Molecular Biology of THE CELL

Fifth Edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

With problems by

John Wilson

Tim Hunt

Garland Science
Taylor & Francis Group

Garland Science
Vice President: Denise Schanck
Assistant Editor: Sigrid Masson
Production Editor and Layout: Emma Jeffcock
Senior Publisher: Jackie Harbor
Illustrator: Nigel Orme
Designer: Matthew McClements, Blink Studio, Ltd.
Editors: Marjorie Anderson and Sherry Granum
Copy Editor: Bruce Goatly
Indexer: Merrall-Ross International, Ltd.
Permissions Coordinator: Mary Dispenza

Cell Biology Interactive
Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production Design and Development: Michael Morales

Bruce Alberts received his Ph.D. from Harvard University and is Professor of Biochemistry and Biophysics at the University of California, San Francisco. For 12 years, he served as President of the U.S. National Academy of Sciences (1993–2005). **Alexander Johnson** received his Ph.D. from Harvard University and is Professor of Microbiology and Immunology and Director of the Biochemistry, Cell Biology, Genetics, and Developmental Biology Graduate Program at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the London Research Institute of Cancer Research UK. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Emeritus Fellow at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

Library of Congress Cataloging-in-Publication Data
Molecular biology of the cell / Bruce Alberts … [et al.].-- 5th ed.
p. cm
ISBN 978-0-8153-4105-5 (hardcover)---ISBN 978-0-8153-4106-2 (paperback)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
QH581.2 .M64 2008
571.6--dc22
2007005475 CIP

Published by Garland Science, Taylor & Francis Group, LLC, an informa business, 270 Madison Avenue, New York NY 10016, USA, and 2 Park Square, Milton Park, Abingdon, OX14 4RN, UK.

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2

3 Proteins

When we look at a cell through a microscope or analyze its electrical or biochemical activity, we are, in essence, observing proteins. Proteins constitute most of a cell's dry mass. They are not only the cell's building blocks; they also execute nearly all the cell's functions. Thus, enzymes provide the intricate molecular surfaces in a cell that promote its many chemical reactions. Proteins embedded in the plasma membrane form channels and pumps that control the passage of small molecules into and out of the cell. Other proteins carry messages from one cell to another, or act as signal integrators that relay sets of signals inward from the plasma membrane to the cell nucleus. Yet others serve as tiny molecular machines with moving parts: *kinesin*, for example, propels organelles through the cytoplasm; *topoisomerase* can untangle knotted DNA molecules. Other specialized proteins act as antibodies, toxins, hormones, antifreeze molecules, elastic fibers, ropes, or sources of luminescence. Before we can hope to understand how genes work, how muscles contract, how nerves conduct electricity, how embryos develop, or how our bodies function, we must attain a deep understanding of proteins.

In This Chapter

THE SHAPE AND STRUCTURE OF PROTEINS

From a chemical point of view, proteins are by far the most structurally complex and functionally sophisticated molecules known. This is perhaps not surprising, once we realize that the structure and chemistry of each protein has been developed and fine-tuned over billions of years of evolutionary history. Yet, even to experts, the remarkable versatility of proteins can seem truly amazing.

In this section, we consider how the location of each amino acid in the long string of amino acids that forms a protein determines its three-dimensional shape. Later in the chapter, we use this understanding of protein structure at the atomic level to describe how the precise shape of each protein molecule determines its function in a cell.

The Shape of a Protein Is Specified by Its Amino Acid Sequence

There are 20 types of amino acids in proteins, each with different chemical properties. A **protein** molecule is made from a long chain of these amino acids, each linked to its neighbor through a covalent peptide bond. Proteins are therefore also known as *polypeptides*. Each type of protein has a unique sequence of amino acids, and there are many thousands of different proteins, each with its own particular amino acid sequence.

The repeating sequence of atoms along the core of the polypeptide chain is referred to as the **polypeptide backbone**. Attached to this repetitive chain are those portions of the amino acids that are not involved in making a peptide bond and that give each amino acid its unique properties: the 20 different amino acid **side chains** (**Figure 3–1**). Some of these side chains are nonpolar and hydrophobic ("water-fearing"), others are negatively or positively charged, some readily form covalent bonds, and so on. **Panel 3–1** (pp. 128–129) shows their atomic structures and **Figure 3–2** lists their abbreviations.

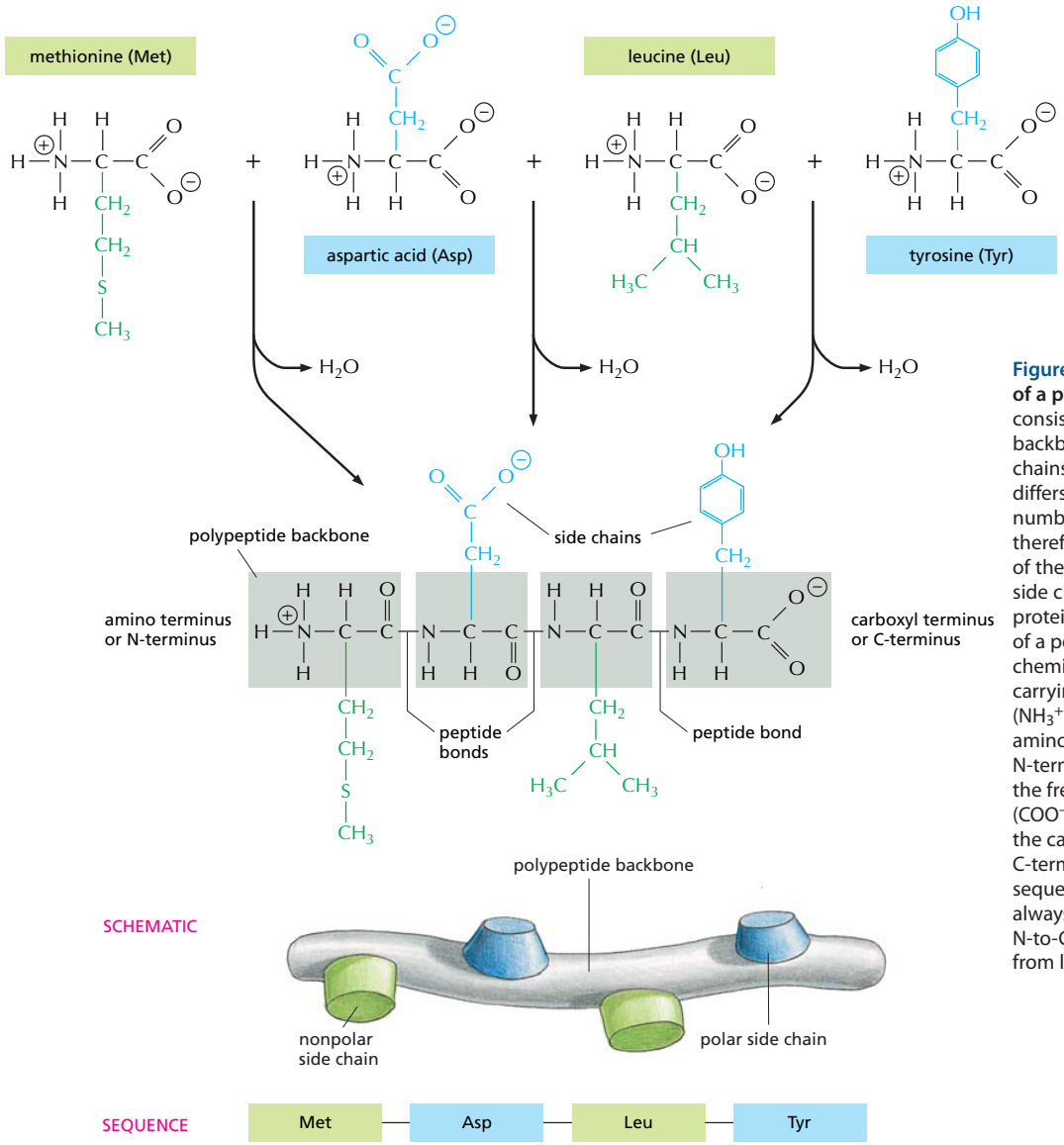


Figure 3–1 The components of a protein. A protein consists of a polypeptide backbone with attached side chains. Each type of protein differs in its sequence and number of amino acids; therefore, it is the sequence of the chemically different side chains that makes each protein distinct. The two ends of a polypeptide chain are chemically different: the end carrying the free amino group (NH_3^+, also written NH_2) is the amino terminus, or N-terminus, and that carrying the free carboxyl group (COO^-, also written COOH) is the carboxyl terminus or C-terminus. The amino acid sequence of a protein is always presented in the N-to-C direction, reading from left to right.

As discussed in Chapter 2, atoms behave almost as if they were hard spheres with a definite radius (their *van der Waals radius*). The requirement that no two atoms overlap limits greatly the possible bond angles in a polypeptide chain (**Figure 3–3**). This constraint and other steric interactions severely restrict the possible three-dimensional arrangements of atoms (or *conformations)*. Nevertheless, a long flexible chain, such as a protein, can still fold in an enormous number of ways.

The folding of a protein chain is, however, further constrained by many different sets of weak *noncovalent bonds* that form between one part of the chain and another. These involve atoms in the polypeptide backbone, as well as atoms in the amino acid side chains. There are three types of weak bonds: *hydrogen bonds*, *electrostatic attractions*, and *van der Waals attractions*, as explained in Chapter 2 (see p. 54). Individual noncovalent bonds are 30–300 times weaker than the typical covalent bonds that create biological molecules. But many weak bonds acting in parallel can hold two regions of a polypeptide chain tightly together. In this way, the combined strength of large numbers of such noncovalent bonds determines the stability of each folded shape (**Figure 3–4**).

AMINO ACID			SIDE CHAIN
Aspartic acid	Asp	D	negative
Glutamic acid	Glu	E	negative
Arginine	Arg	R	positive
Lysine	Lys	K	positive
Histidine	His	H	positive
Asparagine	Asn	N	uncharged polar
Glutamine	Gln	Q	uncharged polar
Serine	Ser	S	uncharged polar
Threonine	Thr	T	uncharged polar
Tyrosine	Tyr	Y	uncharged polar

POLAR AMINO ACIDS

AMINO ACID			SIDE CHAIN
Alanine	Ala	A	nonpolar
Glycine	Gly	G	nonpolar
Valine	Val	V	nonpolar
Leucine	Leu	L	nonpolar
Isoleucine	Ile	I	nonpolar
Proline	Pro	P	nonpolar
Phenylalanine	Phe	F	nonpolar
Methionine	Met	M	nonpolar
Tryptophan	Trp	W	nonpolar
Cysteine	Cys	C	nonpolar

NONPOLAR AMINO ACIDS

Figure 3–2 The 20 amino acids found in proteins. Each amino acid has a three-letter and a one-letter abbreviation. There are equal numbers of polar and nonpolar side chains; however, some side chains listed here as polar are large enough to have some non-polar properties (for example, Tyr, Thr, Arg, Lys). For atomic structures, see Panel 3–1 (pp. 128–129).

A fourth weak force also has a central role in determining the shape of a protein. As described in Chapter 2, hydrophobic molecules, including the nonpolar side chains of particular amino acids, tend to be forced together in an aqueous environment in order to minimize their disruptive effect on the hydrogen-bonded network of water molecules (see p. 54 and Panel 2–2, pp. 108–109). Therefore, an important factor governing the folding of any protein is the distribution of its polar and nonpolar amino acids. The nonpolar (hydrophobic) side chains in a protein—belonging to such amino acids as phenylalanine, leucine, valine, and tryptophan—tend to cluster in the interior of the molecule (just as hydrophobic oil droplets coalesce in water to form one large droplet). This enables them to avoid contact with the water that surrounds them inside a cell. In contrast, polar groups—such as those belonging to arginine, glutamine, and histidine—tend to arrange themselves near the outside of the molecule, where they can form hydrogen bonds with water and with other polar molecules (**Figure 3–5**). Polar amino acids buried within the protein are usually hydrogen-bonded to other polar amino acids or to the polypeptide backbone.

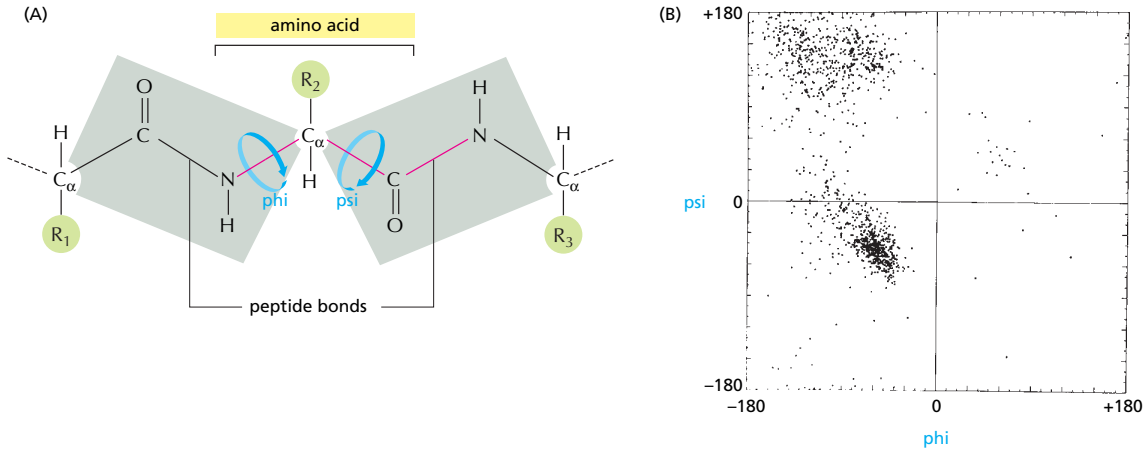
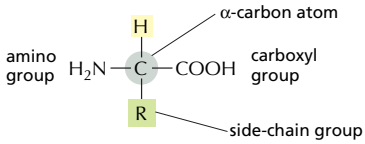


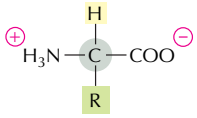
Figure 3–3 Steric limitations on the bond angles in a polypeptide chain. (A) Each amino acid contributes three bonds *(red)* to the backbone of the chain. The peptide bond is planar *(gray shading)* and does not permit rotation. By contrast, rotation can occur about the C_α–C bond, whose angle of rotation is called psi (ψ), and about the N–C_α bond, whose angle of rotation is called phi (ϕ). By convention, an R group is often used to denote an amino acid side chain *(green circles)*. (B) The conformation of the main-chain atoms in a protein is determined by one pair of ϕ and ψ angles for each amino acid; because of steric collisions between atoms within each amino acid, most pairs of ϕ and ψ angles do not occur. In this so-called Ramachandran plot, each dot represents an observed pair of angles in a protein. The cluster of dots in the bottom left quadrant represents all of the amino acids that are located in α-helix structures (see Figure 3–7A). (B, from J. Richardson, *Adv. Prot. Chem.* 34:174–175, 1981. With permission from Academic Press.)

THE AMINO ACID

The general formula of an amino acid is

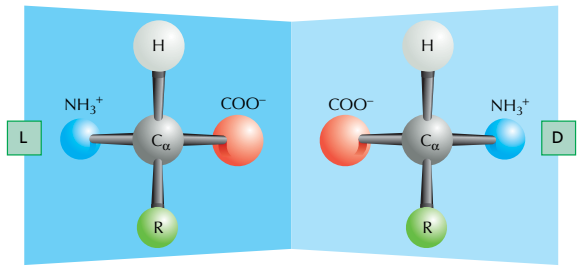


R is commonly one of 20 different side chains. At pH 7 both the amino and carboxyl groups are ionized.



OPTICAL ISOMERS

The α-carbon atom is asymmetric, which allows for two mirror image (or stereo-) isomers, L and D.



Proteins consist exclusively of L-amino acids.

FAMILIES OF AMINO ACIDS

The common amino acids are grouped according to whether their side chains are

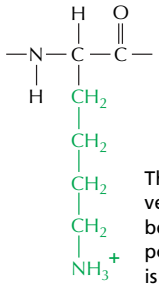
- acidic
- basic
- uncharged polar
- nonpolar

These 20 amino acids are given both three-letter and one-letter abbreviations.

Thus: alanine = Ala = A

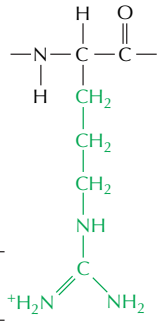
BASIC SIDE CHAINS

lysine (Lys, or K)

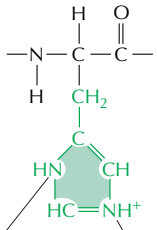


This group is very basic because its positive charge is stabilized by resonance.

arginine (Arg, or R)



histidine (His, or H)

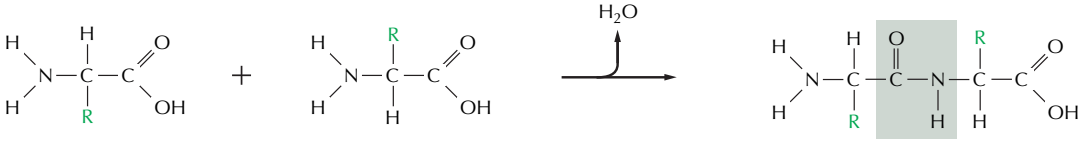


These nitrogens have a relatively weak affinity for an H^+ and are only partly positive at neutral pH.

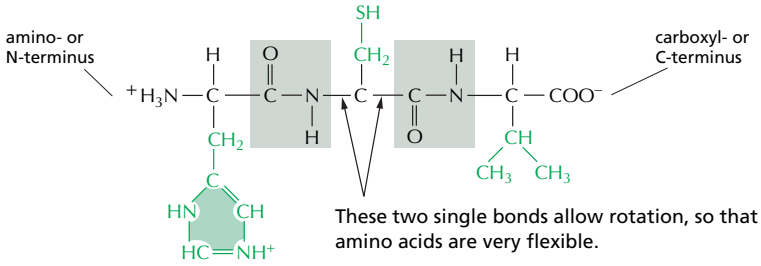
PEPTIDE BONDS

Amino acids are commonly joined together by an amide linkage, called a peptide bond.

Peptide bond: The four atoms in each *gray box* form a rigid planar unit. There is no rotation around the C–N bond.



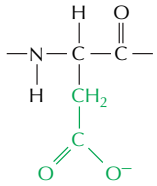
Proteins are long polymers of amino acids linked by peptide bonds, and they are always written with the N-terminus toward the left. The sequence of this tripeptide is histidine-cysteine-valine.



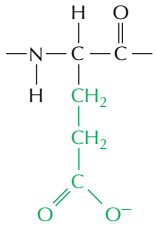
These two single bonds allow rotation, so that long chains of amino acids are very flexible.

ACIDIC SIDE CHAINS

aspartic acid
(Asp, or D)

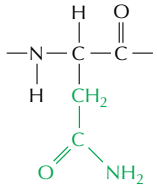


glutamic acid
(Glu, or E)

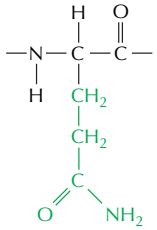


UNCHARGED POLAR SIDE CHAINS

asparagine
(Asn, or N)

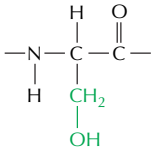


glutamine
(Gln, or Q)

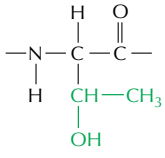


Although the amide N is not charged at neutral pH, it is polar.

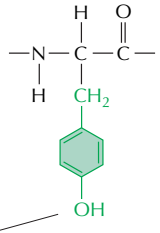
serine
(Ser, or S)



threonine
(Thr, or T)



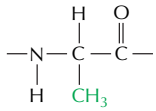
tyrosine
(Tyr, or Y)



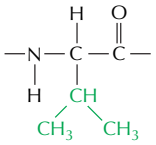
The –OH group is polar.

NONPOLAR SIDE CHAINS

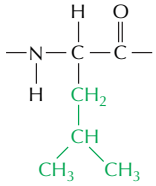
alanine
(Ala, or A)



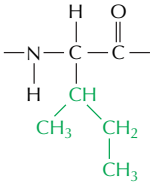
valine
(Val, or V)



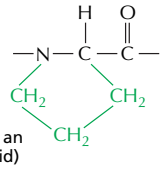
leucine
(Leu, or L)



isoleucine
(Ile, or I)

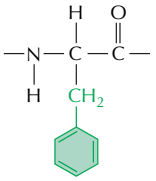


proline
(Pro, or P)

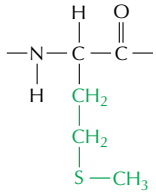


(actually an imino acid)

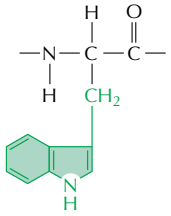
phenylalanine
(Phe, or F)



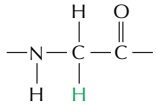
methionine
(Met, or M)



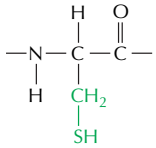
tryptophan
(Trp, or W)



glycine
(Gly, or G)



cysteine
(Cys, or C)



Disulfide bonds can form between two cysteine side chains in proteins.

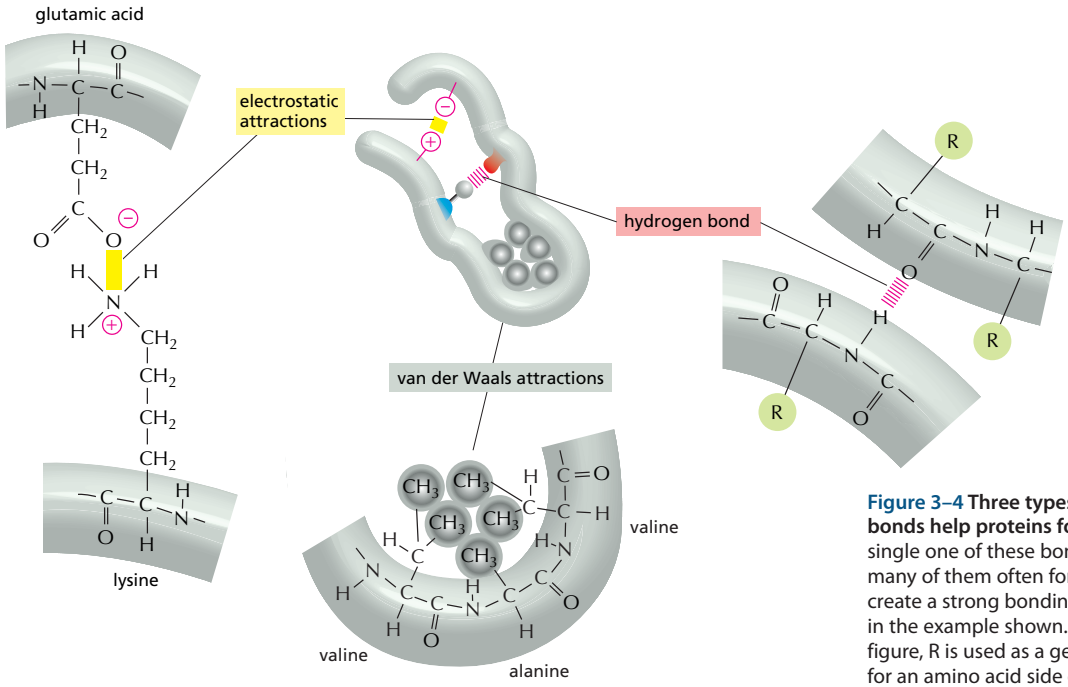


Figure 3–4 Three types of noncovalent bonds help proteins fold. Although a single one of these bonds is quite weak, many of them often form together to create a strong bonding arrangement, as in the example shown. As in the previous figure, R is used as a general designation for an amino acid side chain.

Proteins Fold into a Conformation of Lowest Energy

As a result of all of these interactions, most proteins have a particular three-dimensional structure, which is determined by the order of the amino acids in its chain. The final folded structure, or **conformation**, of any polypeptide chain is generally the one that minimizes its free energy. Biologists have studied protein folding in a test tube by using highly purified proteins. Treatment with certain solvents, which disrupt the noncovalent interactions holding the folded chain together, unfolds, or *denatures*, a protein. This treatment converts the protein into a flexible polypeptide chain that has lost its natural shape. When the denaturing solvent is removed, the protein often refolds spontaneously, or *renatures*, into its original conformation (**Figure 3–6**). This indicates that the amino acid sequence contains all the information needed for specifying the three-dimensional shape of a protein, which is a critical point for understanding cell function.

Each protein normally folds up into a single stable conformation. However, the conformation changes slightly when the protein interacts with other molecules in the cell. This change in shape is often crucial to the function of the protein, as we see later.

Although a protein chain can fold into its correct conformation without outside help, in a living cell special proteins called *molecular chaperones* often assist in protein folding. Molecular chaperones bind to partly folded polypeptide chains and help them progress along the most energetically favorable folding

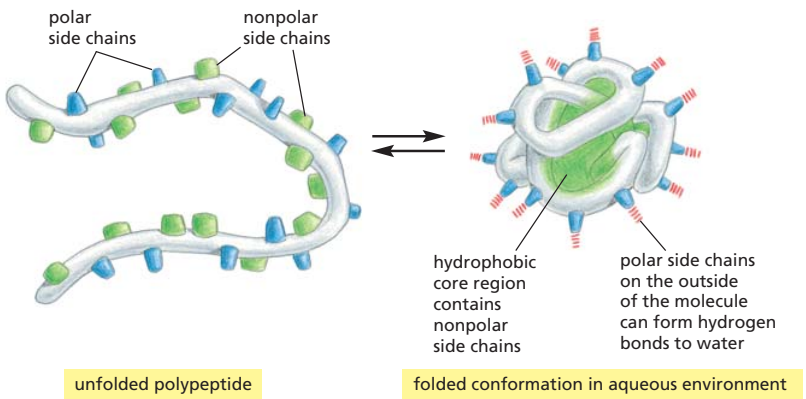


Figure 3–5 How a protein folds into a compact conformation. The polar amino acid side chains tend to gather on the outside of the protein, where they can interact with water; the nonpolar amino acid side chains are buried on the inside to form a tightly packed hydrophobic core of atoms that are hidden from water. In this schematic drawing, the protein contains only about 30 amino acids.

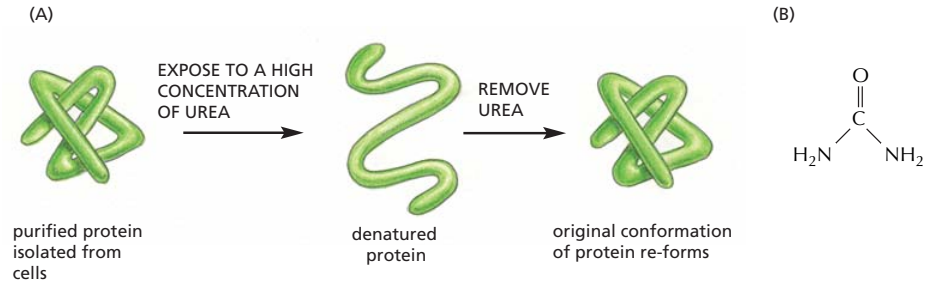


Figure 3–6 The refolding of a denatured protein. (A) This type of experiment, first performed more than 40 years ago, demonstrates that a protein's conformation is determined solely by its amino acid sequence. (B) The structure of urea. Urea is very soluble in water and unfolds proteins at high concentrations, where there is about one urea molecule for every six water molecules.

pathway. In the crowded conditions of the cytoplasm, chaperones prevent the temporarily exposed hydrophobic regions in newly synthesized protein chains from associating with each other to form protein aggregates (see p. 388). However, the final three-dimensional shape of the protein is still specified by its amino acid sequence: chaperones simply make the folding process more reliable.

Proteins come in a wide variety of shapes, and they are generally between 50 and 2000 amino acids long. Large proteins usually consist of several distinct *protein domains*—structural units that fold more or less independently of each other, as we discuss below. Since the detailed structure of any protein is complicated, several different representations are used to depict the protein's structure, each emphasizing different features.

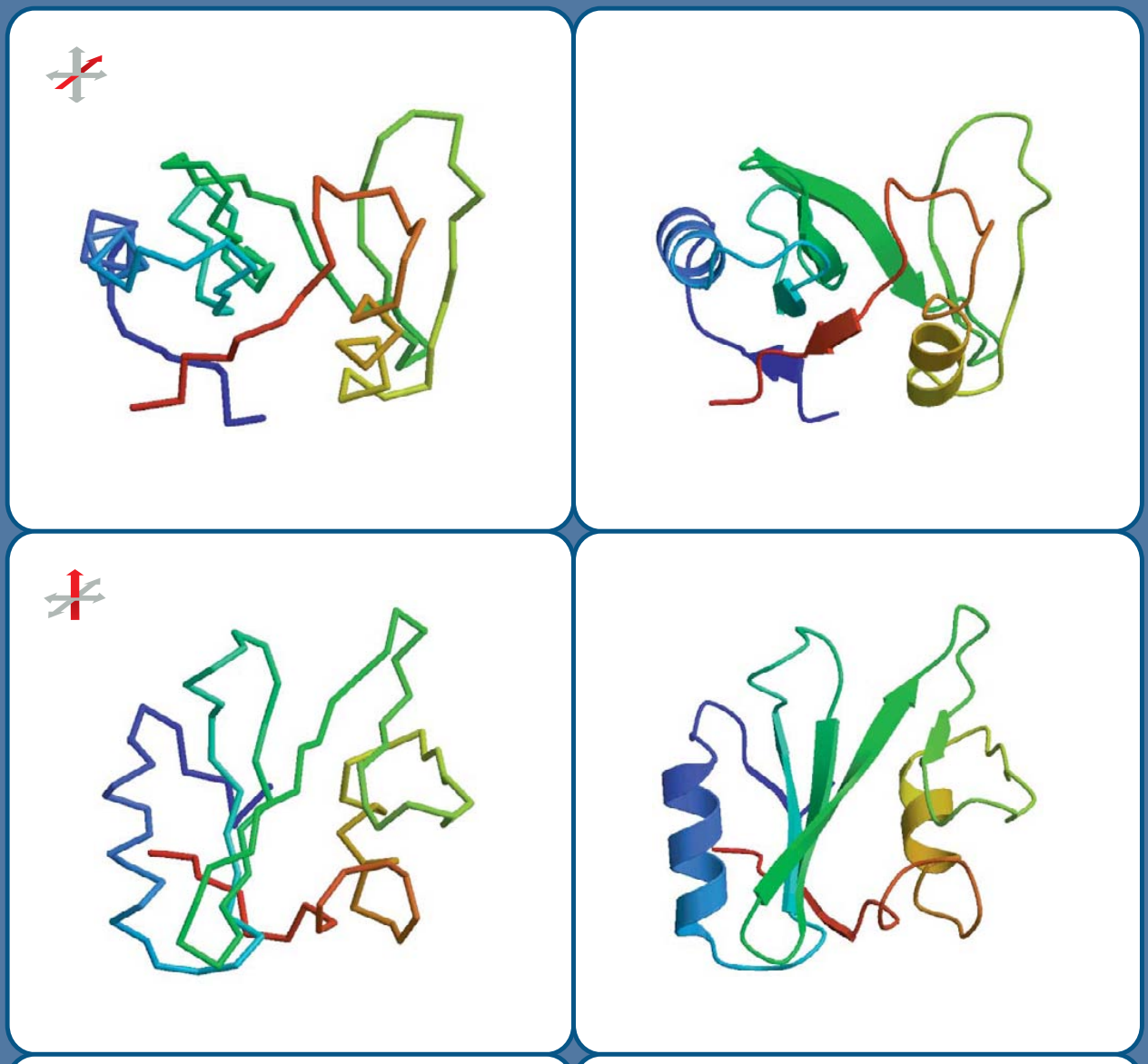
Panel 3–2 (pp. 132–133) presents four different representations of a protein domain called SH2, which has important functions in eucaryotic cells. Constructed from a string of 100 amino acids, the structure is displayed as (A) a polypeptide backbone model, (B) a ribbon model, (C) a wire model that includes the amino acid side chains, and (D) a space-filling model. Each of the three horizontal rows shows the protein in a different orientation, and the image is colored in a way that allows the polypeptide chain to be followed from its N-terminus *(purple)* to its C-terminus *(red)*. <GTGA>

Panel 3–2 shows that a protein's conformation is amazingly complex, even for a structure as small as the SH2 domain. But the description of protein structures can be simplified because they are built up from combinations of several common structural motifs, as we discuss next.

The α Helix and the β Sheet Are Common Folding Patterns

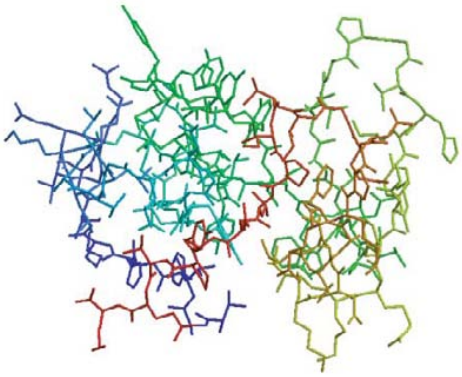
When we compare the three-dimensional structures of many different protein molecules, it becomes clear that, although the overall conformation of each protein is unique, two regular folding patterns are often found in parts of them. Both patterns were discovered more than 50 years ago from studies of hair and silk. The first folding pattern to be discovered, called the **α helix**, was found in the protein *α-keratin*, which is abundant in skin and its derivatives—such as hair, nails, and horns. Within a year of the discovery of the α helix, a second folded structure, called a **β sheet**, was found in the protein *fibroin*, the major constituent of silk. These two patterns are particularly common because they result from hydrogen-bonding between the N–H and C=O groups in the polypeptide backbone, without involving the side chains of the amino acids. Thus, many different amino acid sequences can form them. In each case, the protein chain adopts a regular, repeating conformation. **Figure 3–7** shows these two conformations, as well as the abbreviations that are used to denote them in ribbon models of proteins.

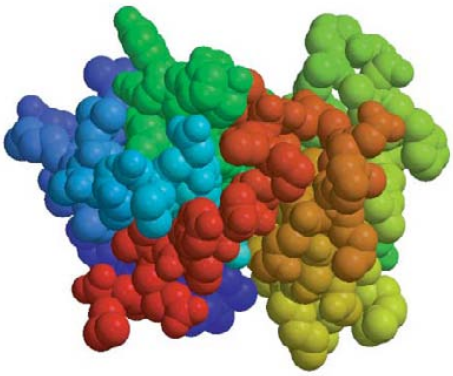
The core of many proteins contains extensive regions of β sheet. As shown in **Figure 3–8**, these β sheets can form either from neighboring polypeptide chains that run in the same orientation (parallel chains) or from a polypeptide chain that folds back and forth upon itself, with each section of the chain running in the direction opposite to that of its immediate neighbors (antiparallel chains). Both types of β sheet produce a very rigid structure, held together by hydrogen bonds that connect the peptide bonds in neighboring chains (see Figure 3–7D).

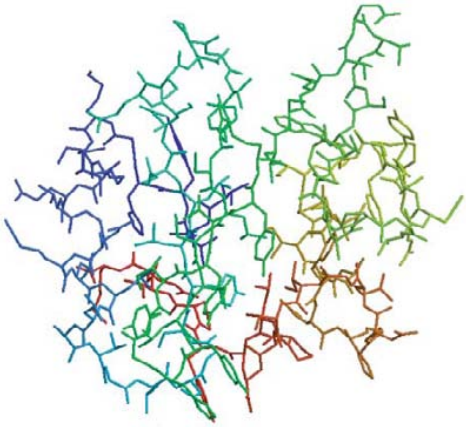


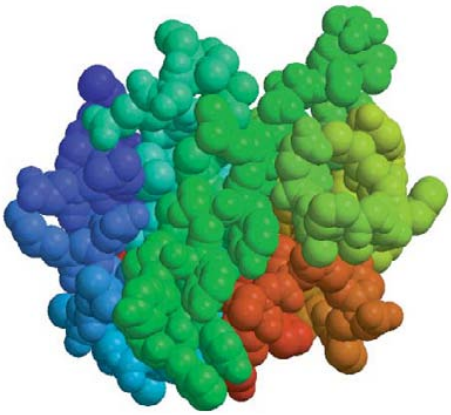
(A) Backbone: Shows the overall organization of the polypeptide chain; a clean way to compare structures of related proteins.

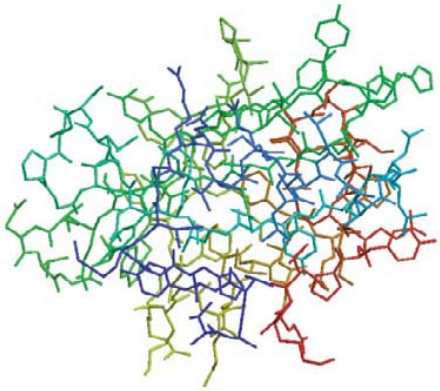
(B) Ribbon: Easy way to visualize secondary structures, such as α helices and β sheets.



