in the large domain of the kinase core. The peptide specificity of protein kinase A is conferred by several glutamic acid residues in the large domain, which form salt bridges with the two arginine residues in the target peptide. Different residues determine the specificity of other protein kinases.

The catalytic core of protein kinase A exists in an “open” and “closed” conformation (Figure 3-17b). In the open conformation, the large and small domains of the core region are separated enough that substrate molecules can enter and bind. When the active site is occupied by substrate, the domains move together into the closed position. This change in tertiary structure, an example of *induced fit*, brings the target peptide sequence sufficiently close to accept a phosphate





▲ **FIGURE 3-17 Protein kinase A and conformational change induced by substrate binding.** (a) Model of the catalytic subunit of protein kinase A with bound substrates; the conserved kinase core is indicated as a molecular surface. An overhanging glycine-rich sequence (blue) traps ATP (green) in a deep cleft between the large and small domains of the core. Residues in the large domain bind the target peptide (red). The structure of the kinase core is largely conserved in other eukaryotic protein kinases. (b) Schematic diagrams of open and closed conformations of the kinase core. In the absence of substrate, the kinase core is in the open conformation. Substrate binding causes a rotation of the large and small domains that brings the ATP- and peptide-binding sites closer together and causes the glycine lid to move over the adenine residue of ATP, thereby trapping the nucleotide in the binding cleft. The model in part (a) is in the closed conformation.

group from the bound ATP. After the phosphorylation reaction has been completed, the presence of the products causes the domains to rotate to the open position, from which the products are released.

The rotation from the open to the closed position also causes movement of the glycine lid over the ATP-binding cleft. The glycine lid controls the entry of ATP and release of ADP at

the active site. In the open position, ATP can enter and bind the active site cleft; in the closed position, the glycine lid prevents ATP from leaving the cleft. Subsequent to phosphoryl transfer from the bound ATP to the bound peptide sequence, the glycine lid must rotate back to the open position before ADP can be released. Kinetic measurements show that the rate of ADP release is 20-fold slower than that of phosphoryl transfer, indicating the influence of the glycine lid on the rate of kinase reactions. Mutations in the glycine lid that inhibit its flexibility slow catalysis by protein kinase A even further.

**Phosphoryl Transfer by Protein Kinases** After substrates have bound and the catalytic core of protein kinase A has assumed the closed conformation, the phosphorylation of a serine or threonine residue on the target peptide can take place (Figure 3-18). As with all chemical reactions, phosphoryl transfer catalyzed by protein kinase A proceeds through a transition state in which the phosphate group to be transferred and the acceptor hydroxyl group are brought into close proximity. Binding and stabilization of the intermediates by protein kinase A reduce the activation energy of the phosphoryl transfer reaction, permitting it to take place at measurable rates under the mild conditions present within cells (see Figure 3-16). Formation of the products induces the enzyme to revert to its open conformational state, allowing ADP and the phosphorylated target peptide to diffuse from the active site.

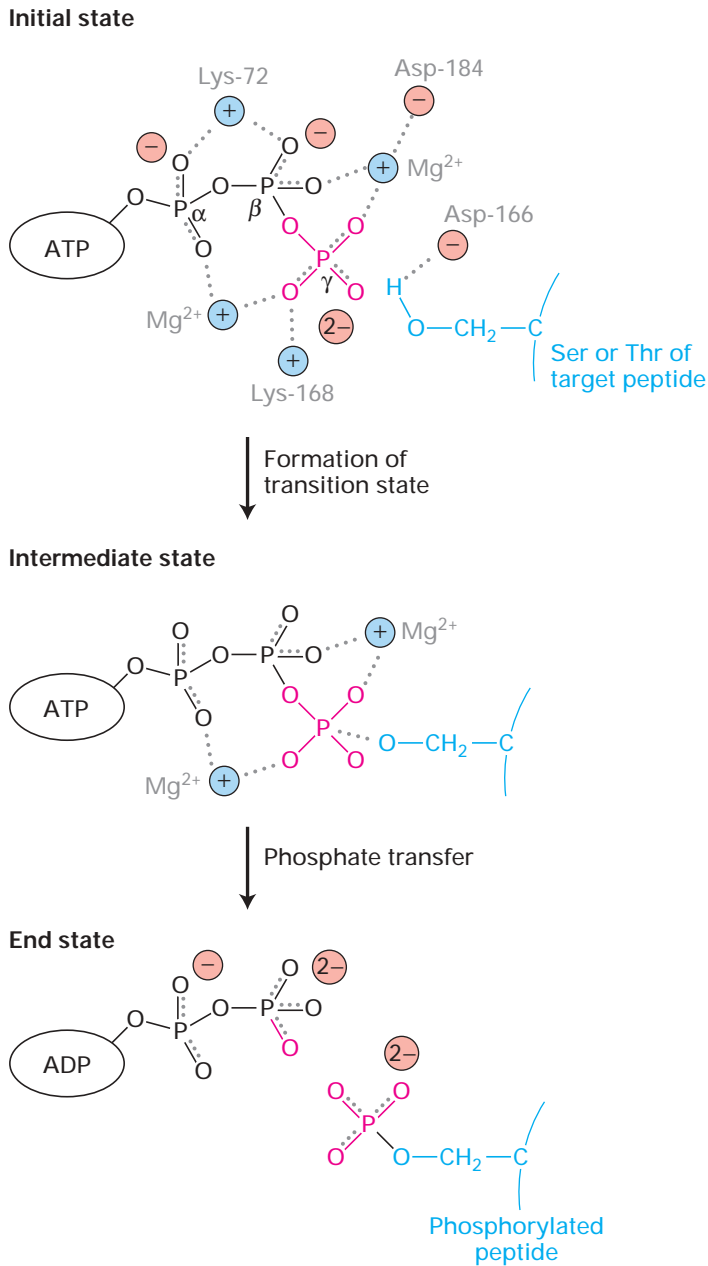
### $V_{\max}$ and $K_m$ Characterize an Enzymatic Reaction

The catalytic action of an enzyme on a given substrate can be described by two parameters:  $V_{\max}$ , the maximal velocity of the reaction at saturating substrate concentrations, and  $K_m$  (the **Michaelis constant**), a measure of the affinity of an enzyme for its substrate (Figure 3-19). The  $K_m$  is defined as the substrate concentration that yields a half-maximal reaction rate (i.e.,  $\frac{1}{2} V_{\max}$ ). The smaller the value of  $K_m$ , the more avidly an enzyme can bind substrate from a dilute solution and the smaller the substrate concentration needed to reach half-maximal velocity.

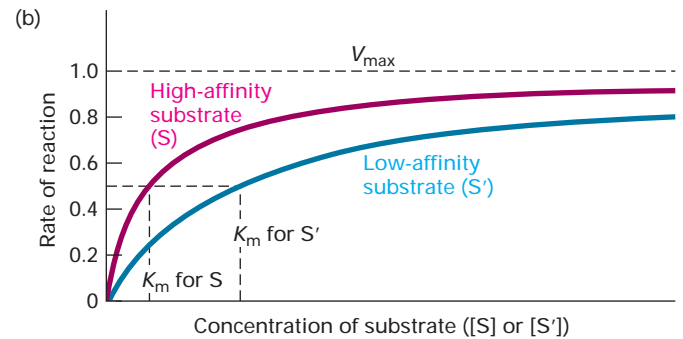
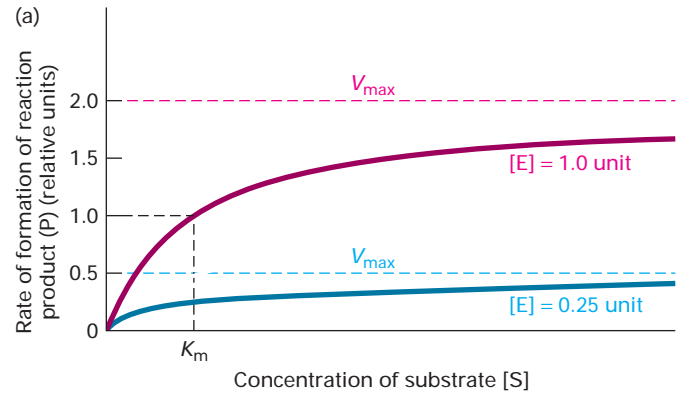
The concentrations of the various small molecules in a cell vary widely, as do the  $K_m$  values for the different enzymes that act on them. Generally, the intracellular concentration of a substrate is approximately the same as or greater than the  $K_m$  value of the enzyme to which it binds.

### Enzymes in a Common Pathway Are Often Physically Associated with One Another

Enzymes taking part in a common metabolic process (e.g., the degradation of glucose to pyruvate) are generally located in the same cellular compartment (e.g., in the cytosol, at a membrane, within a particular organelle). Within a compartment, products from one reaction can move by diffusion to the next enzyme in the pathway. However, diffusion entails random movement and is a slow, inefficient process for



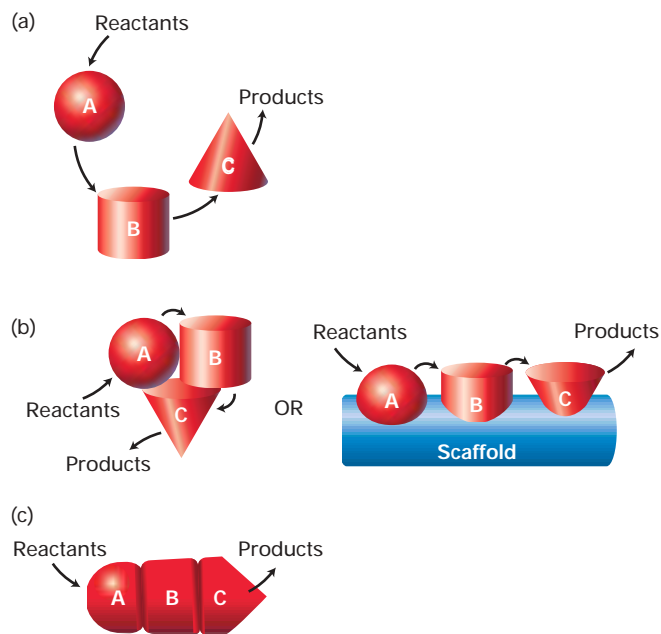
▲ **FIGURE 3-18 Mechanism of phosphorylation by protein kinase A.** (Top) Initially, ATP and the target peptide bind to the active site (see Figure 3-17a). Electrons of the phosphate group are delocalized by interactions with lysine side chains and  $Mg^{2+}$ . Colored circles represent the residues in the kinase core critical to substrate binding and phosphoryl transfer. Note that these residues are not adjacent to one another in the amino acid sequence. (Middle) A new bond then forms between the serine or threonine side-chain oxygen atom and  $\gamma$  phosphate, yielding a pentavalent intermediate. (Bottom) The phosphoester bond between the  $\beta$  and  $\gamma$  phosphates is broken, yielding the products ADP and a peptide with a phosphorylated serine or threonine side chain. The catalytic mechanism of other protein kinases is similar.



▲ **EXPERIMENTAL FIGURE 3-19 The  $K_m$  and  $V_{max}$  for an enzyme-catalyzed reaction are determined from plots of the initial velocity versus substrate concentration.** The shape of these hypothetical kinetic curves is characteristic of a simple enzyme-catalyzed reaction in which one substrate (S) is converted into product (P). The initial velocity is measured immediately after addition of enzyme to substrate before the substrate concentration changes appreciably. (a) Plots of the initial velocity at two different concentrations of enzyme [E] as a function of substrate concentration [S]. The [S] that yields a half-maximal reaction rate is the Michaelis constant  $K_m$ , a measure of the affinity of E for S. Doubling the enzyme concentration causes a proportional increase in the reaction rate, and so the maximal velocity  $V_{max}$  is doubled; the  $K_m$ , however, is unaltered. (b) Plots of the initial velocity versus substrate concentration with a substrate S for which the enzyme has a high affinity and with a substrate S' for which the enzyme has a low affinity. Note that the  $V_{max}$  is the same with both substrates but that  $K_m$  is higher for S', the low-affinity substrate.

moving molecules between widely dispersed enzymes (Figure 3-20a). To overcome this impediment, cells have evolved mechanisms for bringing enzymes in a common pathway into close proximity.

In the simplest such mechanism, polypeptides with different catalytic activities cluster closely together as subunits of a multimeric enzyme or assemble on a common “scaffold” (Figure 3-20b). This arrangement allows the products of one reaction to be channeled directly to the next enzyme in the pathway. The first approach is illustrated by pyruvate



▲ FIGURE 3-20 Evolution of multifunctional enzyme.

In the hypothetical reaction pathways illustrated here the initial reactants are converted into final products by the sequential action of three enzymes: A, B, and C. (a) When the enzymes are free in solution or even constrained within the same cellular compartment, the intermediates in the reaction sequence must diffuse from one enzyme to the next, an inherently slow process. (b) Diffusion is greatly reduced or eliminated when the enzymes associate into multisubunit complexes. (c) The closest integration of different catalytic activities occurs when the enzymes are fused at the genetic level, becoming domains in a single protein.

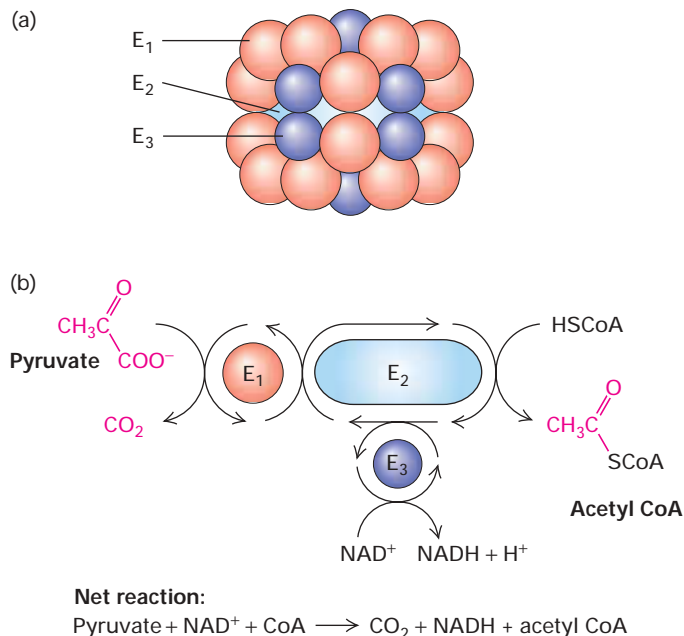
dehydrogenase, a complex of three distinct enzymes that converts pyruvate into acetyl CoA in mitochondria (Figure 3-21). The scaffold approach is employed by MAP kinase signal-transduction pathways, discussed in Chapter 14. In yeast, three protein kinases assembled on the Ste5 scaffold protein form a kinase cascade that transduces the signal triggered by the binding of mating factor to the cell surface.