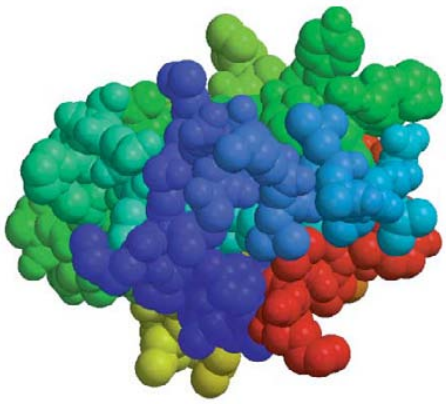








(C) Wire: Highlights side chains and their relative proximities; useful for predicting which amino acids might be involved in a protein's activity, particularly if the protein is an enzyme.



(D) Space-filling: Provides contour map of the protein; gives a feel for the shape of the protein and shows which amino acid side chains are exposed on its surface. Shows how the protein might look to a small molecule, such as water, or to another protein.

(Courtesy of David Lawson.)

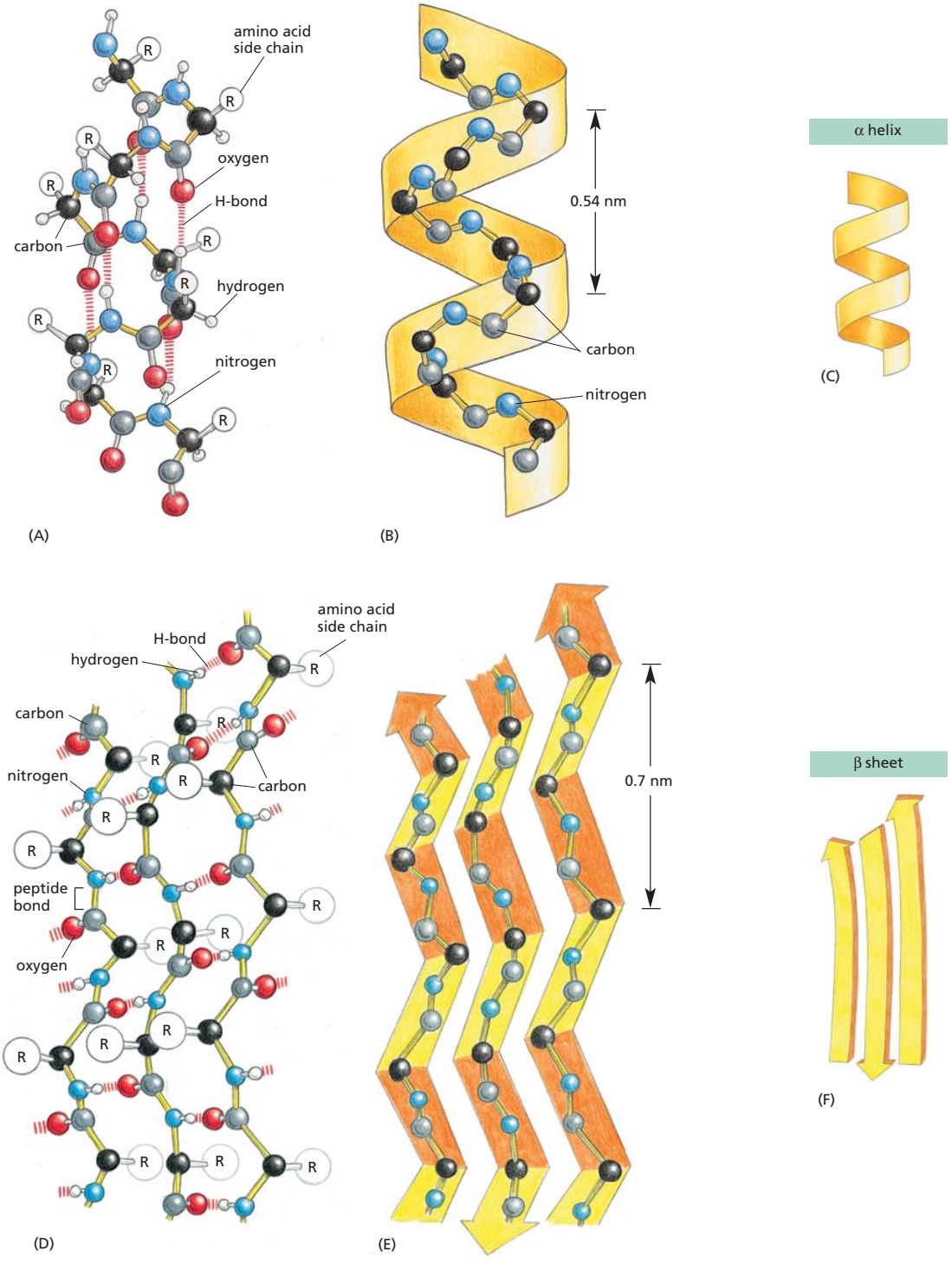


Figure 3–7 The regular conformation of the polypeptide backbone in the α helix and the β sheet. <GTAG> <TGCT>
(A, B, and C) The α helix. The N–H of every peptide bond is hydrogen-bonded to the C=O of a neighboring peptide bond located four peptide bonds away in the same chain. Note that all of the N–H groups point up in this diagram and that all of the C=O groups point down (toward the C-terminus); this gives a polarity to the helix, with the C-terminus having a partial negative and the N-terminus a partial positive charge. (D, E, and F) The β sheet. In this example, adjacent peptide chains run in opposite (antiparallel) directions. Hydrogen-bonding between peptide bonds in different strands holds the individual polypeptide chains (strands) together in a β sheet, and the amino acid side chains in each strand alternately project above and below the plane of the sheet. (A) and (D) show all the atoms in the polypeptide backbone, but the amino acid side chains are truncated and denoted by R. In contrast, (B) and (E) show the backbone atoms only, while (C) and (F) display the shorthand symbols that are used to represent the α helix and the β sheet in ribbon drawings of proteins (see Panel 3–2B).

An α helix is generated when a single polypeptide chain twists around on itself to form a rigid cylinder. A hydrogen bond forms between every fourth peptide bond, linking the C=O of one peptide bond to the N–H of another (see Figure 3–7A). This gives rise to a regular helix with a complete turn every 3.6 amino acids. Note that the protein domain illustrated in Panel 3–2 contains two α helices, as well as a three-stranded antiparallel β sheet.

Regions of α helix are especially abundant in proteins located in cell membranes, such as transport proteins and receptors. As we discuss in Chapter 10, those portions of a transmembrane protein that cross the lipid bilayer usually cross as an α helix composed largely of amino acids with nonpolar side chains. The polypeptide backbone, which is hydrophilic, is hydrogen-bonded to itself in the α helix and shielded from the hydrophobic lipid environment of the membrane by its protruding nonpolar side chains (see also Figure 3–78).

In other proteins, α helices wrap around each other to form a particularly stable structure, known as a **coiled-coil**. This structure can form when the two (or in some cases three) α helices have most of their nonpolar (hydrophobic) side chains on one side, so that they can twist around each other with these side chains facing inward (**Figure 3–9**). Long rodlike coiled-coils provide the structural framework for many elongated proteins. Examples are α-keratin, which forms the intracellular fibers that reinforce the outer layer of the skin and its appendages, and the myosin molecules responsible for muscle contraction.

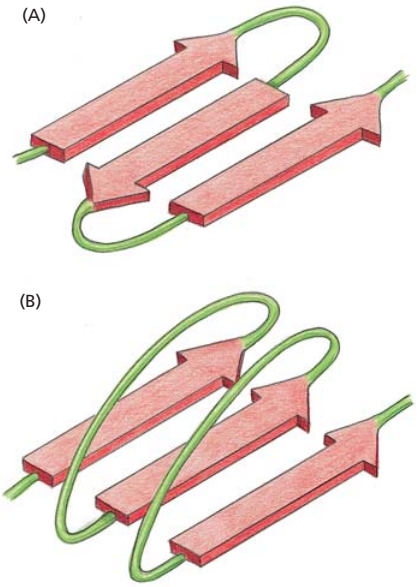


Figure 3–8 Two types of β sheet structures. (A) An antiparallel β sheet (see Figure 3–7D). (B) A parallel β sheet. Both of these structures are common in proteins.

Protein Domains Are Modular Units from which Larger Proteins Are Built

Even a small protein molecule is built from thousands of atoms linked together by precisely oriented covalent and noncovalent bonds, and it is extremely difficult to visualize such a complicated structure without a three-dimensional display. For

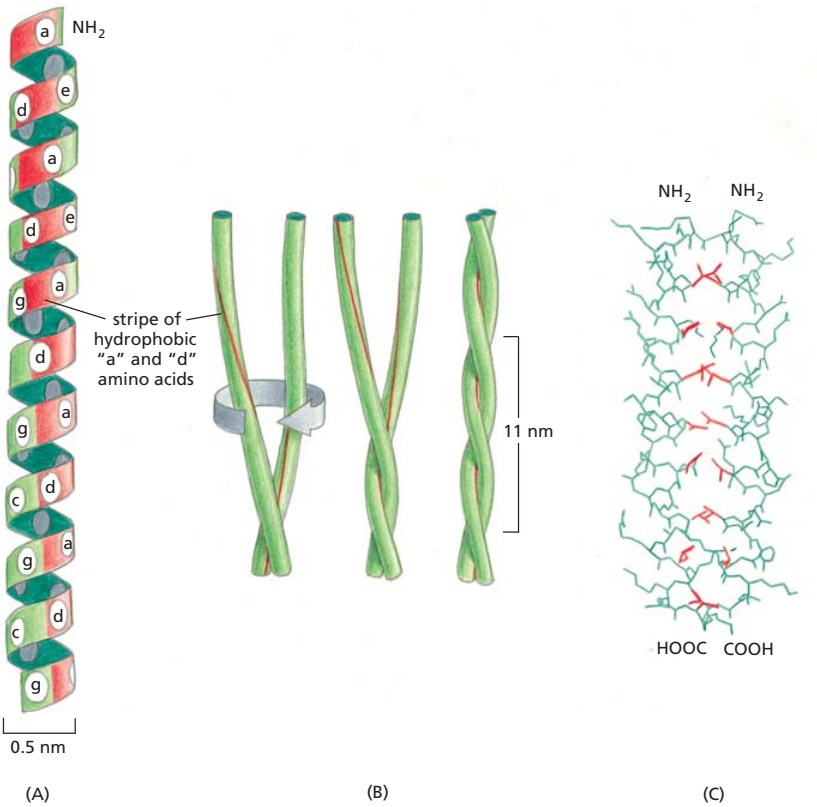


Figure 3–9 A coiled-coil. <CGGA> (A) A single α helix, with successive amino acid side chains labeled in a sevenfold sequence, "abcdefg" (*from bottom to top*). Amino acids "a" and "d" in such a sequence lie close together on the cylinder surface, forming a "stripe" (*red*) that winds slowly around the α helix. Proteins that form coiled-coils typically have nonpolar amino acids at positions "a" and "d." Consequently, as shown in (B), the two α helices can wrap around each other with the nonpolar side chains of one α helix interacting with the nonpolar side chains of the other, while the more hydrophilic amino acid side chains are left exposed to the aqueous environment. (C) The atomic structure of a coiled-coil determined by x-ray crystallography. The *red* side chains are nonpolar.

this reason, biologists use various graphic and computer-based aids. A DVD that accompanies this book contains computer-generated images of selected proteins, displayed and rotated on the screen in a variety of formats.

Biologists distinguish four levels of organization in the structure of a protein. The amino acid sequence is known as the **primary structure**. Stretches of polypeptide chain that form α helices and β sheets constitute the protein's **secondary structure**. The full three-dimensional organization of a polypeptide chain is sometimes referred to as the **tertiary structure**, and if a particular protein molecule is formed as a complex of more than one polypeptide chain, the complete structure is designated as the **quaternary structure**.

Studies of the conformation, function, and evolution of proteins have also revealed the central importance of a unit of organization distinct from these four. This is the **protein domain**, a substructure produced by any part of a polypeptide chain that can fold independently into a compact, stable structure. A domain usually contains between 40 and 350 amino acids, and it is the modular unit from which many larger proteins are constructed.

The different domains of a protein are often associated with different functions. **Figure 3–10** shows an example—the Src protein kinase, which functions in signaling pathways inside vertebrate cells (Src is pronounced "sarc"). This protein is considered to have three domains: the SH2 and SH3 domains have regulatory roles, while the C-terminal domain is responsible for the kinase catalytic activity. Later in the chapter, we shall return to this protein, in order to explain how proteins can form molecular switches that transmit information throughout cells.

Figure 3–11 presents ribbon models of three differently organized protein domains. As these examples illustrate, the polypeptide chain tends to cross the entire domain before making a sharp turn at the surface. The central core of a domain can be constructed from α helices, from β sheets, or from various combinations of these two fundamental folding elements. <CAGT>

The smallest protein molecules contain only a single domain, whereas larger proteins can contain as many as several dozen domains, often connected to each other by short, relatively unstructured lengths of polypeptide chain.

Few of the Many Possible Polypeptide Chains Will Be Useful to Cells

Since each of the 20 amino acids is chemically distinct and each can, in principle, occur at any position in a protein chain, there are $20 \times 20 \times 20 \times 20 = 160{,}000$ different possible polypeptide chains four amino acids long, or 20^n different possible polypeptide chains n amino acids long. For a typical protein length of

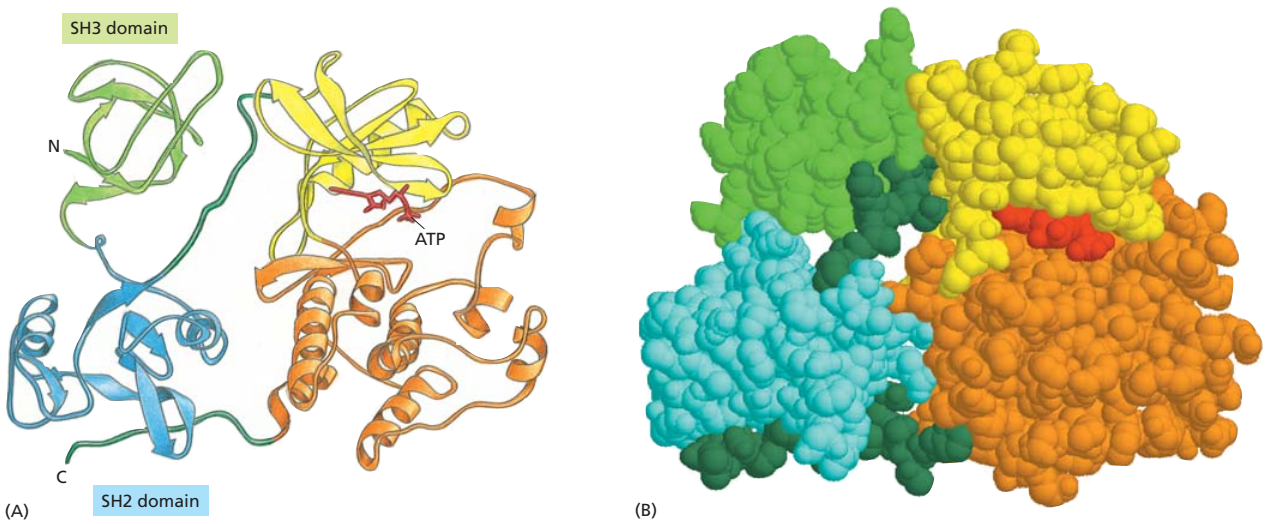


Figure 3–10 A protein formed from multiple domains. In the Src protein shown, a C-terminal domain with two lobes (*yellow* and *orange*) forms a protein kinase enzyme, while the SH2 and SH3 domains perform regulatory functions. (A) A ribbon model, with ATP substrate in *red*. (B) A spacing-filling model, with ATP substrate in *red*. Note that the site that binds ATP is positioned at the interface of the two lobes that form the kinase. The detailed structure of the SH2 domain is illustrated in Panel 3–2 (pp. 132–133).

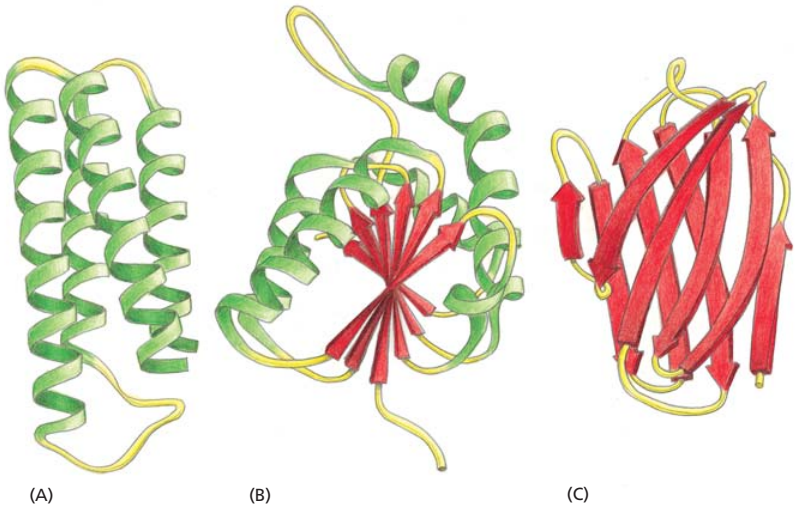


Figure 3–11 Ribbon models of three different protein domains. (A) Cytochrome b_{562}, a single-domain protein involved in electron transport in mitochondria. This protein is composed almost entirely of α helices. (B) The NAD-binding domain of the enzyme lactic dehydrogenase, which is composed of a mixture of α helices and parallel β sheets. (C) The variable domain of an immunoglobulin (antibody) light chain, composed of a sandwich of two antiparallel β sheets. In these examples, the α helices are shown in *green*, while strands organized as β sheets are denoted by *red arrows*.

Note how the polypeptide chain generally traverses back and forth across the entire domain, making sharp turns only at the protein surface. It is the protruding loop regions *(yellow)* that often form the binding sites for other molecules. (Adapted from drawings courtesy of Jane Richardson.)

about 300 amino acids, a cell could theoretically make more than 10^{390} (20^{300}) different polypeptide chains. This is such an enormous number that to produce just one molecule of each kind would require many more atoms than exist in the universe.

Only a very small fraction of this vast set of conceivable polypeptide chains would adopt a single, stable three-dimensional conformation—by some estimates, less than one in a billion. And yet the vast majority of proteins present in cells adopt unique and stable conformations. How is this possible? The answer lies in natural selection. A protein with an unpredictably variable structure and biochemical activity is unlikely to help the survival of a cell that contains it. Such proteins would therefore have been eliminated by natural selection through the enormously long trial-and-error process that underlies biological evolution.

Because evolution has selected for protein function in living organisms, the amino acid sequence of most present-day proteins is such that a single conformation is extremely stable. In addition, this conformation has its chemical properties finely tuned to enable the protein to perform a particular catalytic or structural function in the cell. Proteins are so precisely built that the change of even a few atoms in one amino acid can sometimes disrupt the structure of the whole molecule so severely that all function is lost.

Proteins Can Be Classified into Many Families

Once a protein had evolved that folded up into a stable conformation with useful properties, its structure could be modified during evolution to enable it to perform new functions. This process has been greatly accelerated by genetic mechanisms that occasionally duplicate genes, allowing one gene copy to evolve independently to perform a new function (discussed in Chapter 4). This type of event has occurred quite often in the past; as a result, many present-day proteins can be grouped into protein families, each family member having an amino acid sequence and a three-dimensional conformation that resemble those of the other family members.

Consider, for example, the *serine proteases*, a large family of protein-cleaving (proteolytic) enzymes that includes the digestive enzymes chymotrypsin, trypsin, and elastase, and several proteases involved in blood clotting. When the protease portions of any two of these enzymes are compared, parts of their amino acid sequences are found to match. The similarity of their three-dimensional conformations is even more striking: most of the detailed twists and turns in their polypeptide chains, which are several hundred amino acids long, are virtually identical (**Figure 3–12**). The many different serine proteases nevertheless

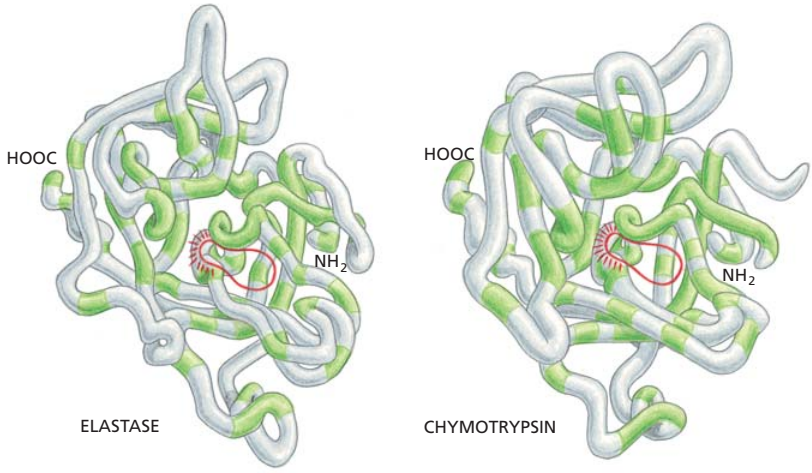


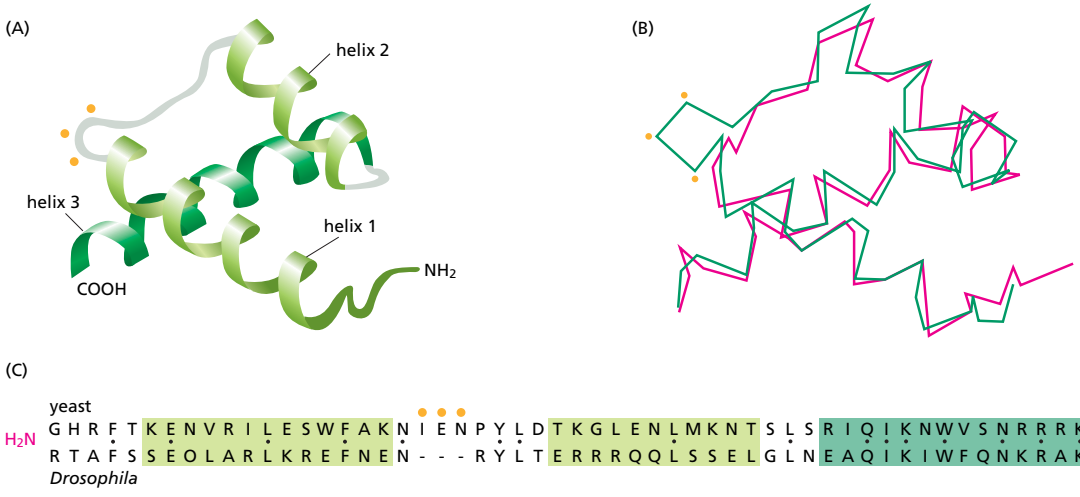
Figure 3–12 A comparison of the conformations of two serine proteases. The backbone conformations of elastase and chymotrypsin. Although only those amino acids in the polypeptide chain shaded in *green* are the same in the two proteins, the two conformations are very similar nearly everywhere. The active site of each enzyme is circled in *red;* this is where the peptide bonds of the proteins that serve as substrates are bound and cleaved by hydrolysis. The serine proteases derive their name from the amino acid serine, whose side chain is part of the active site of each enzyme and directly participates in the cleavage reaction.

have distinct enzymatic activities, each cleaving different proteins or the peptide bonds between different types of amino acids. Each therefore performs a distinct function in an organism.

The story we have told for the serine proteases could be repeated for hundreds of other protein families. In general, the structure of the different members of a protein family has been more highly conserved than has the amino acid sequence. In many cases, the amino acid sequences have diverged so far that we cannot be certain of a family relationship between two proteins without determining their three-dimensional structures. The yeast α2 protein and the *Drosophila* engrailed protein, for example, are both gene regulatory proteins in the homeodomain family (discussed in Chapter 7). Because they are identical in only 17 of their 60 amino acid residues, their relationship became certain only by comparing their three-dimensional structures (**Figure 3–13**). Many similar examples show that two proteins with more than 25% identity in their amino acid sequences usually share the same overall structure.

The various members of a large protein family often have distinct functions. Some of the amino acid changes that make family members different were no doubt selected in the course of evolution because they resulted in useful changes in biological activity, giving the individual family members the different functional properties they have today. But many other amino acid changes are effectively "neutral," having neither a beneficial nor a damaging effect on the basic structure and function of the protein. In addition, since mutation is a random process, there must also have been many deleterious changes that altered the three-dimensional structure of these proteins sufficiently to harm them.

Figure 3–13 A comparison of a class of DNA-binding domains, called homeodomains, in a pair of proteins from two organisms separated by more than a billion years of evolution. (A) A ribbon model of the structure common to both proteins. (B) A trace of the α-carbon positions. The three-dimensional structures shown were determined by x-ray crystallography for the yeast α2 protein *(green)* and the *Drosophila* engrailed protein *(red)*. (C) A comparison of amino acid sequences for the region of the proteins shown in (A) and (B). *Black dots* mark sites with identical amino acids. *Orange dots* indicate the position of a three amino acid insert in the α2 protein. (Adapted from C. Wolberger et al., *Cell* 67:517–528, 1991. With permission from Elsevier.)



Such faulty proteins would have been lost whenever the individual organisms making them were at enough of a disadvantage to be eliminated by natural selection.

Protein families are readily recognized when the genome of any organism is sequenced; for example, the determination of the DNA sequence for the entire human genome has revealed that we contain about 24,000 protein-coding genes. Through sequence comparisons, we can assign the products of about 40 percent of these genes to known protein structures, belonging to more than 500 different protein families. Most of the proteins in each family have evolved to perform somewhat different functions, as for the enzymes elastase and chymotrypsin illustrated previously in Figure 3–12. These are sometimes called *paralogs* to distinguish them from the corresponding proteins in different organisms (*orthologs*, such as mouse and human elastase).

As described in Chapter 8, because of the powerful techniques of x-ray crystallography and nuclear magnetic resonance (NMR), we now know the three-dimensional shapes, or conformations, of more than 20,000 proteins. <GGCC> By carefully comparing the conformations of these proteins, structural biologists (that is, experts on the structure of biological molecules) have concluded that there are a limited number of ways in which protein domains fold up in nature—maybe as few as 2000. The structures for about 800 of these protein folds have thus far been determined. These known folds tend to be those most represented in the universe of protein structures: for example, 50 folds account for nearly three-fourths of the domain families with predicted structures. A complete catalog of the most significant protein folds that exist in living organisms would therefore seem to be within our reach.

Sequence Searches Can Identify Close Relatives

The present database of known protein sequences contains more than ten million entries, and it is growing very rapidly as more and more genomes are sequenced—revealing huge numbers of new genes that encode proteins. Powerful computer search programs are available that allow us to compare each newly discovered protein with this entire database, looking for possible relatives. Many proteins whose genes have evolved from a common ancestral gene can be identified by the discovery of statistically significant similarities in amino acid sequences.

With such a large number of proteins in the database, the search programs find many nonsignificant matches, resulting in a background noise level that makes it very difficult to pick out all but the closest relatives. Generally speaking, one requires a 30% identity in sequence to consider that two proteins match. However, we know the function of many short signature sequences ("fingerprints"), and these are widely used to find more distant relationships (**Figure 3–14**).

Protein comparisons are important because related structures often imply related functions. Many years of experimentation can be saved by discovering that a new protein has an amino acid sequence similarity with a protein of known function. Such sequence relationships, for example, first indicated that certain genes that cause mammalian cells to become cancerous are protein kinases. In the same way, many of the proteins that control pattern formation during the embryonic development of the fruit fly *Drosophila* were quickly recognized to be gene regulatory proteins.

```
WYFGKITRRESERLL         GTFLVRESE                    –   signature sequences
WYFGKITRRESERLLLNAENPRGTFLVRESETTKGAYCLSVSDFDNAKGL –   human
W+F   + R+E+++LLL   ENP GTFLVR SE      Y LSV D+++ +G  –   sequence matches
WFFENVLRKEADKLLLAEENPEGTFLVRPSEHNPNGYSLSVKDWEDGRGY –   Drosophila
1          10          20          30          40          50
```

Figure 3–14 The use of short signature sequences to find related protein domains. The two short sequences of 15 and 9 amino acids shown *(green)* can be used to search large databases for a protein domain that is found in many proteins, the SH2 domain. Here, the first 50 amino acids of the SH2 domain of 100 amino acids is compared for the human and *Drosophila* Src protein (see Figure 3–10). In the computer-generated sequence comparison *(yellow row),* exact matches between the human and *Drosophila* proteins are noted by the one-letter abbreviation for the amino acid; the positions with a similar but nonidentical amino acid are denoted by +, and nonmatches are blank. In this diagram, wherever one or both proteins contain an exact match to a position in the *green* sequences, both aligned sequences are colored *red*.

Some Protein Domains Form Parts of Many Different Proteins

As previously stated, most proteins are composed of a series of protein domains, in which different regions of the polypeptide chain have folded independently to form compact structures. Such multidomain proteins are believed to have originated from the accidental joining of the DNA sequences that encode each domain, creating a new gene. Novel binding surfaces have often been created at the juxtaposition of domains, and many of the functional sites where proteins bind to small molecules are found to be located there. In an evolutionary process called *domain shuffling*, many large proteins have evolved through the joining of preexisting domains in new combinations (**Figure 3–15**).

A subset of protein domains have been especially mobile during evolution; these seem to have particularly versatile structures and are sometimes referred to as *protein modules*. The structure of one such module, the SH2 domain, was illustrated in Panel 3–2 (pp. 132–133). Some other abundant protein domains are illustrated in **Figure 3–16**.

Each of the domains shown has a stable core structure formed from strands of β sheet, from which less-ordered loops of polypeptide chain protrude *(green)*. The loops are ideally situated to form binding sites for other molecules, as most clearly demonstrated for the immunoglobulin fold, which forms the basis for antibody molecules (see Figure 3–41). Most likely, such β-sheet-based domains have achieved their evolutionary success because they provide a convenient framework for the generation of new binding sites for ligands through small changes to their protruding loops.

A second feature of these protein domains that explains their utility is the ease with which they can be integrated into other proteins. Five of the six domains illustrated in Figure 3–16 have their N- and C-terminal ends at opposite poles of the domain. When the DNA encoding such a domain undergoes tandem duplication, which is not unusual in the evolution of genomes (discussed in Chapter 4), the duplicated modules with this "in-line" arrangement

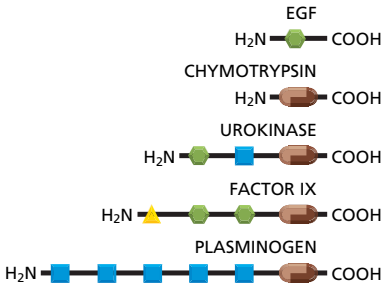


Figure 3–15 Domain shuffling. An extensive shuffling of blocks of protein sequence (protein domains) has occurred during protein evolution. Those portions of a protein denoted by the same shape and color in this diagram are evolutionarily related. Serine proteases like chymotrypsin are formed from two domains *(brown)*. In the three other proteases shown, which are highly regulated and more specialized, these two protease domains are connected to one or more domains that are similar to domains found in epidermal growth factor (EGF; *green)*, to a calcium-binding protein *(yellow)*, or to a "kringle" domain *(blue)* that contains three internal disulfide bridges. Chymotrypsin is illustrated in Figure 3–12.

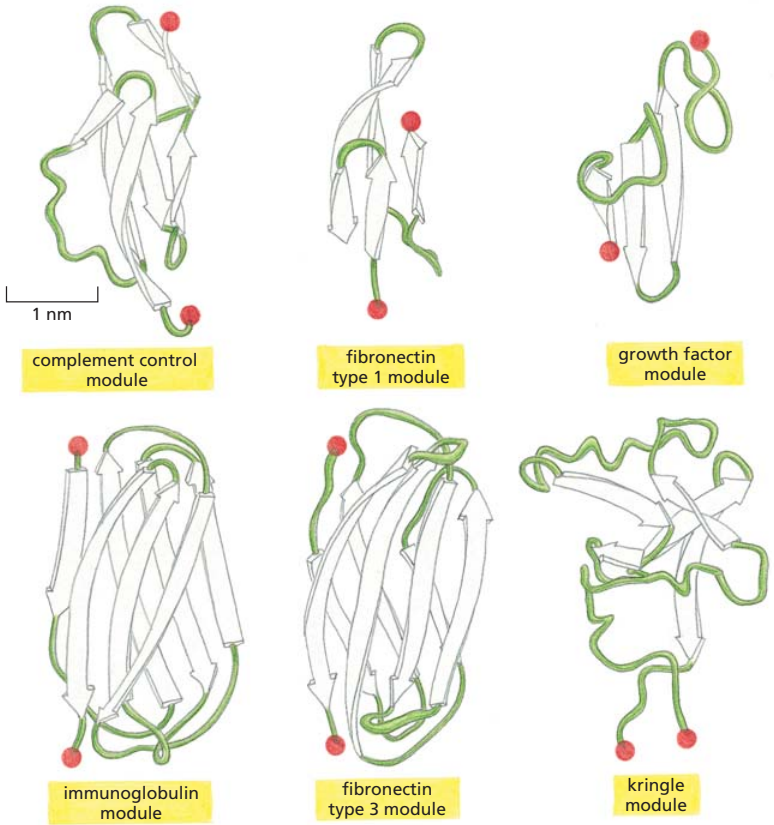


Figure 3–16 The three-dimensional structures of some protein modules. In these ribbon diagrams, β-sheet strands are shown as *arrows*, and the N- and C-termini are indicated by *red spheres*. (Adapted from M. Baron, D.G. Norman and I.D. Campbell, *Trends Biochem. Sci.* 16:13–17, 1991, with permission from Elsevier, and D.J. Leahy et al., *Science* 258:987–991, 1992, with permission from AAAS.)

can be readily linked in series to form extended structures—either with themselves or with other in-line domains (**Figure 3–17**). Stiff extended structures composed of a series of domains are especially common in extracellular matrix molecules and in the extracellular portions of cell-surface receptor proteins. Other modules, including the SH2 domain and the kringle domain illustrated in Figure 3–16, are of a "plug-in" type, with their N- and C-termini close together. After genomic rearrangements, such modules are usually accommodated as an insertion into a loop region of a second protein.

A comparison of the relative frequency of domain utilization in different eucaryotes reveals that, for many common domains, such as protein kinases, this frequency is similar in organisms as diverse as yeast, plants, worms, flies, and humans (**Figure 3–18**). But there are some notable exceptions, such as the Major Histocompatibility Complex (MHC) antigen-recognition domain (see Figure 25–52) that is present in 57 copies in humans, but absent in the other four organisms just mentioned. Such domains presumably have specialized functions that are not shared with the other eucaryotes, being strongly selected for during evolution so as to give rise to the multiple copies observed. Similarly, a domain like SH2 that shows an unusual increase in its numbers in higher eucaryotes might be assumed to be especially useful for multicellularity (compare the multicellular organisms with yeast in Figure 3–18).

Certain Pairs of Domains Are Found Together in Many Proteins

We can construct a large table displaying domain usage for each organism whose genome sequence is known. For example, the human genome is estimated to contain about 1000 immunoglobulin domains, 500 protein kinase domains, 250 DNA-binding homeodomains, 300 SH3 domains, and 120 SH2 domains. Important additional information can be derived by comparing the frequencies and arrangements of domains in the more than 100 eucaryotic, bacterial, and archaeal genomes that have been completely sequenced. For example, we find that more than two-thirds of proteins consist of two or more domains, and that the same pairs of domains occur repeatedly in the same relative arrangement in a protein. Although half of all domain families are common to archaea, bacteria, and eucaryotes, only about 5 percent of the two-domain combinations are similarly shared. This pattern suggests that most proteins containing especially useful two-domain combinations arose relatively late in evolution.

The 200 most abundant two-domain combinations occur in about one-fourth of all of the proteins with recognizable domains in the complete data set. It would therefore be very useful to determine the precise three-dimensional structure for at least one protein from each common two-domain combination, so as to reveal how the domains interact in that type of protein structure.

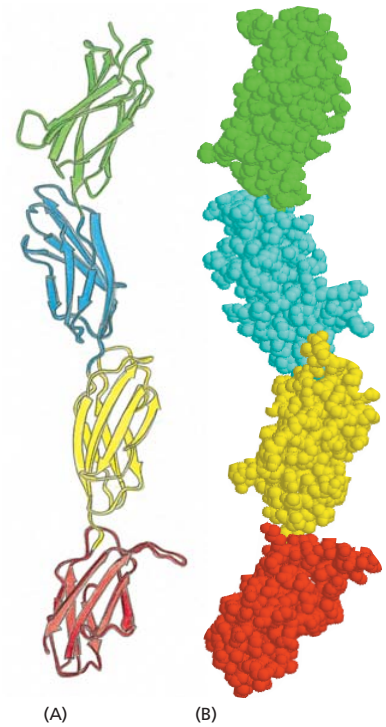


Figure 3–17 An extended structure formed from a series of in-line protein modules. Four fibronectin type 3 modules (see Figure 3–16) from the extracellular matrix molecule fibronectin are illustrated in (A) ribbon and (B) space-filling models. (Adapted from D.J. Leahy, I. Aukhil and H.P. Erickson, *Cell* 84:155–164, 1996. With permission from Elsevier.)