In some cases, separate proteins have been fused together at the genetic level to create a single multidomain, multifunctional enzyme (Figure 3-20c). For instance, the isomerization of citrate to isocitrate in the citric acid cycle is catalyzed by aconitase, a single polypeptide that carries out two separate reactions: (1) the dehydration of citrate to form *cis*-aconitate and then (2) the hydration of *cis*-aconitate to yield isocitrate (see Figure 8-9).

### KEY CONCEPTS OF SECTION 3.3

#### Enzymes and the Chemical Work of Cells

- The function of nearly all proteins depends on their ability to bind other molecules (ligands). Ligand-binding sites



▲ FIGURE 3-21 Structure and function of pyruvate dehydrogenase, a large multimeric enzyme complex that converts pyruvate into acetyl CoA.

(a) The complex consists of 24 copies of pyruvate decarboxylase (E<sub>1</sub>), 24 copies of lipoamide transacetylase (E<sub>2</sub>), and 12 copies of dihydrolipooyl dehydrogenase (E<sub>3</sub>). The E<sub>1</sub> and E<sub>3</sub> subunits are bound to the outside of the core formed by the E<sub>2</sub> subunits. (b) The reactions catalyzed by the complex include several enzyme-bound intermediates (not shown). The tight structural integration of the three enzymes increases the rate of the overall reaction and minimizes possible side reactions.

on proteins and the corresponding ligands are chemically and topologically complementary.

- The affinity of a protein for a particular ligand refers to the strength of binding; its specificity refers to the preferential binding of one or a few closely related ligands.
- Enzymes are catalytic proteins that accelerate the rate of cellular reactions by lowering the activation energy and stabilizing transition-state intermediates (see Figure 3-16).
- An enzyme active site comprises two functional parts: a substrate-binding region and a catalytic region. The amino acids composing the active site are not necessarily adjacent in the amino acid sequence but are brought into proximity in the native conformation.
- From plots of reaction rate versus substrate concentration, two characteristic parameters of an enzyme can be determined: the Michaelis constant  $K_m$ , a measure of the enzyme's affinity for substrate, and the maximal velocity  $V_{max}$ , a measure of its catalytic power (see Figure 3-19).



- Enzymes in a common pathway are located within specific cell compartments and may be further associated as domains of a monomeric protein, subunits of a multimeric protein, or components of a protein complex assembled on a common scaffold (see Figure 3-20).

### 3.4 Molecular Motors and the Mechanical Work of Cells

A common property of all cells is motility, the ability to move in a specified direction. Many cell processes exhibit some type of movement at either the molecular or the cellular level; all movements result from the application of a force. In Brownian motion, for instance, thermal energy constantly buffets molecules and organelles in *random* directions and for very short distances. On the other hand, materials within a cell are transported in specific directions and for longer distances. This type of movement results from the mechanical work carried out by proteins that function as motors. We first briefly describe the types and general properties of molecular motors and then look at how one type of motor protein generates force for movement.

#### Molecular Motors Convert Energy into Motion

At the nanoscale of cells and molecules, movement is effected by much different forces from those in the macroscopic world. For example, the high protein concentration (200–300 mg/ml) of the cytoplasm prevents organelles and vesicles from diffusing faster than 100  $\mu\text{m}^3$  hours. Even a micrometer-sized bacterium experiences a drag force from water that stops its forward movement within a fraction of a nanometer when it stops actively swimming. To generate the forces necessary for many cellular movements, cells depend on specialized enzymes commonly called **motor proteins**. These *mechanochemical enzymes* convert energy released by the hydrolysis of ATP or from ion gradients into a mechanical force.

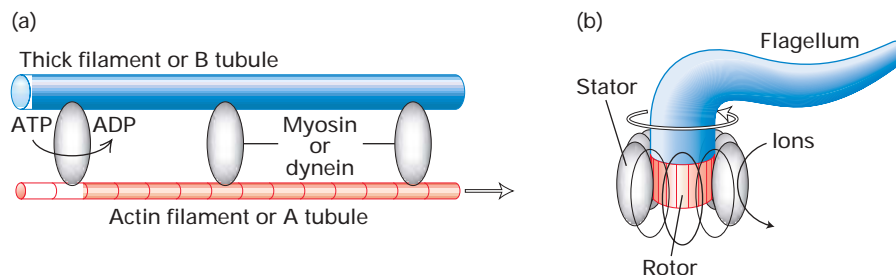
Motor proteins generate either linear or rotary motion (Table 3-2). Some motor proteins are components of macro-

molecular assemblies, but those that move along cytoskeletal fibers are not. This latter group comprises the **myosins**, **kinosins**, and **dyneins**—linear motor proteins that carry attached “cargo” with them as they proceed along either **microfilaments** or **microtubules** (Figure 3-22a). DNA and RNA polymerases also are linear motor proteins because they translocate along DNA during replication and transcription. In contrast, rotary motors revolve to cause the beat of bacterial flagella, to pack DNA into the capsid of a virus, and to synthesize ATP. The propulsive force for bacterial swimming, for instance, is generated by a rotary motor protein complex in the bacterial membrane. Ions flow down an electrochemical gradient through an immobile ring of proteins, the *stator*, which is located in the membrane. Torque generated by the stator rotates an inner ring of proteins and the attached flagellum (Figure 3-22b). Similarly, in the mitochondrial ATP synthase, or  $F_0F_1$  complex, a flux of ions across the inner mitochondrial membrane is transduced by the  $F_0$  part into rotation of the  $\gamma$  subunit, which projects into a surrounding ring of  $\alpha$  and  $\beta$  subunits in the  $F_1$  part. Interactions between the  $\gamma$  subunit and the  $\beta$  subunits directs the synthesis of ATP (Chapter 8).

From the observed activities of motor proteins, we can infer three general properties that they possess:

- The ability to transduce a source of energy, either ATP or an ion gradient, into linear or rotary movement
- The ability to bind and translocate along a cytoskeletal filament, nucleic acid strand, or protein complex
- Net movement in a given direction

The motor proteins that attach to cytoskeletal fibers also bind to and carry along cargo as they translocate. The cargo in muscle cells and eukaryotic flagella consists of thick filaments and B tubules, respectively (see Figure 3-22a). These motor proteins can also transport cargo chromosomes and membrane-limited vesicles as they move along microtubules or microfilaments (Figure 3-23).



▲ **FIGURE 3-22 Comparison of linear and rotary molecular motors.** (a) In muscle and eukaryotic flagella, the head domains of motor proteins (blue) bind to an actin thin filament (muscle) or the A tubule of a doublet microtubule (flagella). ATP hydrolysis in the head causes linear movement of the cytoskeletal fiber (orange) relative to the attached thick filament or B tubule of an

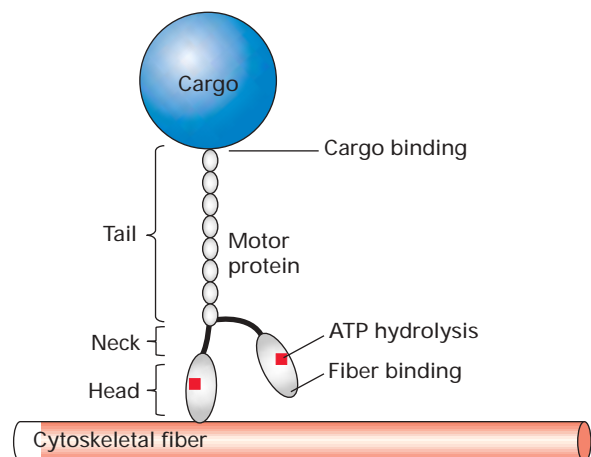
adjacent doublet microtubule. (b) In the rotary motor in the bacterial membrane, the stator (blue) is immobile in the membrane. Ion flow through the stator generates a torque that powers rotation of the rotor (orange) and the flagellum attached to it.

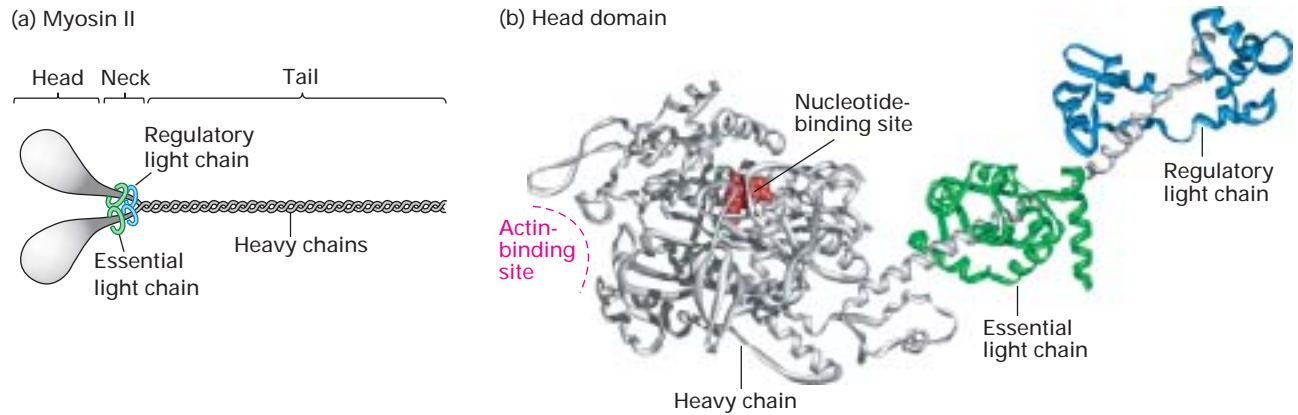
TABLE 3-2 Selected Molecular Motors

Motor*	Energy Source	Structure/Components	Cellular Location	Movement Generated
<b>LINEAR MOTORS</b>				
DNA polymerase (4)	ATP	Multisubunit polymerase $\delta$ within replisome	Nucleus	Translocation along DNA during replication
RNA polymerase (4)	ATP	Multisubunit polymerase within transcription elongation complex	Nucleus	Translocation along DNA during transcription
Ribosome (4)	GTP	Elongation factor 2 (EF2) bound to ribosome	Cytoplasm/ER membrane	Translocation along mRNA during translation
Myosins (3, 19)	ATP	Heavy and light chains; head domains with ATPase activity and microfilament-binding site	Cytoplasm	Transport of cargo vesicles; contraction
Kinesins (20)	ATP	Heavy and light chains; head domains with ATPase activity and microtubule-binding site	Cytoplasm	Transport of cargo vesicles and chromosomes during mitosis
Dyneins (20)	ATP	Multiple heavy, intermediate, and light chains; head domains with ATPase activity and microtubule-binding site	Cytoplasm	Transport of cargo vesicles; beating of cilia and eukaryotic flagella
<b>ROTARY MOTORS</b>				
Bacterial flagellar motor	$H^+/Na^+$ gradient	Stator and rotor proteins, flagellum	Plasma membrane	Rotation of flagellum attached to rotor
ATP synthase, $F_0F_1$ (8)	$H^+$ gradient	Multiple subunits forming $F_0$ and $F_1$ particles	Inner mitochondrial membrane, thylakoid membrane, bacterial plasma membrane	Rotation of $\gamma$ subunit leading to ATP synthesis
Viral capsid motor	ATP	Connector, prohead RNA, ATPase	Capsid	Rotation of connector leading to DNA packaging

\*Numbers in parentheses indicate chapters in which various motors are discussed.

► **FIGURE 3-23 Motor protein-dependent movement of cargo.** The head domains of myosin, dynein, and kinesin motor proteins bind to a cytoskeletal fiber (microfilaments or microtubules), and the tail domain attaches to one of various types of cargo—in this case, a membrane-limited vesicle. Hydrolysis of ATP in the head domain causes the head domain to “walk” along the track in one direction by a repeating cycle of conformational changes.





▲ **FIGURE 3-24 Structure of myosin II.** (a) Myosin II is a dimeric protein composed of two identical heavy chains (white) and four light chains (blue and green). Each of the head domains transduces the energy from ATP hydrolysis into movement. Two light chains are associated with the neck domain of each heavy chain. The coiled-coil sequence of the tail domain organizes myosin II into a dimer. (b) Three-dimensional model of a single

head domain shows that it has a curved, elongated shape and is bisected by a large cleft. The nucleotide-binding pocket lies on one side of this cleft, and the actin-binding site lies on the other side near the tip of the head. Wrapped around the shaft of the  $\alpha$ -helical neck are the two light chains. These chains stiffen the neck so that it can act as a lever arm for the head. Shown here is the ADP-bound conformation.

### All Myosins Have Head, Neck, and Tail Domains with Distinct Functions

To further illustrate the properties of motor proteins, we consider myosin II, which moves along actin filaments in muscle cells during contraction. Other types of myosin can transport vesicles along actin filaments in the cytoskeleton. Myosin II and other members of the myosin superfamily are composed of one or two heavy chains and several light chains. The heavy chains are organized into three structurally and functionally different types of domains (Figure 3-24a).

The two globular *head domains* are specialized ATPases that couple the hydrolysis of ATP with motion. A critical feature of the myosin ATPase activity is that it is *actin activated*. In the absence of actin, solutions of myosin slowly convert ATP into ADP and phosphate. However, when myosin is complexed with actin, the rate of myosin ATPase activity is four to five times as fast as it is in the absence of actin. The actin-activation step ensures that the myosin ATPase operates at its maximal rate only when the myosin head domain is bound to actin. Adjacent to the head domain lies the  $\alpha$ -helical *neck region*, which is associated with the light chains. These light chains are crucial for converting small conformational changes in the head into large movements of the molecule and for regulating the activity of the head domain. The rodlike *tail domain* contains the binding sites that determine the specific activities of a particular myosin.

The results of studies of myosin fragments produced by proteolysis helped elucidate the functions of the domains. X-ray crystallographic analysis of the S1 fragment of myosin II, which consists of the head and neck domains, revealed its shape, the positions of the light chains, and the locations of the ATP-binding and actin-binding sites. The elongated myosin head is attached at one end to the  $\alpha$ -helical neck (Figure 3-24b). Two light-chain molecules lie at the base of the

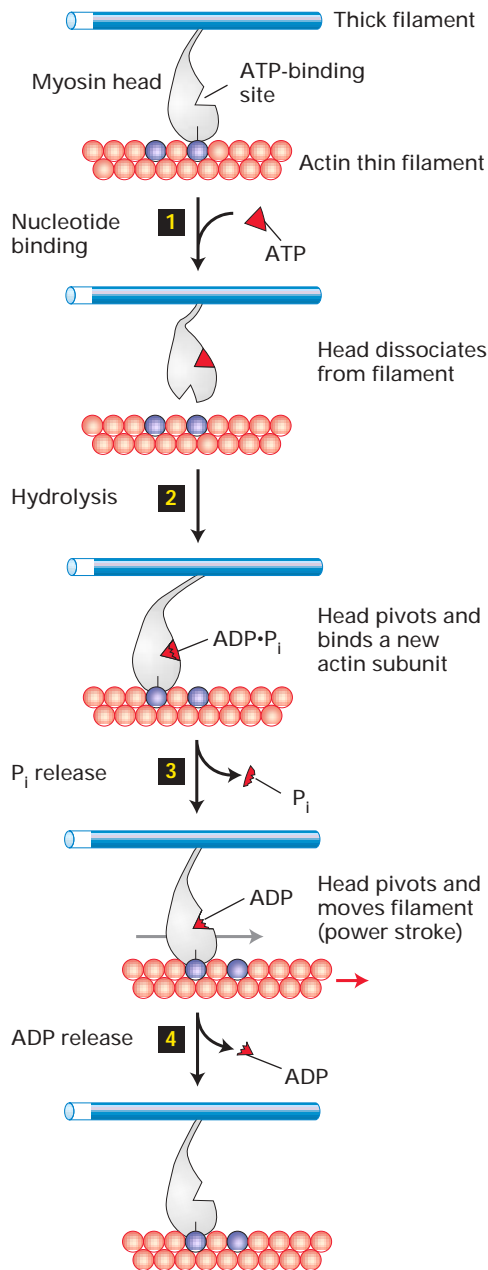
head, wrapped around the neck like C-clamps. In this position, the light chains stiffen the neck region and are therefore able to regulate the activity of the head domain.