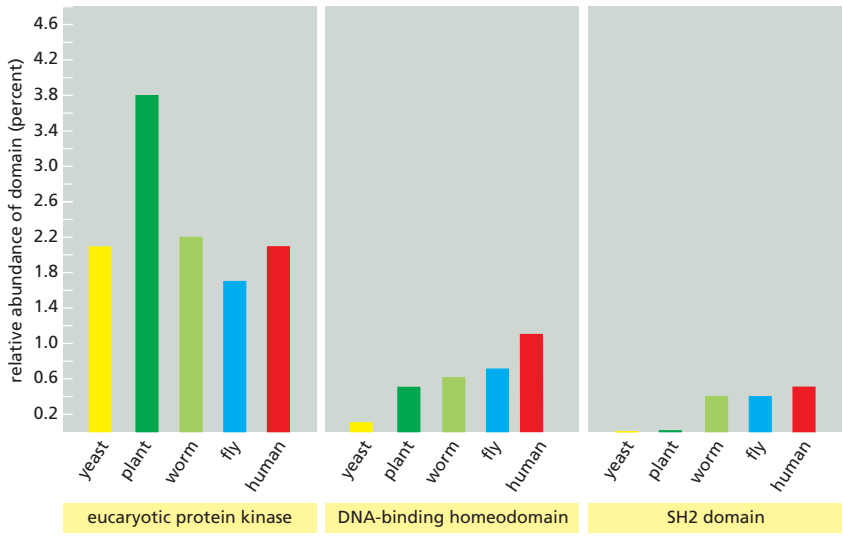


Figure 3–18 The relative frequencies of three protein domains in five eucaryotic organisms. The approximate percentages given have been determined by dividing the number of copies of each domain by the total number of distinct proteins thought to be encoded by each organism, as determined from the sequence of its genome. Thus, for SH2 domains in humans, 120/24,000 = 0.005.

The Human Genome Encodes a Complex Set of Proteins, Revealing Much That Remains Unknown

The result of sequencing the human genome has been surprising, because it reveals that our chromosomes contain only about 25,000 genes. Based on gene number alone, we would appear to be no more complex than the tiny mustard weed, *Arabidopsis,* and only about 1.3-fold more complex than a nematode worm. The genome sequences also reveal that vertebrates have inherited nearly all of their protein domains from invertebrates—with only 7 percent of identified human domains being vertebrate-specific.

Each of our proteins is on average more complicated, however (**Figure 3–19**). Domain shuffling during vertebrate evolution has given rise to many novel combinations of protein domains, with the result that there are nearly twice as many combinations of domains found in human proteins as in a worm or a fly. Thus, for example, the trypsinlike serine protease domain is linked to at least 18 other types of protein domains in human proteins, whereas it is found covalently joined to only 5 different domains in the worm. This extra variety in our proteins greatly increases the range of protein–protein interactions possible (see Figure 3–82), but how it contributes to making us human is not known.

The complexity of living organisms is staggering, and it is quite sobering to note that we currently lack even the tiniest hint of what the function might be for more than 10,000 of the proteins that have thus far been identified in the human genome. There are certainly enormous challenges ahead for the next generation of cell biologists, with no shortage of fascinating mysteries to solve.

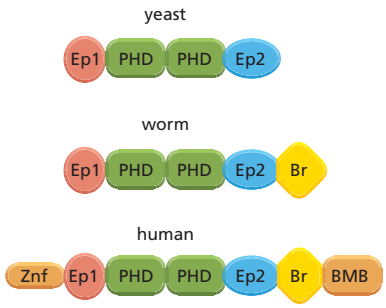


Figure 3–19 Domain structure of a group of evolutionarily related proteins that are thought to have a similar function. In general, there is a tendency for the proteins in more complex organisms, such as humans, to contain additional domains—as is the case for the DNA-binding protein compared here.

Larger Protein Molecules Often Contain More Than One Polypeptide Chain <GCCT>

The same weak noncovalent bonds that enable a protein chain to fold into a specific conformation also allow proteins to bind to each other to produce larger structures in the cell. Any region of a protein's surface that can interact with another molecule through sets of noncovalent bonds is called a **binding site**. A protein can contain binding sites for various large and small molecules. If a binding site recognizes the surface of a second protein, the tight binding of two folded polypeptide chains at this site creates a larger protein molecule with a precisely defined geometry. Each polypeptide chain in such a protein is called a **protein subunit**.

In the simplest case, two identical folded polypeptide chains bind to each other in a "head-to-head" arrangement, forming a symmetric complex of two protein subunits (a *dimer*) held together by interactions between two identical binding sites. The *Cro repressor protein*—a viral gene regulatory protein that binds to DNA to turn viral genes off in an infected bacterial cell—provides an example (**Figure 3–20**). Cells contain many other types of symmetric protein complexes, formed from multiple copies of a single polypeptide chain. The enzyme *neuraminidase*, for example, consists of four identical protein subunits, each bound to the next in a "head-to-tail" arrangement that forms a closed ring (**Figure 3–21**).

Many of the proteins in cells contain two or more types of polypeptide chains. *Hemoglobin*, the protein that carries oxygen in red blood cells, contains

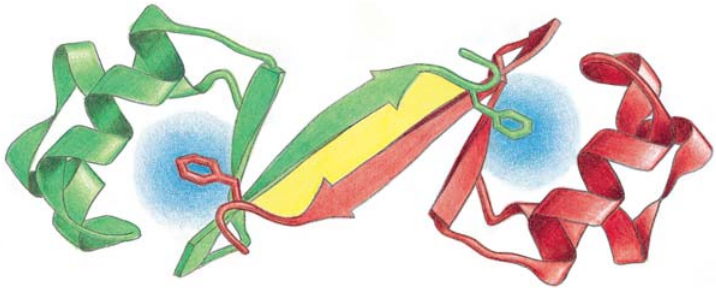


Figure 3–20 Two identical protein subunits binding together to form a symmetric protein dimer. The Cro repressor protein from bacteriophage lambda binds to DNA to turn off viral genes. Its two identical subunits bind head-to-head, held together by a combination of hydrophobic forces *(blue)* and a set of hydrogen bonds *(yellow* region). (Adapted from D.H. Ohlendorf, D.E. Tronrud and B.W. Matthews, *J. Mol. Biol.* 280:129–136, 1998. With permission from Academic Press.)

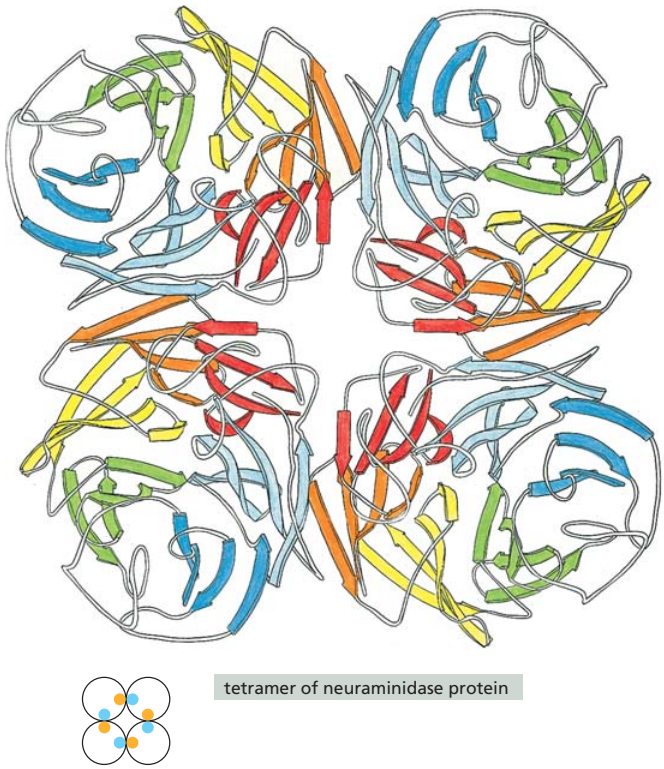


Figure 3–21 A protein molecule containing multiple copies of a single protein subunit. The enzyme neuraminidase exists as a ring of four identical polypeptide chains. Each of these chains is formed from six repeats of a four-stranded β sheet, as indicated by the colored arrows. The small diagram shows how the repeated use of the same binding interaction forms the structure.

two identical α-globin subunits and two identical β-globin subunits, symmetrically arranged (**Figure 3–22**). Such multisubunit proteins are very common in cells, and they can be very large. **Figure 3–23** shows a sample of proteins whose exact structures are known, and it compares the sizes and shapes of a few larger proteins with some of the relatively small proteins that we have thus far presented as models.

Some Proteins Form Long Helical Filaments

Some protein molecules can assemble to form filaments that may span the entire length of a cell. Most simply, a long chain of identical protein molecules can be constructed if each molecule has a binding site complementary to another region of the surface of the same molecule (**Figure 3–24**). An actin filament, for example, is a long helical structure produced from many molecules of the protein *actin* (**Figure 3–25**). Actin is very abundant in eucaryotic cells, where it constitutes one of the major filament systems of the cytoskeleton (discussed in Chapter 16).

Why is a helix such a common structure in biology? As we have seen, biological structures are often formed by linking similar subunits—such as amino acids or protein molecules—into long, repetitive chains. If all the subunits are identical, the neighboring subunits in the chain can often fit together in only one way, adjusting their relative positions to minimize the free energy of the contact between them. As a result, each subunit is positioned in exactly the same way in relation to the next, so that subunit 3 fits onto subunit 2 in the same way that subunit 2 fits onto subunit 1, and so on. Because it is very rare for subunits to join up in a straight line, this arrangement generally results in a helix—a regular structure that resembles a spiral staircase, as illustrated in **Figure 3–26**. Depending on the twist of the staircase, a helix is said to be either right-handed or left-handed (see Figure 3–26E). Handedness is not affected by turning the helix upside down, but it is reversed if the helix is reflected in the mirror.

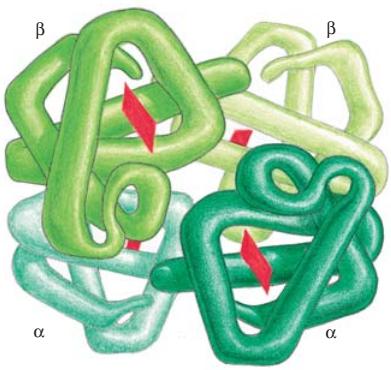


Figure 3–22 A protein formed as a symmetric assembly of two different subunits. Hemoglobin is an abundant protein in red blood cells that contains two copies of α-globin and two copies of β-globin. Each of these four polypeptide chains contains a heme molecule *(red),* which is the site that binds oxygen (O_2). Thus, each molecule of hemoglobin in the blood carries four molecules of oxygen.

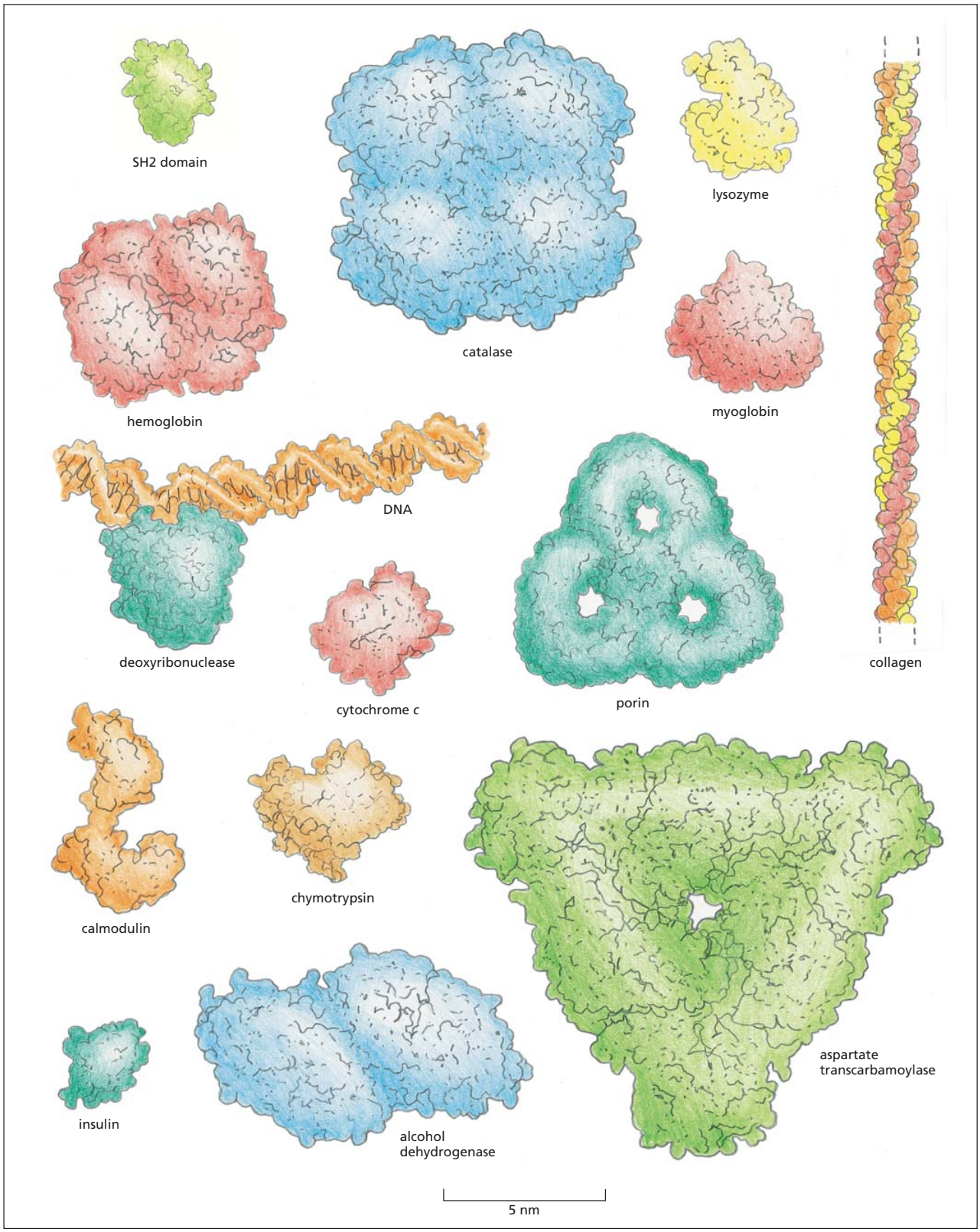


Figure 3–23 A collection of protein molecules, shown at the same scale. For comparison, a DNA molecule bound to a protein is also illustrated. These space-filling models represent a range of sizes and shapes. Hemoglobin, catalase, porin, alcohol dehydrogenase, and aspartate transcarbamoylase are formed from multiple copies of subunits. The SH2 domain *(top left)* is presented in detail in Panel 3–2 (pp. 132–133). (Adapted from David S. Goodsell, Our Molecular Nature. New York: Springer-Verlag, 1996. With permission from Springer Science and Business Media.)

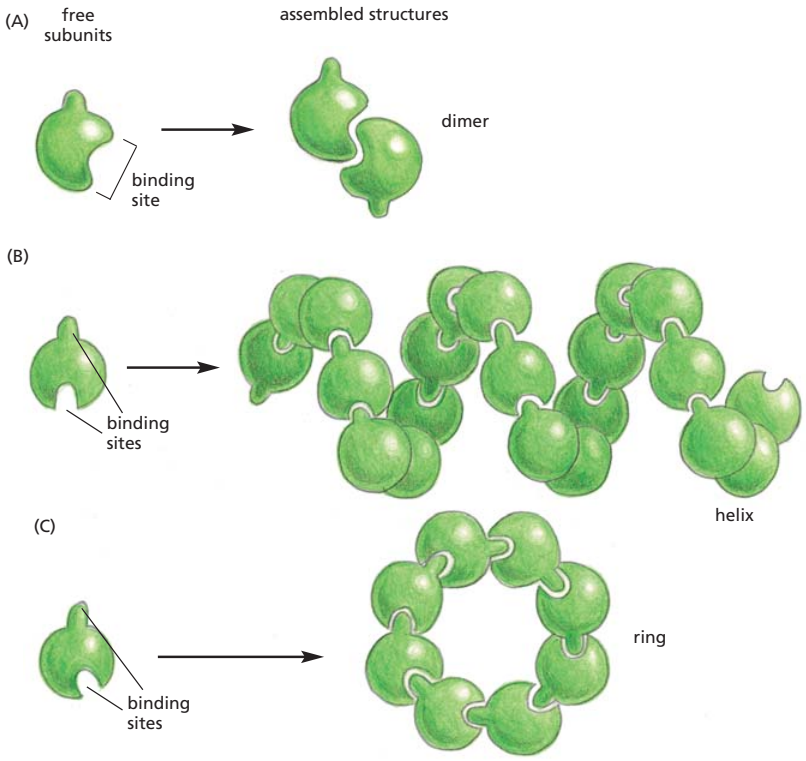


Figure 3–24 Protein assemblies. (A) A protein with just one binding site can form a dimer with another identical protein. (B) Identical proteins with two different binding sites often form a long helical filament. (C) If the two binding sites are disposed appropriately in relation to each other, the protein subunits may form a closed ring instead of a helix. (For an example of A, see Figure 3–20; for an example of C, see Figure 3–21.)

Helices occur commonly in biological structures, whether the subunits are small molecules linked together by covalent bonds (for example, the amino acids in an α helix) or large protein molecules that are linked by noncovalent forces (for example, the actin molecules in actin filaments). This is not surprising. A helix is an unexceptional structure, and it is generated simply by placing many similar subunits next to each other, each in the same strictly repeated relationship to the one before—that is, with a fixed rotation followed by a fixed translation along the helix axis, as in a spiral staircase.

Many Protein Molecules Have Elongated, Fibrous Shapes

Most of the proteins that we have discussed so far are *globular proteins,* in which the polypeptide chain folds up into a compact shape like a ball with an irregular surface. Enzymes tend to be globular proteins: even though many are large and complicated, with multiple subunits, most have an overall rounded shape (see Figure 3–23). In contrast, other proteins have roles in the cell that require each individual protein molecule to span a large distance. These proteins generally have a relatively simple, elongated three-dimensional structure and are commonly referred to as *fibrous proteins*.

One large family of intracellular fibrous proteins consists of α-keratin, introduced when we presented the α helix, and its relatives. Keratin filaments are extremely stable and are the main component in long-lived structures such as hair, horn, and nails. An α-keratin molecule is a dimer of two identical subunits, with the long α helices of each subunit forming a coiled-coil (see Figure 3–9). The coiled-coil regions are capped at each end by globular domains containing binding sites. This enables this class of protein to assemble into ropelike *intermediate filaments*—an important component of the cytoskeleton that creates the cell's internal structural framework (see Figure 16–19).

Fibrous proteins are especially abundant outside the cell, where they are a main component of the gel-like *extracellular matrix* that helps to bind collections of cells together to form tissues. Cells secrete extracellular matrix proteins into their surroundings, where they often assemble into sheets or long fibrils.

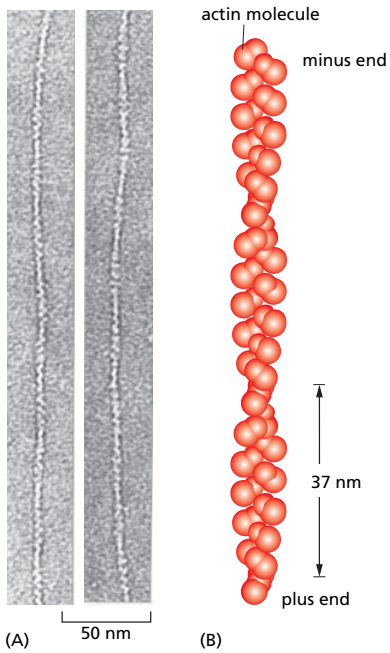
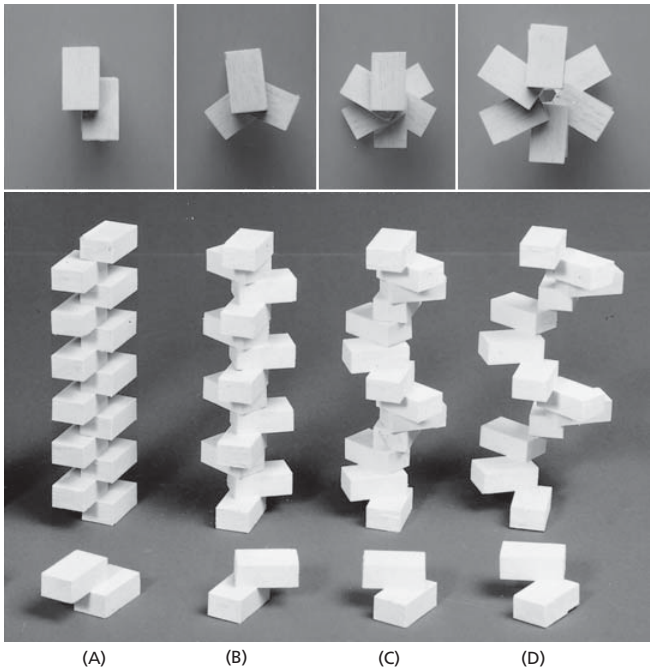


Figure 3–25 Actin filaments. (A) Transmission electron micrographs of negatively stained actin filaments. (B) The helical arrangement of actin molecules in an actin filament. (A, courtesy of Roger Craig.)



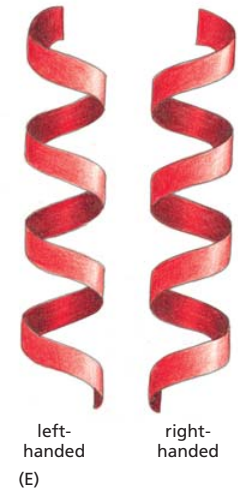
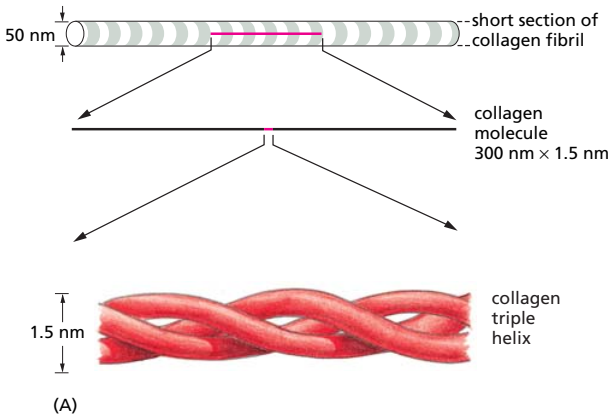


Figure 3–26 Some properties of a helix. (A–D) A helix forms when a series of subunits bind to each other in a regular way. At the bottom, the interaction between two subunits is shown; behind them are the helices that result. These helices have two (A), three (B), and six (C and D) subunits per helical turn. The photographs at the top show the arrangement of subunits viewed from directly above the helix. Note that the helix in (D) has a wider path than that in (C), but the same number of subunits per turn. (E) A helix can be either right-handed or left-handed. As a reference, it is useful to remember that standard metal screws, which insert when turned clockwise, are right-handed. Note that a helix retains the same handedness when it is turned upside down.

Collagen is the most abundant of these proteins in animal tissues. A collagen molecule consists of three long polypeptide chains, each containing the non-polar amino acid glycine at every third position. This regular structure allows the chains to wind around one another to generate a long regular triple helix (**Figure 3–27**A). Many collagen molecules then bind to one another side-by-side and end-to-end to create long overlapping arrays—thereby generating the extremely tough collagen fibrils that give connective tissues their tensile strength, as described in Chapter 19.

Many Proteins Contain a Surprisingly Large Amount of Unstructured Polypeptide Chain

It has been well known for a long time that, in complete contrast to collagen, another abundant protein in the extracellular matrix, *elastin,* is formed as a highly disordered polypeptide. This disorder is essential for elastin's function. Its



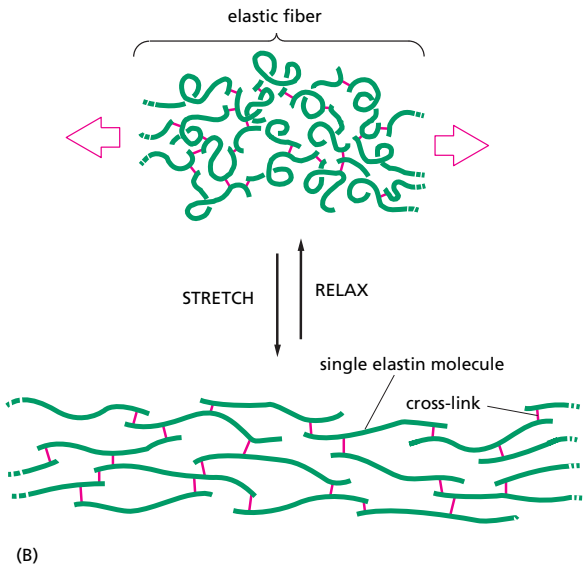


Figure 3–27 Collagen and elastin. (A) Collagen is a triple helix formed by three extended protein chains that wrap around one another *(bottom).* Many rodlike collagen molecules are cross-linked together in the extracellular space to form unextendable collagen fibrils *(top)* that have the tensile strength of steel. The striping on the collagen fibril is caused by the regular repeating arrangement of the collagen molecules within the fibril. (B) Elastin polypeptide chains are cross-linked together to form rubberlike, elastic fibers. Each elastin molecule uncoils into a more extended conformation when the fiber is stretched and recoils spontaneously as soon as the stretching force is relaxed.

relatively loose and unstructured polypeptide chains are covalently cross-linked to produce a rubberlike elastic meshwork that can be reversibly pulled from one conformation to another, as illustrated in Figure 3–27B. The elastic fibers that result enable skin and other tissues, such as arteries and lungs, to stretch and recoil without tearing.

Intrinsically unstructured regions of proteins are quite frequent in nature, having important functions in the interior of cells. As we have already seen, proteins use the short loops of polypeptide chain that generally protrude from the core region of protein domains to bind other molecules. Similarly, many proteins have much longer regions of unstructured amino acid sequences that interact with another molecule (often DNA or a protein), undergoing a structural transition to a specific folded conformation when the other molecule is bound. Other proteins appear to resemble elastin, in so far as their function requires that they remain largely unstructured. For example, the abundant nucleoporins that coat the inner surface of the nuclear pore complex form a random coil meshwork that is intimately involved in nuclear transport (see Figure 12–10). Finally, as will be discussed later in this chapter (see Figure 3–80C), unstructured regions of polypeptide chain are often used to connect the binding sites for proteins that function together to catalyze a biological reaction. Thus, for example, in facilitating cell signaling, large *scaffold proteins* use such flexible regions as "tethers" that concentrate sets of interacting proteins, often confining them to particular sites in the cell (discussed in Chapter 15).

We can recognize the unstructured regions in many proteins by their biased amino acid composition: they contain very few of the bulky hydrophobic amino acids that normally form the core of a folded protein, being composed instead of a high proportion of the amino acids Gln, Ser, Pro, Glu, and Lys. Such "natively unfolded" regions also frequently contain repeated sequences of amino acids.

Covalent Cross-Linkages Often Stabilize Extracellular Proteins

Many protein molecules are either attached to the outside of a cell's plasma membrane or secreted as part of the extracellular matrix. All such proteins are directly exposed to extracellular conditions. To help maintain their structures, the polypeptide chains in such proteins are often stabilized by covalent cross-linkages. These linkages can either tie two amino acids in the same protein together, or connect different polypeptide chains in a multisubunit protein. The most common cross-linkages in proteins are covalent sulfur–sulfur bonds. These *disulfide bonds* (also called *S–S bonds*) form as cells prepare newly synthesized proteins for export. As described in Chapter 12, their formation is catalyzed in the endoplasmic reticulum by an enzyme that links together two pairs of –SH groups of cysteine side chains that are adjacent in the folded protein (**Figure 3–28**). Disulfide bonds do not change the conformation of a protein but instead act as atomic staples to reinforce its most favored conformation. For

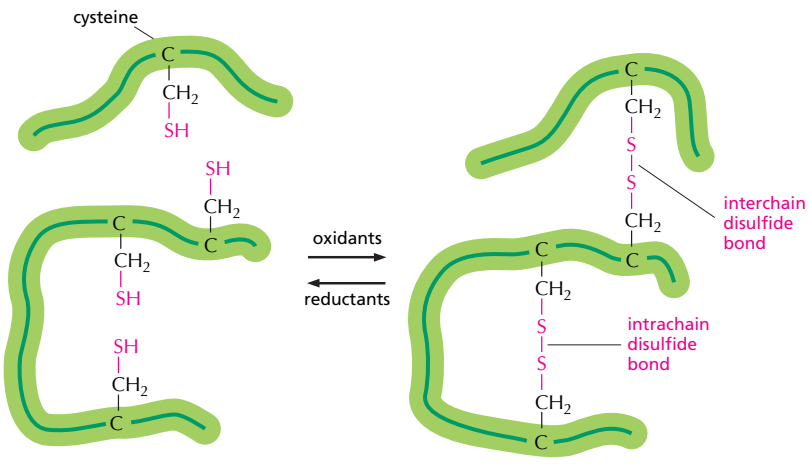


Figure 3–28 Disulfide bonds. <ATAC> This diagram illustrates how covalent disulfide bonds form between adjacent cysteine side chains. As indicated, these cross-linkages can join either two parts of the same polypeptide chain or two different polypeptide chains. Since the energy required to break one covalent bond is much larger than the energy required to break even a whole set of noncovalent bonds (see Table 2–1, p. 53), a disulfide bond can have a major stabilizing effect on a protein.

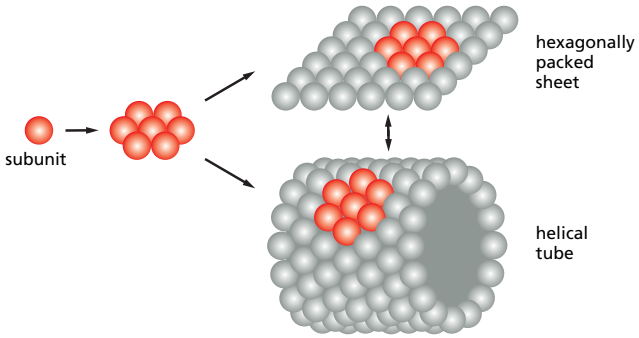


Figure 3–29 An example of the assembly of a single protein subunit requiring multiple protein–protein contacts. Hexagonally packed globular protein subunits can form either a flat sheet or a tube.

example, lysozyme—an enzyme in tears that dissolves bacterial cell walls—retains its antibacterial activity for a long time because it is stabilized by such cross-linkages.

Disulfide bonds generally fail to form in the cell cytosol, where a high concentration of reducing agents converts S–S bonds back to cysteine –SH groups. Apparently, proteins do not require this type of reinforcement in the relatively mild environment inside the cell.

Protein Molecules Often Serve as Subunits for the Assembly of Large Structures

The same principles that enable a protein molecule to associate with itself to form rings or filaments also operate to generate much larger structures in the cell—supramolecular structures such as enzyme complexes, ribosomes, protein filaments, viruses, and membranes. These large objects are not made as single, giant, covalently linked molecules. Instead they are formed by the noncovalent assembly of many separately manufactured molecules, which serve as the subunits of the final structure.

The use of smaller subunits to build larger structures has several advantages:

1. A large structure built from one or a few repeating smaller subunits requires only a small amount of genetic information.
2. Both assembly and disassembly can be readily controlled, reversible processes, because the subunits associate through multiple bonds of relatively low energy.
3. Errors in the synthesis of the structure can be more easily avoided, since correction mechanisms can operate during the course of assembly to exclude malformed subunits.

Some protein subunits assemble into flat sheets in which the subunits are arranged in hexagonal patterns. Specialized membrane proteins are sometimes arranged this way in lipid bilayers. With a slight change in the geometry of the individual subunits, a hexagonal sheet can be converted into a tube (**Figure 3–29**) or, with more changes, into a hollow sphere. Protein tubes and spheres that bind specific RNA and DNA molecules in their interior form the coats of viruses.

The formation of closed structures, such as rings, tubes, or spheres, provides additional stability because it increases the number of bonds between the protein subunits. Moreover, because such a structure is created by mutually dependent, cooperative interactions between subunits, a relatively small change that affects each subunit individually can cause the structure to assemble or disassemble. These principles are dramatically illustrated in the protein coat or *capsid* of many simple viruses, which takes the form of a hollow sphere based on an icosahedron (**Figure 3–30**). Capsids are often made of hundreds of identical protein subunits that enclose and protect the viral nucleic acid (**Figure 3–31**). The protein in such a capsid must have a particularly adaptable structure: not only must it make several different kinds of contacts to create the sphere, it must also change this arrangement to let the nucleic acid out to initiate viral replication once the virus has entered a cell.

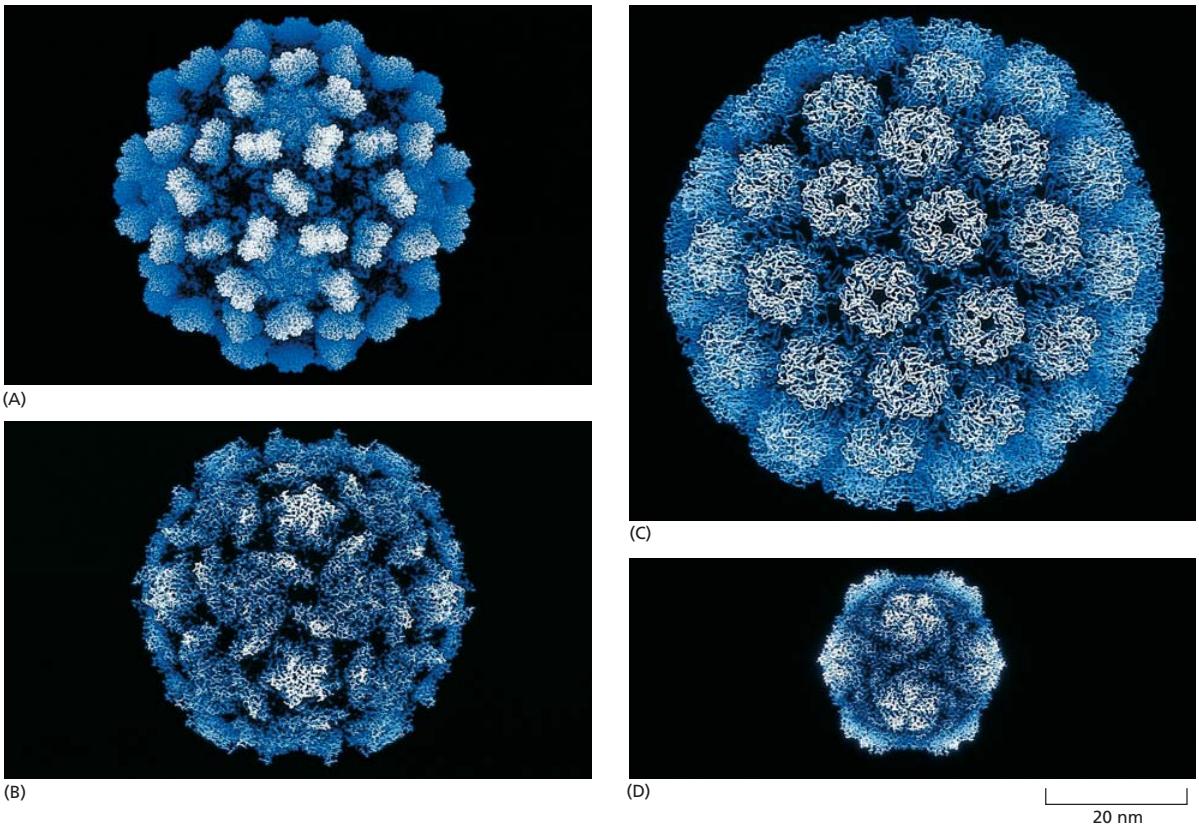


Figure 3–30 The capsids of some viruses, all shown at the same scale. (A) Tomato bushy stunt virus; (B) poliovirus; (C) simian virus 40 (SV40); (D) satellite tobacco necrosis virus. The structures of all of these capsids have been determined by x-ray crystallography and are known in atomic detail. (Courtesy of Robert Grant, Stephan Crainic, and James M. Hogle.)

Many Structures in Cells Are Capable of Self-Assembly

The information for forming many of the complex assemblies of macromolecules in cells must be contained in the subunits themselves, because purified subunits can spontaneously assemble into the final structure under the appropriate conditions. The first large macromolecular aggregate shown to be capable of self-assembly from its component parts was *tobacco mosaic virus (TMV)*. This virus is a long rod in which a cylinder of protein is arranged around a helical RNA core (**Figure 3–32**). If the dissociated RNA and protein subunits are mixed together in solution, they recombine to form fully active viral particles. The assembly process is unexpectedly complex and includes the formation of double rings of protein, which serve as intermediates that add to the growing viral coat.