### Conformational Changes in the Myosin Head Couple ATP Hydrolysis to Movement

The results of studies of muscle contraction provided the first evidence that myosin heads slide or walk along actin filaments. Unraveling the mechanism of muscle contraction was greatly aided by the development of *in vitro* motility assays and single-molecule force measurements. On the basis of information obtained with these techniques and the three-dimensional structure of the myosin head, researchers developed a general model for how myosin harnesses the energy released by ATP hydrolysis to move along an actin filament. Because all myosins are thought to use the same mechanism to generate movement, we will ignore whether the myosin tail is bound to a vesicle or is part of a thick filament as it is in muscle. One assumption in this model is that the hydrolysis of a single ATP molecule is coupled to each step taken by a myosin molecule along an actin filament. Evidence supporting this assumption is discussed in Chapter 19.

As shown in Figure 3-25, myosin undergoes a series of events during each step of movement. In the course of one cycle, myosin must exist in at least three conformational states: an ATP state unbound to actin, an ADP- $P_i$  state bound to actin, and a state after the power-generating stroke has been completed. The major question is how the nucleotide-binding pocket and the distant actin-binding site are mutually influenced and how changes at these sites are converted into force. The results of structural studies of myosin in the presence of nucleotides and nucleotide analogs that mimic the various steps in the cycle indicate that the binding and hydrolysis of a nucleotide cause a





◀ **FIGURE 3-25 Operational model for the coupling of ATP hydrolysis to movement of myosin along an actin filament.**

Shown here is the cycle for a myosin II head that is part of a thick filament in muscle, but other myosins that attach to other cargo (e.g., the membrane of a vesicle) are thought to operate according to the same cyclical mechanism. In the absence of bound nucleotide, a myosin head binds actin tightly in a “rigor” state. Step **1**: Binding of ATP opens the cleft in the myosin head, disrupting the actin-binding site and weakening the interaction with actin. Step **2**: Freed of actin, the myosin head hydrolyzes ATP, causing a conformational change in the head that moves it to a new position, closer to the (+) end of the actin filament, where it rebinds to the filament. Step **3**: As phosphate (P<sub>i</sub>) dissociates from the ATP-binding pocket, the myosin head undergoes a second conformational change—the power stroke—which restores myosin to its rigor conformation. Because myosin is bound to actin, this conformational change exerts a force that causes myosin to move the actin filament. Step **4**: Release of ADP completes the cycle. [Adapted from R. D. Vale and R. A. Milligan, 2002, *Science* **288**:88.]

## KEY CONCEPTS OF SECTION 3.4

### Molecular Motors and the Mechanical Work of Cells

- Motor proteins are mechanochemical enzymes that convert energy released by ATP hydrolysis into either linear or rotary movement (see Figure 3-22).
- Linear motor proteins (myosins, kinesins, and dyneins) move along cytoskeletal fibers carrying bound cargo, which includes vesicles, chromosomes, thick filaments in muscle, and microtubules in eukaryotic flagella.
- Myosin II consists of two heavy chains and several light chains. Each heavy chain has a head (motor) domain, which is an actin-activated ATPase; a neck domain, which is associated with light chains; and a long rodlike tail domain that organizes the dimeric molecule and binds to thick filaments in muscle cells (see Figure 3-24).
- Movement of myosin relative to an actin filament results from the attachment of the myosin head to an actin filament, rotation of the neck region, and detachment in a cyclical ATP-dependent process (see Figure 3-25). The same general mechanism is thought to account for all myosin- and kinesin-mediated movement.

small conformational change in the head domain that is amplified into a large movement of the neck region. The small conformational change in the head domain is localized to a “switch” region consisting of the nucleotide- and actin-binding sites. A “converter” region at the base of the head acts like a fulcrum that causes the leverlike neck to bend and rotate.

Homologous switch, converter, and lever arm structures in kinesin are responsible for the movement of kinesin motor proteins along microtubules. The structural basis for dynein movement is unknown because the three-dimensional structure of dynein has not been determined.

## 3.5 Common Mechanisms for Regulating Protein Function

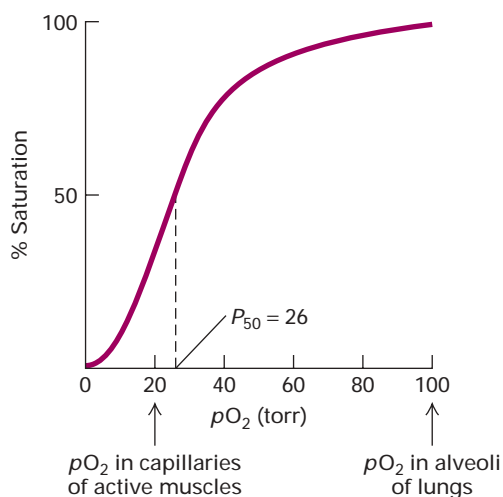
Most processes in cells do not take place independently of one another or at a constant rate. Instead, the catalytic activity of enzymes or the assembly of a macromolecular complex is so regulated that the amount of reaction product or the appearance of the complex is just sufficient to meet the needs of the cell. As a result, the steady-state concentrations

of substrates and products will vary, depending on cellular conditions. The flow of material in an enzymatic pathway is controlled by several mechanisms, some of which also regulate the functions of nonenzymatic proteins.

One of the most important mechanisms for regulating protein function entails **allostery**. Broadly speaking, allostery refers to any change in a protein's tertiary or quaternary structure or both induced by the binding of a ligand, which may be an activator, inhibitor, substrate, or all three. Allosteric regulation is particularly prevalent in multimeric enzymes and other proteins. We first explore several ways in which allostery influences protein function and then consider other mechanisms for regulating proteins.

### Cooperative Binding Increases a Protein's Response to Small Changes in Ligand Concentration

In many cases, especially when a protein binds several molecules of one ligand, the binding is graded; that is, the binding of one ligand molecule affects the binding of subsequent ligand molecules. This type of allostery, often called **cooper-**



▲ **EXPERIMENTAL FIGURE 3-26 Sequential binding of oxygen to hemoglobin exhibits positive cooperativity.** Each hemoglobin molecule has four oxygen-binding sites; at saturation all the sites are loaded with oxygen. The oxygen concentration is commonly measured as the partial pressure ( $pO_2$ ).  $P_{50}$  is the  $pO_2$  at which half the oxygen-binding sites at a given hemoglobin concentration are occupied; it is equivalent to the  $K_m$  for an enzymatic reaction. The large change in the amount of oxygen bound over a small range of  $pO_2$  values permits efficient unloading of oxygen in peripheral tissues such as muscle. The sigmoidal shape of a plot of percent saturation versus ligand concentration is indicative of cooperative binding. In the absence of cooperative binding, a binding curve is a hyperbola, similar to the simple kinetic curves in Figure 3-19. [Adapted from L. Stryer, *Biochemistry*, 4th ed., 1995, W. H. Freeman and Company.]

**activity**, permits many multisubunit proteins to respond more efficiently to small changes in ligand concentration than would otherwise be possible. In positive cooperativity, sequential binding is enhanced; in negative cooperativity, sequential binding is inhibited.

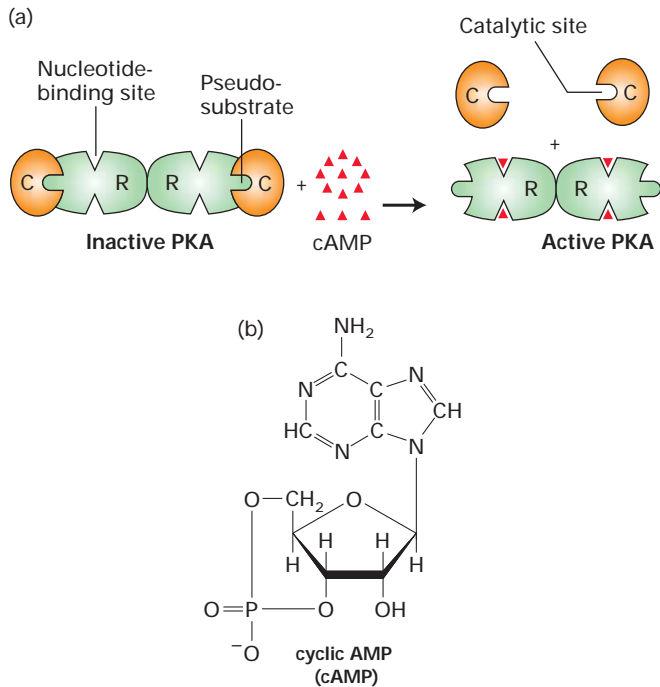
Hemoglobin presents a classic example of positive cooperative binding. Each of the four subunits in hemoglobin contains one heme molecule, which consists of an iron atom held within a porphyrin ring (see Figure 8-16a). The heme groups are the oxygen-binding components of hemoglobin (see Figure 3-10). The binding of oxygen to the heme molecule in one of the four hemoglobin subunits induces a local conformational change whose effect spreads to the other subunits, lowering the  $K_m$  for the binding of additional oxygen molecules and yielding a sigmoidal oxygen-binding curve (Figure 3-26). Consequently, the sequential binding of oxygen is facilitated, permitting hemoglobin to load more oxygen in peripheral tissues than it otherwise could at normal oxygen concentrations.

### Ligand Binding Can Induce Allosteric Release of Catalytic Subunits or Transition to a State with Different Activity

Previously, we looked at protein kinase A to illustrate binding and catalysis by the active site of an enzyme. This enzyme can exist as an inactive tetrameric protein composed of two catalytic subunits and two regulatory subunits. Each regulatory subunit contains a *pseudosubstrate* sequence that binds to the active site in a catalytic subunit. By blocking substrate binding, the regulatory subunit inhibits the activity of the catalytic subunit.

Inactive protein kinase A is turned on by **cyclic AMP (cAMP)**, a small second-messenger molecule. The binding of cAMP to the regulatory subunits induces a conformational change in the pseudosubstrate sequence so that it can no longer bind the catalytic subunit. Thus, in the presence of cAMP, the inactive tetramer dissociates into two monomeric active catalytic subunits and a dimeric regulatory subunit (Figure 3-27). As discussed in Chapter 13, the binding of various hormones to cell-surface receptors induces a rise in the intracellular concentration of cAMP, leading to the activation of protein kinase A. When the signaling ceases and the cAMP level decreases, the activity of protein kinase A is turned off by reassembly of the inactive tetramer. The binding of cAMP to the regulatory subunits exhibits positive cooperativity; thus small changes in the concentration of this allosteric molecule produce a large change in the activity of protein kinase A.

Many multimeric enzymes undergo allosteric transitions that alter the relation of the subunits to one another but do not cause dissociation as in protein kinase A. In this type of allostery, the activity of a protein in the ligand-bound state differs from that in the unbound state. An example is the GroEL chaperonin discussed earlier. This barrel-shaped



▲ FIGURE 3-27 **Ligand-induced activation of protein kinase A (PKA).** At low concentrations of cyclic AMP (cAMP), the PKA is an inactive tetramer. Binding of cAMP to the regulatory (R) subunits causes a conformational change in these subunits that permits release of the active, monomeric catalytic (C) subunits. (b) Cyclic AMP is a derivative of adenosine monophosphate. This intracellular signaling molecule, whose concentration rises in response to various extracellular signals, can modulate the activity of many proteins.

protein-folding machine comprises two back-to-back multi-subunit rings, which can exist in a “tight” peptide-binding state and a “relaxed” peptide-releasing state (see Figure 3-11). The binding of ATP and the co-chaperonin GroES to one of the rings in the tight state causes a twofold expansion of the GroEL cavity, shifting the equilibrium toward the relaxed peptide-folding state.

### Calcium and GTP Are Widely Used to Modulate Protein Activity

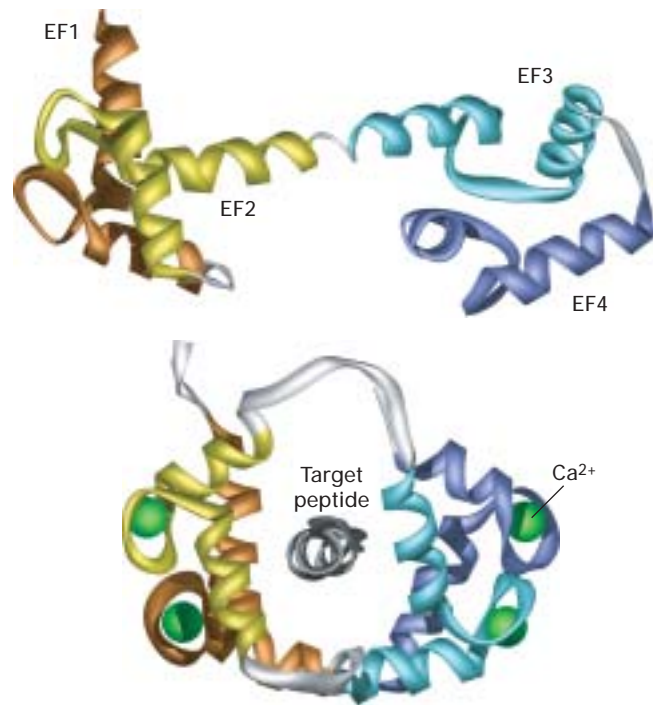
In the preceding examples, oxygen, cAMP, and ATP cause allosteric changes in the activity of their target proteins (hemoglobin, protein kinase A, and GroEL, respectively). Two additional allosteric ligands, Ca<sup>2+</sup> and GTP, act through two types of ubiquitous proteins to regulate many cellular processes.