Another complex macromolecular aggregate that can reassemble from its component parts is the bacterial ribosome. This structure is composed of about 55 different protein molecules and 3 different rRNA molecules. Incubating the individual components under appropriate conditions in a test tube causes them to spontaneously re-form the original structure. Most importantly, such reconstituted ribosomes are able to catalyze protein synthesis. As might be expected, the reassembly of ribosomes follows a specific pathway: after certain proteins have bound to the RNA, this complex is then recognized by other proteins, and so on, until the structure is complete.

It is still not clear how some of the more elaborate self-assembly processes are regulated. Many structures in the cell, for example, seem to have a precisely

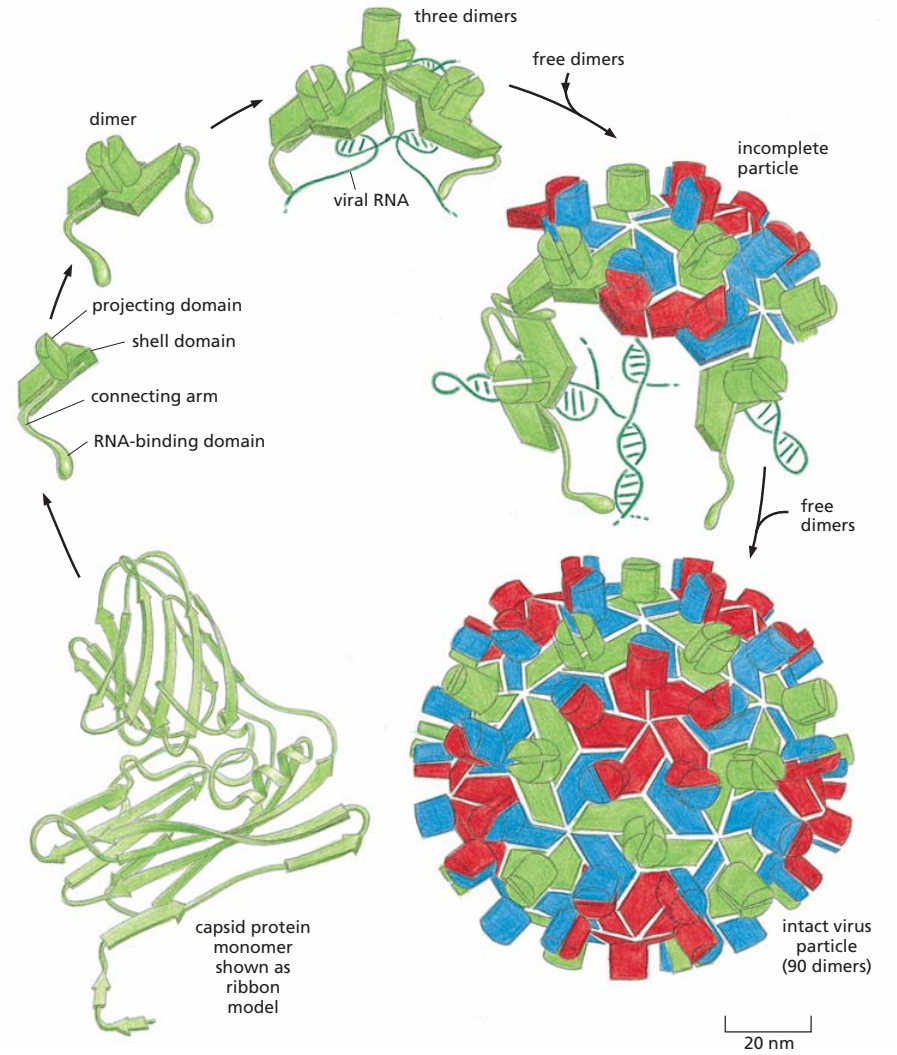
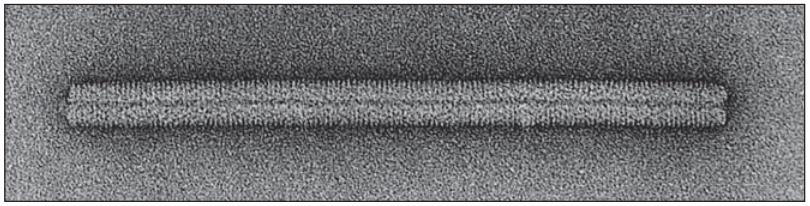


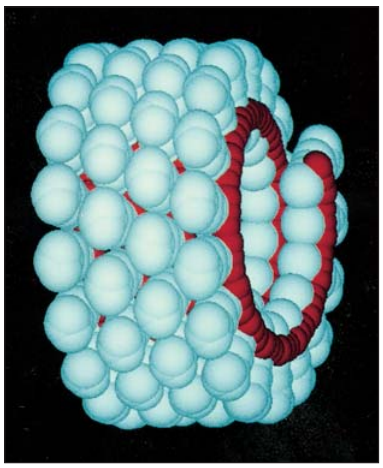
Figure 3–31 The structure of a spherical virus. In many viruses, identical protein subunits pack together to create a spherical shell (a capsid) that encloses the viral genome, composed of either RNA or DNA (see also Figure 3–30). For geometric reasons, no more than 60 identical subunits can pack together in a precisely symmetric way. If slight irregularities are allowed, however, more subunits can be used to produce a larger capsid that retains icosahedral symmetry. The tomato bushy stunt virus (TBSV) shown here, for example, is a spherical virus about 33 nm in diameter formed from 180 identical copies of a 386 amino acid capsid protein plus an RNA genome of 4500 nucleotides. To construct such a large capsid, the protein must be able to fit into three somewhat different environments, each of which is differently colored in the virus particle shown here. The postulated pathway of assembly is shown; the precise three-dimensional structure has been determined by x-ray diffraction. (Courtesy of Steve Harrison.)

defined length that is many times greater than that of their component macromolecules. How such length determination is achieved is in many cases a mystery. Three possible mechanisms are illustrated in **Figure 3–33**. In the simplest

Figure 3–32 The structure of tobacco mosaic virus (TMV). (A) An electron micrograph of the viral particle, which consists of a single long RNA molecule enclosed in a cylindrical protein coat composed of identical protein subunits. (B) A model showing part of the structure of TMV. A single-stranded RNA molecule of 6395 nucleotides is packaged in a helical coat constructed from 2130 copies of a coat protein 158 amino acids long. Fully infective viral particles can self-assemble in a test tube from purified RNA and protein molecules. (A, courtesy of Robley Williams; B, courtesy of Richard J. Feldmann.)



(A)



(B)

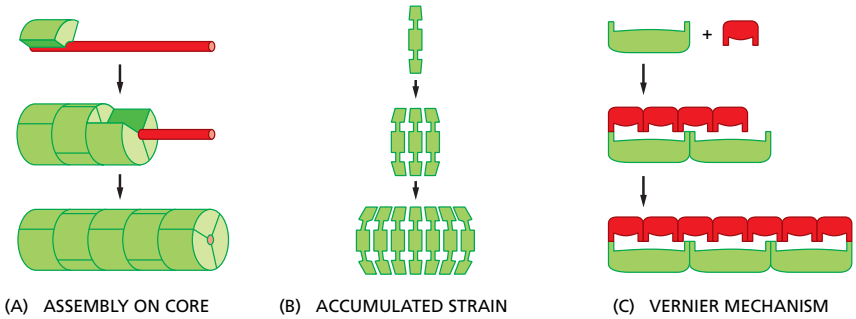


Figure 3–33 Three mechanisms of length determination for large protein assemblies. (A) Coassembly along an elongated core protein or other macromolecule that acts as a measuring device. (B) Termination of assembly because of strain that accumulates in the polymeric structure as additional subunits are added, so that beyond a certain length the energy required to fit another subunit onto the chain becomes excessively large. (C) A vernier type of assembly, in which two sets of rodlike molecules differing in length form a staggered complex that grows until their ends exactly match. The name derives from a measuring device based on the same principle, used in mechanical instruments.

case, a long core protein or other macromolecule provides a scaffold that determines the extent of the final assembly. This is the mechanism that determines the length of the TMV particle, where the RNA chain provides the core. Similarly, a core protein is thought to determine the length of the thin filaments in muscle, as well as the length of the long tails of some bacterial viruses (**Figure 3–34**).

Assembly Factors Often Aid the Formation of Complex Biological Structures

Not all cellular structures held together by noncovalent bonds self-assemble. A mitochondrion, a cilium, or a myofibril of a muscle cell, for example, cannot form spontaneously from a solution of its component macromolecules. In these cases, part of the assembly information is provided by special enzymes and other proteins that perform the function of templates, guiding construction but taking no part in the final assembled structure.

Even relatively simple structures may lack some of the ingredients necessary for their own assembly. In the formation of certain bacterial viruses, for example, the head, which is composed of many copies of a single protein subunit, is assembled on a temporary scaffold composed of a second protein. Because the second protein is absent from the final viral particle, the head structure cannot spontaneously reassemble once it has been taken apart. Other examples are known in which proteolytic cleavage is an essential and irreversible step in the normal assembly process. This is even the case for some small protein assemblies, including the structural protein collagen and the hormone insulin (**Figure 3–35**). From these relatively simple examples, it seems certain that the assembly of a structure as complex as a mitochondrion or a cilium will involve temporal and spatial ordering imparted by numerous other cell components.

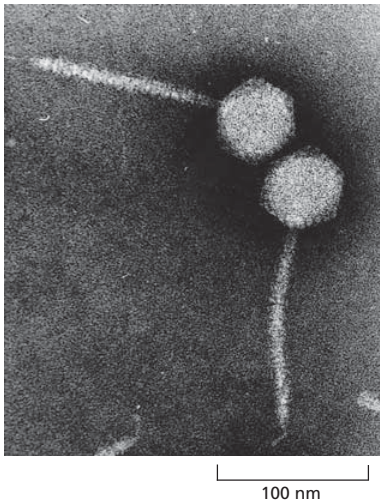


Figure 3–34 An electron micrograph of bacteriophage lambda. The tip of the virus tail attaches to a specific protein on the surface of a bacterial cell, after which the tightly packaged DNA in the head is injected through the tail into the cell. The tail has a precise length, determined by the mechanism shown in Figure 3–33A.

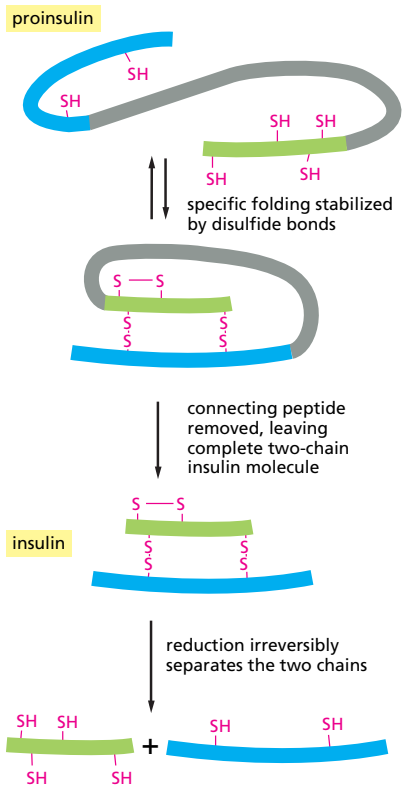


Figure 3–35 Proteolytic cleavage in insulin assembly. The polypeptide hormone insulin cannot spontaneously re-form efficiently if its disulfide bonds are disrupted. It is synthesized as a larger protein *(proinsulin)* that is cleaved by a proteolytic enzyme after the protein chain has folded into a specific shape. Excision of part of the proinsulin polypeptide chain removes some of the information needed for the protein to fold spontaneously into its normal conformation. Once insulin has been denatured and its two polypeptide chains have separated, its ability to reassemble is lost.

Summary

A protein molecule's amino acid sequence determines its three-dimensional conformation. Noncovalent interactions between different parts of the polypeptide chain stabilize its folded structure. The amino acids with hydrophobic side chains tend to cluster in the interior of the molecule, and local hydrogen-bond interactions between neighboring peptide bonds give rise to α helices and β sheets.

Globular regions, known as domains, are the modular units from which many proteins are constructed; such domains generally contain 40–350 amino acids. Small proteins typically consist of only a single domain, while large proteins are formed from several domains linked together by various lengths of polypeptide chain, some of which can be relatively disordered. As proteins have evolved, domains have been modified and combined with other domains to construct new proteins. Thus far, about 800 different ways of folding up a domain have been observed, among more than 20,000 known protein structures.

Proteins are brought together into larger structures by the same noncovalent forces that determine protein folding. Proteins with binding sites for their own surface can assemble into dimers, closed rings, spherical shells, or helical polymers. Although mixtures of proteins and nucleic acids can assemble spontaneously into complex structures in a test tube, many biological assembly processes involve irreversible steps. Consequently, not all structures in the cell are capable of spontaneous reassembly after they have been dissociated into their component parts.

PROTEIN FUNCTION

We have seen that each type of protein consists of a precise sequence of amino acids that allows it to fold up into a particular three-dimensional shape, or conformation. But proteins are not rigid lumps of material. They often have precisely engineered moving parts whose mechanical actions are coupled to chemical events. It is this coupling of chemistry and movement that gives proteins the extraordinary capabilities that underlie the dynamic processes in living cells.

In this section, we explain how proteins bind to other selected molecules and how their activity depends on such binding. We show that the ability to bind to other molecules enables proteins to act as catalysts, signal receptors, switches, motors, or tiny pumps. The examples we discuss in this chapter by no means exhaust the vast functional repertoire of proteins. You will encounter the specialized functions of many other proteins elsewhere in this book, based on similar principles.

All Proteins Bind to Other Molecules

A protein molecule's physical interaction with other molecules determines its biological properties. Thus, antibodies attach to viruses or bacteria to mark them for destruction, the enzyme hexokinase binds glucose and ATP so as to catalyze a reaction between them, actin molecules bind to each other to assemble into actin filaments, and so on. Indeed, all proteins stick, or *bind*, to other molecules. In some cases, this binding is very tight; in others it is weak and short-lived. But the binding always shows great *specificity*, in the sense that each protein molecule can usually bind just one or a few molecules out of the many thousands of different types it encounters. The substance that is bound by the protein—whether it is an ion, a small molecule, or a macromolecule such as another protein—is referred to as a **ligand** for that protein (from the Latin word *ligare*, meaning "to bind").

The ability of a protein to bind selectively and with high affinity to a ligand depends on the formation of a set of weak, noncovalent bonds—hydrogen bonds, electrostatic attractions, and van der Waals attractions—plus favorable hydrophobic interactions (see Panel 2–3, pp. 110–111). Because each individual bond is weak, effective binding occurs only when many of these bonds form simultaneously. Such binding is possible only if the surface contours of the ligand molecule fit very closely to the protein, matching it like a hand in a glove (**Figure 3–36**).

The region of a protein that associates with a ligand, known as the ligand's *binding site,* usually consists of a cavity in the protein surface formed by a particular arrangement of amino acids. These amino acids can belong to different portions of the polypeptide chain that are brought together when the protein folds (**Figure 3–37**). Separate regions of the protein surface generally provide binding sites for different ligands, allowing the protein's activity to be regulated, as we shall see later. And other parts of the protein act as a handle to position the protein in the cell—an example is the SH2 domain discussed previously, which often moves a protein containing it to particular intracellular sites in response to particular signals.

Although the atoms buried in the interior of the protein have no direct contact with the ligand, they form the framework that gives the surface its contours and its chemical and mechanical properties. Even small changes to the amino acids in the interior of a protein molecule can change its three-dimensional shape enough to destroy a binding site on the surface.

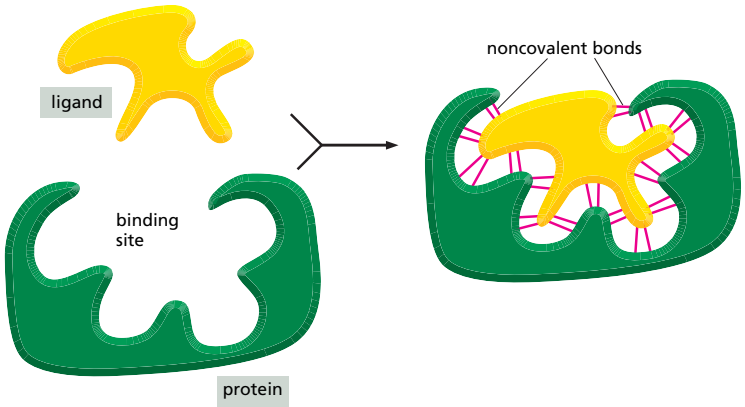


Figure 3–36 The selective binding of a protein to another molecule. Many weak bonds are needed to enable a protein to bind tightly to a second molecule, which is called a *ligand* for the protein. A ligand must therefore fit precisely into a protein's binding site, like a hand into a glove, so that a large number of noncovalent bonds form between the protein and the ligand.

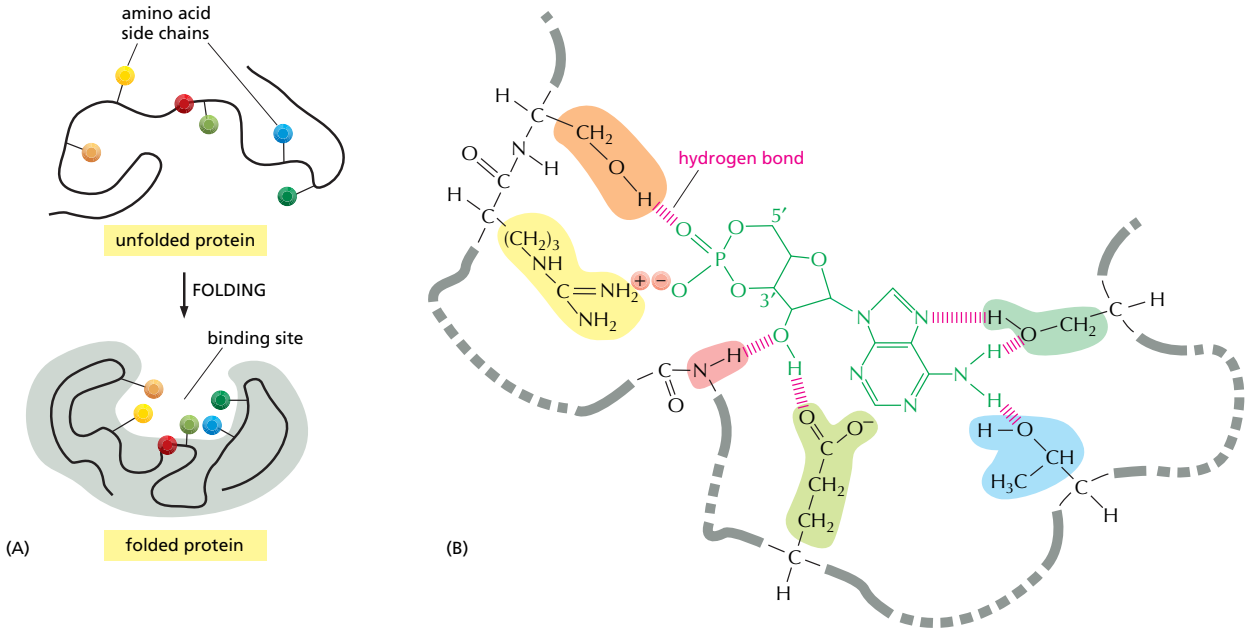


Figure 3–37 The binding site of a protein. (A) The folding of the polypeptide chain typically creates a crevice or cavity on the protein surface. This crevice contains a set of amino acid side chains disposed in such a way that they can form noncovalent bonds only with certain ligands. (B) A close-up of an actual binding site showing the hydrogen bonds and electrostatic interactions formed between a protein and its ligand. In this example, cyclic AMP is the bound ligand.

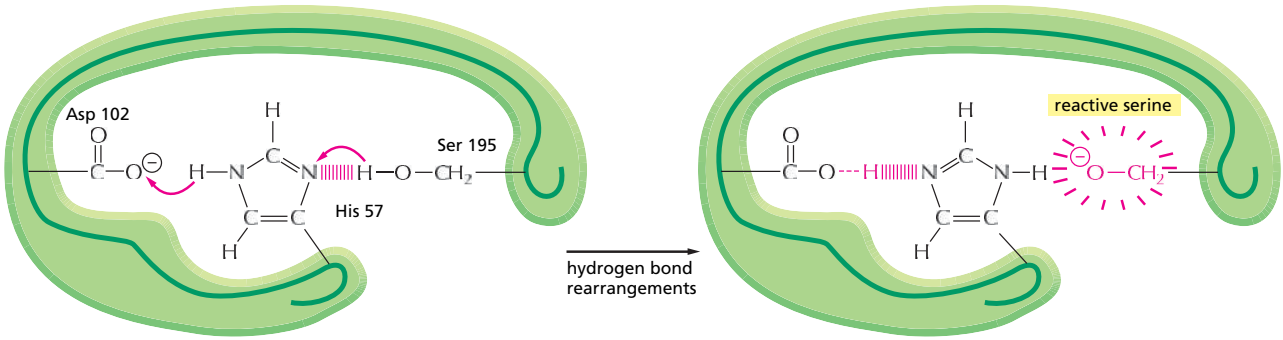
The Surface Conformation of a Protein Determines Its Chemistry

Proteins have impressive chemical capabilities because the neighboring chemical groups on their surface often interact in ways that enhance the chemical reactivity of amino acid side chains. These interactions fall into two main categories.

First, the interaction of neighboring parts of the polypeptide chain may restrict the access of water molecules to that protein's ligand-binding sites. This is important because water molecules readily form hydrogen bonds that can compete with ligands for sites on the protein surface. Proteins and their ligands form tighter hydrogen bonds (and electrostatic interactions) if the protein can exclude water molecules from its binding sites. It might be hard to imagine a mechanism that would exclude a molecule as small as water from a protein surface without affecting the access of the ligand itself. However, because of the strong tendency of water molecules to form water–water hydrogen bonds, water molecules exist in a large hydrogen-bonded network (see Panel 2–2, pp. 108–109). In effect, a protein can keep a ligand-binding site dry because it is energetically unfavorable for individual water molecules to break away from this network, as they must do to reach into a crevice on a protein's surface.

Second, the clustering of neighboring polar amino acid side chains can alter their reactivity. If protein folding forces together a number of negatively charged side chains against their mutual repulsion, for example, the affinity of the site for a positively charged ion is greatly increased. In addition, when amino acid side chains interact with one another through hydrogen bonds, normally unreactive side groups (such as the $-CH_2OH$ on the serine shown in **Figure 3–38**) can become reactive, enabling them to be used to make or break selected covalent bonds.

The surface of each protein molecule therefore has a unique chemical reactivity that depends not only on which amino acid side chains are exposed, but



also on their exact orientation relative to one another. For this reason, two even slightly different conformations of the same protein molecule may differ greatly in their chemistry.

Figure 3–38 An unusually reactive amino acid at the active site of an enzyme. This example is the "catalytic triad" found in chymotrypsin, elastase, and other serine proteases (see Figure 3–12). The aspartic acid side chain (Asp 102) induces the histidine (His 57) to remove the proton from serine 195. This activates the serine to form a covalent bond with the enzyme substrate, hydrolyzing a peptide bond. The many convolutions of the polypeptide chain are omitted here.

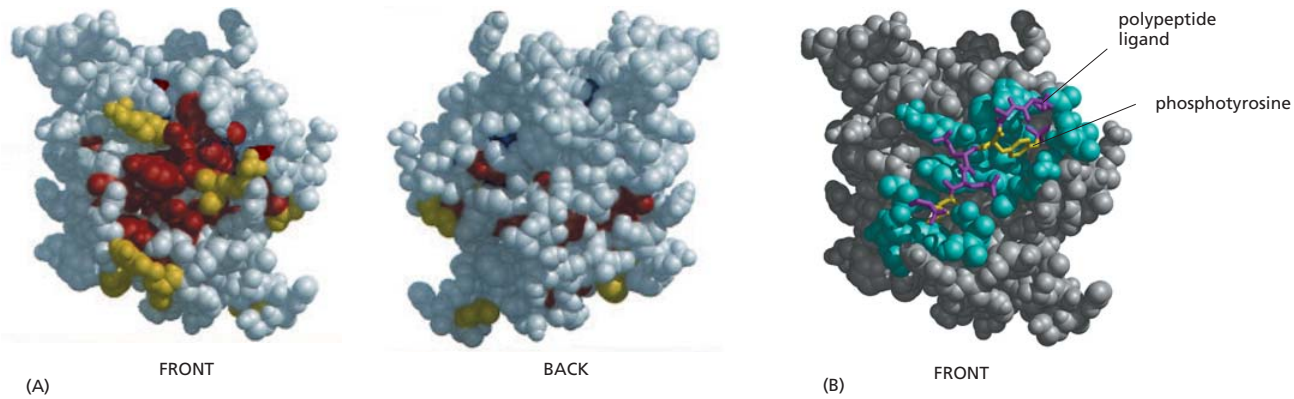
Sequence Comparisons Between Protein Family Members Highlight Crucial Ligand-Binding Sites

As we have described previously, genome sequences allow us to group many of the domains in proteins into families that show clear evidence of their evolution from a common ancestor. The three-dimensional structures of the members of the same domain family are remarkably similar. For example, even when the amino acid sequence identity falls to 25%, the backbone atoms in a domain follow a common protein fold within 0.2 nanometers (2 Å).

We can therefore use a method called evolutionary tracing to identify those sites in a protein domain that are the most crucial to the domain's function. For this purpose, those amino acids that are unchanged, or nearly unchanged, in all of the known protein family members are mapped onto a model of the three-dimensional structure of one family member. When this is done, the most invariant positions often form one or more clusters on the protein surface, as illustrated in **Figure 3–39**A for the SH2 domain described previously (see Panel 3–2, pp. 132–133). These clusters generally correspond to ligand binding sites.

The SH2 domain is a module that functions in protein–protein interactions. It binds the protein containing it to a second protein that contains a phosphorylated tyrosine side chain in a specific amino acid sequence context, as shown in Figure 3–39B. The amino acids located at the binding site for the phosphorylated polypeptide have been the slowest to change during the long evolutionary process that produced the large SH2 family of peptide recognition domains. Because mutation is a random process, this result is attributed to the preferential elimination during evolution of all organisms whose SH2 domains became altered in a way that inactivated the SH2-binding site, thereby destroying the function of the SH2 domain.

Figure 3–39 The evolutionary trace method applied to the SH2 domain. (A) Front and back views of a space-filling model of the SH2 domain, with evolutionarily conserved amino acids on the protein surface colored *yellow*, and those more toward the protein interior colored *red.* (B) The structure of the SH2 domain with its bound polypeptide. Here, those amino acids located within 0.4 nm of the bound ligand are colored *blue*. The two key amino acids of the ligand are *yellow,* and the others are *purple*. Note the high degree of correspondance between (A) and (B). (Adapted from O. Lichtarge, H.R. Bourne and F.E. Cohen, *J. Mol. Biol.* 257:342–358, 1996. With permission from Elsevier.)



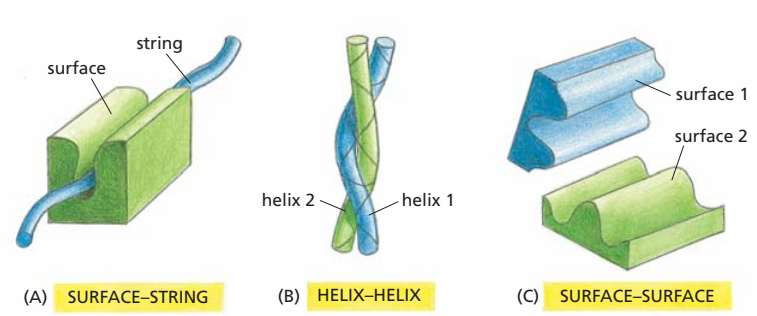


Figure 3–40 Three ways in which two proteins can bind to each other. Only the interacting parts of the two proteins are shown. (A) A rigid surface on one protein can bind to an extended loop of polypeptide chain (a "string") on a second protein. (B) Two α helices can bind together to form a coiled-coil. (C) Two complementary rigid surfaces often link two proteins together.

In this era of extensive genome sequencing, many new protein families have been discovered whose functions are unknown. Once a three-dimensional structure has been determined for one family member, evolutionary tracing allows biologists to determine binding sites for the members of that family, thereby helping to decipher protein function.

Proteins Bind to Other Proteins Through Several Types of Interfaces

Proteins can bind to other proteins in at least three ways. In many cases, a portion of the surface of one protein contacts an extended loop of polypeptide chain (a "string") on a second protein (**Figure 3–40**A). Such a surface–string interaction, for example, allows the SH2 domain to recognize a phosphorylated polypeptide loop on a second protein, as just described, and it also enables a protein kinase to recognize the proteins that it will phosphorylate (see below).

A second type of protein–protein interface forms when two α helices, one from each protein, pair together to form a coiled-coil (Figure 3–40B). This type of protein interface is found in several families of gene regulatory proteins, as discussed in Chapter 7.

The most common way for proteins to interact, however, is by the precise matching of one rigid surface with that of another (Figure 3–40C). Such interactions can be very tight, since a large number of weak bonds can form between two surfaces that match well. For the same reason, such surface–surface interactions can be extremely specific, enabling a protein to select just one partner from the many thousands of different proteins found in a cell.

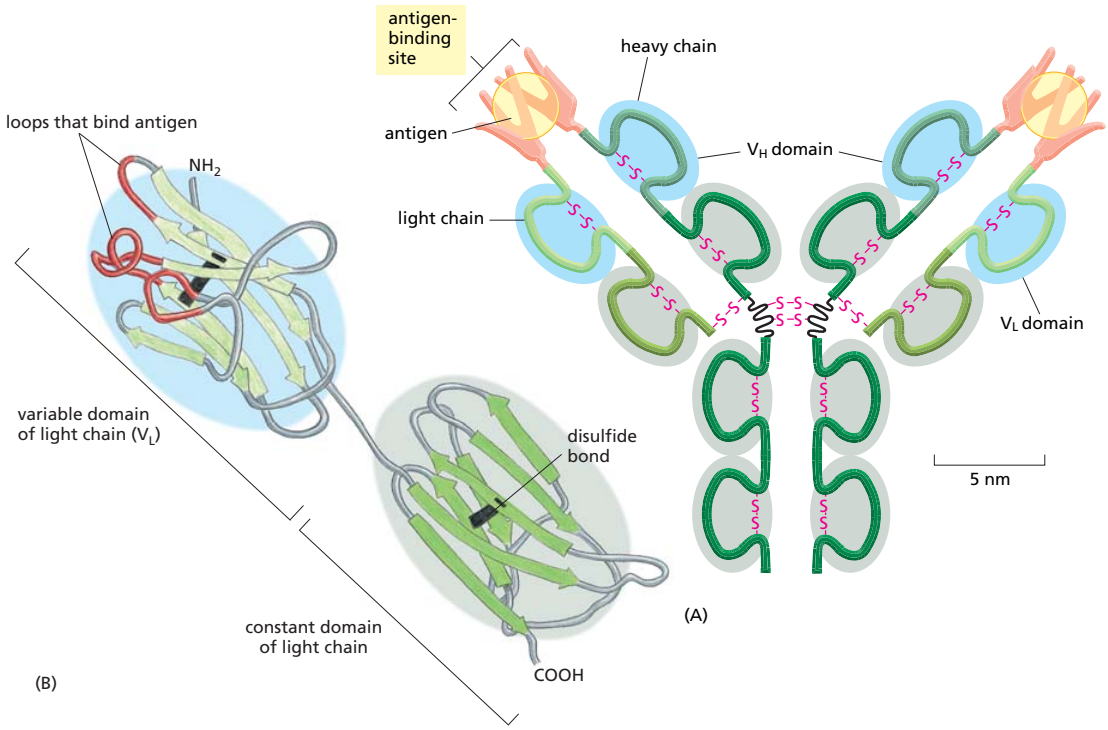
Antibody Binding Sites Are Especially Versatile <GCCG>

All proteins must bind to particular ligands to carry out their various functions. The antibody family is notable for its capacity for tight selective binding (discussed in detail in Chapter 25).

Antibodies, or immunoglobulins, are proteins produced by the immune system in response to foreign molecules, such as those on the surface of an invading microorganism. Each antibody binds tightly to a particular target molecule, thereby either inactivating the target molecule directly or marking it for destruction. An antibody recognizes its target (called an **antigen**) with remarkable specificity. Because there are potentially billions of different antigens that humans might encounter, we have to be able to produce billions of different antibodies.

Antibodies are Y-shaped molecules with two identical binding sites that are complementary to a small portion of the surface of the antigen molecule. A detailed examination of the antigen-binding sites of antibodies reveals that they are formed from several loops of polypeptide chain that protrude from the ends of a pair of closely juxtaposed protein domains (**Figure 3–41**). Different antibodies generate an enormous diversity of antigen-binding sites by changing only the length and amino acid sequence of these loops, without altering the basic protein structure.

Loops of this kind are ideal for grasping other molecules. They allow a large number of chemical groups to surround a ligand so that the protein can link to



it with many weak bonds. For this reason, loops often form the ligand-binding sites in proteins.

The Equilibrium Constant Measures Binding Strength

Molecules in the cell encounter each other very frequently because of their continual random thermal movements. Colliding molecules with poorly matching surfaces form few noncovalent bonds with one another, and the two molecules dissociate as rapidly as they come together. At the other extreme, when many noncovalent bonds form between two colliding molecules, the association can persist for a very long time (**Figure 3–42**). Strong interactions occur in cells whenever a biological function requires that molecules remain associated for a long time—for example, when a group of RNA and protein molecules come together to make a subcellular structure such as a ribosome.

We can measure the strength with which any two molecules bind to each other. As an example, consider a population of identical antibody molecules that suddenly encounters a population of ligands diffusing in the fluid surrounding them. At frequent intervals, one of the ligand molecules will bump into the binding site of an antibody and form an antibody–ligand complex. The population of antibody–ligand complexes will therefore increase, but not without limit: over time, a second process, in which individual complexes break apart because of thermally induced motion, will become increasingly important. Eventually, any population of antibody molecules and ligands will reach a steady state, or equilibrium, in which the number of binding (association) events per second is precisely equal to the number of "unbinding" (dissociation) events (see Figure 2–52).

From the concentrations of the ligand, antibody, and antibody–ligand complex at equilibrium, we can calculate a convenient measure—the **equilibrium constant** ***(K)***—of the strength of binding (**Figure 3–43**A). The equilibrium constant for a reaction in which two molecules (A and B) bind to each other to form a complex (AB) has units of liters/mole, and half of the binding sites will be occupied by ligand when that ligand's concentration (in moles/liter) reaches a value that is equal to $1/K$. This equilibrium constant is larger the greater the binding strength, and it is a direct measure of the free-energy difference

Figure 3–41 An antibody molecule. (A) A typical antibody molecule is Y-shaped and has two identical binding sites for its antigen, one on each arm of the Y. The protein is composed of four polypeptide chains (two identical heavy chains and two identical and smaller light chains) held together by disulfide bonds. Each chain is made up of several different immunoglobulin domains, here shaded either *blue* or *gray*. The antigen-binding site is formed where a heavy-chain variable domain (V_H) and a light-chain variable domain (V_L) come close together. These are the domains that differ most in their sequence and structure in different antibodies. Each domain at the end of the two arms of the antibody molecule forms loops that bind to the antigen. In (B) we can see these fingerlike loops *(red)* contributed by the V_L domain.

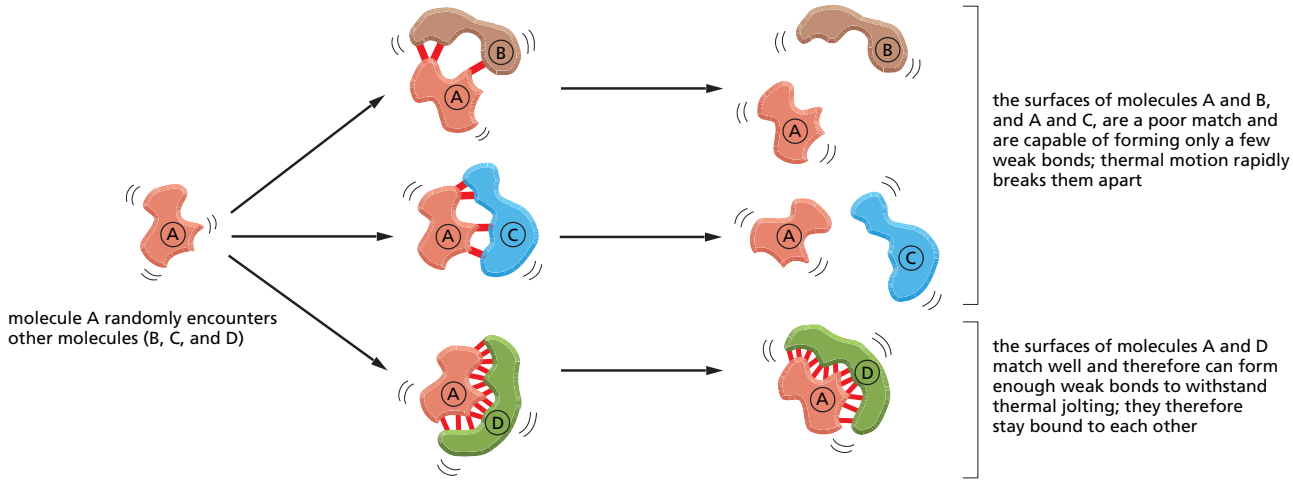


Figure 3–42 How noncovalent bonds mediate interactions between macromolecules.

between the bound and free states (Figure 3–43B and C). Even a change of a few noncovalent bonds can have a striking effect on a binding interaction, as shown by the example in **Figure 3–44**. (Note that the equilibrium constant, as defined here is also known as the association or affinity constant, K_a.)

We have used the case of an antibody binding to its ligand to illustrate the effect of binding strength on the equilibrium state, but the same principles apply to any molecule and its ligand. Many proteins are enzymes, which, as we now discuss, first bind to their ligands and then catalyze the breakage or formation of covalent bonds in these molecules.

Enzymes Are Powerful and Highly Specific Catalysts

Many proteins can perform their function simply by binding to another molecule. An actin molecule, for example, need only associate with other actin

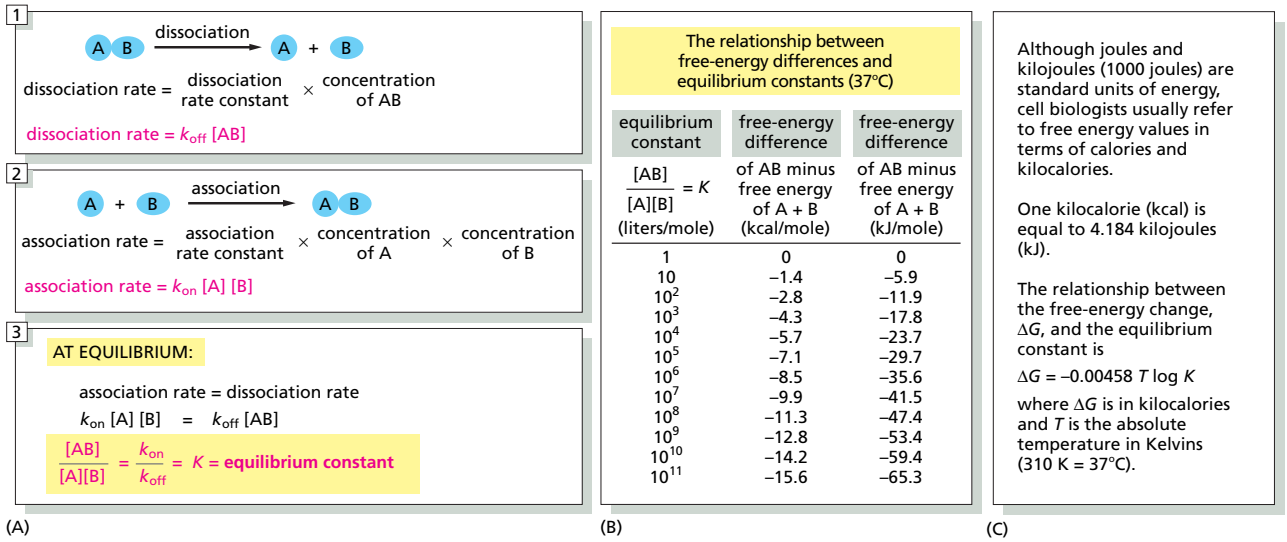


Figure 3–43 Relating binding energies to the equilibrium constant for an association reaction. (A) The equilibrium between molecules A and B and the complex AB is maintained by a balance between the two opposing reactions shown in panels 1 and 2. Molecules A and B must collide if they are to react, and the association rate is therefore proportional to the product of their individual concentrations [A] × [B]. (Square brackets indicate concentration.) As shown in panel 3, the ratio of the rate constants for the association and the dissociation reactions is equal to the equilibrium constant (*K*) for the reaction. (B) The equilibrium constant in panel 3 is that for the reaction A + B ⇌ AB, and the larger its value, the stronger the binding between A and B. Note that for every 1.41 kcal/mole (5.91 kJ/mole) decrease in free energy the equilibrium constant increases by a factor of 10 at 37°C.

The equilibrium constant here has units of liters/mole: for simple binding interactions it is also called the *affinity constant* or *association constant,* denoted K_a. The reciprocal of K_a is called the dissociation constant, K_d (in units of moles/liter).

molecules to form a filament. There are other proteins, however, for which ligand binding is only a necessary first step in their function. This is the case for the large and very important class of proteins called **enzymes**. As described in Chapter 2, enzymes are remarkable molecules that determine all the chemical transformations that make and break covalent bonds in cells. They bind to one or more ligands, called **substrates**, and convert them into one or more chemically modified *products*, doing this over and over again with amazing rapidity. Enzymes speed up reactions, often by a factor of a million or more, without themselves being changed—that is, they act as **catalysts** that permit cells to make or break covalent bonds in a controlled way. It is the catalysis of organized sets of chemical reactions by enzymes that creates and maintains the cell, making life possible.

We can group enzymes into functional classes that perform similar chemical reactions (**Table 3–1**). Each type of enzyme within such a class is highly specific, catalyzing only a single type of reaction. Thus, *hexokinase* adds a phosphate group to D-glucose but ignores its optical isomer L-glucose; the blood-clotting enzyme *thrombin* cuts one type of blood protein between a particular arginine and its adjacent glycine and nowhere else, and so on. As discussed in detail in Chapter 2, enzymes work in teams, with the product of one enzyme becoming the substrate for the next. The result is an elaborate network of metabolic pathways that provides the cell with energy and generates the many large and small molecules that the cell needs (see Figure 2–35).

Consider 1000 molecules of A and 1000 molecules of B in a eucaryotic cell. The concentration of both will be about 10^{-9} M.

If the equilibrium constant (K) for A + B ⇌ AB is 10^{10}, then one can calculate that at equilibrium there will be

270	270	730
A molecules	B molecules	AB molecules

If the equilibrium constant is a little weaker at 10^{8}, which represents a loss of 2.8 kcal/mole of binding energy from the example above, or 2–3 fewer hydrogen bonds, then there will be

915	915	85
A molecules	B molecules	AB molecules

Figure 3–44 Small changes in the number of weak bonds can have drastic effects on a binding interaction. This example illustrates the dramatic effect of the presence or absence of a few weak noncovalent bonds in a biological context.

Substrate Binding Is the First Step in Enzyme Catalysis

For a protein that catalyzes a chemical reaction (an enzyme), the binding of each substrate molecule to the protein is an essential prelude. In the simplest case, if we denote the enzyme by E, the substrate by S, and the product by P, the basic reaction path is E + S → ES → EP → E + P. From this reaction path, we see that there is a limit to the amount of substrate that a single enzyme molecule can process in a given time. An increase in the concentration of substrate also increases the rate at which product is formed, up to a maximum value (**Figure 3–45**). At that point the enzyme molecule is saturated with substrate, and the rate of reaction (V_{max}) depends only on how rapidly the enzyme can process the substrate molecule. This maximum rate divided by the enzyme concentration is

Table 3–1 Some Common Types of Enzymes

ENZYME	REACTION CATALYZED
Hydrolases	general term for enzymes that catalyze a hydrolytic cleavage reaction; *nucleases* and *proteases* are more specific names for subclasses of these enzymes.
Nucleases	break down nucleic acids by hydrolyzing bonds between nucleotides.
Proteases	break down proteins by hydrolyzing bonds between amino acids.
Synthases	synthesize molecules in anabolic reactions by condensing two smaller molecules together.
Isomerases	catalyze the rearrangement of bonds within a single molecule.
Polymerases	catalyze polymerization reactions such as the synthesis of DNA and RNA.
Kinases	catalyze the addition of phosphate groups to molecules. Protein kinases are an important group of kinases that attach phosphate groups to proteins.
Phosphatases	catalyze the hydrolytic removal of a phosphate group from a molecule.
Oxido-Reductases	general name for enzymes that catalyze reactions in which one molecule is oxidized while the other is reduced. Enzymes of this type are often more specifically named either *oxidases, reductases,* or *dehydrogenases.*
ATPases	hydrolyze ATP. Many proteins with a wide range of roles have an energy-harnessing ATPase activity as part of their function, for example, motor proteins such as *myosin* and membrane transport proteins such as the *sodium–potassium pump.*

Enzyme names typically end in "-ase," with the exception of some enzymes, such as pepsin, trypsin, thrombin and lysozyme that were discovered and named before the convention became generally accepted at the end of the nineteenth century. The common name of an enzyme usually indicates the substrate and the nature of the reaction catalyzed. For example, citrate synthase catalyzes the synthesis of citrate by a reaction between acetyl CoA and oxaloacetate.

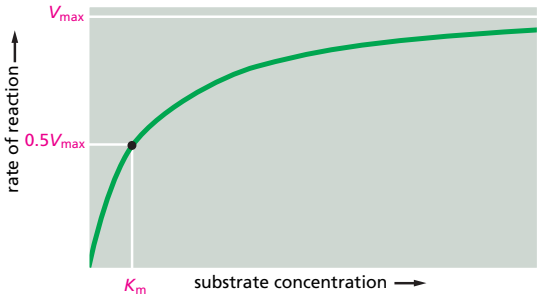


Figure 3–45 Enzyme kinetics. The rate of an enzyme reaction *(V)* increases as the substrate concentration increases until a maximum value (V_{max}) is reached. At this point all substrate-binding sites on the enzyme molecules are fully occupied, and the rate of reaction is limited by the rate of the catalytic process on the enzyme surface. For most enzymes, the concentration of substrate (K_m) at which the reaction rate is half-maximal (*black dot*) is a measure of how tightly the substrate is bound, with a large value of K_m corresponding to weak binding.

called the *turnover number*. The turnover number is often about 1000 substrate molecules processed per second per enzyme molecule, although turnover numbers between 1 and 10,000 are known.

The other kinetic parameter frequently used to characterize an enzyme is its K_m, the concentration of substrate that allows the reaction to proceed at one-half its maximum rate (0.5 V_{max}) (see Figure 3–45). A *low* K_m value means that the enzyme reaches its maximum catalytic rate at a *low concentration* of substrate and generally indicates that the enzyme binds to its substrate very tightly, whereas a *high* K_m value corresponds to weak binding. The methods used to characterize enzymes in this way are explained in **Panel 3–3** (pp. 162–163).

Enzymes Speed Reactions by Selectively Stabilizing Transition States

Enzymes achieve extremely high rates of chemical reaction—rates that are far higher than for any synthetic catalysts. There are several reasons for this efficiency. First, the enzyme increases the local concentration of substrate molecules at the catalytic site and holds all the appropriate atoms in the correct orientation for the reaction that is to follow. More importantly, however, some of the binding energy contributes directly to the catalysis. Substrate molecules must pass through a series of intermediate states of altered geometry and electron distribution before they form the ultimate products of the reaction. The free energy required to attain the most unstable **transition state** is called the *activation energy* for the reaction, and it is the major determinant of the reaction rate. Enzymes have a much higher affinity for the transition state of the substrate than they have for the stable form. Because this tight binding greatly lowers the energies of the transition state, the enzyme greatly accelerates a particular reaction by lowering the activation energy that is required (**Figure 3–46**).

By intentionally producing antibodies that act like enzymes, we can demonstrate that stabilizing a transition state can greatly increase a reaction rate. Consider, for example, the hydrolysis of an amide bond, which is similar to the peptide bond that joins two adjacent amino acids in a protein. In an aqueous solution, an amide bond hydrolyzes very slowly by the mechanism shown in **Figure 3–47**A. In the central intermediate, or transition state, the carbonyl carbon is bonded to four atoms arranged at the corners of a tetrahedron. By generating monoclonal antibodies that bind tightly to a stable analog of this very unstable *tetrahedral intermediate*, one can obtain an antibody that functions like an enzyme (Figure 3–47B). Because this *catalytic antibody* binds to and stabilizes the tetrahedral intermediate, it increases the spontaneous rate of amide-bond hydrolysis more than 10,000-fold.

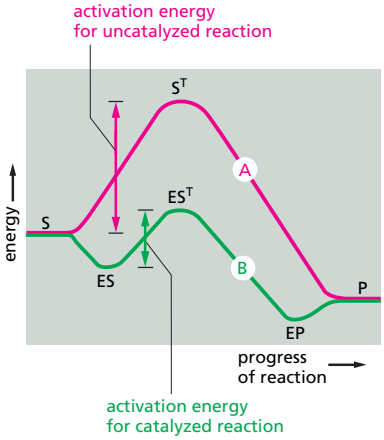


Figure 3–46 Enzymatic acceleration of chemical reactions by decreasing the activation energy. Often both the uncatalyzed reaction (A) and the enzyme-catalyzed reaction (B) can go through several transition states. It is the transition state with the highest energy (S^T and ES^T) that determines the activation energy and limits the rate of the reaction. (S = substrate; P = product of the reaction; ES = enzyme–substrate complex; EP = enzyme–product complex.)

Enzymes Can Use Simultaneous Acid and Base Catalysis

Figure 3–48 compares the spontaneous reaction rates and the corresponding enzyme-catalyzed rates for five enzymes. Rate accelerations range from 10^9 to 10^{23}. Clearly, enzymes are much better catalysts than catalytic antibodies.

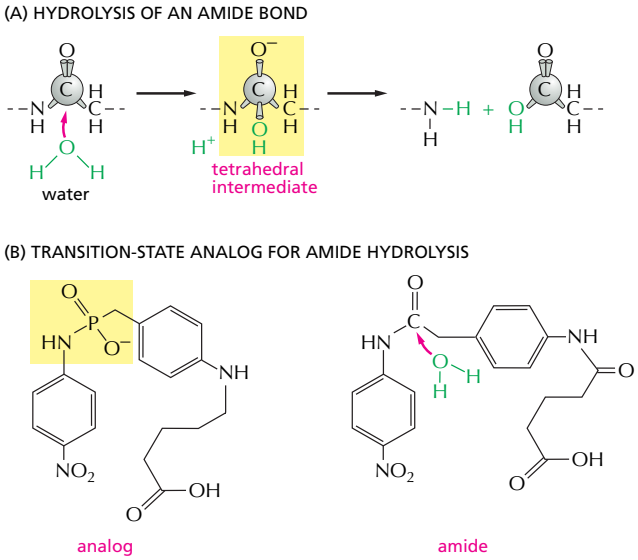


Figure 3–47 Catalytic antibodies. The stabilization of a transition state by an antibody creates an enzyme. (A) The reaction path for the hydrolysis of an amide bond goes through a tetrahedral intermediate, the high-energy transition state for the reaction. (B) The molecule on the left was covalently linked to a protein and used as an antigen to generate an antibody that binds tightly to the region of the molecule shown in *yellow*. Because this antibody also bound tightly to the transition state in (A), it was found to function as an enzyme that efficiently catalyzed the hydrolysis of the amide bond in the molecule on the right.

Enzymes not only bind tightly to a transition state, they also contain precisely positioned atoms that alter the electron distributions in those atoms that participate directly in the making and breaking of covalent bonds. Peptide bonds, for example, can be hydrolyzed in the absence of an enzyme by exposing a polypeptide to either a strong acid or a strong base, as illustrated in **Figure 3–49**. Enzymes are unique, however, in being able to use acid and base catalysis simultaneously, since the rigid framework of the protein binds the acidic and basic residues and prevents them from combining with each other (as they would do in solution) (Figure 3–49D).

The fit between an enzyme and its substrate needs to be precise. A small change introduced by genetic engineering in the active site of an enzyme can have a profound effect. Replacing a glutamic acid with an aspartic acid in one enzyme, for example, shifts the position of the catalytic carboxylate ion by only 1 Å (about the radius of a hydrogen atom); yet this is enough to decrease the activity of the enzyme a thousandfold.

Lysozyme Illustrates How an Enzyme Works <AGCA>

To demonstrate how enzymes catalyze chemical reactions, we examine an enzyme that acts as a natural antibiotic in egg white, saliva, tears, and other secretions. **Lysozyme** catalyzes the cutting of polysaccharide chains in the cell walls of bacteria. Because the bacterial cell is under pressure from osmotic forces, cutting even a small number of polysaccharide chains causes the cell wall to rupture and the cell to burst. Lysozyme is a relatively small and stable protein

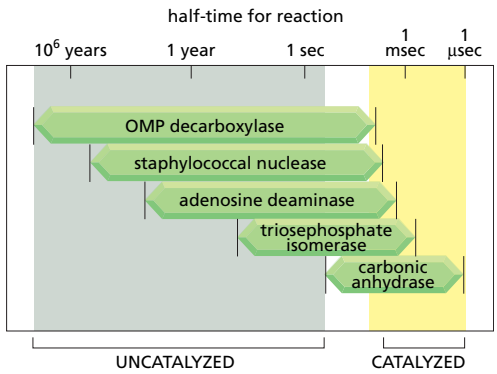


Figure 3–48 The rate accelerations caused by five different enzymes. (Adapted from A. Radzicka and R. Wolfenden, *Science* 267:90–93, 1995. With permission from AAAS.)

WHY ANALYZE THE KINETICS OF ENZYMES?

Enzymes are the most selective and powerful catalysts known. An understanding of their detailed mechanisms provides a critical tool for the discovery of new drugs, for the large-scale industrial synthesis of useful chemicals, and for appreciating the chemistry of cells and organisms. A detailed study of the rates of the chemical reactions that are catalyzed by a purified enzyme—more specifically how these rates change with changes in conditions such as the concentrations of substrates, products, inhibitors, and regulatory ligands—allows biochemists to figure out exactly how each enzyme works. For example, this is the way that the ATP-producing reactions of glycolysis, shown previously in Figure 2–72, were deciphered—allowing us to appreciate the rationale for this critical enzymatic pathway.

In this Panel, we introduce the important field of enzyme kinetics, which has been indispensable for deriving much of the detailed knowledge that we now have about cell chemistry.

STEADY-STATE ENZYME KINETICS

Many enzymes have only one substrate, which they bind and then process to produce products according to the scheme outlined in Figure 3–50A. In this case, the reaction is written as

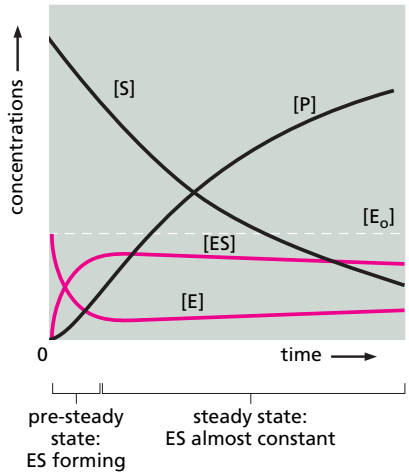
$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P$$

Here we have assumed that the reverse reaction, in which E + P recombine to form EP and then ES, occurs so rarely that we can ignore it. In this case, EP need not be represented, and we can express the rate of the reaction — known as its velocity, V, as

$$V = k_{cat}\,[ES]$$

where [ES] is the concentration of the enzyme–substrate complex, and k_{cat} is the turnover number, a rate constant that has a value equal to the number of substrate molecules processed per enzyme molecule each second.

But how does the value of [ES] relate to the concentrations that we know directly, which are the total concentration of the enzyme, $[E_o]$, and the concentration of the substrate, [S]? When enzyme and substrate are first mixed, the concentration [ES] will rise rapidly from zero to a so-called steady-state level, as illustrated below.



At this steady state, [ES] is nearly constant, so that

$$\text{rate of ES breakdown } k_{-1}\,[ES] + k_{cat}\,[ES] = \text{rate of ES formation } k_1\,[E][S]$$

or, since the concentration of the free enzyme, [E], is equal to $[E_o] - [ES]$,

$$[ES] = \left(\frac{k_1}{k_{-1} + k_{cat}}\right)[E][S] = \left(\frac{k_1}{k_{-1} + k_{cat}}\right)\left([E_o] - [ES]\right)[S]$$

Rearranging, and defining the constant K_m as

$$\frac{k_{-1} + k_{cat}}{k_1}$$

we get

$$[ES] = \frac{[E_o][S]}{K_m + [S]}$$

or, remembering that $V = k_{cat}\,[ES]$, we obtain the famous Michaelis–Menten equation

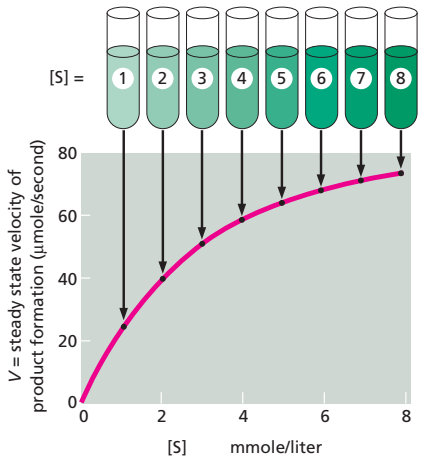
$$V = \frac{k_{cat}\,[E_o][S]}{K_m + [S]}$$

As [S] is increased to higher and higher levels, essentially all of the enzyme will be bound to substrate at steady state; at this point, a maximum rate of reaction, V_{max}, will be reached where $V = V_{max} = k_{cat}\,[E_o]$. Thus, it is convenient to rewrite the Michaelis–Menten equation as

$$V = \frac{V_{max}\,[S]}{K_m + [S]}$$

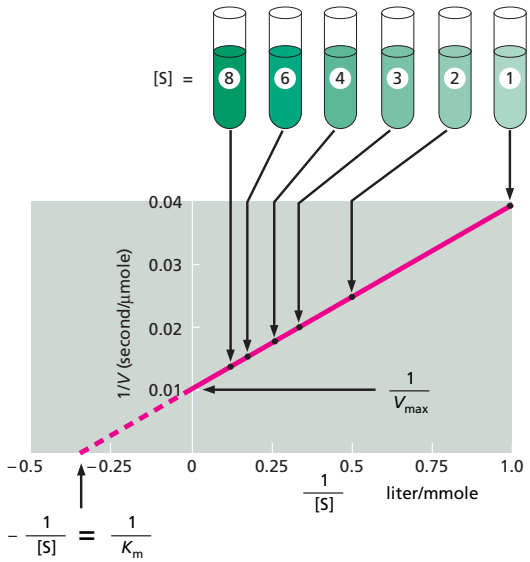
THE DOUBLE-RECIPROCAL PLOT

A typical plot of V versus [S] for an enzyme that follows Michaelis–Menten kinetics is shown below. From this plot, neither the value of V_{max} nor of K_m is immediately clear.



To obtain V_{max} and K_m from such data, a double-reciprocal plot is often used, in which the Michaelis–Menten equation has merely been rearranged, so that $1/V$ can be plotted versus $1/[S]$.

$$1/V = \left(\frac{K_m}{V_{max}}\right)\left(\frac{1}{[S]}\right) + 1/V_{max}$$



THE SIGNIFICANCE OF K_m, k_{cat}, and k_{cat}/K_m

As described in the text, K_m is an approximate measure of substrate affinity for the enzyme: it is numerically equal to the concentration of [S] at $V = 0.5\ V_{max}$. In general, a lower value of K_m means tighter substrate binding. In fact, for those cases where k_{cat} is much smaller than k_{-1}, the K_m will be equal to K_d, the dissociation constant for substrate binding to the enzyme ($K_d = 1/K_a$).

We have seen that k_{cat} is the turnover number for the enzyme. At very low substrate concentrations, where $[S] << K_m$, most of the enzyme is free. Thus we can think of $[E] = [E_o]$, so that the Michaelis–Menten equation becomes $V = k_{cat}/K_m\ [E][S]$. Thus, the ratio k_{cat}/K_m is equivalent to the rate constant for the reaction between free enzyme and free substrate.

A comparison of k_{cat}/K_m for the same enzyme with different substrates, or for two enzymes with their different substrates, is widely used as a measure of enzyme effectiveness.

For simplicity, in this Panel we have discussed enzymes that have only one substrate, such as the lysozyme enzyme described in the text (see p. 164). Most enzymes have two substrates, one of which is often an active carrier molecule—such as NADH or ATP.

A similar, but more complex, analysis is used to determine the kinetics of such enzymes—allowing the order of substrate binding and the presence of covalent intermediates along the pathway to be revealed.

SOME ENZYMES ARE DIFFUSION LIMITED

The values of k_{cat}, K_m, and k_{cat}/K_m for some selected enzymes are given below:

enzyme	substrate	k_{cat} (sec^{-1})	K_m (M)	k_{cat}/K_m ($sec^{-1}M^{-1}$)
acetylcholinesterase	acetylcholine	$1.4x10^4$	$9x10^{-5}$	$1.6x10^8$
catalase	H_2O_2	$4x10^7$	1	$4x10^7$
fumarase	fumarate	$8x10^2$	$5x10^{-6}$	$1.6x10^8$

Because an enzyme and its substrate must collide before they can react, k_{cat}/K_m has a maximum possible value that is limited by collision rates. If every collision forms an enzyme–substrate complex, one can calculate from diffusion theory that k_{cat}/K_m will be between 10^8 and $10^9\ sec^{-1}M^{-1}$, in the case where all subsequent steps proceed immediately. Thus, it is claimed that enzymes like acetylcholinesterase and fumarase are "perfect enzymes," each enzyme having evolved to the point where nearly every collision with its substrate converts the substrate to a product.

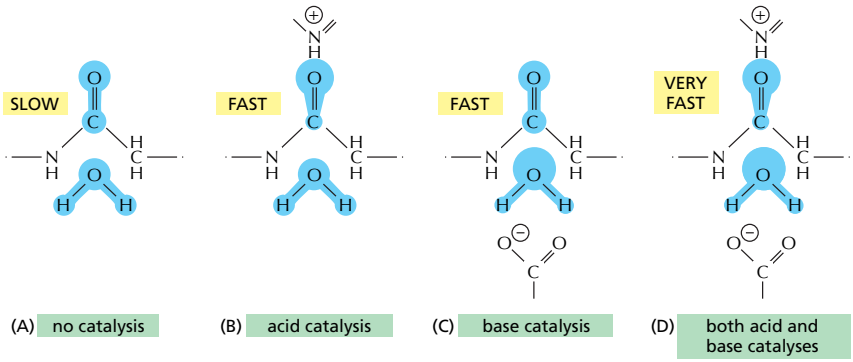


Figure 3–49 Acid catalysis and base catalysis. (A) The start of the uncatalyzed reaction shown in Figure 3–47A, with *blue* indicating electron distribution in the water and carbonyl bonds. (B) An acid likes to donate a proton (H^+) to other atoms. By pairing with the carbonyl oxygen, an acid causes electrons to move away from the carbonyl carbon, making this atom much more attractive to the electronegative oxygen of an attacking water molecule. (C) A base likes to take up H^+. By pairing with a hydrogen of the attacking water molecule, a base causes electrons to move toward the water oxygen, making it a better attacking group for the carbonyl carbon. (D) By having appropriately positioned atoms on its surface, an enzyme can perform both acid catalysis and base catalysis at the same time.

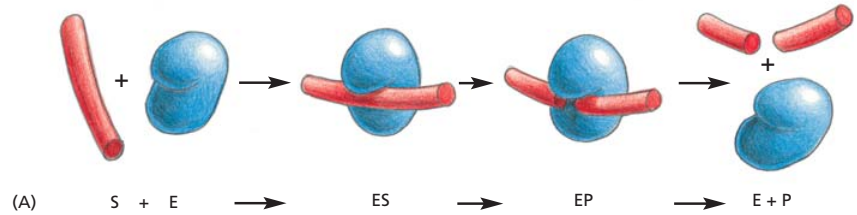
that can be easily isolated in large quantities. For these reasons, it has been intensively studied, and it was the first enzyme to have its structure worked out in atomic detail by x-ray crystallography.

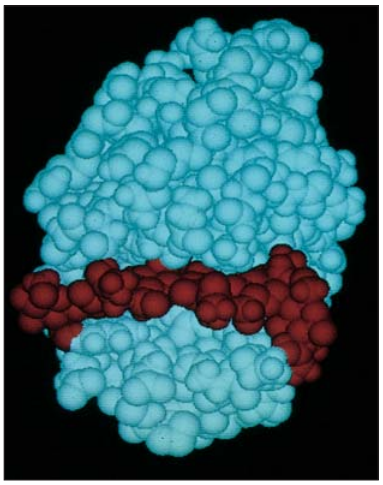
The reaction that lysozyme catalyzes is a hydrolysis: it adds a molecule of water to a single bond between two adjacent sugar groups in the polysaccharide chain, thereby causing the bond to break (see Figure 2–19). The reaction is energetically favorable because the free energy of the severed polysaccharide chain is lower than the free energy of the intact chain. However, the pure polysaccharide can remain for years in water without being hydrolyzed to any detectable degree. This is because there is an energy barrier to the reaction, as discussed in Chapter 2 (see Figure 2–46). A colliding water molecule can break a bond linking two sugars only if the polysaccharide molecule is distorted into a particular shape—the transition state—in which the atoms around the bond have an altered geometry and electron distribution. Because of this distortion, random collisions must supply a very large activation energy for the reaction to take place. In an aqueous solution at room temperature, the energy of collisions almost never exceeds the activation energy. Consequently, hydrolysis occurs extremely slowly, if at all.

This situation changes drastically when the polysaccharide binds to lysozyme. The active site of lysozyme, because its substrate is a polymer, is a long groove that holds six linked sugars at the same time. As soon as the polysaccharide binds to form an enzyme–substrate complex, the enzyme cuts the polysaccharide by adding a water molecule across one of its sugar–sugar bonds. The product chains are then quickly released, freeing the enzyme for further cycles of reaction (**Figure 3–50**).

The chemistry of the binding of lysozyme to its substrate is the same as that for antibody binding to its antigen—the formation of multiple noncovalent

Figure 3–50 The reaction catalyzed by lysozyme. (A) The enzyme lysozyme (E) catalyzes the cutting of a polysaccharide chain, which is its substrate (S). The enzyme first binds to the chain to form an enzyme–substrate complex (ES) and then catalyzes the cleavage of a specific covalent bond in the backbone of the polysaccharide, forming an enzyme–product complex (EP) that rapidly dissociates. Release of the severed chain (the products P) leaves the enzyme free to act on another substrate molecule. (B) A space-filling model of the lysozyme molecule bound to a short length of polysaccharide chain before cleavage. (B, courtesy of Richard J. Feldmann.)





(B)

bonds. However, lysozyme holds its polysaccharide substrate in a particular way, so that it distorts one of the two sugars in the bond to be broken from its normal, most stable conformation. The bond to be broken is also held close to two amino acids with acidic side chains (a glutamic acid and an aspartic acid) within the active site.

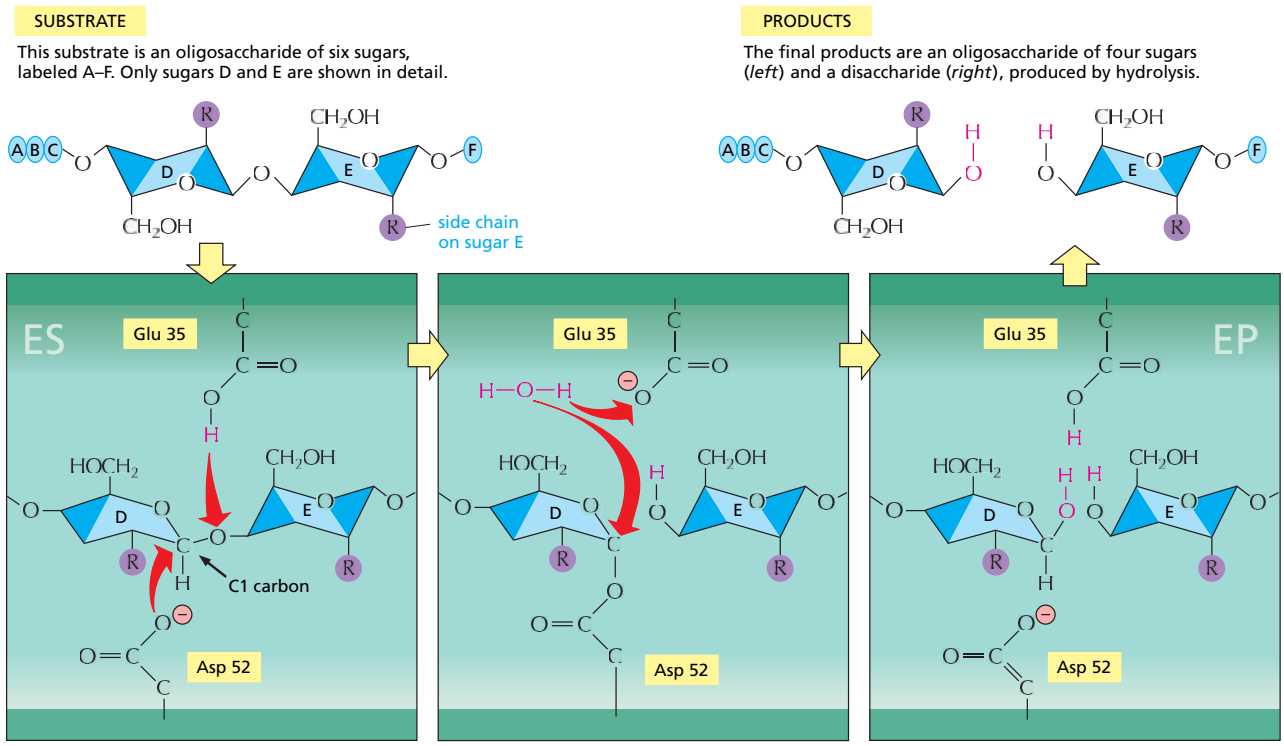
Conditions are thereby created in the microenvironment of the lysozyme active site that greatly reduce the activation energy necessary for the hydrolysis to take place. **Figure 3–51** shows three central steps in this enzymatically catalyzed reaction.

1. The enzyme stresses its bound substrate, so that the shape of one sugar more closely resembles the shape of high-energy transition states formed during the reaction.
2. The negatively charged aspartic acid reacts with the C1 carbon atom on the distorted sugar, and the glutamic acid donates its proton to the oxygen that links this sugar to its neighbor. This breaks the sugar–sugar bond and leaves the aspartic acid side chain covalently linked to the site of bond cleavage.
3. Aided by the negatively charged glutamic acid, a water molecule reacts with the C1 carbon atom, displacing the aspartic acid side chain and completing the process of hydrolysis.

The overall chemical reaction, from the initial binding of the polysaccharide on the surface of the enzyme through the final release of the severed chains, occurs many millions of times faster than it would in the absence of enzyme.

Other enzymes use similar mechanisms to lower activation energies and speed up the reactions they catalyze. In reactions involving two or more reactants, the active site also acts like a template, or mold, that brings the substrates together in the proper orientation for a reaction to occur between them (**Figure**

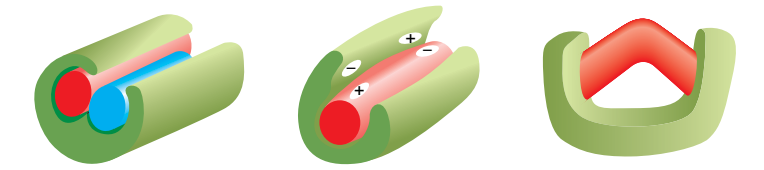
Figure 3–51 Events at the active site of lysozyme. <TGGT> The top left and top right drawings show the free substrate and the free products, respectively, whereas the other three drawings show the sequential events at the enzyme active site. Note the change in the conformation of sugar D in the enzyme–substrate complex; this shape change stabilizes the oxocarbenium ion-like transition states required for formation and hydrolysis of the covalent intermediate shown in the middle panel. It is also possible that a carbonium ion intermediate forms in step 2, as the covalent intermediate shown in the middle panel has been detected only with a synthetic substrate. (See D.J. Vocadlo et al., *Nature* 412:835–838, 2001.)



In the enzyme–substrate complex (ES), the enzyme forces sugar D into a strained conformation, with Glu 35 positioned to serve as an acid that attacks the adjacent sugar–sugar bond by donating a proton (H^+) to sugar E, and Asp 52 poised to attack the C1 carbon atom.

The Asp 52 has formed a covalent bond between the enzyme and the C1 carbon atom of sugar D. The Glu 35 then polarizes a water molecule (*red*), so that its oxygen can readily attack the C1 carbon atom and displace Asp 52.

The reaction of the water molecule (*red*) completes the hydrolysis and returns the enzyme to its initial state, forming the final enzyme–product complex (EP).



(A) enzyme binds to two substrate molecules and orients them precisely to encourage a reaction to occur between them

(B) binding of substrate to enzyme rearranges electrons in the substrate, creating partial negative and positive charges that favor a reaction

(C) enzyme strains the bound substrate molecule, forcing it toward a transition state to favor a reaction

Figure 3–52 Some general strategies of enzyme catalysis. (A) Holding substrates together in a precise alignment. (B) Charge stabilization of reaction intermediates. (C) Applying forces that distort bonds in the substrate to increase the rate of a particular reaction.