**Calmodulin-Mediated Switching** The concentration of Ca<sup>2+</sup> free in the cytosol is kept very low ( $\approx 10^{-7}$  M) by membrane transport proteins that continually pump Ca<sup>2+</sup> out of the cell or into the endoplasmic reticulum. As we learn in Chapter 7, the cytosolic Ca<sup>2+</sup> level can increase from 10- to

100-fold by the release of Ca<sup>2+</sup> from ER stores or by its import from the extracellular environment. This rise in cytosolic Ca<sup>2+</sup> is sensed by Ca<sup>2+</sup>-binding proteins, particularly those of the *EF hand family*, all of which contain the helix-loop-helix motif discussed earlier (see Figure 3-6a).

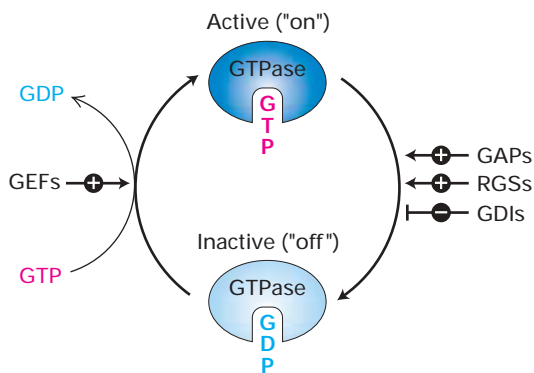
The prototype EF hand protein, **calmodulin**, is found in all eukaryotic cells and may exist as an individual monomeric protein or as a subunit of a multimeric protein. A dumbbell-shaped molecule, calmodulin contains four Ca<sup>2+</sup>-binding sites with a  $K_D$  of  $\approx 10^{-6}$  M. The binding of Ca<sup>2+</sup> to calmodulin causes a conformational change that permits Ca<sup>2+</sup>/calmodulin to bind various target proteins, thereby switching their activity on or off (Figure 3-28). Calmodulin and similar EF hand proteins thus function as *switch proteins*, acting in concert with Ca<sup>2+</sup> to modulate the activity of other proteins.

**Switching Mediated by Guanine Nucleotide-Binding Proteins** Another group of intracellular switch proteins constitutes the **GTPase superfamily**. These proteins include monomeric **Ras protein** (see Figure 3-5) and the G<sub>α</sub> subunit of the trimeric **G proteins**. Both Ras and G<sub>α</sub> are bound to the plasma membrane, function in cell signaling, and play a key role in cell proliferation and differentiation. Other members



▲ FIGURE 3-28 **Switching mediated by Ca<sup>2+</sup>/calmodulin.** Calmodulin is a widely distributed cytosolic protein that contains four Ca<sup>2+</sup>-binding sites, one in each of its EF hands. Each EF hand has a helix-loop-helix motif. At cytosolic Ca<sup>2+</sup> concentrations above about  $5 \times 10^{-7}$  M, binding of Ca<sup>2+</sup> to calmodulin changes the protein's conformation. The resulting Ca<sup>2+</sup>/calmodulin wraps around exposed helices of various target proteins, thereby altering their activity.





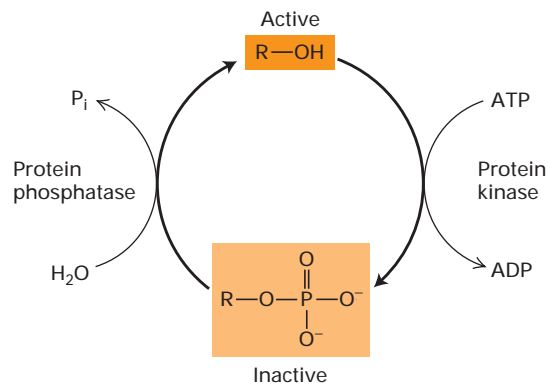
▲ FIGURE 3-29 **Cycling of GTPase switch proteins between the active and inactive forms.** Conversion of the active into the inactive form by hydrolysis of the bound GTP is accelerated by GAPs (GTPase-accelerating proteins) and RGSs (regulators of G protein–signaling) and inhibited by GDIs (guanine nucleotide dissociation inhibitors). Reactivation is promoted by GEFs (guanine nucleotide–exchange factors).

of the GTPase superfamily function in protein synthesis, the transport of proteins between the nucleus and the cytoplasm, the formation of coated vesicles and their fusion with target membranes, and rearrangements of the actin cytoskeleton.

All the GTPase switch proteins exist in two forms (Figure 3-29): (1) an active (“on”) form with bound GTP (guanosine triphosphate) that modulates the activity of specific target proteins and (2) an inactive (“off”) form with bound GDP (guanosine diphosphate). The GTPase activity of these switch proteins hydrolyzes bound GTP to GDP slowly, yielding the inactive form. The subsequent exchange of GDP with GTP to regenerate the active form occurs even more slowly. Activation is temporary and is enhanced or depressed by other proteins acting as allosteric regulators of the switch protein. We examine the role of various GTPase switch proteins in regulating intracellular signaling and other processes in several later chapters.

### Cyclic Protein Phosphorylation and Dephosphorylation Regulate Many Cellular Functions

As noted earlier, one of the most common mechanisms for regulating protein activity is phosphorylation, the addition and removal of phosphate groups from serine, threonine, or tyrosine residues. Protein **kinases** catalyze phosphorylation, and **phosphatases** catalyze dephosphorylation. Although both reactions are essentially irreversible, the counteracting activities of kinases and phosphatases provide cells with a “switch” that can turn on or turn off the function of various proteins (Figure 3-30). Phosphorylation changes a protein’s charge and generally leads to a conformational change; these effects can significantly alter ligand binding by a protein, leading to an increase or decrease in its activity.



▲ FIGURE 3-30 **Regulation of protein activity by kinase/phosphatase switch.** The cyclic phosphorylation and dephosphorylation of a protein is a common cellular mechanism for regulating protein activity. In this example, the target protein R is inactive (light orange) when phosphorylated and active (dark orange) when dephosphorylated; some proteins have the opposite pattern.

Nearly 3 percent of all yeast proteins are protein kinases or phosphatases, indicating the importance of phosphorylation and dephosphorylation reactions even in simple cells. All classes of proteins—including structural proteins, enzymes, membrane channels, and signaling molecules—are regulated by kinase/phosphatase switches. Different protein kinases and phosphatases are specific for different target proteins and can thus regulate a variety of cellular pathways, as discussed in later chapters. Some of these enzymes act on one or a few target proteins, whereas others have multiple targets. The latter are useful in integrating the activities of proteins that are coordinately controlled by a single kinase/phosphatase switch. Frequently, another kinase or phosphatase is a target, thus creating a web of interdependent controls.

### Proteolytic Cleavage Irreversibly Activates or Inactivates Some Proteins

The regulatory mechanisms discussed so far act as switches, reversibly turning proteins on and off. The regulation of some proteins is by a distinctly different mechanism: the irreversible activation or inactivation of protein function by proteolytic cleavage. This mechanism is most common in regard to some hormones (e.g., insulin) and digestive proteases. Good examples of such enzymes are trypsin and chymotrypsin, which are synthesized in the pancreas and secreted into the small intestine as the inactive zymogens *trypsinogen* and *chymotrypsinogen*, respectively. Enterokinase, an aminopeptidase secreted from cells lining the small intestine, converts trypsinogen into trypsin, which in turn cleaves chymotrypsinogen to form chymotrypsin. The delay in the activation of these proteases until they reach the intestine prevents them from digesting the pancreatic tissue in which they are made.

## Higher-Order Regulation Includes Control of Protein Location and Concentration

The activities of proteins are extensively regulated in order that the numerous proteins in a cell can work together harmoniously. For example, all metabolic pathways are closely controlled at all times. Synthetic reactions take place when the products of these reactions are needed; degradative reactions take place when molecules must be broken down. All the regulatory mechanisms heretofore described affect a protein locally at its site of action, turning its activity on or off.

Normal functioning of a cell, however, also requires the segregation of proteins to particular compartments such as the mitochondria, nucleus, and lysosomes. In regard to enzymes, compartmentation not only provides an opportunity for controlling the delivery of substrate or the exit of product but also permits competing reactions to take place simultaneously in different parts of a cell. We describe the mechanisms that cells use to direct various proteins to different compartments in Chapters 16 and 17.

In addition to compartmentation, cellular processes are regulated by protein synthesis and degradation. For example, proteins are often synthesized at low rates when a cell has little or no need for their activities. When the cell faces increased demand (e.g., appearance of substrate in the case of enzymes, stimulation of B lymphocytes by antigen), the cell responds by synthesizing new protein molecules. Later, the protein pool is lowered when levels of substrate decrease or the cell becomes inactive. Extracellular signals are often instrumental in inducing changes in the rates of protein synthesis and degradation (Chapters 13–15). Such regulated changes play a key role in the cell cycle (Chapter 21) and in cell differentiation (Chapter 22).

### KEY CONCEPTS OF SECTION 3.5

#### Common Mechanisms for Regulating Protein Function

- In allostery, the binding of one ligand molecule (a substrate, activator, or inhibitor) induces a conformational change, or allosteric transition, that alters a protein's activity or affinity for other ligands.
- In multimeric proteins, such as hemoglobin, that bind multiple ligand molecules, the binding of one ligand molecule may modulate the binding affinity for subsequent ligand molecules. Enzymes that cooperatively bind substrates exhibit sigmoidal kinetics similar to the oxygen-binding curve of hemoglobin (see Figure 3-26).
- Several allosteric mechanisms act as switches, turning protein activity on and off in a reversible fashion.
- The binding of allosteric ligand molecules may lead to the conversion of a protein from one conformational/

activity state into another or to the release of active subunits (see Figure 3-27).

- Two classes of intracellular switch proteins regulate a variety of cellular processes: (1) calmodulin and related  $\text{Ca}^{2+}$ -binding proteins in the EF hand family and (2) members of the GTPase superfamily (e.g., Ras and  $\text{G}_\alpha$ ), which cycle between active GTP-bound and inactive GDP-bound forms (see Figure 3-29).
- The phosphorylation and dephosphorylation of amino acid side chains by protein kinases and phosphatases provide reversible on/off regulation of numerous proteins.
- Nonallosteric mechanisms for regulating protein activity include proteolytic cleavage, which irreversibly converts inactive zymogens into active enzymes, compartmentation of proteins, and signal-induced modulation of protein synthesis and degradation.

## 3.6 Purifying, Detecting, and Characterizing Proteins

A protein must be purified before its structure and the mechanism of its action can be studied. However, because proteins vary in size, charge, and water solubility, no single method can be used to isolate all proteins. To isolate one particular protein from the estimated 10,000 different proteins in a cell is a daunting task that requires methods both for separating proteins and for detecting the presence of specific proteins.

Any molecule, whether protein, carbohydrate, or nucleic acid, can be separated, or *resolved*, from other molecules on the basis of their differences in one or more physical or chemical characteristics. The larger and more numerous the differences between two proteins, the easier and more efficient their separation. The two most widely used characteristics for separating proteins are *size*, defined as either length or mass, and *binding affinity* for specific ligands. In this section, we briefly outline several important techniques for separating proteins; these techniques are also useful for the separation of nucleic acids and other biomolecules. (Specialized methods for removing membrane proteins from membranes are described in the next chapter after the unique properties of these proteins are discussed.) We then consider general methods for detecting, or *assaying*, specific proteins, including the use of radioactive compounds for tracking biological activity. Finally, we consider several techniques for characterizing a protein's mass, sequence, and three-dimensional structure.

### Centrifugation Can Separate Particles and Molecules That Differ in Mass or Density

The first step in a typical protein purification scheme is centrifugation. The principle behind centrifugation is that

two particles in suspension (cells, organelles, or molecules) with different masses or densities will settle to the bottom of a tube at different rates. Remember, *mass* is the weight of a sample (measured in grams), whereas *density* is the ratio of its weight to volume (grams/liter). Proteins vary greatly in mass but not in density. Unless a protein has an attached lipid or carbohydrate, its density will not vary by more than 15 percent from 1.37 g/cm<sup>3</sup>, the average protein density. Heavier or more dense molecules settle, or sediment, more quickly than lighter or less dense molecules.

A centrifuge speeds sedimentation by subjecting particles in suspension to centrifugal forces as great as 1,000,000 times the force of gravity *g*, which can sediment particles as small as 10 kDa. Modern ultracentrifuges achieve these forces by reaching speeds of 150,000 revolutions per minute (rpm) or greater. However, small particles with masses of 5 kDa or less will not sediment uniformly even at such high rotor speeds.

Centrifugation is used for two basic purposes: (1) as a preparative technique to separate one type of material from others and (2) as an analytical technique to measure physical properties (e.g., molecular weight, density, shape, and equilibrium binding constants) of macromolecules. The *sedimentation constant*, *s*, of a protein is a measure of its sedimentation rate. The sedimentation constant is commonly expressed in svedbergs (S): 1 S = 10<sup>-13</sup> seconds.

**Differential Centrifugation** The most common initial step in protein purification is the separation of soluble proteins from insoluble cellular material by *differential centrifugation*. A starting mixture, commonly a cell homogenate, is poured into a tube and spun at a rotor speed and for a period of time that forces cell organelles such as nuclei to collect as a pellet at the bottom; the soluble proteins remain in the supernatant (Figure 3-31a). The supernatant fraction then is poured off and can be subjected to other purification methods to separate the many different proteins that it contains.

**Rate-Zonal Centrifugation** On the basis of differences in their masses, proteins can be separated by centrifugation through a solution of increasing density called a *density gradient*. A concentrated sucrose solution is commonly used to form density gradients. When a protein mixture is layered on top of a sucrose gradient in a tube and subjected to centrifugation, each protein in the mixture migrates down the tube at a rate controlled by the factors that affect the sedimentation constant. All the proteins start from a thin zone at the top of the tube and separate into bands, or zones (actually disks), of proteins of different masses. In this separation technique, called *rate-zonal centrifugation*, samples are centrifuged just long enough to separate the molecules of interest into discrete zones (Figure 3-31b). If a sample is centrifuged for too short a time, the different protein molecules will not separate sufficiently. If a sample is centrifuged much

longer than necessary, all the proteins will end up in a pellet at the bottom of the tube.

Although the sedimentation rate is strongly influenced by particle mass, rate-zonal centrifugation is seldom effective in determining *precise* molecular weights because variations in shape also affect sedimentation rate. The exact effects of shape are hard to assess, especially for proteins and single-stranded nucleic acid molecules that can assume many complex shapes. Nevertheless, rate-zonal centrifugation has proved to be the most practical method for separating many different types of polymers and particles. A second density-gradient technique, called *equilibrium density-gradient centrifugation*, is used mainly to separate DNA or organelles (see Figure 5-37).

### Electrophoresis Separates Molecules on the Basis of Their Charge : Mass Ratio

**Electrophoresis** is a technique for separating molecules in a mixture under the influence of an applied electric field. Dissolved molecules in an electric field move, or migrate, at a speed determined by their charge:mass ratio. For example, if two molecules have the same mass and shape, the one with the greater net charge will move faster toward an electrode.