3–52A). As we saw for lysozyme, the active site of an enzyme contains precisely positioned atoms that speed up a reaction by using charged groups to alter the distribution of electrons in the substrates (Figure 3–52B). In addition, when a substrate binds to an enzyme, bonds in the substrate often bend, changing the substrate shape. These changes, along with mechanical forces, drive a substrate toward a particular transition state (Figure 3–52C). Finally, like lysozyme, many enzymes participate intimately in the reaction by briefly forming a covalent bond between the substrate and a side chain of the enzyme. Subsequent steps in the reaction restore the side chain to its original state, so that the enzyme remains unchanged after the reaction (see also Figure 2–72).

Tightly Bound Small Molecules Add Extra Functions to Proteins

Although we have emphasized the versatility of proteins as chains of amino acids that perform different functions, there are many instances in which the amino acids by themselves are not enough. Just as humans employ tools to enhance and extend the capabilities of their hands, proteins often use small nonprotein molecules to perform functions that would be difficult or impossible to do with amino acids alone. Thus, the signal receptor protein *rhodopsin*, which is made by the photoreceptor cells in the retina, detects light by means of a small molecule, *retinal*, embedded in the protein (**Figure 3–53**A). Retinal changes its shape when it absorbs a photon of light, and this change causes the protein to trigger a cascade of enzymatic reactions that eventually lead to an electrical signal being carried to the brain.

Another example of a protein that contains a nonprotein portion is hemoglobin (see Figure 3–22). A molecule of hemoglobin carries four *heme* groups, ring-shaped molecules each with a single central iron atom (Figure 3–53B). Heme gives hemoglobin (and blood) its red color. By binding reversibly to oxygen gas through its iron atom, heme enables hemoglobin to pick up oxygen in the lungs and release it in the tissues.

Sometimes these small molecules are attached covalently and permanently to their protein, thereby becoming an integral part of the protein molecule itself. We shall see in Chapter 10 that proteins are often anchored to cell membranes through covalently attached lipid molecules. And membrane proteins exposed

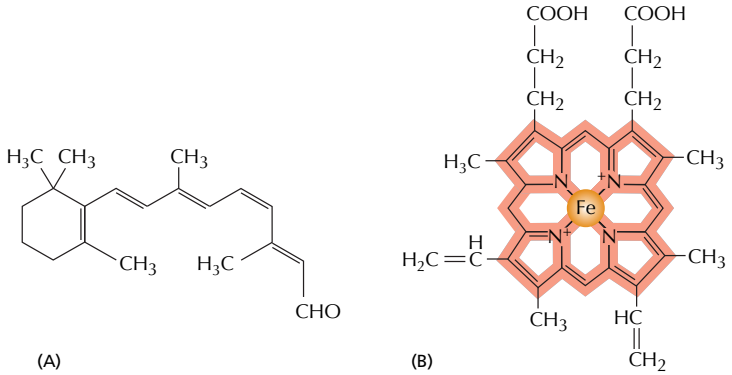


Figure 3–53 Retinal and heme. (A) The structure of retinal, the light-sensitive molecule attached to rhodopsin in the eye. (B) The structure of a heme group. The carbon-containing heme ring is *red* and the iron atom at its center is *orange*. A heme group is tightly bound to each of the four polypeptide chains in hemoglobin, the oxygen-carrying protein whose structure is shown in Figure 3–22.

Table 3–2 Many Vitamins Provide Critical Coenzymes for Human Cells

VITAMIN	COENZYME	ENZYME-CATALYZED REACTIONS REQUIRING THESE COENZYMES
Thiamine (vitamin B_1)	thiamine pyrophosphate	activation and transfer of aldehydes
Riboflavin (vitamin B_2)	FADH	oxidation–reduction
Niacin	NADH, NADPH	oxidation–reduction
Pantothenic acid	coenzyme A	acyl group activation and transfer
Pyridoxine	pyridoxal phosphate	amino acid activation; also glycogen phosphorylase
Biotin	biotin	CO_2 activation and transfer
Lipoic acid	lipoamide	acyl group activation; oxidation–reduction
Folic acid	tetrahydrofolate	activation and transfer of single carbon groups
Vitamin B_{12}	cobalamin coenzymes	isomerization and methyl group transfers

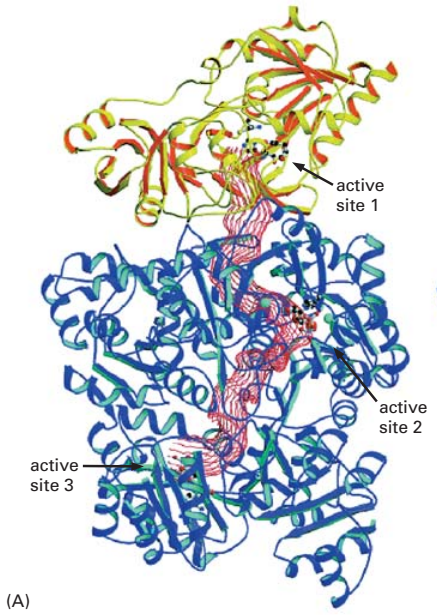
on the surface of the cell, as well as proteins secreted outside the cell, are often modified by the covalent addition of sugars and oligosaccharides.

Enzymes frequently have a small molecule or metal atom tightly associated with their active site that assists with their catalytic function. *Carboxypeptidase*, for example, an enzyme that cuts polypeptide chains, carries a tightly bound zinc ion in its active site. During the cleavage of a peptide bond by carboxypeptidase, the zinc ion forms a transient bond with one of the substrate atoms, thereby assisting the hydrolysis reaction. In other enzymes, a small organic molecule serves a similar purpose. Such organic molecules are often referred to as **coenzymes**. An example is *biotin*, which is found in enzymes that transfer a carboxylate group ($-COO^-$) from one molecule to another (see Figure 2–63). Biotin participates in these reactions by forming a transient covalent bond to the $-COO^-$ group to be transferred, being better suited to this function than any of the amino acids used to make proteins. Because it cannot be synthesized by humans, and must therefore be supplied in small quantities in our diet, biotin is a *vitamin*. Many other coenzymes are produced from vitamins (**Table 3–2**). Vitamins are also needed to make other types of small molecules that are essential components of our proteins; vitamin A, for example, is needed in the diet to make retinal, the light-sensitive part of rhodopsin.

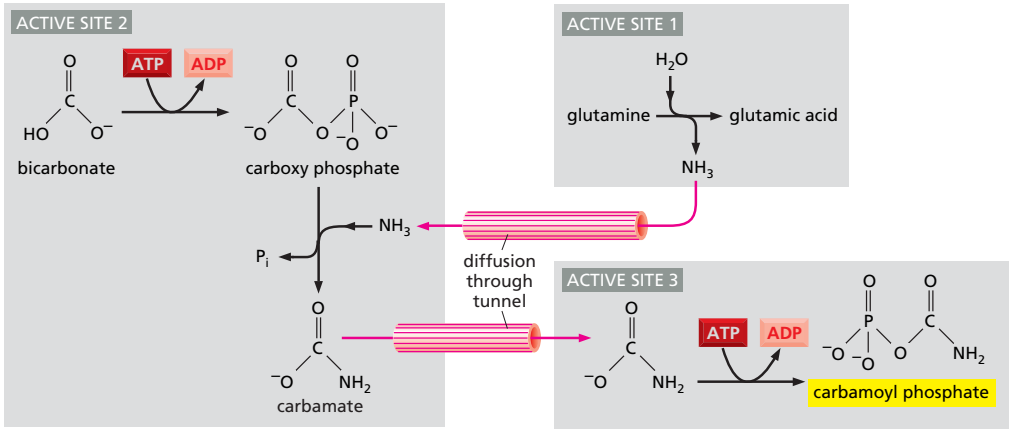
Molecular Tunnels Channel Substrates in Enzymes with Multiple Catalytic Sites

Some of the chemical reactions catalyzed by enzymes in cells produce intermediates that are either very unstable or that could readily diffuse out of the cell through the plasma membrane if released into the cytosol. To preserve these intermediates, enzymes have evolved *molecular tunnels* that connect two or more active sites, allowing the intermediate to be rapidly processed to a final product—without ever leaving the enzyme.

Consider, for example, the enzyme carbamoyl phosphate synthetase, which uses ammonia derived from glutamine plus two molecules of ATP to convert bicarbonate (HCO_3^-) to carbamoyl phosphate—an important intermediate in several metabolic pathways (**Figure 3–54**). This enzyme contains three widely separated active sites that are connected to each other by a tunnel. The reaction starts at active site 2, located in the middle of the tunnel, where ATP is used to phosphorylate (add a phosphate group to) bicarbonate, forming carboxy phosphate. This event triggers the hydrolysis of glutamine to glutamic acid at active site 1, releasing ammonia into the tunnel. The ammonia immediately diffuses through the first half of the tunnel to active site 2, where it reacts with the carboxyphosphate to form carbamate. This unstable intermediate then diffuses through the second half of the tunnel to active site 3, where it is phosphorylated by ATP to the final product, carbamoyl phosphate.



(A)



(B)

Figure 3–54 The tunneling of reaction intermediates in the enzyme carbamoyl phosphate synthetase. (A) Diagram of the structure of the enzyme, in which a *red* ribbon has been used to outline the tunnel on the inside of the protein connecting its three active sites. The small and large subunits of this dimeric enzyme are color coded *yellow* and *blue*, respectively. (B) The path of the reaction. As indicated, active site 1 produces ammonia, which diffuses through the tunnel to active site 2, where it combines with carboxy phosphate to form carbamate. This highly unstable intermediate then diffuses through the tunnel to active site 3, where it is phosphorylated by ATP to produce the final product, carbamoyl phosphate. (A, modified from F.M. Raushel, J.B. Thoden, and H.M. Holden, *Acc. Chem. Res.* 36:539–548, 2003. With permission from American Chemical Society.)

Several other well characterized enzymes contain similar molecular tunnels. Ammonia, a readily diffusable intermediate that might otherwise be lost from the cell, is the substrate most frequently channeled in the examples thus far known.

Multienzyme Complexes Help to Increase the Rate of Cell Metabolism

The efficiency of enzymes in accelerating chemical reactions is crucial to the maintenance of life. Cells, in effect, must race against the unavoidable processes of decay, which—if left unattended—cause macromolecules to run downhill toward greater and greater disorder. If the rates of desirable reactions were not greater than the rates of competing side reactions, a cell would soon die. We can get some idea of the rate at which cell metabolism proceeds by measuring the rate of ATP utilization. A typical mammalian cell "turns over" (i.e., hydrolyzes and restores by phosphorylation) its entire ATP pool once every 1 or 2 minutes. For each cell, this turnover represents the utilization of roughly 10^7 molecules of ATP per second (or, for the human body, about 1 gram of ATP every minute).

The rates of reactions in cells are rapid because enzyme catalysis is so effective. Many important enzymes have become so efficient that there is no possibility of further useful improvement. The factor that limits the reaction rate is no longer the enzyme's intrinsic speed of action; rather, it is the frequency with which the enzyme collides with its substrate. Such a reaction is said to be *diffusion-limited* (see Panel 3–3, p. 162–163).

If an enzyme-catalyzed reaction is diffusion-limited, its rate depends on the concentration of both the enzyme and its substrate. If a sequence of reactions is to occur extremely rapidly, each metabolic intermediate and enzyme involved must be present in high concentration. However, given the enormous number of different reactions performed by a cell, there are limits to the concentrations that can be achieved. In fact, most metabolites are present in micromolar (10^{-6} M) concentrations, and most enzyme concentrations are much lower. How is it possible, therefore, to maintain very fast metabolic rates?

The answer lies in the spatial organization of cell components. The cell can increase reaction rates without raising substrate concentrations by bringing the various enzymes involved in a reaction sequence together to form a large protein assembly known as a *multienzyme complex* (**Figure 3–55**). Because this allows the product of enzyme A to be passed directly to enzyme B, and so on, diffusion rates need not be limiting, even when the concentrations of the substrates in the cell as a whole are very low. It is perhaps not surprising, therefore, that such enzyme complexes are very common, and they are involved in nearly all aspects of metabolism—including the central genetic processes of DNA, RNA, and protein synthesis. In fact, few enzymes in eucaryotic cells diffuse freely in solution; instead, most seem to have evolved binding sites that concentrate them with other proteins of related function in particular regions of the cell, thereby increasing the rate and efficiency of the reactions that they catalyze.

Eucaryotic cells have yet another way of increasing the rate of metabolic reactions: using their intracellular membrane systems. These membranes can segregate particular substrates and the enzymes that act on them into the same membrane-enclosed compartment, such as the endoplasmic reticulum or the cell nucleus. If, for example, a compartment occupies a total of 10% of the volume of the cell, the concentration of reactants in that compartment may be increased by 10 times compared with a cell with the same number of enzymes and substrate molecules, but no compartmentalization. Reactions limited by the speed of diffusion can thereby be speeded up by a factor of 10.

The Cell Regulates the Catalytic Activities of its Enzymes

A living cell contains thousands of enzymes, many of which operate at the same time and in the same small volume of the cytosol. By their catalytic action, these enzymes generate a complex web of metabolic pathways, each composed of chains of chemical reactions in which the product of one enzyme becomes the substrate of the next. In this maze of pathways, there are many branch points (nodes) where different enzymes compete for the same substrate. The system is so complex (see Figure 2–88) that elaborate controls are required to regulate when and how rapidly each reaction occurs.

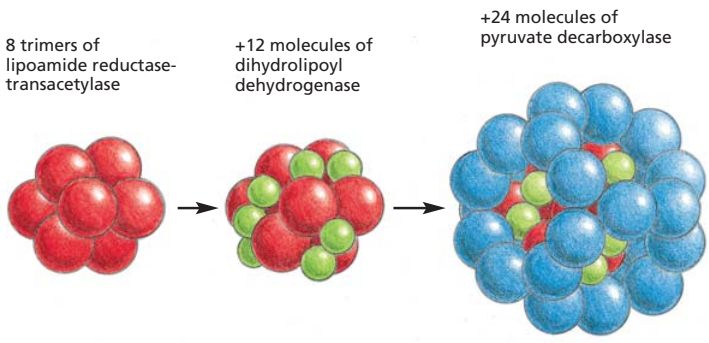


Figure 3–55 The structure of pyruvate dehydrogenase. This enzyme complex catalyzes the conversion of pyruvate to acetyl CoA, as part of the pathway that oxidizes sugars to CO_2 and H_2O (see Figure 2–79). It is an example of a large multienzyme complex in which reaction intermediates are passed directly from one enzyme to another.

Regulation occurs at many levels. At one level, the cell controls how many molecules of each enzyme it makes by regulating the expression of the gene that encodes that enzyme (discussed in Chapter 7). The cell also controls enzymatic activities by confining sets of enzymes to particular subcellular compartments, enclosed by distinct membranes (discussed in Chapters 12 and 14). As will be discussed later in this chapter, enzymes are frequently covalently modified to control their activity. The rate of protein destruction by targeted proteolysis represents yet another important regulatory mechanism (see p. 395). But the most general process that adjusts reaction rates operates through a direct, reversible change in the activity of an enzyme in response to the specific small molecules that it encounters.

The most common type of control occurs when a molecule other than one of the substrates binds to an enzyme at a special regulatory site outside the active site, thereby altering the rate at which the enzyme converts its substrates to products. For example, in **feedback inhibition** a product produced late in a reaction pathway inhibits an enzyme that acts earlier in the pathway. Thus, whenever large quantities of the final product begin to accumulate, this product binds to the enzyme and slows down its catalytic action, thereby limiting the further entry of substrates into that reaction pathway (**Figure 3–56**). Where pathways branch or intersect, there are usually multiple points of control by different final products, each of which works to regulate its own synthesis (**Figure 3–57**). Feedback inhibition can work almost instantaneously, and it is rapidly reversed when the level of the product falls.

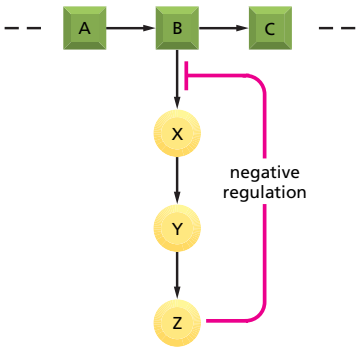


Figure 3–56 Feedback inhibition of a single biosynthetic pathway. The end-product Z inhibits the first enzyme that is unique to its synthesis and thereby controls its own level in the cell. This is an example of negative regulation.

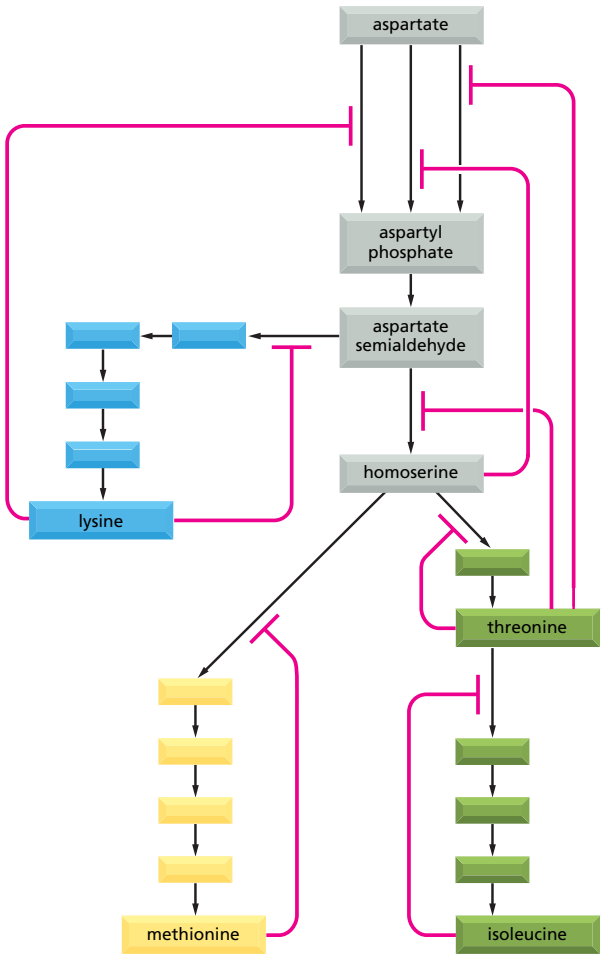


Figure 3–57 Multiple feedback inhibition. In this example, which shows the biosynthetic pathways for four different amino acids in bacteria, the *red arrows* indicate positions at which products feed back to inhibit enzymes. Each amino acid controls the first enzyme specific to its own synthesis, thereby controlling its own levels and avoiding a wasteful, or even dangerous, buildup of intermediates. The products can also separately inhibit the initial set of reactions common to all the syntheses; in this case, three different enzymes catalyze the initial reaction, each inhibited by a different product.

Feedback inhibition is *negative regulation:* it prevents an enzyme from acting. Enzymes can also be subject to *positive regulation*, in which a regulatory molecule stimulates the enzyme's activity rather than shutting the enzyme down. Positive regulation occurs when a product in one branch of the metabolic network stimulates the activity of an enzyme in another pathway. As one example, the accumulation of ADP activates several enzymes involved in the oxidation of sugar molecules, thereby stimulating the cell to convert more ADP to ATP.

Allosteric Enzymes Have Two or More Binding Sites That Interact

A striking feature of both positive and negative feedback regulation is that the regulatory molecule often has a shape totally different from the shape of the substrate of the enzyme. This is why the effect on a protein is termed *allostery* (from the Greek words *allos*, meaning "other," and *stereos*, meaning "solid" or "three-dimensional"). As biologists learned more about feedback regulation, they recognized that the enzymes involved must have at least two different binding sites on their surface—an **active site** that recognizes the substrates, and a **regulatory site** that recognizes a regulatory molecule. These two sites must somehow communicate so that the catalytic events at the active site can be influenced by the binding of the regulatory molecule at its separate site on the protein's surface.

The interaction between separated sites on a protein molecule is now known to depend on a *conformational change* in the protein: binding at one of the sites causes a shift from one folded shape to a slightly different folded shape. During feedback inhibition, for example, the binding of an inhibitor at one site on the protein causes the protein to shift to a conformation that incapacitates its active site, located elsewhere in the protein.

It is thought that most protein molecules are allosteric. They can adopt two or more slightly different conformations, and a shift from one to another caused by the binding of a ligand can alter their activity. This is true not only for enzymes but also for many other proteins, including receptors, structural proteins, and motor proteins. In all instances of allosteric regulation, each conformation of the protein has somewhat different surface contours, and the protein's binding sites for ligands are altered when the protein changes shape. Moreover as we discuss next, each ligand will stabilize the conformation that it binds to most strongly, and thus—at high enough concentrations—will tend to "switch" the protein toward the conformation that the ligand prefers.

Two Ligands Whose Binding Sites Are Coupled Must Reciprocally Affect Each Other's Binding

The effects of ligand binding on a protein follow from a fundamental chemical principle known as **linkage**. Suppose, for example, that a protein that binds glucose also binds another molecule, X, at a distant site on the protein's surface. If the binding site for X changes shape as part of the conformational change induced by glucose binding, the binding sites for X and for glucose are said to be *coupled*. Whenever two ligands prefer to bind to the *same* conformation of an allosteric protein, it follows from basic thermodynamic principles that each ligand must increase the affinity of the protein for the other. Thus, if the shift of the protein in **Figure 3–58** to the closed conformation that binds glucose best also causes the binding site for X to fit X better, then the protein will bind glucose more tightly when X is present than when X is absent.

Conversely, linkage operates in a negative way if two ligands prefer to bind to *different* conformations of the same protein. In this case, the binding of the first ligand discourages the binding of the second ligand. Thus, if a shape change caused by glucose binding decreases the affinity of a protein for molecule X, the binding of X must also decrease the protein's affinity for glucose (**Figure 3–59**). The linkage relationship is quantitatively reciprocal, so that, for example, if glucose has a very large effect on the binding of X, X has a very large effect on the binding of glucose.

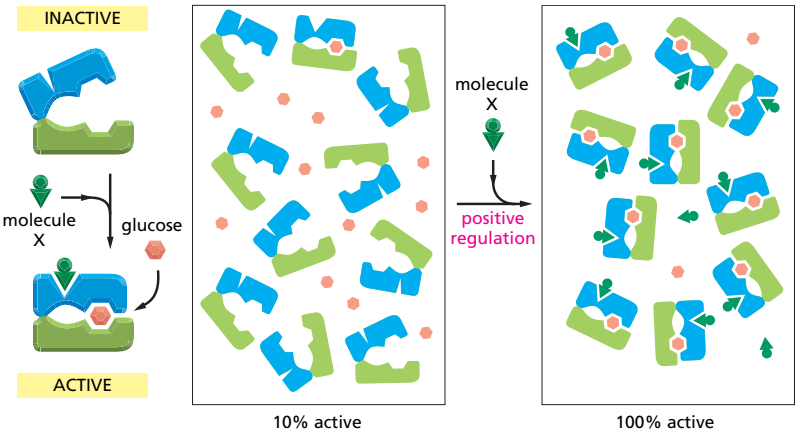


Figure 3–58 Positive regulation caused by conformational coupling between two distant binding sites. In this example, both glucose and molecule X bind best to the *closed* conformation of a protein with two domains. Because both glucose and molecule X drive the protein toward its closed conformation, each ligand helps the other to bind. Glucose and molecule X are therefore said to bind *cooperatively* to the protein.

The relationships shown in Figures 3–58 and 3–59 apply to all proteins, and they underlie all of cell biology. They seem so obvious in retrospect that we now take it for granted. But the discovery of linkage in studies of a few enzymes in the 1950s, followed by an extensive analysis of allosteric mechanisms in proteins in the early 1960s, had a revolutionary effect on our understanding of biology. Since molecule X in these examples binds at a site on the enzyme that is distinct from the site where catalysis occurs, it need have no chemical relationship to glucose or to any other ligand that binds at the active site. Moreover, as we have just seen, for enzymes that are regulated in this way, molecule X can either turn the enzyme on (positive regulation) or turn it off (negative regulation). By such a mechanism, **allosteric proteins** serve as general switches that, in principle, allow one molecule in a cell to affect the fate of any other.

Symmetric Protein Assemblies Produce Cooperative Allosteric Transitions

A single-subunit enzyme that is regulated by negative feedback can at most decrease from 90% to about 10% activity in response to a 100-fold increase in the concentration of an inhibitory ligand that it binds (**Figure 3–60**, *red line).* Responses of this type are apparently not sharp enough for optimal cell regulation, and most enzymes that are turned on or off by ligand binding consist of symmetric assemblies of identical subunits. With this arrangement, the binding of a molecule of ligand to a single site on one subunit can promote an allosteric change in the entire assembly that helps the neighboring subunits bind the same ligand. As a result, a *cooperative allosteric transition* occurs (Figure 3–60, *blue line),* allowing a relatively small change in ligand concentration in the cell to switch the whole assembly from an almost fully active to an almost fully inactive conformation (or vice versa).

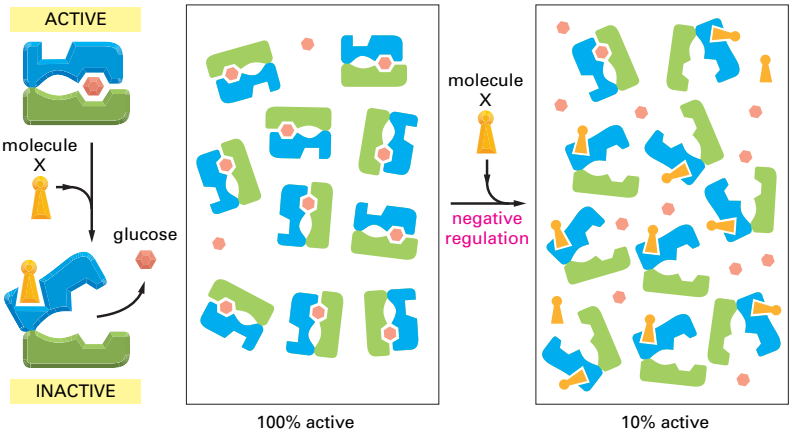


Figure 3–59 Negative regulation caused by conformational coupling between two distant binding sites. The scheme here resembles that in the previous figure, but here molecule X prefers the *open* conformation, while glucose prefers the *closed* conformation. Because glucose and molecule X drive the protein toward opposite conformations (closed and open, respectively), the presence of either ligand interferes with the binding of the other.

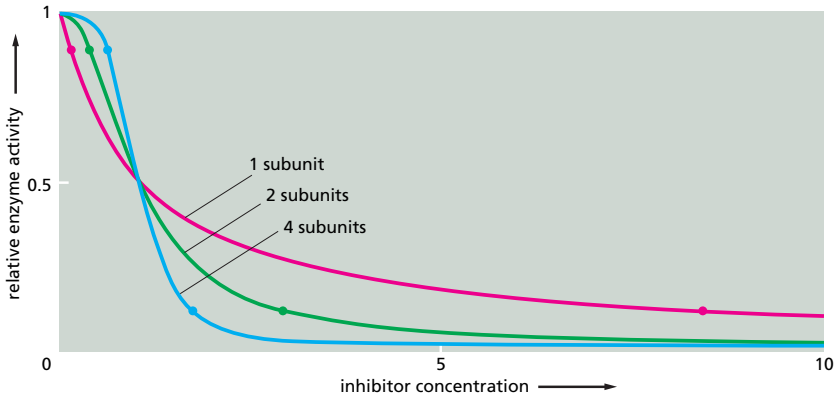


Figure 3–60 Enzyme activity versus the concentration of inhibitory ligand for single-subunit and multisubunit allosteric enzymes. For an enzyme with a single subunit *(red line),* a drop from 90% enzyme activity to 10% activity (indicated by the two dots on the curve) requires a 100-fold increase in the concentration of inhibitor. The enzyme activity is calculated from the simple equilibrium relationship $K = [IP]/[I][P]$, where P is active protein, I is inhibitor, and IP is the inactive protein bound to inhibitor. An identical curve applies to any simple binding interaction between two molecules, A and B. In contrast, a multisubunit allosteric enzyme can respond in a switchlike manner to a change in ligand concentration: the steep response is caused by a cooperative binding of the ligand molecules, as explained in Figure 3–61. Here, the *green line* represents the idealized result expected for the cooperative binding of two inhibitory ligand molecules to an allosteric enzyme with two subunits, and the *blue line* shows the idealized response of an enzyme with four subunits. As indicated by the two dots on each of these curves, the more complex enzymes drop from 90% to 10% activity over a much narrower range of inhibitor concentration than does the enzyme composed of a single subunit.

The principles involved in a cooperative "all-or-none" transition are the same for all proteins, whether or not they are enzymes. But they are perhaps easiest to visualize for an enzyme that forms a symmetric dimer. In the example shown in **Figure 3–61**, the first molecule of an inhibitory ligand binds with great difficulty since its binding disrupts an energetically favorable interaction between the two identical monomers in the dimer. A second molecule of inhibitory ligand now binds more easily, however, because its binding restores the energetically favorable monomer–monomer contacts of a symmetric dimer (this also completely inactivates the enzyme).

As an alternative to this *induced fit* model for a cooperative allosteric transition, we can view such a symmetrical enzyme as having only two possible conformations, corresponding to the "enzyme on" and "enzyme off" structures in Figure 3–61. In this view, ligand binding perturbs an all-or-none equilibrium between these two states, thereby changing the proportion of active molecules. Both models represent true and useful concepts; it is the second model that we shall describe next.

The Allosteric Transition in Aspartate Transcarbamoylase Is Understood in Atomic Detail

One enzyme used in the early studies of allosteric regulation was aspartate transcarbamoylase from *E. coli*. It catalyzes the important reaction that begins the synthesis of the pyrimidine ring of C, U, and T nucleotides: carbamoyl phosphate + aspartate → *N*-carbamoylaspartate. One of the final products of this pathway, cytosine triphosphate (CTP), binds to the enzyme to turn it off whenever CTP is plentiful.

Aspartate transcarbamoylase is a large complex of six regulatory and six catalytic subunits. The catalytic subunits form two trimers, each arranged in the shape of an equilateral triangle; the two trimers face each other and are held

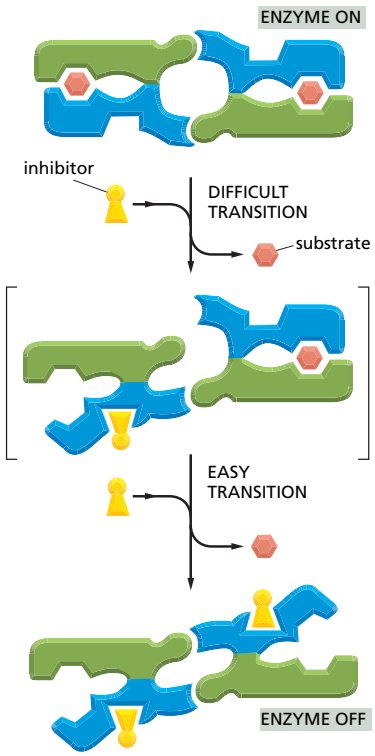


Figure 3–61 A cooperative allosteric transition in an enzyme composed of two identical subunits. This diagram illustrates how the conformation of one subunit can influence that of its neighbor. The binding of a single molecule of an inhibitory ligand *(yellow)* to one subunit of the enzyme occurs with difficulty because it changes the conformation of this subunit and thereby disrupts the symmetry of the enzyme. Once this conformational change has occurred, however, the energy gained by restoring the symmetric pairing interaction between the two subunits makes it especially easy for the second subunit to bind the inhibitory ligand and undergo the same conformational change. Because the binding of the first molecule of ligand increases the affinity with which the other subunit binds the same ligand, the response of the enzyme to changes in the concentration of the ligand is much steeper than the response of an enzyme with only one subunit (see Figure 3–60).

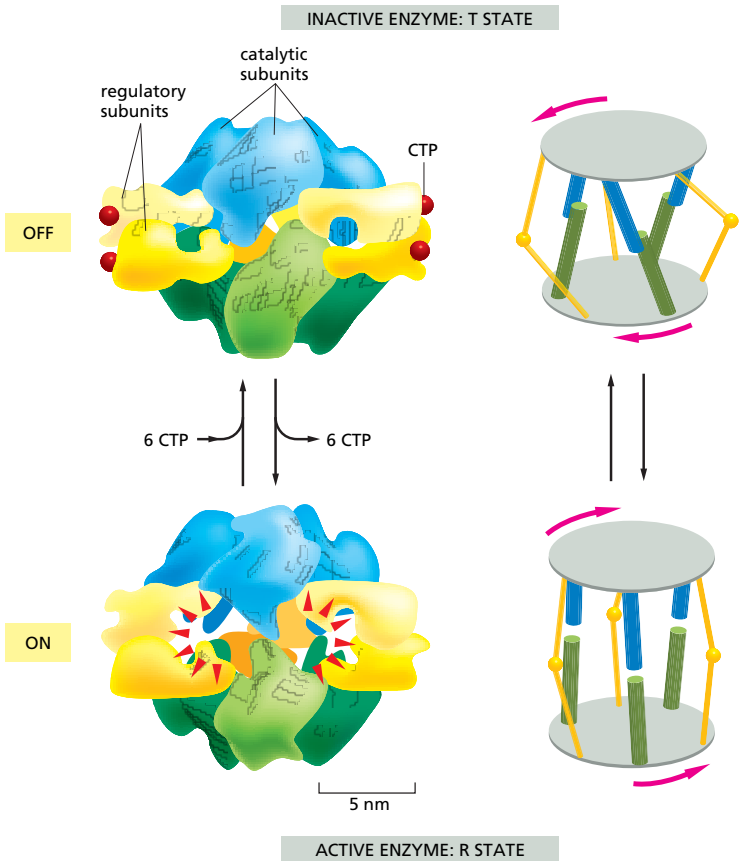


Figure 3–62 The transition between R and T states in the enzyme aspartate transcarbamoylase. <CTAA> The enzyme consists of a complex of six catalytic subunits and six regulatory subunits, and the structures of its inactive (T state) and active (R state) forms have been determined by x-ray crystallography. The enzyme is turned off by feedback inhibition when CTP concentrations rise. Each regulatory subunit can bind one molecule of CTP, which is one of the final products in the pathway. By means of this negative feedback regulation, the pathway is prevented from producing more CTP than the cell needs. (Based on K.L. Krause, K.W. Volz and W.N. Lipscomb, *Proc. Natl Acad. Sci. U.S.A.* 82:1643–1647, 1985. With permission from National Academy of Sciences.)

together by three regulatory dimers that form a bridge between them. The entire molecule is poised to undergo a concerted, all-or-none, allosteric transition between two conformations, designated as T (tense) and R (relaxed) states (**Figure 3–62**).

The binding of substrates (carbamoyl phosphate and aspartate) to the catalytic trimers drives aspartate transcarbamoylase into its catalytically active R state, from which the regulatory CTP molecules dissociate. By contrast, the binding of CTP to the regulatory dimers converts the enzyme to the inactive T state, from which the substrates dissociate. This tug-of-war between CTP and substrates is identical in principle to that described previously in Figure 3–59 for a simpler allosteric protein. But because the tug-of-war occurs in a symmetric molecule with multiple binding sites, the enzyme undergoes a cooperative allosteric transition that will turn it on suddenly as substrates accumulate (forming the R state) or shut it off rapidly when CTP accumulates (forming the T state).

A combination of biochemistry and x-ray crystallography has revealed many fascinating details of this allosteric transition. Each regulatory subunit has two domains, and the binding of CTP causes the two domains to move relative to each other, so that they function like a lever that rotates the two catalytic trimers and pulls them closer together into the T state (see Figure 3–62). When this occurs, hydrogen bonds form between opposing catalytic subunits. This helps widen the cleft that forms the active site within each catalytic subunit, thereby disrupting the binding sites for the substrates (**Figure 3–63**). Adding large amounts of substrate has the opposite effect, favoring the R state by binding in the cleft of each catalytic subunit and opposing the above conformational change. Conformations that are intermediate between R and T are unstable, so that the enzyme mostly clicks back and forth between its R and T forms, producing a mixture of these two species in proportions that depend on the relative concentrations of CTP and substrates.

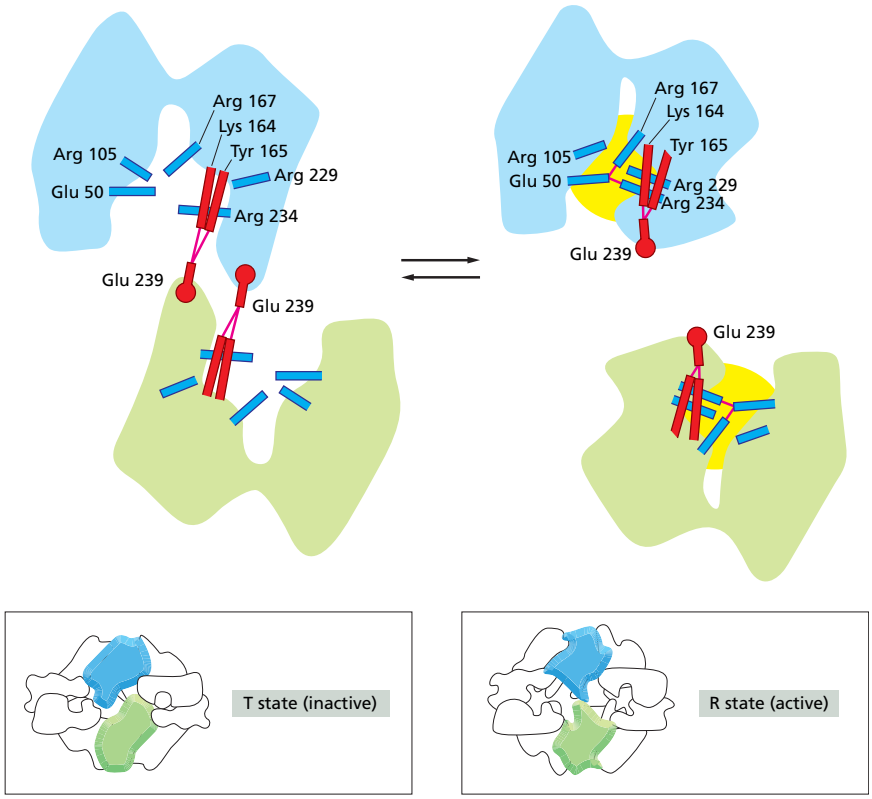


Figure 3–63 Part of the on–off switch in the catalytic subunits of aspartate transcarbamoylase. Changes in the indicated hydrogen-bonding interactions are partly responsible for switching this enzyme's active site between active *(yellow)* and inactive conformations. Hydrogen bonds are indicated by *thin red lines*. The amino acids involved in the subunit–subunit interaction in the T state are shown in *red*, while those that form the active site of the enzyme in the R state are shown in *blue*. The large drawings show the catalytic site in the interior of the enzyme; the boxed sketches show the same subunits viewed from the enzyme's external surface. (Adapted from E.R. Kantrowitz and W.N. Lipscomb, *Trends Biochem. Sci.* 15:53–59, 1990. With permission from Elsevier.)

Many Changes in Proteins Are Driven by Protein Phosphorylation

Proteins are regulated by more than the reversible binding of other molecules. A second method that eucaryotic cells use to regulate a protein's function is the covalent addition of a smaller molecule to one or more of its amino acid side chains. The most common such regulatory modification in higher eucaryotes is the addition of a phosphate group. We shall therefore use protein phosphorylation to illustrate some of the general principles involved in the control of protein function through the modification of amino acid side chains.

A phosphorylation event can affect the protein that is modified in two important ways. First, because each phosphate group carries two negative charges, the enzyme-catalyzed addition of a phosphate group to a protein can cause a major conformational change in the protein by, for example, attracting a cluster of positively charged amino acid side chains. This can, in turn, affect the binding of ligands elsewhere on the protein surface, dramatically changing the protein's activity. When a second enzyme removes the phosphate group, the protein returns to its original conformation and restores its initial activity.

Second, an attached phosphate group can form part of a structure that the binding sites of other proteins recognize. As previously discussed, certain protein domains, sometimes referred to as modules, appear very frequently as parts of larger proteins. One such module is the SH2 domain, described earlier, which binds to a short peptide sequence containing a phosphorylated tyrosine side chain (see Figure 3–39B). More than ten other common domains provide binding sites for attaching their protein to phosphorylated peptides in other protein molecules, each recognizing a phosphorylated amino acid side chain in a different protein context. As a result, protein phosphorylation and dephosphorylation very often drive the regulated assembly and disassembly of protein complexes (see Figure 15–22).

Reversible protein phosphorylation controls the activity, structure, and cellular localization of both enzymes and many other types of proteins in

eucaryotic cells. In fact, this regulation is so extensive that more than one-third of the 10,000 or so proteins in a typical mammalian cell are thought to be phosphorylated at any given time—many with more than one phosphate. As might be expected, the addition and removal of phosphate groups from specific proteins often occur in response to signals that specify some change in a cell's state. For example, the complicated series of events that takes place as a eucaryotic cell divides is largely timed in this way (discussed in Chapter 17), and many of the signals mediating cell–cell interactions are relayed from the plasma membrane to the nucleus by a cascade of protein phosphorylation events (discussed in Chapter 15).

A Eucaryotic Cell Contains a Large Collection of Protein Kinases and Protein Phosphatases

Protein phosphorylation involves the enzyme-catalyzed transfer of the terminal phosphate group of an ATP molecule to the hydroxyl group on a serine, threonine, or tyrosine side chain of the protein (**Figure 3–64**). A **protein kinase** catalyzes this reaction, and the reaction is essentially unidirectional because of the large amount of free energy released when the phosphate–phosphate bond in ATP is broken to produce ADP (discussed in Chapter 2). A **protein phosphatase** catalyzes the reverse reaction of phosphate removal, or *dephosphorylation*. Cells contain hundreds of different protein kinases, each responsible for phosphorylating a different protein or set of proteins. There are also many different protein phosphatases; some are highly specific and remove phosphate groups from only one or a few proteins, whereas others act on a broad range of proteins and are targeted to specific substrates by regulatory subunits. The state of phosphorylation of a protein at any moment, and thus its activity, depends on the relative activities of the protein kinases and phosphatases that modify it.

The protein kinases that phosphorylate proteins in eucaryotic cells belong to a very large family of enzymes, which share a catalytic (kinase) sequence of about 290 amino acids. The various family members contain different amino acid sequences on either end of the kinase sequence (for example, see Figure 3–10), and often have short amino acid sequences inserted into loops within it *(red arrowheads* in **Figure 3–65**). Some of these additional amino acid sequences enable each kinase to recognize the specific set of proteins it phosphorylates, or to bind to structures that localize it in specific regions of the cell. Other parts of the protein regulate the activity of each kinase, so it can be turned on and off in response to different specific signals, as described below.

By comparing the number of amino acid sequence differences between the various members of a protein family, we can construct an "evolutionary tree" that is thought to reflect the pattern of gene duplication and divergence that gave rise to the family. **Figure 3–66** shows an evolutionary tree of protein kinases. Kinases with related functions are often located on nearby branches of the tree: the protein kinases involved in cell signaling that phosphorylate tyrosine side chains, for example, are all clustered in the top left corner of the tree. The other kinases shown phosphorylate either a serine or a threonine side chain, and many are organized into clusters that seem to reflect their function—in transmembrane signal transduction, intracellular signal amplification, cell-cycle control, and so on.

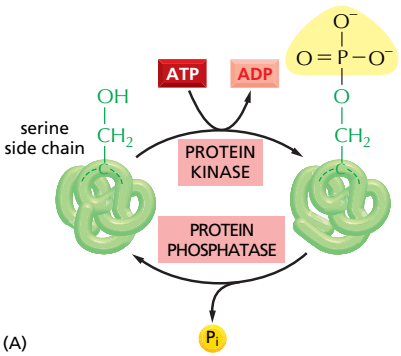


Figure 3–64 Protein phosphorylation. Many thousands of proteins in a typical eucaryotic cell are modified by the covalent addition of a phosphate group. (A) The general reaction, shown here, transfers a phosphate group from ATP to an amino acid side chain of the target protein by a protein kinase. Removal of the phosphate group is catalyzed by a second enzyme, a protein phosphatase. In this example, the phosphate is added to a serine side chain; in other cases, the phosphate is instead linked to the –OH group of a threonine or a tyrosine in the protein. (B) The phosphorylation of a protein by a protein kinase can either increase or decrease the protein's activity, depending on the site of phosphorylation and the structure of the protein.

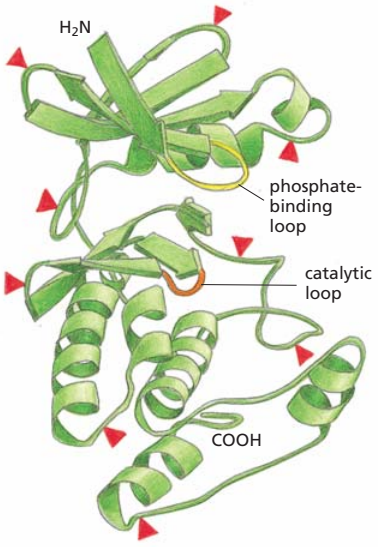


Figure 3–65 The three-dimensional structure of a protein kinase. Superimposed on this structure are *red arrowheads* to indicate sites where insertions of 5–100 amino acids are found in some members of the protein kinase family. These insertions are located in loops on the surface of the enzyme where other ligands interact with the protein. Thus, they distinguish different kinases and confer on them distinctive interactions with other proteins. The ATP (which donates a phosphate group) and the peptide to be phosphorylated are held in the active site, which extends between the phosphate-binding loop *(yellow)* and the catalytic loop *(orange)*. See also Figure 3–10. (Adapted from D.R. Knighton et al., *Science* 253:407–414, 1991. With permission from AAAS.)

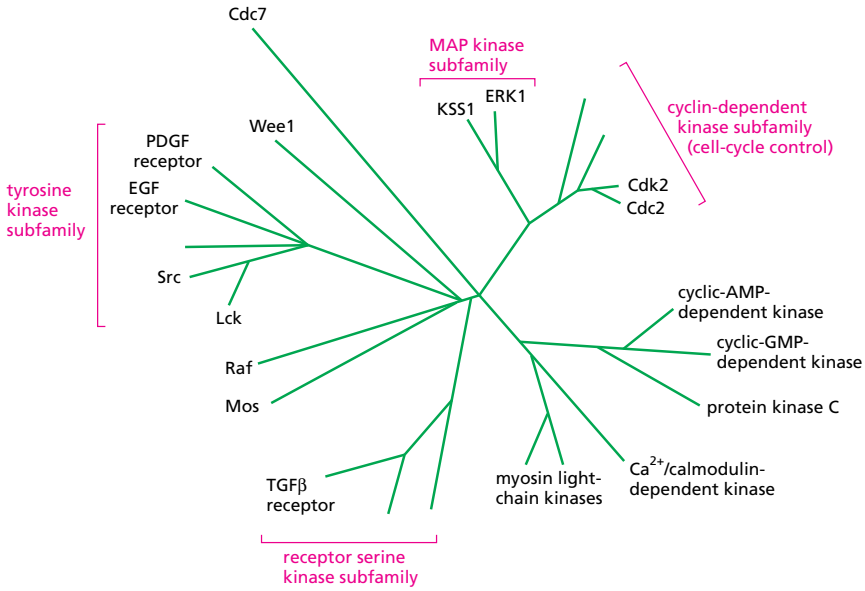


Figure 3–66 An evolutionary tree of selected protein kinases. Although a higher eucaryotic cell contains hundreds of such enzymes, and the human genome codes for more than 500, only some of those discussed in this book are shown.

As a result of the combined activities of protein kinases and protein phosphatases, the phosphate groups on proteins are continually turning over—being added and then rapidly removed. Such phosphorylation cycles may seem wasteful, but they are important in allowing the phosphorylated proteins to switch rapidly from one state to another: the more rapid the cycle, the faster a population of protein molecules can change its state of phosphorylation in response to a sudden change in the phosphorylation rate (see Figure 15–11). The energy required to drive this phosphorylation cycle is derived from the free energy of ATP hydrolysis, one molecule of which is consumed for each phosphorylation event.

The Regulation of Cdk and Src Protein Kinases Shows How a Protein Can Function as a Microchip

The hundreds of different protein kinases in a eucaryotic cell are organized into complex networks of signaling pathways that help to coordinate the cell's activities, drive the cell cycle, and relay signals into the cell from the cell's environment. Many of the extracellular signals involved need to be both integrated and amplified by the cell. Individual protein kinases (and other signaling proteins) serve as input–output devices, or "microchips," in the integration process. An important part of the input to these signal processing proteins comes from the control that is exerted by phosphates added and removed from them by protein kinases and protein phosphatases, respectively.

In general, specific sets of phosphate groups serve to activate the protein, while other sets can inactivate it. A cyclin-dependent protein kinase (Cdk) provides a good example. Kinases in this class phosphorylate serines and threonines, and they are central components of the cell-cycle control system in eucaryotic cells, as discussed in detail in Chapter 17. In a vertebrate cell, individual Cdk proteins turn on and off in succession, as a cell proceeds through the different phases of its division cycle. When a particular kinase is on, it influences various aspects of cell behavior through effects on the proteins it phosphorylates.

A Cdk protein becomes active as a serine/threonine protein kinase only when it is bound to a second protein called a *cyclin*. But, as **Figure 3–67** shows, the binding of cyclin is only one of three distinct "inputs" required to activate the Cdk. In addition to cyclin binding, a phosphate must be added to a specific threonine side chain, and a phosphate elsewhere in the protein (covalently bound to a specific tyrosine side chain) must be removed. Cdk thus monitors a specific set

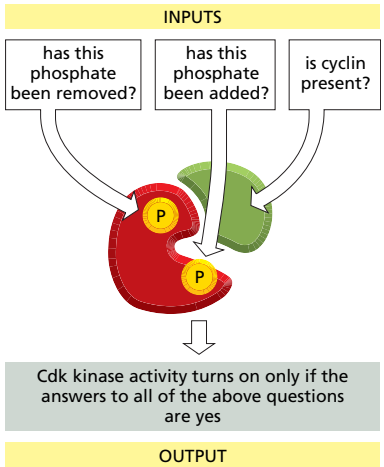


Figure 3–67 How a Cdk protein acts as an integrating device. <TAGA> The function of these central regulators of the cell cycle is discussed in Chapter 17.

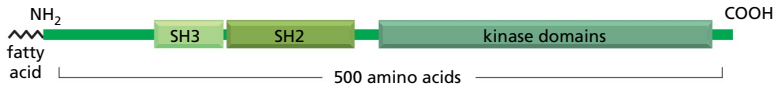


Figure 3–68 The domain structure of the Src family of protein kinases, mapped along the amino acid sequence. For the three-dimensional structure of Src, see Figure 3–10.

of cell components—a cyclin, a protein kinase, and a protein phosphatase—and it acts as an input–output device that turns on if, and only if, each of these components has attained its appropriate activity state. Some cyclins rise and fall in concentration in step with the cell cycle, increasing gradually in amount until they are suddenly destroyed at a particular point in the cycle. The sudden destruction of a cyclin (by targeted proteolysis) immediately shuts off its partner Cdk enzyme, and this triggers a specific step in the cell cycle.

The Src family of protein kinases (see Figure 3–10) exhibits a similar type of microchip behavior. The *Src protein* (pronounced "sarc" and named for the type of tumor, a sarcoma, that its deregulation can cause) was the first tyrosine kinase to be discovered. It is now known to be part of a subfamily of nine very similar protein kinases, which are found only in multicellular animals. As indicated by the evolutionary tree in Figure 3–66, sequence comparisons suggest that tyrosine kinases as a group were a relatively late innovation that branched off from the serine/threonine kinases, with the Src subfamily being only one subgroup of the tyrosine kinases created in this way.

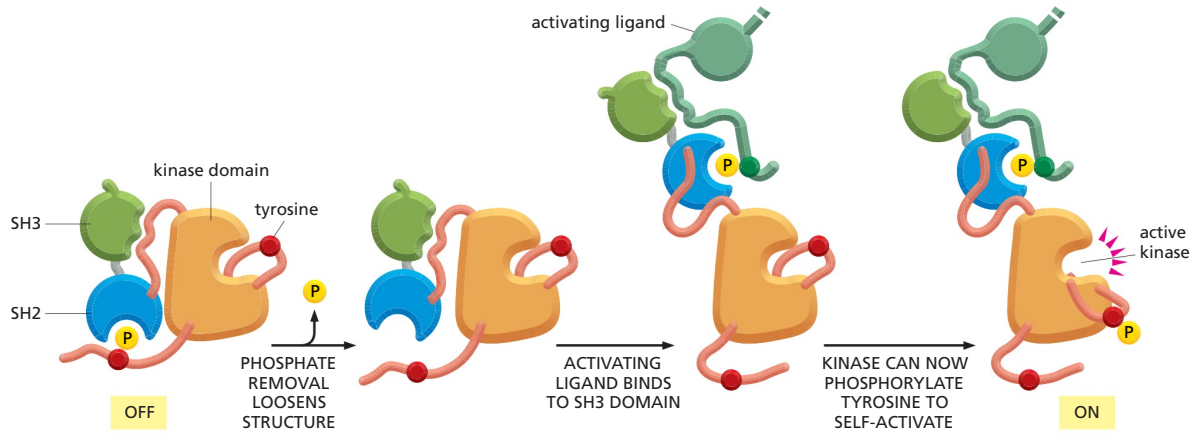
The Src protein and its relatives contain a short N-terminal region that becomes covalently linked to a strongly hydrophobic fatty acid, which holds the kinase at the cytoplasmic face of the plasma membrane. Next come two peptide-binding modules, a Src homology 3 (SH3) domain and a SH2 domain, followed by the kinase catalytic domain (**Figure 3–68**). These kinases normally exist in an inactive conformation, in which a phosphorylated tyrosine near the C-terminus is bound to the SH2 domain, and the SH3 domain is bound to an internal peptide in a way that distorts the active site of the enzyme and helps to render it inactive.

Turning the kinase on involves at least two specific inputs: removal of the C-terminal phosphate and the binding of the SH3 domain by a specific activating protein (**Figure 3–69**). Like the activation of the Cdk protein, the activation of the Src kinase signals the completion of a particular set of separate upstream events (**Figure 3–70**). Thus, both the Cdk and Src families of proteins serve as specific signal integrators, helping to generate the complex web of information-processing events that enable the cell to compute logical responses to a complex set of conditions.

Proteins That Bind and Hydrolyze GTP Are Ubiquitous Cellular Regulators

We have described how the addition or removal of phosphate groups on a protein can be used by a cell to control the protein's activity. In the examples discussed so

Figure 3–69 The activation of a Src-type protein kinase by two sequential events. (Adapted from S.C. Harrison et al., *Cell* 112:737–740, 2003. With permission from Elsevier.)



far, the phosphate is transferred from an ATP molecule to an amino acid side chain of the protein in a reaction catalyzed by a specific protein kinase. Eucaryotic cells also have another way to control protein activity by phosphate addition and removal. In this case, the phosphate is not attached directly to the protein; instead, it is a part of the guanine nucleotide GTP, which binds very tightly to the protein. In general, proteins regulated in this way are in their active conformations with GTP bound. The loss of a phosphate group occurs when the bound GTP is hydrolyzed to GDP in a reaction catalyzed by the protein itself, and in its GDP-bound state the protein is inactive. In this way, GTP-binding proteins act as on–off switches whose activity is determined by the presence or absence of an additional phosphate on a bound GDP molecule (**Figure 3–71**).

GTP-binding proteins (also called **GTPases** because of the GTP hydrolysis they catalyze) comprise a large family of proteins that all contain variations on the same GTP-binding globular domain. When the tightly bound GTP is hydrolyzed to GDP, this domain undergoes a conformational change that inactivates it. The three-dimensional structure of a prototypical member of this family, the monomeric GTPase called Ras, is shown in **Figure 3–72**.

The *Ras protein* has an important role in cell signaling (discussed in Chapter 15). In its GTP-bound form, it is active and stimulates a cascade of protein phosphorylations in the cell. Most of the time, however, the protein is in its inactive, GDP-bound form. It becomes active when it exchanges its GDP for a GTP molecule in response to extracellular signals, such as growth factors, that bind to receptors in the plasma membrane (see Figure 15–58).

Regulatory Proteins Control the Activity of GTP-Binding Proteins by Determining Whether GTP or GDP Is Bound

GTP-binding proteins are controlled by regulatory proteins that determine whether GTP or GDP is bound, just as phosphorylated proteins are turned on and off by protein kinases and protein phosphatases. Thus, Ras is inactivated by a *GTPase-activating protein (GAP)*, which binds to the Ras protein and induces it to hydrolyze its bound GTP molecule to GDP—which remains tightly bound—and inorganic phosphate (P_i), which is rapidly released. The Ras protein stays in its inactive, GDP-bound conformation until it encounters a *guanine nucleotide exchange factor (GEF)*, which binds to GDP-Ras and causes it to release its GDP. Because the empty nucleotide-binding site is immediately filled by a GTP molecule (GTP is present in large excess over GDP in cells), the GEF activates Ras by *indirectly* adding back the phosphate removed by GTP hydrolysis. Thus, in a sense, the roles of GAP and GEF are analogous to those of a protein phosphatase and a protein kinase, respectively (**Figure 3–73**).

Large Protein Movements Can Be Generated From Small Ones

The Ras protein belongs to a large superfamily of *monomeric GTPases*, each of which consists of a single GTP-binding domain of about 200 amino acids. Over the course of evolution, this domain has also become joined to larger proteins with additional domains, creating a large family of GTP-binding proteins. Family members include the receptor-associated trimeric G proteins involved in cell signaling (discussed in Chapter 15), proteins regulating the traffic of vesicles between intracellular compartments (discussed in Chapter 13), and proteins that bind to transfer RNA and are required as assembly factors for protein

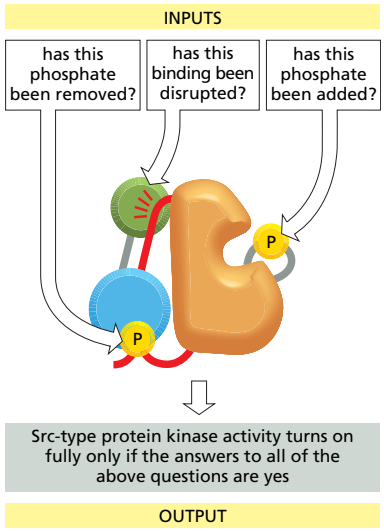


Figure 3–70 How a Src-type protein kinase acts as an integrating device. The disruption of the SH3 domain interaction *(green)* involves replacing its binding to the indicated *red* linker region with a tighter interaction with an activating ligand, as illustrated in Figure 3–69.

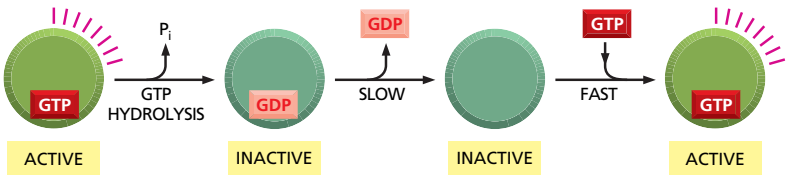


Figure 3–71 GTP-binding proteins as molecular switches. The activity of a GTP-binding protein (also called a GTPase) generally requires the presence of a tightly bound GTP molecule (switch "on"). Hydrolysis of this GTP molecule produces GDP and inorganic phosphate (P_i), and it causes the protein to convert to a different, usually inactive, conformation (switch "off"). As shown here, resetting the switch requires the tightly bound GDP to dissociate, a slow step that is greatly accelerated by specific signals; once the GDP has dissociated, a molecule of GTP is quickly rebound.

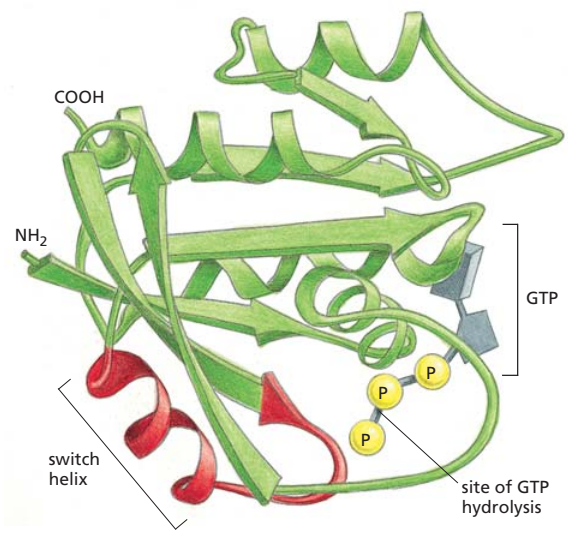


Figure 3–72 The structure of the Ras protein in its GTP-bound form. <GAAC> This monomeric GTPase illustrates the structure of a GTP-binding domain, which is present in a large family of GTP-binding proteins. The *red* regions change their conformation when the GTP molecule is hydrolyzed to GDP and inorganic phosphate by the protein; the GDP remains bound to the protein, while the inorganic phosphate is released. The special role of the "switch helix" in proteins related to Ras is explained next (see Figure 3–75).

synthesis on the ribosome (discussed in Chapter 6). In each case, an important biological activity is controlled by a change in the protein's conformation that is caused by GTP hydrolysis in a Ras-like domain.

The *EF-Tu protein* provides a good example of how this family of proteins works. EF-Tu is an abundant molecule that serves as an elongation factor (hence the EF) in protein synthesis, loading each aminoacyl tRNA molecule onto the ribosome. The tRNA molecule forms a tight complex with the GTP-bound form of EF-Tu (**Figure 3–74**). In this complex, the amino acid attached to the tRNA is improperly positioned for protein synthesis. The tRNA can transfer its amino acid only after the GTP bound to EF-Tu is hydrolyzed on the ribosome, allowing the EF-Tu to dissociate. Since the GTP hydrolysis is triggered by a proper fit of the tRNA to the mRNA molecule on the ribosome, the EF-Tu serves as a factor that discriminates between correct and incorrect mRNA–tRNA pairings (see Figure 6–67 for a further discussion of this function of EF-Tu).

By comparing the three-dimensional structure of EF-Tu in its GTP-bound and GDP-bound forms, we can see how the repositioning of the tRNA occurs. The dissociation of the inorganic phosphate group (P_i), which follows the reaction GTP → GDP + P_i, causes a shift of a few tenths of a nanometer at the GTP-binding site, just as it does in the Ras protein. This tiny movement, equivalent to

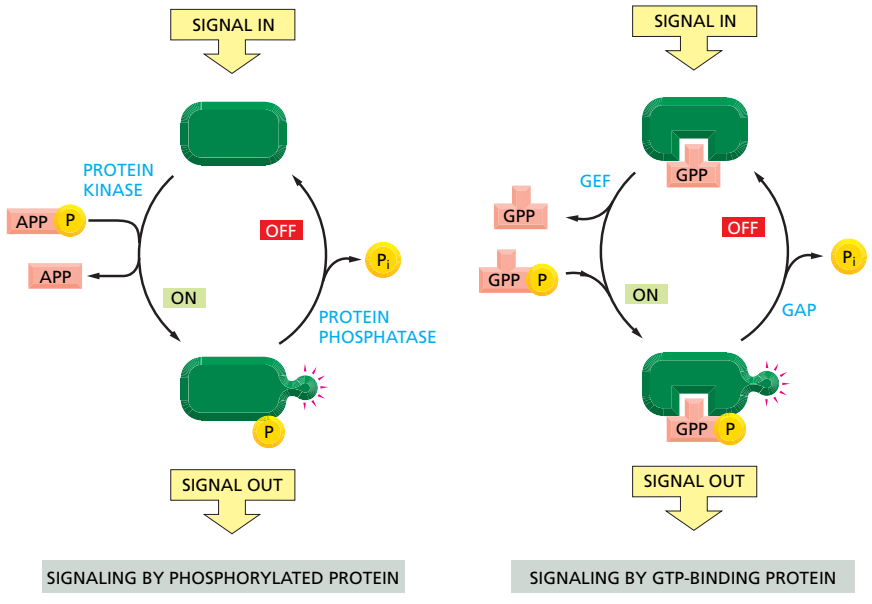


Figure 3–73 A comparison of the two major intracellular signaling mechanisms in eucaryotic cells. In both cases, a signaling protein is activated by the addition of a phosphate group and inactivated by the removal of this phosphate. To emphasize the similarities in the two pathways, ATP and GTP are drawn as APPP and GPPP, and ADP and GDP as APP and GPP, respectively. As shown in Figure 3–64, the addition of a phosphate to a protein can also be inhibitory.

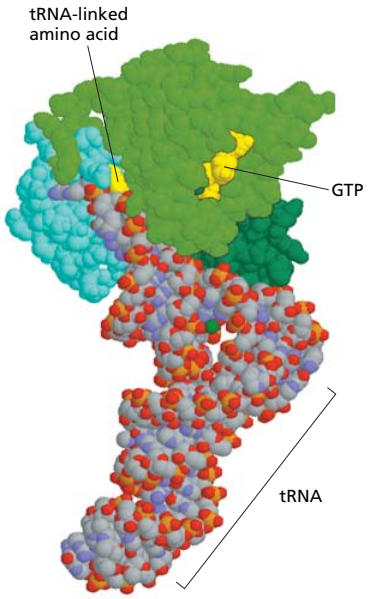


Figure 3–74 An aminoacyl tRNA molecule bound to EF-Tu. The three domains of the EF-Tu protein are colored differently, to match Figure 3–75. This is a bacterial protein; however, a very similar protein exists in eucaryotes, where it is called EF-1. (Coordinates determined by P. Nissen et al., *Science* 270:1464–1472, 1995. With permission from AAAS.)

a few times the diameter of a hydrogen atom, causes a conformational change to propagate along a crucial piece of α helix, called the *switch helix*, in the Ras-like domain of the protein. The switch helix seems to serve as a latch that adheres to a specific site in another domain of the molecule, holding the protein in a "shut" conformation. The conformational change triggered by GTP hydrolysis causes the switch helix to detach, allowing separate domains of the protein to swing apart, through a distance of about 4 nm. This releases the bound tRNA molecule, allowing its attached amino acid to be used (**Figure 3–75**).

Notice in this example how cells have exploited a simple chemical change that occurs on the surface of a small protein domain to create a movement 50 times larger. Dramatic shape changes of this type also cause the very large movements that occur in motor proteins, as we discuss next.

Motor Proteins Produce Large Movements in Cells

We have seen that conformational changes in proteins have a central role in enzyme regulation and cell signaling. We now discuss proteins whose major function is to move other molecules. These **motor proteins** generate the forces responsible for muscle contraction and the crawling and swimming of cells. Motor proteins also power smaller-scale intracellular movements: they help to move chromosomes to opposite ends of the cell during mitosis (discussed in Chapter 17), to move organelles along molecular tracks within the cell (discussed

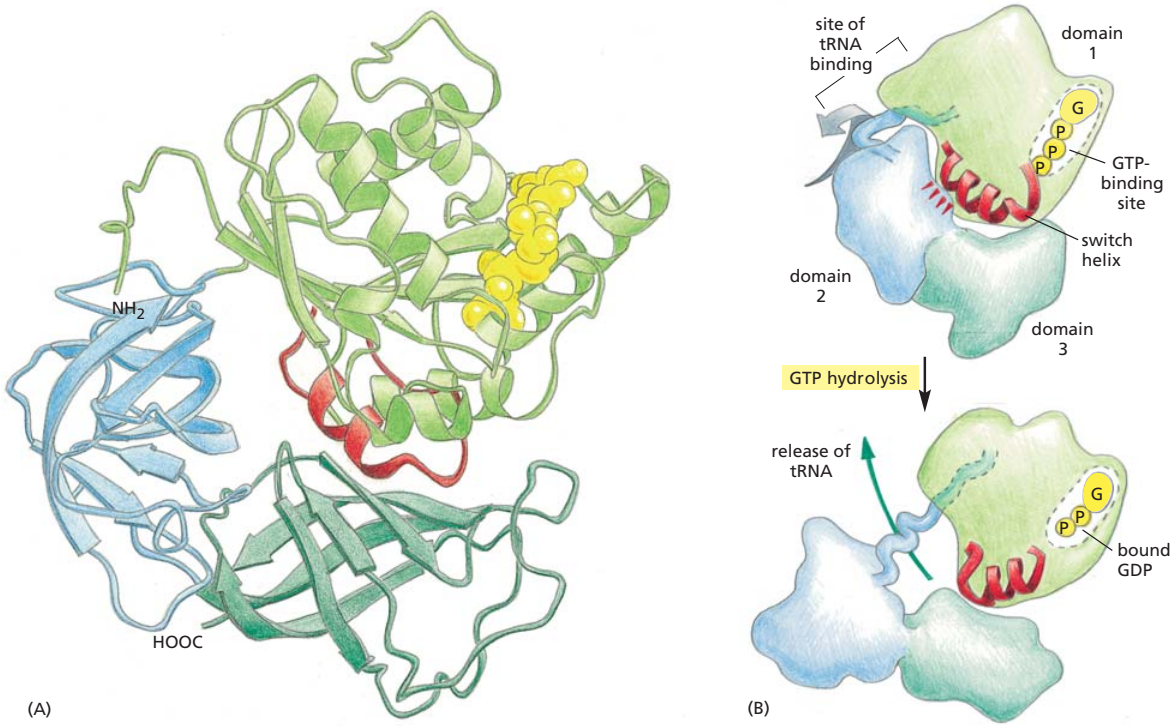


Figure 3–75 The large conformational change in EF-Tu caused by GTP hydrolysis. <GTAA> (A) The three-dimensional structure of EF-Tu with GTP bound. The domain at the top has a structure similar to the Ras protein, and its *red* α helix is the switch helix, which moves after GTP hydrolysis. (B) The change in the conformation of the switch helix in domain 1 causes domains 2 and 3 to rotate as a single unit by about 90° toward the viewer, which releases the tRNA that was shown bound to this structure in Figure 3–74. (A, adapted from H. Berchtold et al., *Nature* 365:126–132, 1993. With permission from Macmillan Publishers Ltd. B, courtesy of Mathias Sprinzl and Rolf Hilgenfeld.)

in Chapter 16), and to move enzymes along a DNA strand during the synthesis of a new DNA molecule (discussed in Chapter 5). All these fundamental processes depend on proteins with moving parts that operate as force-generating machines.

How do these machines work? In other words, how do cells use shape changes in proteins to generate directed movements? If, for example, a protein is required to walk along a narrow thread such as a DNA molecule, it can do this by undergoing a series of conformational changes, such as those shown in **Figure 3–76**. But with nothing to drive these changes in an orderly sequence, they are perfectly reversible, and the protein can only wander randomly back and forth along the thread. We can look at this situation in another way. Since the directional movement of a protein does work, the laws of thermodynamics (discussed in Chapter 2) demand that such movement use free energy from some other source (otherwise the protein could be used to make a perpetual motion machine). Therefore, without an input of energy, the protein molecule can only wander aimlessly.

How can the cell make such a series of conformational changes unidirectional? To force the entire cycle to proceed in one direction, it is enough to make any one of the changes in shape irreversible. Most proteins that are able to walk in one direction for long distances achieve this motion by coupling one of the conformational changes to the hydrolysis of an ATP molecule bound to the protein. The mechanism is similar to the one just discussed that drives allosteric protein shape changes by GTP hydrolysis. Because ATP (or GTP) hydrolysis releases a great deal of free energy, it is very unlikely that the nucleotide-binding protein will undergo the reverse shape change needed for moving backward—since this would require that it also reverse the ATP hydrolysis by adding a phosphate molecule to ADP to form ATP.

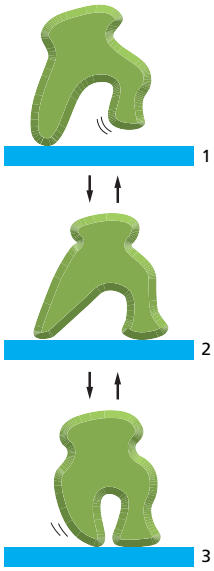


Figure 3–76 An allosteric "walking" protein. Although its three different conformations allow it to wander randomly back and forth while bound to a thread or a filament, the protein cannot move uniformly in a single direction.

In the model shown in **Figure 3–77**, ATP binding shifts a motor protein from conformation 1 to conformation 2. The bound ATP is then hydrolyzed to produce ADP and inorganic phosphate (P_i), causing a change from conformation 2 to conformation 3. Finally, the release of the bound ADP and P_i drives the protein back to conformation 1. Because the energy provided by ATP hydrolysis drives the transition 2 → 3, this series of conformational changes is effectively irreversible. Thus, the entire cycle goes in only one direction, causing the protein molecule to walk continuously to the right in this example.

Many motor proteins generate directional movement in this general way, including the muscle motor protein *myosin*, which walks along actin filaments to generate muscle contraction, and the *kinesin* proteins that walk along microtubules (both discussed in Chapter 16). These movements can be rapid: some of the motor proteins involved in DNA replication (the DNA helicases) propel themselves along a DNA strand at rates as high as 1000 nucleotides per second.

Membrane-Bound Transporters Harness Energy to Pump Molecules Through Membranes

We have thus far seen how allosteric proteins can act as microchips (Cdk and Src kinases), as assembly factors (EF-Tu), and as generators of mechanical force and motion (motor proteins). Allosteric proteins can also harness energy derived from ATP hydrolysis, ion gradients, or electron transport processes to pump specific ions or small molecules across a membrane. We consider one example here; others will be discussed in Chapter 11.

The ABC transporters constitute an important class of membrane-bound pump proteins. In humans at least 48 different genes encode them. These transporters mostly function to export hydrophobic molecules from the cytoplasm,

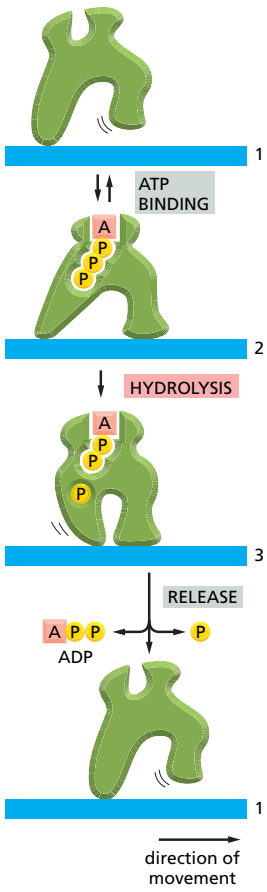


Figure 3–77 An allosteric motor protein. The transition between three different conformations includes a step driven by the hydrolysis of a bound ATP molecule, and this makes the entire cycle essentially irreversible. By repeated cycles, the protein therefore moves continuously to the right along the thread.