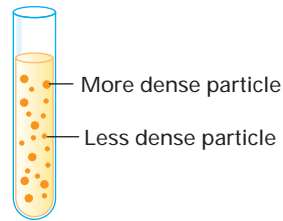
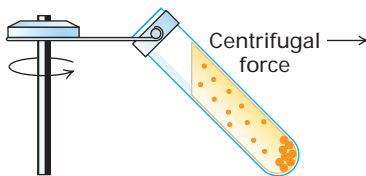
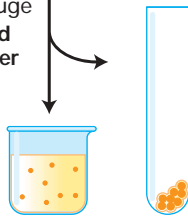
**SDS-Polyacrylamide Gel Electrophoresis** Because many proteins or nucleic acids that differ in size and shape have nearly identical charge:mass ratios, electrophoresis of these macromolecules in solution results in little or no separation of molecules of different lengths. However, successful separation of proteins and nucleic acids can be accomplished by electrophoresis in various gels (semisolid suspensions in water) rather than in a liquid solution. Electrophoretic separation of proteins is most commonly performed in *polyacrylamide gels*. When a mixture of proteins is applied to a gel and an electric current is applied, smaller proteins migrate faster through the gel than do larger proteins.

Gels are cast between a pair of glass plates by polymerizing a solution of acrylamide monomers into polyacrylamide chains and simultaneously cross-linking the chains into a semisolid matrix. The *pore size* of a gel can be varied by adjusting the concentrations of polyacrylamide and the cross-linking reagent. The rate at which a protein moves through a gel is influenced by the gel's pore size and the strength of the electric field. By suitable adjustment of these parameters, proteins of widely varying sizes can be separated.

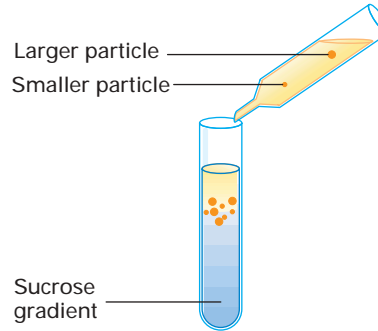
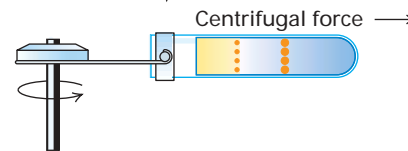
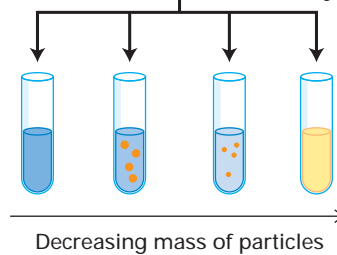
In the most powerful technique for resolving protein mixtures, proteins are exposed to the ionic detergent SDS (sodium dodecylsulfate) before and during gel electrophoresis (Figure 3-32). SDS denatures proteins, causing multimeric proteins to dissociate into their subunits, and all polypeptide chains are forced into extended conformations with similar charge:mass ratios. SDS treatment thus



(a) Differential centrifugation

**1** Sample is poured into tube**2** Centrifuge  
Particles settle  
according to  
mass**3** Stop centrifuge  
Decant liquid  
into container

(b) Rate-zonal centrifugation

**1** Sample is layered on top of gradient**2** Centrifuge  
Particles settle  
according to  
mass**3** Stop centrifuge  
Collect fractions  
and do assay

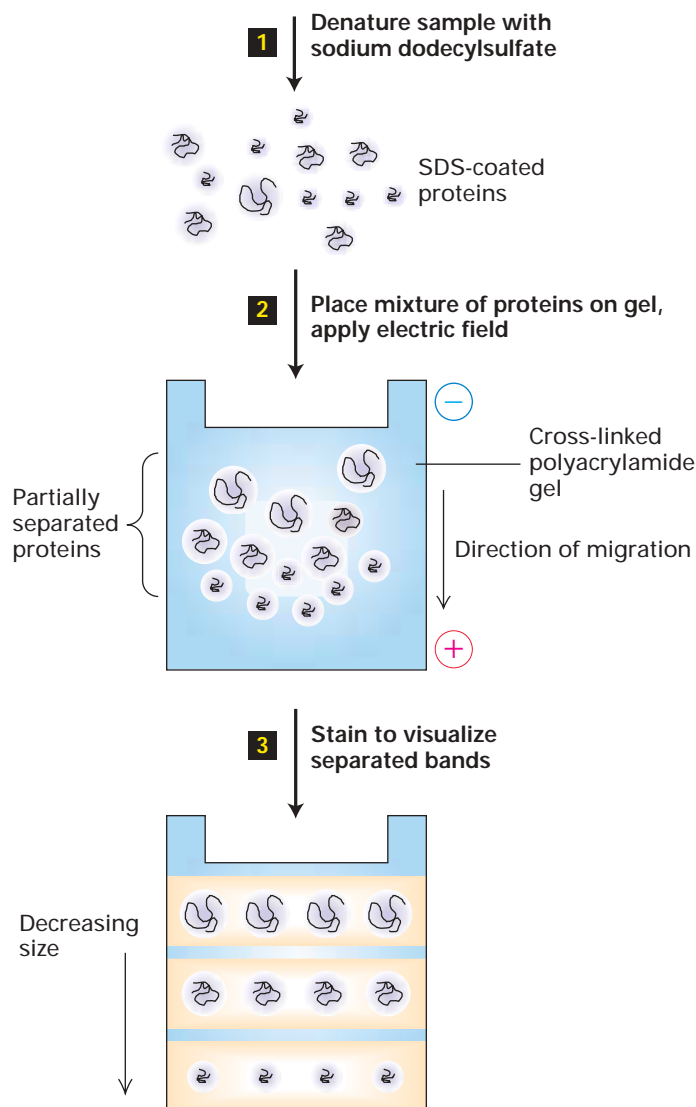
▲ **EXPERIMENTAL FIGURE 3-31 Centrifugation techniques separate particles that differ in mass or density.**

(a) In differential centrifugation, a cell homogenate or other mixture is spun long enough to sediment the denser particles (e.g., cell organelles, cells), which collect as a pellet at the bottom of the tube (step **2**). The less dense particles (e.g., soluble proteins, nucleic acids) remain in the liquid supernatant, which can be

transferred to another tube (step **3**). (b) In rate-zonal centrifugation, a mixture is spun just long enough to separate molecules that differ in mass but may be similar in shape and density (e.g., globular proteins, RNA molecules) into discrete zones within a density gradient commonly formed by a concentrated sucrose solution (step **2**). Fractions are removed from the bottom of the tube and assayed (step **3**).

eliminates the effect of differences in shape, and so chain length, which corresponds to mass, is the sole determinant of the migration rate of proteins in SDS-polyacrylamide electrophoresis. Even chains that differ in molecular weight by less than 10 percent can be separated by this technique. Moreover, the molecular weight of a protein can be estimated by comparing the distance that it migrates through a gel with the distances that proteins of known molecular weight migrate.

**Two-Dimensional Gel Electrophoresis** Electrophoresis of all cellular proteins through an SDS gel can separate proteins having relatively large differences in mass but cannot resolve proteins having similar masses (e.g., a 41-kDa protein from a 42-kDa protein). To separate proteins of similar masses, another physical characteristic must be exploited. Most commonly, this characteristic is electric charge, which is determined by the number of acidic and basic residues in a protein. Two unrelated proteins having similar masses are



◀ **EXPERIMENTAL FIGURE 3-32 SDS-polyacrylamide gel electrophoresis separates proteins solely on the basis of their masses.**

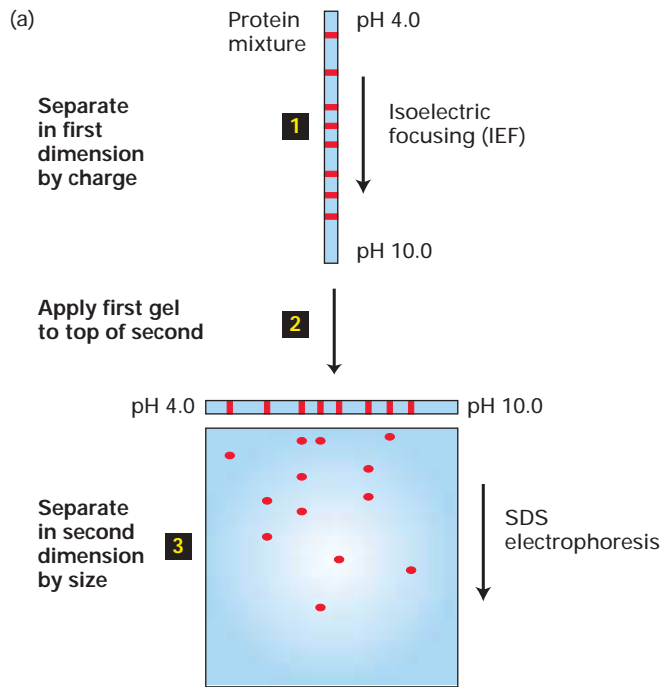
Initial treatment with SDS, a negatively charged detergent, dissociates multimeric proteins and denatures all the polypeptide chains (step **1**). During electrophoresis, the SDS-protein complexes migrate through the polyacrylamide gel (step **2**). Small proteins are able to move through the pores more easily, and faster, than larger proteins. Thus the proteins separate into bands according to their sizes as they migrate through the gel. The separated protein bands are visualized by staining with a dye (step **3**).

unlikely to have identical net charges because their sequences, and thus the number of acidic and basic residues, are different.

In two-dimensional electrophoresis, proteins are separated sequentially, first by their charges and then by their masses (Figure 3-33a). In the first step, a cell extract is fully denatured by high concentrations (8 M) of urea and then layered on a gel strip that contains a continuous pH gradient. The gradient is formed by *ampholytes*, a mixture of polyanionic and polycationic molecules, that are cast into the gel, with the most acidic ampholyte at one end and the most basic ampholyte at the opposite end. A charged protein will migrate through the gradient until it reaches its **isoelectric point (pI)**, the pH at which the net charge of the protein is zero. This technique, called **iso-**

**electric focusing (IEF)**, can resolve proteins that differ by only one charge unit. Proteins that have been separated on an IEF gel can then be separated in a second dimension on the basis of their molecular weights. To accomplish this separation, the IEF gel strip is placed lengthwise on a polyacrylamide slab gel, this time saturated with SDS. When an electric field is imposed, the proteins will migrate from the IEF gel into the SDS slab gel and then separate according to their masses.

The sequential resolution of proteins by charge and mass can achieve excellent separation of cellular proteins (Figure 3-33b). For example, two-dimensional gels have been very useful in comparing the proteomes in undifferentiated and differentiated cells or in normal and cancer cells because as many as 1000 proteins can be resolved simultaneously.



▲ **EXPERIMENTAL FIGURE 3-33 Two-dimensional gel electrophoresis can separate proteins of similar mass.** (a) In this technique, proteins are first separated on the basis of their charges by isoelectric focusing (step 1). The resulting gel strip is applied to an SDS-polyacrylamide gel and the proteins are separated into bands by mass (step 3). (b) In this two-

dimensional gel of a protein extract from cultured cells, each spot represents a single polypeptide. Polypeptides can be detected by dyes, as here, or by other techniques such as autoradiography. Each polypeptide is characterized by its isoelectric point (pI) and molecular weight. [Part (b) courtesy of J. Cells.]

## Liquid Chromatography Resolves Proteins by Mass, Charge, or Binding Affinity

A third common technique for separating mixtures of proteins, as well as other molecules, is based on the principle that molecules dissolved in a solution will interact (bind and dissociate) with a solid surface. If the solution is allowed to flow across the surface, then molecules that interact frequently with the surface will spend more time bound to the surface and thus move more slowly than molecules that interact infrequently with the surface. In this technique, called **liquid chromatography**, the sample is placed on top of a tightly packed column of spherical beads held within a glass cylinder. The nature of these beads determines whether the separation of proteins depends on differences in mass, charge, or binding affinity.

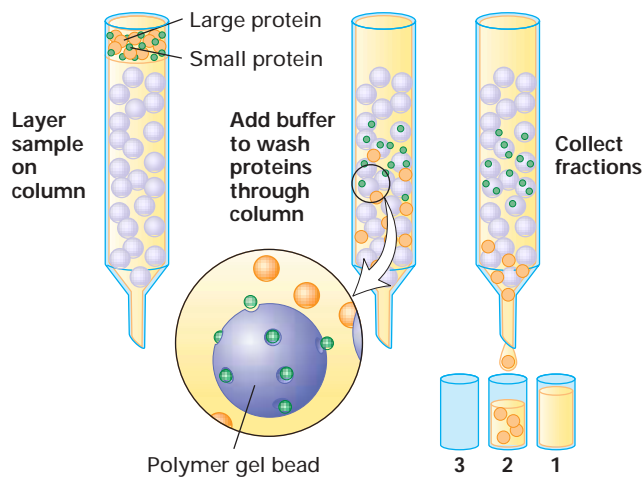
**Gel Filtration Chromatography** Proteins that differ in mass can be separated on a column composed of porous beads made from polyacrylamide, dextran (a bacterial polysaccharide), or agarose (a seaweed derivative), a technique called *gel filtration chromatography*. Although proteins *flow around* the spherical beads in gel filtration chromatography, they spend some time within the large depressions that cover a bead's surface. Because smaller proteins can penetrate into these depres-

sions more easily than can larger proteins, they travel through a gel filtration column more slowly than do larger proteins (Figure 3-34a). (In contrast, proteins migrate *through* the pores in an electrophoretic gel; thus smaller proteins move faster than larger ones.) The total volume of liquid required to elute a protein from a gel filtration column depends on its mass: the smaller the mass, the greater the elution volume. By use of proteins of known mass, the elution volume can be used to estimate the mass of a protein in a mixture.

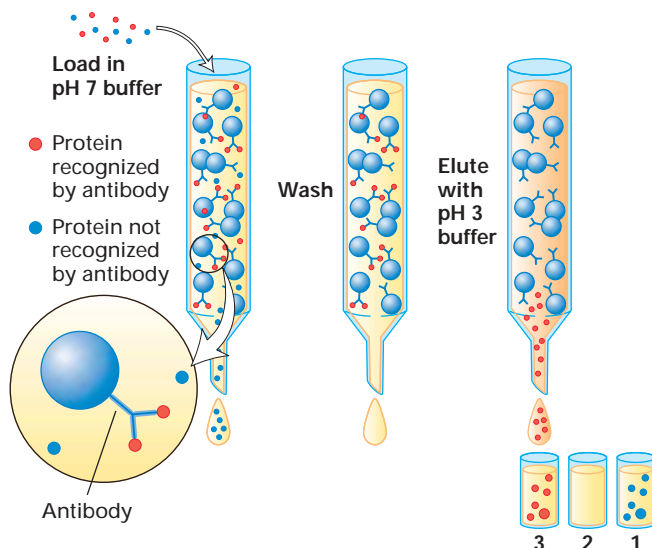
**Ion-Exchange Chromatography** In a second type of liquid chromatography, called *ion-exchange chromatography*, proteins are separated on the basis of differences in their charges. This technique makes use of specially modified beads whose surfaces are covered by amino groups or carboxyl groups and thus carry either a positive charge ( $\text{NH}_3^+$ ) or a negative charge ( $\text{COO}^-$ ) at neutral pH.

The proteins in a mixture carry various net charges at any given pH. When a solution of a protein mixture flows through a column of positively charged beads, only proteins with a net negative charge (acidic proteins) adhere to the beads; neutral and positively charged (basic) proteins flow unimpeded through the column (Figure 3-34b). The acidic proteins are then eluted selectively by passing a gradient of increasing concentrations of salt through the column. At low

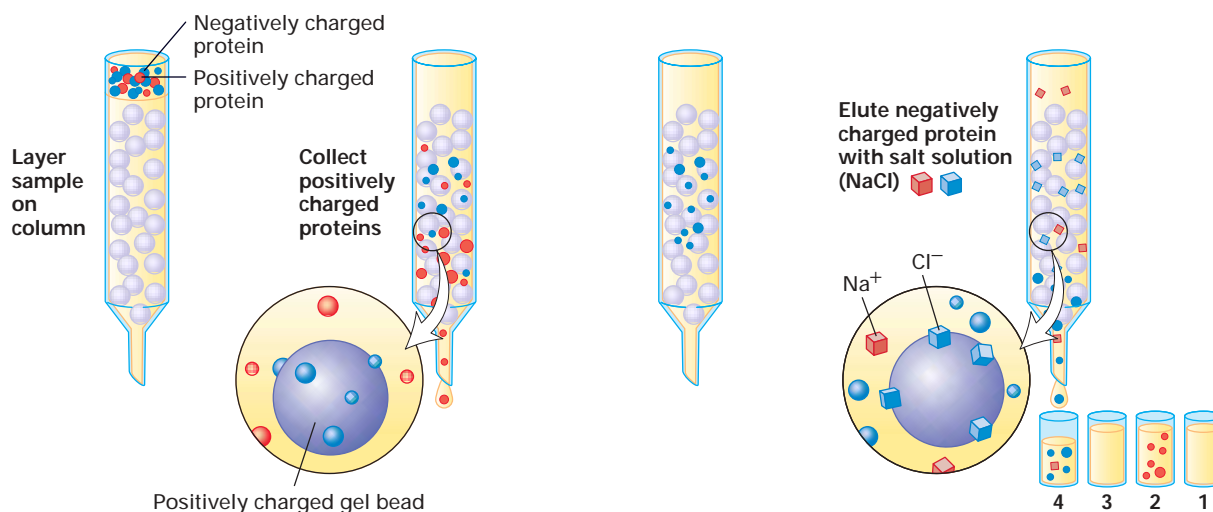
(a) Gel filtration chromatography



(c) Antibody-affinity chromatography



(b) Ion-exchange chromatography



▲ **EXPERIMENTAL FIGURE 3-34** Three commonly used liquid chromatographic techniques separate proteins on the basis of mass, charge, or affinity for a specific ligand. (a) Gel filtration chromatography separates proteins that differ in size. A mixture of proteins is carefully layered on the top of a glass cylinder packed with porous beads. Smaller proteins travel through the column more slowly than larger proteins. Thus different proteins have different elution volumes and can be collected in separate liquid fractions from the bottom. (b) Ion-exchange chromatography separates proteins that differ in net charge in columns packed with special beads that carry either a positive charge (shown here) or a negative charge. Proteins

having the same net charge as the beads are repelled and flow through the column, whereas proteins having the opposite charge bind to the beads. Bound proteins—in this case, negatively charged—are eluted by passing a salt gradient (usually of NaCl or KCl) through the column. As the ions bind to the beads, they desorb the protein. (c) In antibody-affinity chromatography, a specific antibody is covalently attached to beads packed in a column. Only protein with high affinity for the antibody is retained by the column; all the nonbinding proteins flow through. The bound protein is eluted with an acidic solution, which disrupts the antigen-antibody complexes.

salt concentrations, protein molecules and beads are attracted by their opposite charges. At higher salt concentrations, negative salt ions bind to the positively charged beads, displacing the negatively charged proteins. In a gradient of

increasing salt concentration, weakly charged proteins are eluted first and highly charged proteins are eluted last. Similarly, a negatively charged column can be used to retain and fractionate basic proteins.



**Affinity Chromatography** The ability of proteins to bind specifically to other molecules is the basis of *affinity chromatography*. In this technique, ligand molecules that bind to the protein of interest are covalently attached to the beads used to form the column. Ligands can be enzyme substrates or other small molecules that bind to specific proteins. In a widely used form of this technique, *antibody-affinity chromatography*, the attached ligand is an antibody specific for the desired protein (Figure 3-34c).

An affinity column will retain only those proteins that bind the ligand attached to the beads; the remaining proteins, regardless of their charges or masses, will pass through the column without binding to it. However, if a retained protein interacts with other molecules, forming a complex, then the entire complex is retained on the column. The proteins bound to the affinity column are then eluted by adding an excess of ligand or by changing the salt concentration or pH. The ability of this technique to separate particular proteins depends on the selection of appropriate ligands.

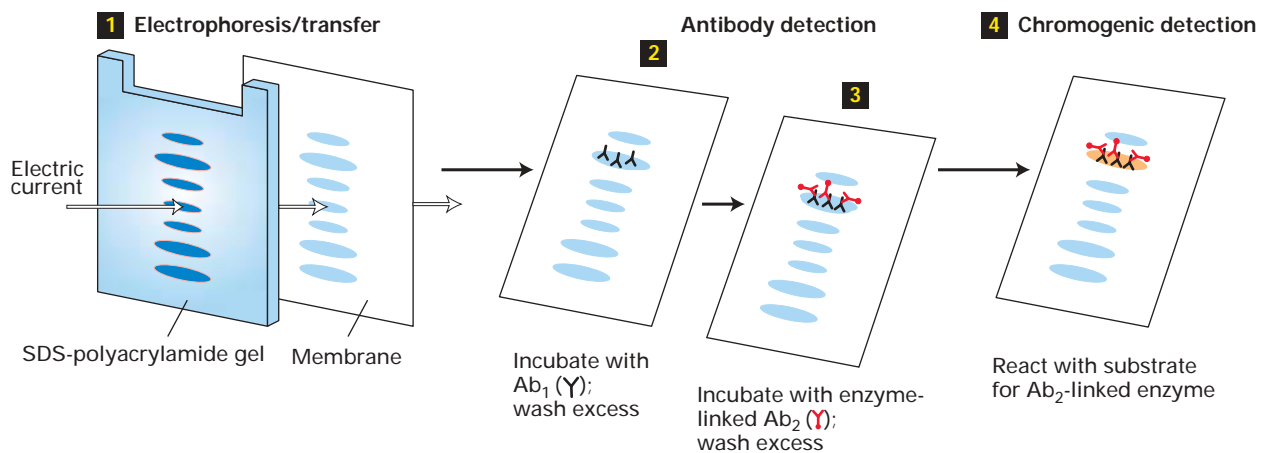
### Highly Specific Enzyme and Antibody Assays Can Detect Individual Proteins

The purification of a protein, or any other molecule, requires a specific assay that can detect the molecule of interest in column fractions or gel bands. An assay capitalizes on some highly distinctive characteristic of a protein: the ability to bind a particular ligand, to catalyze a particular reaction, or to be recognized by a specific antibody. An assay must also

be simple and fast to minimize errors and the possibility that the protein of interest becomes denatured or degraded while the assay is performed. The goal of any purification scheme is to isolate sufficient amounts of a given protein for study; thus a useful assay must also be sensitive enough that only a small proportion of the available material is consumed. Many common protein assays require just from  $10^{-9}$  to  $10^{-12}$  g of material.

**Chromogenic and Light-Emitting Enzyme Reactions** Many assays are tailored to detect some functional aspect of a protein. For example, enzyme assays are based on the ability to detect the loss of substrate or the formation of product. Some enzyme assays utilize *chromogenic* substrates, which change color in the course of the reaction. (Some substrates are naturally chromogenic; if they are not, they can be linked to a chromogenic molecule.) Because of the specificity of an enzyme for its substrate, only samples that contain the enzyme will change color in the presence of a chromogenic substrate and other required reaction components; the rate of the reaction provides a measure of the quantity of enzyme present.

Such chromogenic enzymes can also be fused or chemically linked to an antibody and used to “report” the presence or location of the antigen. Alternatively, *luciferase*, an enzyme present in fireflies and some bacteria, can be linked to an antibody. In the presence of ATP and luciferin, luciferase catalyzes a light-emitting reaction. In either case, after the antibody binds to the protein of interest, substrates of the linked enzyme are added and the appearance of color or



▲ **EXPERIMENTAL FIGURE 3-35 Western blotting (immunoblotting) combines several techniques to resolve and detect a specific protein.** Step **1**: After a protein mixture has been electrophoresed through an SDS gel, the separated bands are transferred (blotted) from the gel onto a porous membrane. Step **2**: The membrane is flooded with a solution of antibody (Ab<sub>1</sub>) specific for the desired protein. Only the band containing this protein binds the antibody, forming a layer of antibody molecules (although their position

cannot be seen at this point). After sufficient time for binding, the membrane is washed to remove unbound Ab<sub>1</sub>. Step **3**: The membrane is incubated with a second antibody (Ab<sub>2</sub>) that binds to the bound Ab<sub>1</sub>. This second antibody is covalently linked to alkaline phosphatase, which catalyzes a chromogenic reaction. Step **4**: Finally, the substrate is added and a deep purple precipitate forms, marking the band containing the desired protein.

emitted light is monitored. A variation of this technique, particularly useful in detecting specific proteins within living cells, makes use of green fluorescent protein (GFP), a naturally fluorescent protein found in jellyfish (see Figure 5-46).

**Western Blotting** A powerful method for detecting a particular protein in a complex mixture combines the superior resolving power of gel electrophoresis, the specificity of antibodies, and the sensitivity of enzyme assays. Called **Western blotting**, or immunoblotting, this multistep procedure is commonly used to separate proteins and then identify a specific protein of interest. As shown in Figure 3-35, two different antibodies are used in this method, one specific for the desired protein and the other linked to a reporter enzyme.

### Radioisotopes Are Indispensable Tools for Detecting Biological Molecules

A sensitive method for tracking a protein or other biological molecule is by detecting the radioactivity emitted from radioisotopes introduced into the molecule. At least one atom in a radiolabeled molecule is present in a radioactive form, called a **radioisotope**.

**Radioisotopes Useful in Biological Research** Hundreds of biological compounds (e.g., amino acids, nucleosides, and numerous metabolic intermediates) labeled with various radioisotopes are commercially available. These preparations vary considerably in their *specific activity*, which is the amount of radioactivity per unit of material, measured in disintegrations per minute (dpm) per millimole. The specific activity of a labeled compound depends on the probability of decay of the radioisotope, indicated by its *half-life*, which is the time required for half the atoms to undergo radioactive decay. In general, the shorter the half-life of a radioisotope, the higher its specific activity (Table 3-3).

The specific activity of a labeled compound must be high enough that sufficient radioactivity is incorporated into cellular molecules to be accurately detected. For example, methionine and cysteine labeled with sulfur-35 ( $^{35}\text{S}$ ) are widely used to label cellular proteins because preparations of these

amino acids with high specific activities ( $>10^{15}$  dpm/mmol) are available. Likewise, commercial preparations of  $^3\text{H}$ -labeled nucleic acid precursors have much higher specific activities than those of the corresponding  $^{14}\text{C}$ -labeled preparations. In most experiments, the former are preferable because they allow RNA or DNA to be adequately labeled after a shorter time of incorporation or require a smaller cell sample. Various phosphate-containing compounds in which every phosphorus atom is the radioisotope phosphorus-32 are readily available. Because of their high specific activity,  $^{32}\text{P}$ -labeled nucleotides are routinely used to label nucleic acids in cell-free systems.

Labeled compounds in which a radioisotope replaces atoms normally present in the molecule have the same chemical properties as the corresponding nonlabeled compounds. Enzymes, for instance, cannot distinguish between substrates labeled in this way and their nonlabeled substrates. In contrast, labeling with the radioisotope iodine-125 ( $^{125}\text{I}$ ) requires the covalent addition of  $^{125}\text{I}$  to a protein or nucleic acid. Because this labeling procedure modifies the chemical structure of a protein or nucleic acid, the biological activity of the labeled molecule may differ somewhat from that of the nonlabeled form.

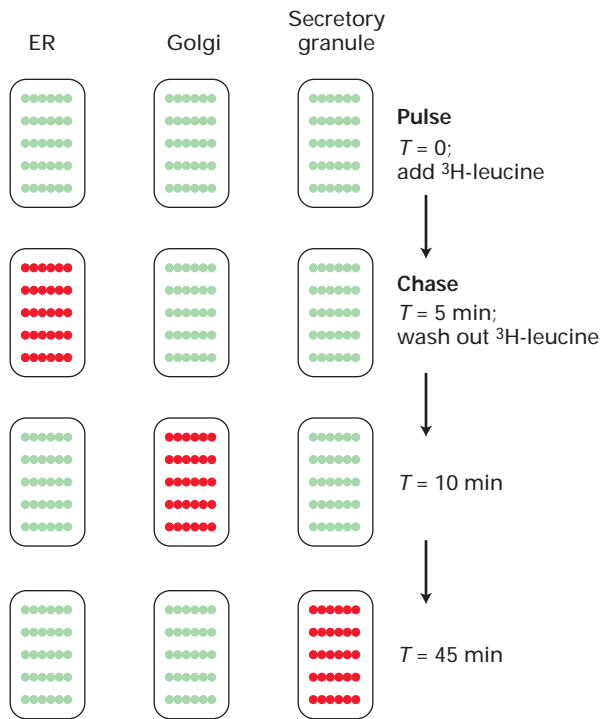
**Labeling Experiments and Detection of Radiolabeled Molecules** Whether labeled compounds are detected by **autoradiography**, a semiquantitative visual assay, or their radioactivity is measured in an appropriate “counter,” a highly quantitative assay that can determine the concentration of a radiolabeled compound in a sample, depends on the nature of the experiment. In some experiments, both types of detection are used.

In one use of autoradiography, a cell or cell constituent is labeled with a radioactive compound and then overlaid with a photographic emulsion sensitive to radiation. Development of the emulsion yields small silver grains whose distribution corresponds to that of the radioactive material. Autoradiographic studies of whole cells were crucial in determining the intracellular sites where various macromolecules are synthesized and the subsequent movements of these macromolecules within cells. Various techniques employing fluorescent microscopy, which we describe in the next chapter, have largely supplanted autoradiography for studies of this type. However, autoradiography is commonly used in various assays for detecting specific isolated DNA or RNA sequences (Chapter 9).

Quantitative measurements of the amount of radioactivity in a labeled material are performed with several different instruments. A *Geiger counter* measures ions produced in a gas by the  $\beta$  particles or  $\gamma$  rays emitted from a radioisotope. In a *scintillation counter*, a radiolabeled sample is mixed with a liquid containing a fluorescent compound that emits a flash of light when it absorbs the energy of the  $\beta$  particles or  $\gamma$  rays released in the decay of the radioisotope; a phototube in the instrument detects and counts these light flashes. *Phosphorimagers* are used to detect radiolabeled compounds on a surface, storing digital data on the number of decays in

**TABLE 3-3** Radioisotopes Commonly Used in Biological Research

Isotope	Half-Life
Phosphorus-32	14.3 days
Iodine-125	60.4 days
Sulfur-35	87.5 days
Tritium (hydrogen-3)	12.4 years
Carbon-14	5730.4 years



**▲ EXPERIMENTAL FIGURE 3-36 Pulse-chase experiments can track the pathway of protein movement within cells.**

To determine the pathway traversed by secreted proteins subsequent to their synthesis on the rough endoplasmic reticulum (ER), cells are briefly incubated in a medium containing a radiolabeled amino acid (e.g., [ $^3\text{H}$ ]leucine), the pulse, which will label any protein synthesized during this period. The cells are then washed with buffer to remove the pulse and transferred to medium lacking a radioactive precursor, the chase. Samples taken periodically are analyzed by autoradiography to determine the cellular location of labeled protein. At the beginning of the experiment ( $t = 0$ ), no protein is labeled, as indicated by the green dotted lines. At the end of the pulse ( $t = 5$  minutes), all the labeled protein (red lines) appears in the ER. At subsequent times, this newly synthesized labeled protein is visualized first in the Golgi complex and then in secretory vesicles. Because any protein synthesized during the chase period is not labeled, the movement of the labeled protein can be defined quite precisely.

disintegrations per minute per small pixel of surface area. These instruments, which can be thought of as a kind of reusable electronic film, are commonly used to quantitate radioactive molecules separated by gel electrophoresis and are replacing photographic emulsions for this purpose.

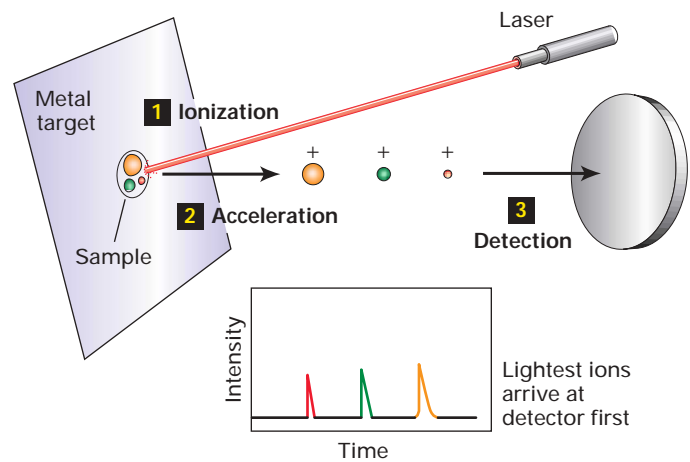
A combination of labeling and biochemical techniques and of visual and quantitative detection methods is often employed in labeling experiments. For instance, to identify the major proteins synthesized by a particular cell type, a sample of the cells is incubated with a radioactive amino acid (e.g., [ $^{35}\text{S}$ ]methionine) for a few minutes. The mixture of cellular proteins is then resolved by gel electrophoresis, and the gel

is subjected to autoradiography or phosphorimager analysis. The radioactive bands correspond to newly synthesized proteins, which have incorporated the radiolabeled amino acid. Alternatively, the proteins can be resolved by liquid chromatography, and the radioactivity in the eluted fractions can be determined quantitatively with a counter.

**Pulse-chase** experiments are particularly useful for tracing changes in the intracellular location of proteins or the transformation of a metabolite into others over time. In this experimental protocol, a cell sample is exposed to a radiolabeled compound—the “pulse”—for a brief period of time, then washed with buffer to remove the labeled pulse, and finally incubated with a nonlabeled form of the compound—the “chase” (Figure 3-36). Samples taken periodically are assayed to determine the location or chemical form of the radiolabel. A classic use of the pulse-chase technique was in studies to elucidate the pathway traversed by secreted proteins from their site of synthesis in the endoplasmic reticulum to the cell surface (Chapter 17).

### Mass Spectrometry Measures the Mass of Proteins and Peptides

A powerful technique for measuring the mass of molecules such as proteins and peptides is *mass spectrometry*. This



**▲ EXPERIMENTAL FIGURE 3-37 The molecular weight of proteins and peptides can be determined by time-of-flight mass spectrometry.** In a laser-desorption mass spectrometer, pulses of light from a laser ionize a protein or peptide mixture that is absorbed on a metal target (1). An electric field accelerates the molecules in the sample toward the detector (2 and 3). The time to the detector is inversely proportional to the mass of a molecule. For molecules having the same charge, the time to the detector is inversely proportional to the mass. The molecular weight is calculated using the time of flight of a standard.

technique requires a method for ionizing the sample, usually a mixture of peptides or proteins, accelerating the molecular ions, and then detecting the ions. In a laser desorption mass spectrometer, the protein sample is mixed with an organic acid and then dried on a metal target. Energy from a laser ionizes the proteins, and an electric field accelerates the ions down a tube to a detector (Figure 3-37). Alternatively, in an electrospray mass spectrometer, a fine mist containing the sample is ionized and then introduced into a separation chamber where the positively charged molecules are accelerated by an electric field. In both instruments, the time of flight is inversely proportional to a protein's mass and directly proportional to its charge. As little as  $1 \times 10^{-15}$  mol (1 femtomole) of a protein as large as 200,000 MW can be measured with an error of 0.1 percent.

### Protein Primary Structure Can Be Determined by Chemical Methods and from Gene Sequences

The classic method for determining the amino acid sequence of a protein is *Edman degradation*. In this procedure, the free amino group of the N-terminal amino acid of a polypeptide is labeled, and the labeled amino acid is then cleaved from the polypeptide and identified by high-pressure liquid chromatography. The polypeptide is left one residue shorter, with a new amino acid at the N-terminus. The cycle is repeated on the ever shortening polypeptide until all the residues have been identified.

Before about 1985, biologists commonly used the Edman chemical procedure for determining protein sequences. Now, however, protein sequences are determined primarily by analysis of genome sequences. The complete genomes of several organisms have already been sequenced, and the database of genome sequences from humans and numerous model organisms is expanding rapidly. As discussed in Chapter 9, the sequences of proteins can be deduced from DNA sequences that are predicted to encode proteins.

A powerful approach for determining the primary structure of an isolated protein combines mass spectroscopy and the use of sequence databases. First, mass spectrometry is used to determine the *peptide mass fingerprint* of the protein. A peptide mass fingerprint is a compilation of the molecular weights of peptides that are generated by a specific protease. The molecular weights of the parent protein and its proteolytic fragments are then used to search genome databases for any similarly sized protein with identical or similar peptide mass maps.

### Peptides with a Defined Sequence Can Be Synthesized Chemically

Synthetic peptides that are identical with peptides synthesized *in vivo* are useful experimental tools in studies of proteins and cells. For example, short synthetic peptides of 10–15 residues can function as antigens to trigger the production of antibodies in animals. A synthetic peptide, when

coupled to a large protein carrier, can trick an animal into producing antibodies that bind the full-sized, natural protein antigen. As we'll see throughout this book, antibodies are extremely versatile reagents for isolating proteins from mixtures by affinity chromatography (see Figure 3-34c), for separating and detecting proteins by Western blotting (see Figure 3-35), and for localizing proteins in cells by microscopic techniques described in Chapter 5.

Peptides are routinely synthesized in a test tube from monomeric amino acids by condensation reactions that form peptide bonds. Peptides are constructed sequentially by coupling the C-terminus of a monomeric amino acid with the N-terminus of the growing peptide. To prevent unwanted reactions entailing the amino groups and carboxyl groups of the side chains during the coupling steps, a protecting (blocking) group is attached to the side chains. Without these protecting groups, branched peptides would be generated. In the last steps of synthesis, the side chain-protecting groups are removed and the peptide is cleaved from the resin on which synthesis takes place.

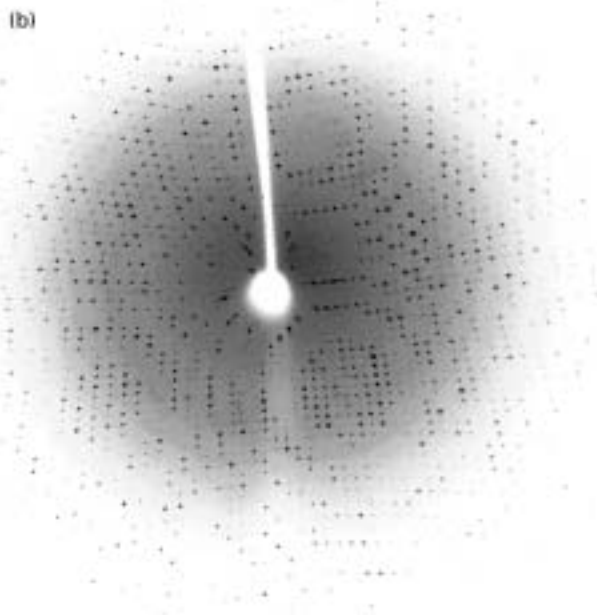
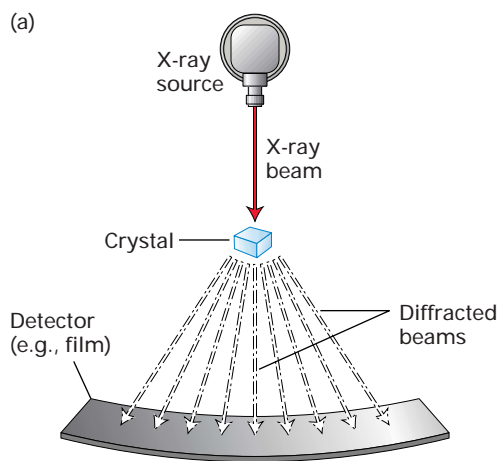
### Protein Conformation Is Determined by Sophisticated Physical Methods

In this chapter, we have emphasized that protein function is dependent on protein structure. Thus, to figure out how a protein works, its three-dimensional structure must be known. Determining a protein's conformation requires sophisticated physical methods and complex analyses of the experimental data. We briefly describe three methods used to generate three-dimensional models of proteins.

**X-Ray Crystallography** The use of **x-ray crystallography** to determine the three-dimensional structures of proteins was pioneered by Max Perutz and John Kendrew in the 1950s. In this technique, beams of x-rays are passed through a protein crystal in which millions of protein molecules are precisely aligned with one another in a rigid array characteristic of the protein. The wavelengths of x-rays are about 0.1–0.2 nm, short enough to resolve the atoms in the protein crystal. Atoms in the crystal scatter the x-rays, which produce a diffraction pattern of discrete spots when they are intercepted by photographic film (Figure 3-38). Such patterns are extremely complex—composed of as many as 25,000 diffraction spots for a small protein. Elaborate calculations and modifications of the protein (such as the binding of heavy metals) must be made to interpret the diffraction pattern and to solve the structure of the protein. The process is analogous to reconstructing the precise shape of a rock from the ripples that it creates in a pond. To date, the detailed three-dimensional structures of more than 10,000 proteins have been established by x-ray crystallography.

**Cryoelectron Microscopy** Although some proteins readily crystallize, obtaining crystals of others—particularly large multisubunit proteins—requires a time-consuming trial-and-





▲ **EXPERIMENTAL FIGURE 3-38 X-ray crystallography provides diffraction data from which the three-dimensional structure of a protein can be determined.** (a) Basic components of an x-ray crystallographic determination. When a narrow beam of x-rays strikes a crystal, part of it passes straight through and the rest is scattered (diffracted) in various directions. The intensity of the diffracted waves is recorded on an x-ray film or with a solid-state electronic detector. (b) X-ray diffraction pattern for a topoisomerase crystal collected on a solid-state detector. From complex analyses of patterns like this one, the location of every atom in a protein can be determined. [Part (a) adapted from L. Stryer, 1995, *Biochemistry*, 4th ed., W. H. Freeman and Company, p. 64; part (b) courtesy of J. Berger.]

error effort to find just the right conditions. The structures of such difficult-to-crystallize proteins can be obtained by *cryo-electron microscopy*. In this technique, a protein sample is rapidly frozen in liquid helium to preserve its structure and then examined in the frozen, hydrated state in a cryoelectron microscope. Pictures are recorded on film by using a low dose

of electrons to prevent radiation-induced damage to the structure. Sophisticated computer programs analyze the images and reconstruct the protein's structure in three dimensions. Recent advances in cryoelectron microscopy permit researchers to generate molecular models that compare with those derived from x-ray crystallography. The use of cryoelectron microscopy and other types of electron microscopy for visualizing cell structures are discussed in Chapter 5.

**NMR Spectroscopy** The three-dimensional structures of small proteins containing about as many as 200 amino acids can be studied with nuclear magnetic resonance (NMR) spectroscopy. In this technique, a concentrated protein solution is placed in a magnetic field and the effects of different radio frequencies on the spin of different atoms are measured. The behavior of any atom is influenced by neighboring atoms in adjacent residues, with closely spaced residues being more perturbed than distant residues. From the magnitude of the effect, the distances between residues can be calculated; these distances are then used to generate a model of the three-dimensional structure of the protein.

Although NMR does not require the crystallization of a protein, a definite advantage, this technique is limited to proteins smaller than about 20 kDa. However, NMR analysis can also be applied to protein domains, which tend to be small enough for this technique and can often be obtained as stable structures.

## KEY CONCEPTS OF SECTION 3.6

### Purifying, Detecting, and Characterizing Proteins

- Proteins can be separated from other cell components and from one another on the basis of differences in their physical and chemical properties.
- Centrifugation separates proteins on the basis of their rates of sedimentation, which are influenced by their masses and shapes.
- Gel electrophoresis separates proteins on the basis of their rates of movement in an applied electric field. SDS-polyacrylamide gel electrophoresis can resolve polypeptide chains differing in molecular weight by 10 percent or less (see Figure 3-32).
- Liquid chromatography separates proteins on the basis of their rates of movement through a column packed with spherical beads. Proteins differing in mass are resolved on gel filtration columns; those differing in charge, on ion-exchange columns; and those differing in ligand-binding properties, on affinity columns (see Figure 3-34).
- Various assays are used to detect and quantify proteins. Some assays use a light-producing reaction or radioactivity to generate a signal. Other assays produce an amplified colored signal with enzymes and chromogenic substrates.
- Antibodies are powerful reagents used to detect, quantify, and isolate proteins. They are used in affinity chromatography and combined with gel electrophoresis in

Western blotting, a powerful method for separating and detecting a protein in a mixture (see Figure 3-35).

- Autoradiography is a semiquantitative technique for detecting radioactively labeled molecules in cells, tissues, or electrophoretic gels.
- Pulse-chase labeling can determine the intracellular fate of proteins and other metabolites (see Figure 3-36).
- Three-dimensional structures of proteins are obtained by x-ray crystallography, cryoelectron microscopy, and NMR spectroscopy. X-ray crystallography provides the most detailed structures but requires protein crystallization. Cryoelectron microscopy is most useful for large protein complexes, which are difficult to crystallize. Only relatively small proteins are amenable to NMR analysis.

## PERSPECTIVES FOR THE FUTURE

Impressive expansion of the computational power of computers is at the core of advances in determining the three-dimensional structures of proteins. For example, vacuum tube computers running on programs punched on cards were used to solve the first protein structures on the basis of x-ray crystallography. In the future, researchers aim to predict the structures of proteins only on the basis of amino acid sequences deduced from gene sequences. This computationally challenging problem requires supercomputers or large clusters of computers working in synchrony. Currently, only the structures of very small domains containing 100 residues or fewer can be predicted at a low resolution. However, continued developments in computing and models of protein folding, combined with large-scale efforts to solve the structures of all protein motifs by x-ray crystallography, will allow the prediction of the structures of larger proteins. With an exponentially expanding database of motifs, domains, and proteins, scientists will be able to identify the motifs in an unknown protein, match the motif to the sequence, and use this head start in predicting the three-dimensional structure of the entire protein.

New combined approaches will also help in determining high-resolution structures of molecular machines such as those listed in Table 3-1. Although these very large macromolecular assemblies usually are difficult to crystallize and thus to solve by x-ray crystallography, they can be imaged in a cryoelectron microscope at liquid helium temperatures and high electron energies. From millions of individual “particles,” each representing a random view of the protein complex, the three-dimensional structure can be built. Because subunits of the complex may already be solved by crystallography, a composite structure consisting of the x-ray-derived subunit structures fitted to the EM-derived model will be generated. An interesting application of this type of study would be the solution of the structures of amyloid and prion pro-

teins, especially in the early stages in the formation of insoluble filaments.

Understanding the operation of protein machines will require the measurement of many new characteristics of proteins. For example, because many machines do nonchemical work of some type, biologists will have to identify the energy sources (mechanical, electrical, or thermal) and measure the amounts of energy to determine the limits of a particular machine. Because most activities of machines include movement of one type or another, the force powering the movement and its relation to biological activity can be a source of insight into how force generation is coupled to chemistry. Improved tools such as optical traps and atomic force microscopes will enable detailed studies of the forces and chemistry pertinent to the operation of individual protein machines.

## KEY TERMS

$\alpha$ helix 61	molecular machine 59
activation energy 74	motif 63
active site 75	motor protein 79
allostery 83	peptide bond 60
amyloid filament 73	polypeptide 61
autoradiography 93	primary structure 61
$\beta$ sheet 61	proteasome 71
chaperone 69	protein 61
conformation 60	proteome 60
cooperativity 83	quaternary structure 66
domain 63	rate-zonal centrifugation 87
electrophoresis 87	secondary structure 61
homology 68	tertiary structure 62
$K_m$ 76	ubiquitin 71
ligand 73	$V_{max}$ 76
liquid chromatography 90	x-ray crystallography 95

## REVIEW THE CONCEPTS

1. The three-dimensional structure of a protein is determined by its primary, secondary, and tertiary structures. Define the primary, secondary, and tertiary structures. What are some of the common secondary structures? What are the forces that hold together the secondary and tertiary structures? What is the quaternary structure?
2. Proper folding of proteins is essential for biological activity. Describe the roles of molecular chaperones and chaperonins in the folding of proteins.
3. Proteins are degraded in cells. What is ubiquitin, and what role does it play in tagging proteins for degradation? What is the role of proteasomes in protein degradation?

4. Enzymes can catalyze chemical reactions. How do enzymes increase the rate of a reaction? What constitutes the active site of an enzyme? For an enzyme-catalyzed reaction, what are  $K_m$  and  $V_{max}$ ? For enzyme X, the  $K_m$  for substrate A is 0.4 mM and for substrate B is 0.01 mM. Which substrate has a higher affinity for enzyme X?

5. Motor proteins, such as myosin, convert energy into a mechanical force. Describe the three general properties characteristic of motor proteins. Describe the biochemical events that occur during one cycle of movement of myosin relative to an actin filament.

6. The function of proteins can be regulated in a number of ways. What is cooperativity, and how does it influence protein function? Describe how protein phosphorylation and proteolytic cleavage can modulate protein function.

7. A number of techniques can separate proteins on the basis of their differences in mass. Describe the use of two of these techniques, centrifugation and gel electrophoresis. The blood proteins transferrin (MW 76 kDa) and lysozyme (MW 15 kDa) can be separated by rate zonal centrifugation or SDS polyacrylamide gel electrophoresis. Which of the two proteins will sediment faster during centrifugation? Which will migrate faster during electrophoresis?

8. Chromatography is an analytical method used to separate proteins. Describe the principles for separating proteins by gel filtration, ion-exchange, and affinity chromatography.

9. Various methods have been developed for detecting proteins. Describe how radioisotopes and autoradiography can be used for labeling and detecting proteins. How does Western blotting detect proteins?

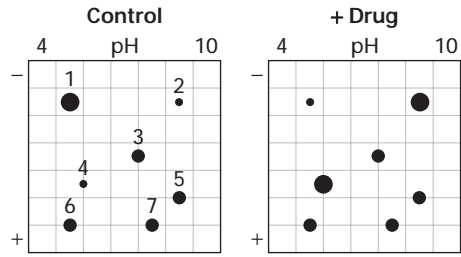
10. Physical methods are often used to determine protein conformation. Describe how x-ray crystallography, cryoelectron microscopy, and NMR spectroscopy can be used to determine the shape of proteins.

**ANALYZE THE DATA**

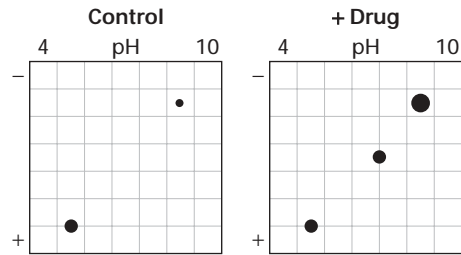
Proteomics involves the global analysis of protein expression. In one approach, all the proteins in control cells and treated cells are extracted and subsequently separated using two-dimensional gel electrophoresis. Typically, hundreds or thousands of protein spots are resolved and the steady-state levels of each protein spots are compared between control and treated cells. In the following example, only a few protein spots are shown for simplicity. Proteins are separated in the first dimension on the basis of charge by isoelectric focusing (pH 4–10) and then separated by size by SDS polyacrylamide gel electrophoresis. Proteins are detected with a stain such as Coomassie blue and assigned numbers for identification.

a. Cells are treated with a drug (“+ Drug”) or left untreated (“Control”) and then proteins are extracted and separated by two-dimensional gel electrophoresis. The stained gels are

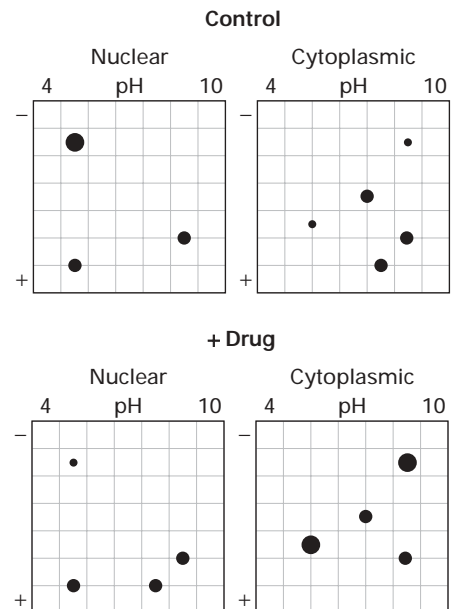
shown below. What do you conclude about the effect of the drug on the steady-state levels of proteins 1–7?



b. You suspect that the drug may be inducing a protein kinase and so repeat the experiment in part a in the presence of  $^{32}\text{P}$ -labeled inorganic phosphate. In this experiment the two-dimensional gels are exposed to x-ray film to detect the presence of  $^{32}\text{P}$ -labeled proteins. The x-ray films are shown below. What do you conclude from this experiment about the effect of the drug on proteins 1–7?



c. To determine the cellular localization of proteins 1–7, the cells from part a were separated into nuclear and cytoplasmic fractions by differential centrifugation. Two-dimensional gels were run and the stained gels are shown below. What do you conclude about the cellular localization of proteins 1–7?



d. Summarize the overall properties of proteins 1–7, combining the data from parts a, b, and c. Describe how you could determine the identity of any one of the proteins.

## REFERENCES

### General References

Berg, J. M., J. L. Tymoczko, and L. Stryer. 2002. *Biochemistry*, 5th ed. W. H. Freeman and Company, chaps. 2–4, 7–10.

Nelson, D. L., and M. M. Cox. 2000. *Lehninger Principles of Biochemistry*, 3d ed. Worth Publishers, chaps. 5–8.

### Web Sites

Entry site into the proteins, structures, genomes, and taxonomy: <http://www.ncbi.nlm.nih.gov/Entrez/>

The protein 3D structure database: <http://www.rcsb.org/>

Structural classifications of proteins: <http://scop.mrcmb.cam.ac.uk/scop/>

Sites containing general information about proteins: <http://www.expaty.ch/>; <http://www.proweb.org/>

Sites for specific protein families: <http://www.pkr.sdsc.edu/html/index.shtml> The protein kinase resource; <http://www.mrc-lmb.cam.ac.uk/myosin/myosin.html> The myosin home page; <http://www.proweb.org/kinesin/> The kinesin home page

### Hierarchical Structure of Proteins

Branden, C., and J. Tooze. 1999. *Introduction to Protein Structure*. Garland.

Creighton, T. E. 1993. *Proteins: Structures and Molecular Properties*, 2d ed. W. H. Freeman and Company.

Hardison, R. 1998. Hemoglobins from bacteria to man: Evolution of different patterns of gene expression. *J. Exp. Biol.* **201**: 1099.

Lesk, A. M. 2001. *Introduction to Protein Architecture*. Oxford.

Macromolecular Machines. 1998. *Cell* **92**:291–423. A special review issue on protein machines.

Patthy, L. 1999. *Protein Evolution*. Blackwell Science.

### Folding, Modification, and Degradation of Proteins

Cohen, F. E. 1999. Protein misfolding and prion diseases. *J. Mol. Biol.* **293**:313–320.

Dobson, C. M. 1999. Protein misfolding, evolution, and disease. *Trends Biochem. Sci.* **24**:329–332.

Hartl, F. U., and M. Hayer-Hartl. 2002. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* **295**:1852–1858.

Kirschner, M. 1999. Intracellular proteolysis. *Trends Cell Biol.* **9**:M42–M45.

Kornitzer, D., and A. Ciechanover. 2000. Modes of regulation of ubiquitin-mediated protein degradation. *J. Cell Physiol.* **182**:1–11.

Laney, J. D., and M. Hochstrasser. 1999. Substrate targeting in the ubiquitin system. *Cell* **97**:427–430.

Rochet, J.-C., and P. T. Lansbury. 2000. Amyloid fibrillogenesis: Themes and variations. *Curr. Opin. Struct. Biol.* **10**:60–68.

Weissman, A. M. 2001. Themes and variations on ubiquitylation. *Nature Cell Biol.* **2**:169–177.

Zhang, X., F. Beuron, and P. S. Freemont. 2002. Machinery of protein folding and unfolding. *Curr. Opin. Struct. Biol.* **12**:231–238.

Zwickil, P., W. Baumeister, and A. Steven. 2000. Dis-assembly lines: The proteasome and related ATPase-assisted proteases. *Curr. Opin. Struct. Biol.* **10**:242–250.

### Enzymes and the Chemical Work of Cells

Dressler, D. H., and H. Potter. 1991. *Discovering Enzymes*. Scientific American Library.

Fersht, A. 1999. *Enzyme Structure and Mechanism*, 3d ed. W. H. Freeman and Company.

Smith, C. M., et al. 1997. The protein kinase resource. *Trends Biochem. Sci.* **22**:444–446.

Taylor, S. S., and E. Radzio-Andzelm. 1994. Three protein kinase structures define a common motif. *Structure* **2**:345–355.

### Molecular Motors and the Mechanical Work of Cells

Cooke, R. 2001. Motor proteins. *Encyclopedia Life Sciences*. Nature Publishing Group.

Spudich, J. A. 2001. The myosin swinging cross-bridge model. *Nature Rev. Mol. Cell Biol.* **2**:387–392.

Vale, R. D., and R. A. Milligan. 2000. The way things move: Looking under the hood of molecular motor proteins. *Science* **288**:88–95.

### Common Mechanisms for Regulating Protein Function

Ackers, G. K. 1998. Deciphering the molecular code of hemoglobin allostery. *Adv. Protein Chem.* **51**:185–253.

Austin, D. J., G. R. Crabtree, and S. L. Schreiber. 1994. Proximity versus allostery: The role of regulated protein dimerization in biology. *Chem. Biol.* **1**:131–136.

Burack, W. R., and A. S. Shaw. 2000. Signal transduction: Hanging on a scaffold. *Curr. Opin. Cell Biol.* **12**:211–216.

Cox, S., E. Radzio-Andzelm, and S. S. Taylor. 1994. Domain movements in protein kinases. *Curr. Opin. Struct. Biol.* **4**:893–901.

Horovitz, A., Y. Fridmann, G. Kafri, and O. Yifrach. 2001. Review: Allostery in chaperonins. *J. Struct. Biol.* **135**:104–114.

Kawasaki, H., S. Nakayama, and R. H. Kretsinger. 1998. Classification and evolution of EF-hand proteins. *Biometals* **11**:277–295.

Lim, W. A. 2002. The modular logic of signaling proteins: Building allosteric switches from simple binding domains. *Curr. Opin. Struct. Biol.* **12**:61–68.

Ptashne, M., and A. Gann. 1998. Imposing specificity by localization: Mechanism and evolvability. *Curr. Biol.* **8**:R812–R822.

Saibil, H. R., A. L. Horwich, and W. A. Fenton. 2001. Allostery and protein substrate conformational change during GroEL/GroES-mediated protein folding. *Adv. Protein Chem.* **59**:45–72.

Yap, K. L., J. A. B. Ames, M. B. Sindells, and M. Ikura. 1999. Diversity of conformational states and changes within the EF-hand protein superfamily. *Proteins* **37**:499–507.

### Purifying, Detecting, and Characterizing Proteins

Hames, B. D. *A Practical Approach*. Oxford University Press. A methods series that describes protein purification methods and assays.