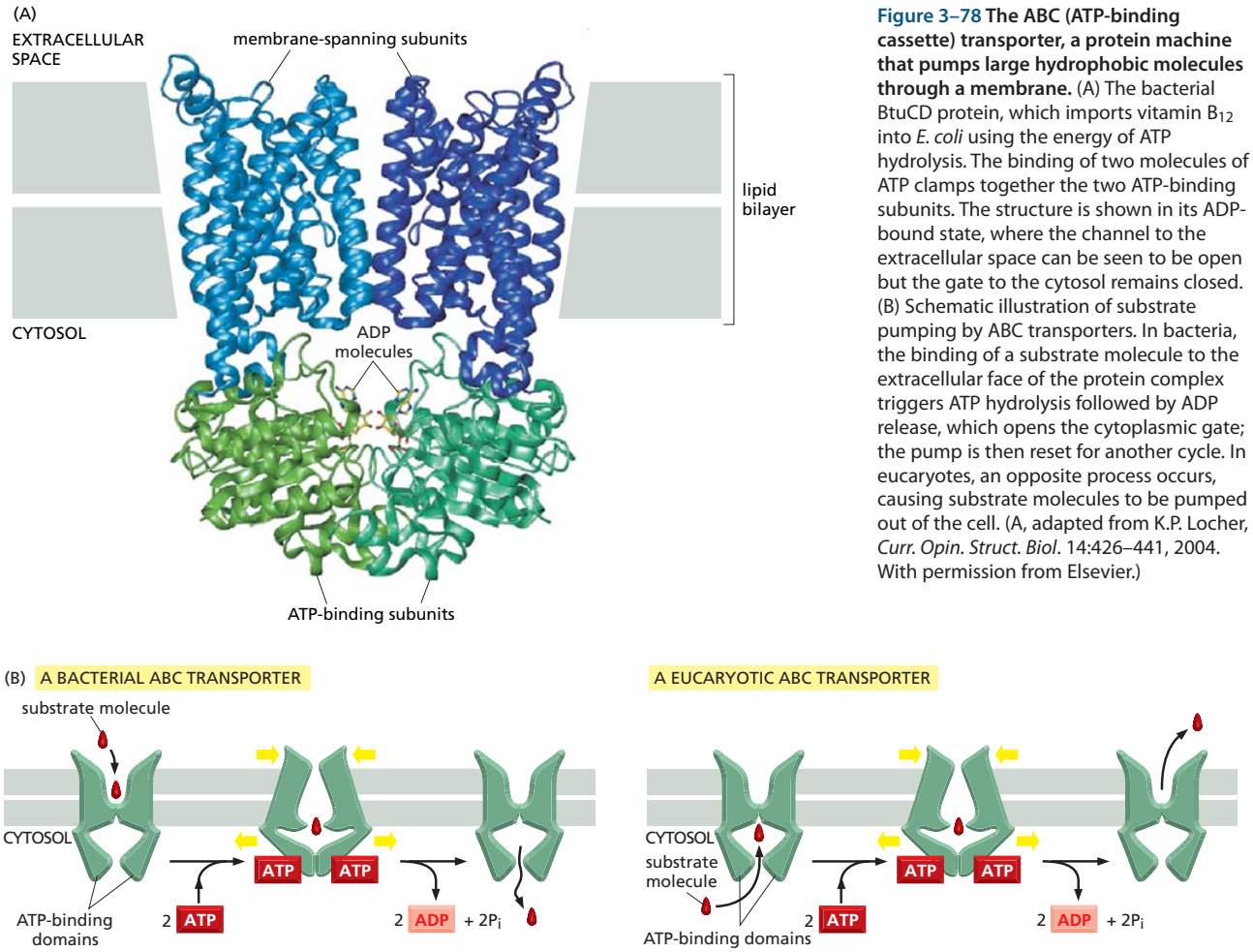


Figure 3–78 The ABC (ATP-binding cassette) transporter, a protein machine that pumps large hydrophobic molecules through a membrane. (A) The bacterial BtuCD protein, which imports vitamin B_{12} into *E. coli* using the energy of ATP hydrolysis. The binding of two molecules of ATP clamps together the two ATP-binding subunits. The structure is shown in its ADP-bound state, where the channel to the extracellular space can be seen to be open but the gate to the cytosol remains closed. (B) Schematic illustration of substrate pumping by ABC transporters. In bacteria, the binding of a substrate molecule to the extracellular face of the protein complex triggers ATP hydrolysis followed by ADP release, which opens the cytoplasmic gate; the pump is then reset for another cycle. In eucaryotes, an opposite process occurs, causing substrate molecules to be pumped out of the cell. (A, adapted from K.P. Locher, *Curr. Opin. Struct. Biol.* 14:426–441, 2004. With permission from Elsevier.)

serving to remove toxic molecules at the mucosal surface of the intestinal tract, for example, or at the blood–brain barrier. The study of ABC transporters is of intense interest in clinical medicine, because the overproduction of proteins in this class contributes to the resistance of tumor cells to chemotherapeutic drugs. And in bacteria, the same type of proteins primarily function to import essential nutrients into the cell.

The ABC transporter is a tetramer, with a pair of membrane-spanning subunits linked to a pair of ATP binding subunits located just below the plasma membrane (**Figure 3–78**A). As in other examples we have discussed, the hydrolysis of the bound ATP molecules drives conformational changes in the protein, transmitting forces that cause the membrane-spanning subunits to move their bound molecules across the lipid bilayer (Figure 3–78B).

Humans have invented many different types of mechanical pumps, and it should not be surprising that cells also contain membrane-bound pumps that function in other ways. Among the most notable are the rotary pumps that couple the hydrolysis of ATP to the transport of H^+ ions (protons). These pumps resemble miniature turbines, and they are used to acidify the interior of lysosomes and other eucaryotic organelles. Like other ion pumps that create ion gradients, they can function in reverse to catalyze the reaction ADP + P_i → ATP, if the gradient across their membrane of the ion that they transport is steep enough.

One such pump, the ATP synthase, harnesses a gradient of proton concentration produced by electron transport processes to produce most of the ATP used in the living world. This ubiquitous pump has a central role in energy conversion, and we shall discuss its three-dimensional structure and mechanism in Chapter 14.

Proteins Often Form Large Complexes That Function as Protein Machines <ACTT> <ATCG>

Large proteins formed from many domains are able to perform more elaborate functions than small, single-domain proteins. But large protein assemblies formed from many protein molecules perform the most impressive tasks. Now that it is possible to reconstruct most biological processes in cell-free systems in the laboratory, it is clear that each of the central processes in a cell—such as DNA replication, protein synthesis, vesicle budding, or transmembrane signaling—is catalyzed by a highly coordinated, linked set of 10 or more proteins. In most such *protein machines*, an energetically favorable reaction such as the hydrolysis of bound nucleoside triphosphates (ATP or GTP) drives an ordered series of conformational changes in one or more of the individual protein subunits, enabling the ensemble of proteins to move coordinately. In this way, each enzyme can be moved directly into position, as the machine catalyzes successive reactions in a series. This is what occurs, for example, in protein synthesis on a ribosome (discussed in Chapter 6)—or in DNA replication, where a large multiprotein complex moves rapidly along the DNA (discussed in Chapter 5).

Cells have evolved protein machines for the same reason that humans have invented mechanical and electronic machines. For accomplishing almost any task, manipulations that are spatially and temporally coordinated through linked processes are much more efficient than the use of individual tools.

Protein Machines with Interchangeable Parts Make Efficient Use of Genetic Information

To probe more deeply into the nature of protein machines, we shall consider a relatively simple one: the **SCF ubiquitin ligase**. This protein complex binds different "target proteins" at different times in the cell cycle, and it covalently adds multiubiquitin polypeptide chains to these proteins. Its C-shaped structure is formed from five protein subunits, the largest of which is a molecule that serves as a *scaffold protein* on which the rest of the structure is built. The structure underlies a remarkable mechanism (**Figure 3–79**). At one end of the C is an E2 ubiquitin-conjugating enzyme. At the other end is a substrate-binding arm, a subunit known as an *F-box protein*. These two subunits are separated by a gap of about 5 nm. When this protein complex is activated, the F-box protein binds to a specific site on a target protein, positioning the protein in the gap so that some of its lysine side chains contact the ubiquitin-conjugating enzyme. This enzyme can then catalyze the repeated addition of a ubiquitin polypeptide to these lysines (see Figure 3–79C), producing a polyubiquitin chain that marks the target protein for rapid destruction in a proteasome (see p. 393).

In this manner, specific proteins are targeted for rapid destruction in response to specific signals, thereby helping to drive the cell cycle (discussed in Chapter 17). The timing of the destruction often involves creating a specific pattern of phosphorylation on the target protein that is required for its recognition by the F-box subunit. It also requires the activation of an SCF ubiquitin ligase that carries the appropriate substrate-binding arm. Many of these arms (the F-box subunits) are interchangeable in the protein complex (see Figure 3–79B), and there are more than 70 human genes that encode them.

As emphasized previously, once a successful protein has evolved, its genetic information tends to be duplicated to produce a family of related proteins. Thus, for example, not only are there many F-box proteins—making possible the recognition of different sets of target proteins—but there is also a family of scaffolds (known as cullins) that give rise to a family of SCF-like ubiquitin ligases.

The pressure on organisms to minimize the number of genes (see p. 265) presumably helps to explain why RNA splicing is so prevalent in higher eucaryotes, allowing multiple related proteins to be synthesized from a single gene (discussed in Chapter 6). A protein machine like the SCF ubiquitin ligase, with its interchangeable parts, likewise makes economical use of the genetic information

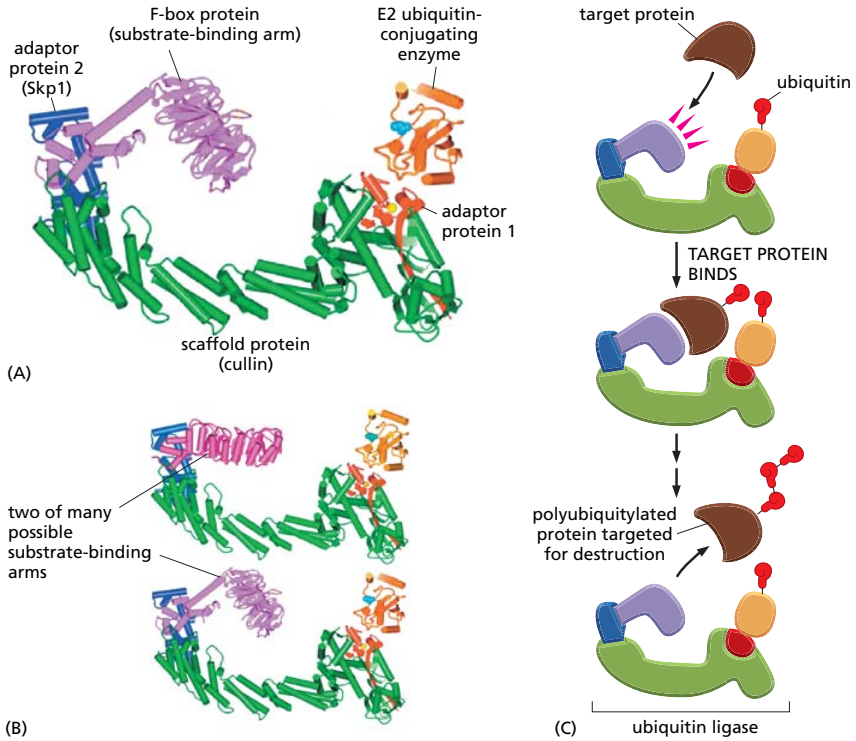


Figure 3–79 The structure and mode of action of a SCF ubiquitin ligase. (A) The structure of the five-protein complex that includes an E2 ubiquitin ligase. The protein denoted here as adapter protein 1 is the Rbx1/Hrt1 protein, adaptor protein 2 is the Skp1 protein, and the cullin is the Cul1 protein. (B) Comparison of the same complex with two different substrate-binding arms, the F-box proteins Skp2 *(top)* and β-trCP1 *(bottom),* respectively. (C) The binding and ubiquitylation of a target protein by the SCF ubiquitin ligase. If, as indicated, a chain of ubiquitin molecules is added to the same lysine of the target protein, that protein is marked for rapid destruction by the proteasome. (A and B, adapted from G. Wu et al., *Mol. Cell* 11:1445–1456, 2003. With permission from Elsevier.)

in cells, inasmuch as new functions can evolve for the entire complex simply by producing an alternative version of one of its subunits.

The Activation of Protein Machines Often Involves Positioning Them at Specific Sites

As scientists have learned more of the details of cell biology, they have recognized increasing degrees of sophistication in cell chemistry. Thus, not only do we now know that protein machines play a predominant role, but it has recently become clear that most of these machines form at specific sites in the cell, being activated only where and when they are needed. Using fluorescent, GFP-tagged fusion proteins in living cells (see p. 593), cell biologists are able to follow the repositioning of individual proteins that occurs in response to specific signals. Thus, when certain extracellular signaling molecules bind to receptor proteins in the plasma membrane, they often recruit a set of other proteins to the inside surface of the plasma membrane to form protein machines that pass the signal on. As an example, **Figure 3–80**A illustrates the rapid movement of a protein kinase C (PKC) enzyme to a complex in the plasma membrane, where it associates with specific substrate proteins that it phosphorylates.

There are more than 10 distinct PKC enzymes in human cells, which differ both in their mode of regulation and in their functions. When activated, these enzymes move from the cytoplasm to different intracellular locations, forming specific complexes with other proteins that allow them to phosphorylate different protein substrates (Figure 3–80B). The SCF ubiquitin ligases can also move to specific sites of function at appropriate times. As will be explained when we discuss cell signaling in Chapter 15, the mechanisms frequently involve protein phosphorylation, as well as **scaffold proteins** that link together a set of activating, inhibiting, adaptor, and substrate proteins at a specific location in a cell.

This general phenomenon is known as *induced proximity*, and it explains the otherwise puzzling observation that slightly different forms of enzymes with the same catalytic site will often have very different biological functions. Cells change the locations of their proteins by covalently modifying them in a variety of different ways, as part of a "regulatory code" to be described next.

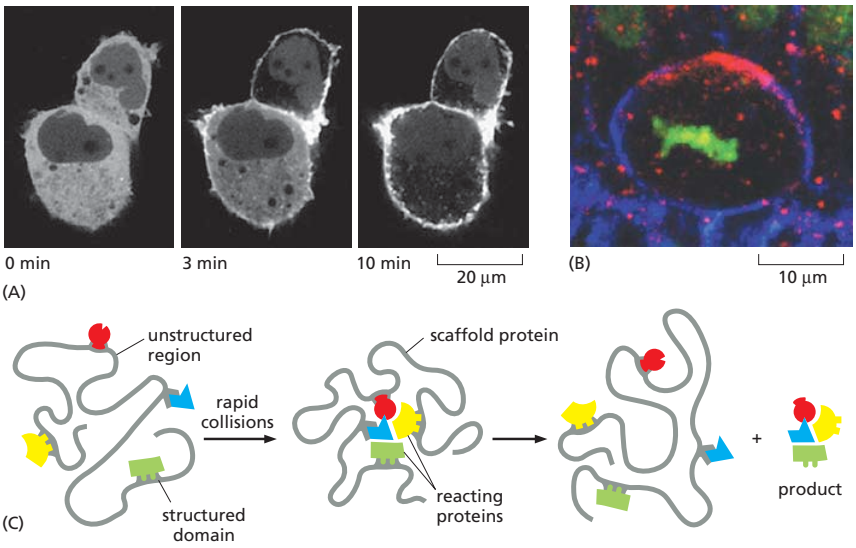


Figure 3–80 The assembly of protein machines at specific sites in a cell. (A) In response to a signal (here a phorbol ester), the gamma subspecies of protein kinase C moves rapidly from the cytosol to the plasma membrane. The protein kinase is fluorescent in these living cells because an engineered gene inside the cell encodes a fusion protein that links the kinase to green fluorescent protein (GFP). (B) The specific association of a different subspecies of protein kinase C (aPKC) with the apical tip of a differentiating neuroblast in an early *Drosophila* embryo. The kinase is stained *red*, and the cell nucleus *green*. (C) Diagram illustrating how a simple proximity created by scaffold proteins can greatly speed reactions in a cell. In this example, long unstructured regions of polypeptide chain in a large scaffold protein connect a series of structured domains that bind a set of reacting proteins. The unstructured regions serve as flexible "tethers" that greatly speed reaction rates by causing a rapid, random collision of all of the proteins that are bound to the scaffold. (For a simple example of tethering, see Figure 16–38.) (A, from N. Sakai et al, *J. Cell Biol.* 139:1465–1476, 1997. With permission from The Rockefeller University Press. B, courtesy of Andreas Wodarz, Institute of Genetics, University of Düsseldorf, Germany.)

These modifications create sites on proteins that bind them to particular scaffold proteins, thereby clustering the proteins required for particular reactions in specific regions of the cell. Most biological reactions are catalyzed by sets of 5 or more proteins, and such a clustering of proteins is often required for the reaction to occur. Scaffolds thereby allow cells to compartmentalize reactions even in the absence of membranes. Although only recently recognized as a widespread phenomenon, this type of clustering is particularly obvious in the cell nucleus (see Figure 4–69).

Many scaffolds appear to be quite different from the cullin illustrated previously in Figure 3–79: rather than holding their bound proteins in precise positions relative to each other, the interacting proteins are linked by unstructured regions of polypeptide chain. This tethers the proteins together, causing them to collide frequently with each other in random orientations—some of which will lead to a productive reaction (Figure 3–80C). In essence, this mechanism greatly speeds reactions by creating a very high local concentration of the reacting species. For this reason, the use of scaffold proteins represents an especially versatile way of controlling cell chemistry (see also Figure 15–61).

Many Proteins Are Controlled by Multisite Covalent Modification

We have thus far described only one type of posttranslational modification of proteins—that in which a phosphate is covalently attached to an amino acid side chain (see Figure 3–64). But a large number of other such modifications also occur, more than 200 distinct types being known. To give a sense of the variety, **Table 3–3** presents a subset of modifying groups with known regulatory roles. As

Table 3–3 Some Molecules Covalently Attached to Proteins Regulate Protein Function

MODIFYING GROUP	SOME PROMINENT FUNCTIONS
Phosphate on Ser, Thr, or Tyr	Drives the assembly of a protein into larger complexes (see Figure 15–19).
Methyl on Lys	Helps to creates histone code in chromatin through forming either mono-, di-, or tri-methyl lysine (see Figure 4–38).
Acetyl on Lys	Helps to creates histone code in chromatin (see Figure 4–38).
Palmityl group on Cys	This fatty acid addition drives protein association with membranes (see Figure 10–20).
***N*-acetylglucosamine** on Ser or Thr	Controls enzyme activity and gene expression in glucose homeostasis.
Ubiquitin on Lys	Monoubiquitin addition regulates the transport of membrane proteins in vesicles (see Figure 13–58).
	A polyubiquitin chain targets a protein for degradation (see Figure 3–79).

Ubiquitin is a 76 amino acid polypeptide; there are at least 10 other ubiquitin-related proteins, such as SUMO, that modify proteins in similar ways.

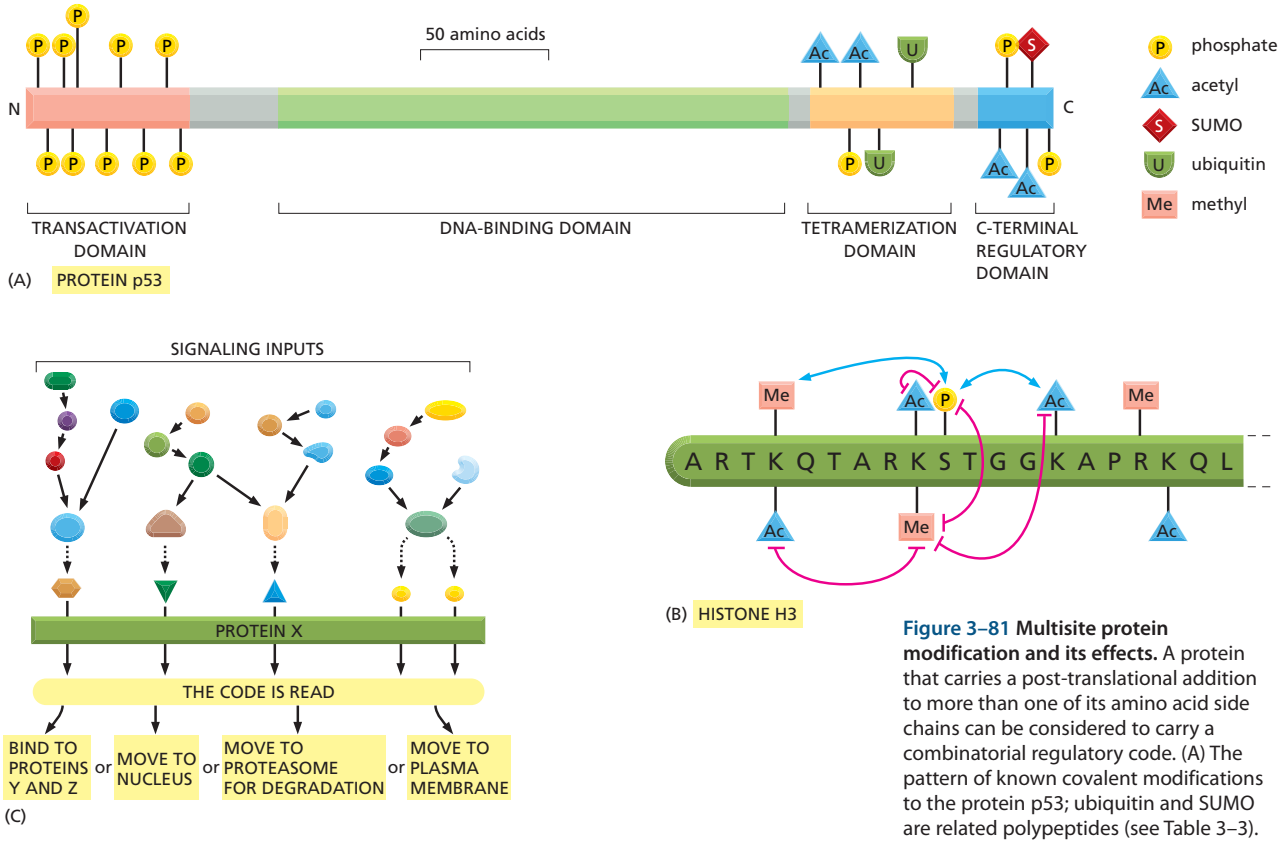


Figure 3–81 Multisite protein modification and its effects. A protein that carries a post-translational addition to more than one of its amino acid side chains can be considered to carry a combinatorial regulatory code. (A) The pattern of known covalent modifications to the protein p53; ubiquitin and SUMO are related polypeptides (see Table 3–3). (B) The possible modifications on the first 20 amino acids at the N-terminus of histone H3, showing not only their locations but also their activating *(blue)* and inhibiting *(red)* effects on the addition of neighboring covalent modifications. In addition to the effects shown, the acetylation and methylation of a lysine are mutually exclusive reactions (see Figure 4–38). (C) Diagram showing the general manner in which multisite modifications are added to (and removed from) a protein through signaling networks, and how the resulting combinatorial regulatory code on the protein is read to alter its behavior in the cell.

in phosphate addition, these groups are added and then removed from proteins according to the needs of the cell.

A large number of proteins are now known to be modified on more than one amino acid side chain, with different regulatory events producing a different pattern of such modifications. A striking example is the protein p53, which plays a central part in controlling a cell's response to adverse circumstances (see p. 1105). Through one of four different types of molecular additions, this protein can be modified at 20 different sites (**Figure 3–81**A). Because an enormous number of different combinations of these 20 modifications are possible, the protein's behavior can in principle be altered in a huge number of ways. Moreover, the pattern of modifications on a protein can determine its susceptibility to further modification, as illustrated by histone H3 in Figure 3–81B.

Cell biologists have only recently come to recognize that each protein's set of covalent modifications constitutes an important *combinatorial regulatory code.* As specific modifying groups are added to or removed from a protein, this code causes a different set of protein behaviors—changing the activity or stability of the protein, its binding partners, and its specific location within the cell (Figure 3–81C). This helps the cell respond rapidly and with great versatility to changes in its condition or environment.

A Complex Network of Protein Interactions Underlies Cell Function

There are many challenges facing cell biologists in this "post-genome" era when complete genome sequences are known. One is the need to dissect and reconstruct each one of the thousands of protein machines that exist in an organism such as ourselves. To understand these remarkable protein complexes, each must be reconstituted from its purified protein parts, so that we can study its detailed mode of operation under controlled conditions in a test tube, free from

all other cell components. This alone is a massive task. But we now know that each of these subcomponents of a cell also interacts with other sets of macromolecules, creating a large network of protein–protein and protein–nucleic acid interactions throughout the cell. To understand the cell, therefore, we need to analyze most of these other interactions as well.

We can gain some idea of the complexity of intracellular protein networks from a particularly well-studied example described in Chapter 16: the many dozens of proteins that interact with the actin cytoskeleton in the yeast *Saccharomyces cerevisiae* (see Figure 16–18). The extent of such protein–protein interactions can also be estimated more generally. An enormous amount of valuable information is now freely available in protein databases on the Internet: tens of thousands of three-dimensional protein structures plus tens of millions of protein sequences derived from the nucleotide sequences of genes. Scientists have been developing new methods for mining this great resource to increase our understanding of cells. In particular, computer-based bioinformatics tools are being combined with robotics and microarray technologies (see p. 574) to allow thousands of proteins to be investigated in a single set of experiments. **Proteomics** is a term that is often used to describe such research focused on the large-scale analysis of proteins, analogous to the term *genomics* describing the large-scale analysis of DNA sequences and genes.

Biologists use two different large-scale methods to map the direct binding interactions between the many different proteins in a cell. The initial method of choice was based on genetics: through an ingenious technique known as the yeast two-hybrid screen (see Figure 8–24), tens of thousands of interactions between thousands of proteins have been mapped in yeast, a nematode, and the fruit fly *Drosophila*. More recently, a biochemical method based on affinity tagging and mass spectroscopy has gained favor (discussed in Chapter 8), because it appears to produce fewer spurious results. The results of these and other analyses that predict protein binding interactions have been tabulated and organized in Internet databases. This allows a cell biologist studying a small set of proteins to readily discover which other proteins in the same cell are thought to bind to, and thus interact with, that set of proteins. When displayed graphically as a *protein interaction map*, each protein is represented by a box or dot in a two-dimensional network, with a straight line connecting those proteins that have been found to bind to each other.

When hundreds or thousands of proteins are displayed on the same map, the network diagram becomes bewilderingly complicated, serving to illustrate how much more we have to learn before we can claim to really understand the cell. Much more useful are small subsections of these maps, centered on a few proteins of interest. Thus, **Figure 3–82** shows a network of protein–protein interactions for the five proteins that form the SCF ubiquitin ligase in a yeast cell (see Figure 3–79). Four of the subunits of this ligase are located at the bottom right of Figure 3–82. The remaining subunit, the F-box protein that serves as its substrate-binding arm, appears as a set of 15 different gene products that bind to adaptor protein 2 (the Skp1 protein). Along the top and left of the figure are sets of additional protein interactions marked with *yellow* and *green* shading: as indicated, these protein sets function at the origin of DNA replication, in cell cycle regulation, in methionine synthesis, in the kinetochore, and in vacuolar H^+-ATPase assembly. We shall use this figure to explain how such protein interaction maps are used, and what they do and do not mean.

1. Protein interaction maps are useful for identifying the likely function of previously uncharacterized proteins. Examples are the products of the genes that have thus far only been inferred to exist from the yeast genome sequence, which are the six proteins in the figure that lack a simple three-letter abbreviation (*white letters* beginning with Y). One, the product of so-called *open reading frame* YDR196C, is located in the origin of replication group, and it is therefore likely to have a role in starting new replication forks. The remaining five in this diagram are F-box proteins that bind to Skp1; these are therefore likely to function as part of the ubiquitin ligase, serving as substrate-binding arms that recognize different target proteins.

However, as we discuss next, neither assignment can be considered certain without additional data.

2. Protein interaction networks need to be interpreted with caution because, as a result of evolution making efficient use of each organism's genetic information, the same protein can be used as part of two different protein complexes that have different types of functions. Thus, although protein A binds to protein B and protein B binds to protein C, proteins A and C need not function in the same process. For example, we know from detailed biochemical studies that the functions of Skp1 in the kinetochore and in vacuolar H^+-ATPase assembly *(yellow shading)* are separate from its function in the SCF ubiquitin ligase. In fact, only the remaining three functions of Skp1 illustrated in the diagram—methionine synthesis, cell cycle regulation, and origin of replication *(green shading)*—involve ubiquitylation.
3. In cross-species comparisons, those proteins displaying similar patterns of interactions in the two protein interaction maps are likely to have the same function in the cell. Thus, as scientists generate more and more highly detailed maps for multiple organisms, the results will become increasingly useful for inferring protein function. These map comparisons are a particularly powerful tool for deciphering the functions of human proteins. There is a vast amount of direct information about protein function that can be obtained from genetic engineering, mutational, and

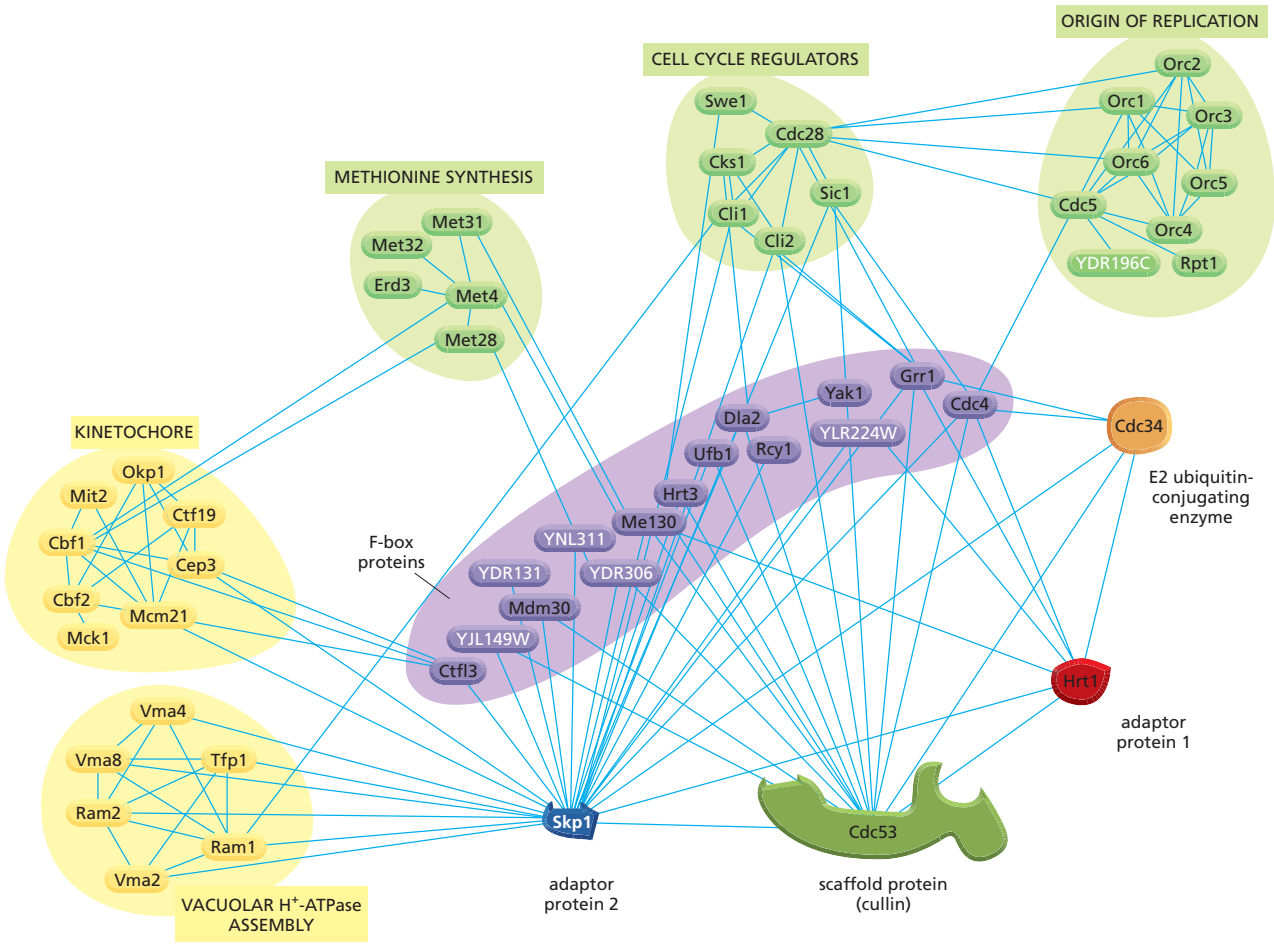


Figure 3–82 A map of some protein– protein interactions of the SCF ubiquitin ligase and other proteins in the yeast *S. cerevisiae*. The symbols and/or colors used for the 5 proteins of the ligase are those in Figure 3–79. Note that 15 different F-box proteins are shown (*purple*); those with *white* lettering (beginning with Y) are only known from the genome sequence as open reading frames. For additional details, see text. (Courtesy of Peter Bowers and David Eisenberg, UCLA-DOE Institute for Genomics and Proteomics, UCLA.)

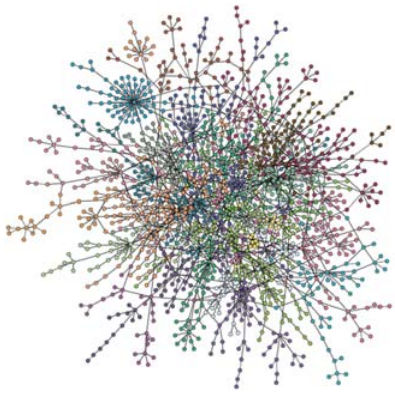


Figure 3–83 A network of protein-binding interactions in a yeast cell. Each line connecting a pair of dots (proteins) indicates a protein–protein interaction. (From A. Guimerá and M. Sales–Pardo, *Mol. Syst. Biol.* 2:42, 2006. With permission from Macmillan Publishers Ltd.)

genetic analyses in model organisms—such as yeast, worms, and flies—that is not available in humans.

The available data suggest that a typical protein in a human cell may interact with between 5 and 15 different partners. Often, each of the different domains in a multidomain protein binds to a different set of partners; in fact, we can speculate that the unusually extensive multidomain structures observed for human proteins may have evolved to facilitate these interactions. Given the enormous complexity of the interacting networks of macromolecules in cells (**Figure 3–83**), deciphering their full functional meaning may well keep scientists busy for centuries.

Summary

Proteins can form enormously sophisticated chemical devices, whose functions largely depend on the detailed chemical properties of their surfaces. Binding sites for ligands are formed as surface cavities in which precisely positioned amino acid side chains are brought together by protein folding. In this way, normally unreactive amino acid side chains can be activated to make and break covalent bonds. Enzymes are catalytic proteins that greatly speed up reaction rates by binding the high-energy transition states for a specific reaction path; they also perform acid catalysis and base catalysis simultaneously. The rates of enzyme reactions are often so fast that they are limited only by diffusion; rates can be further increased if enzymes that act sequentially on a substrate are joined into a single multienzyme complex, or if the enzymes and their substrates are confined to the same compartment of the cell.

Proteins reversibly change their shape when ligands bind to their surface. The allosteric changes in protein conformation produced by one ligand affect the binding of a second ligand, and this linkage between two ligand-binding sites provides a crucial mechanism for regulating cell processes. Metabolic pathways, for example, are controlled by feedback regulation: some small molecules inhibit and other small molecules activate enzymes early in a pathway. Enzymes controlled in this way generally form symmetric assemblies, allowing cooperative conformational changes to create a steep response to changes in the concentrations of the ligands that regulate them.

The expenditure of chemical energy can drive unidirectional changes in protein shape. By coupling allosteric shape changes to ATP hydrolysis, for example, proteins can do useful work, such as generating a mechanical force or moving for long distances in a single direction. The three-dimensional structures of proteins, determined by x-ray crystallography, have revealed how a small local change caused by nucleoside triphosphate hydrolysis is amplified to create major changes elsewhere in the protein. By such means, these proteins can serve as input–output devices that transmit information, as assembly factors, as motors, or as membrane-bound pumps. Highly efficient protein machines are formed by incorporating many different protein molecules into larger assemblies that coordinate the allosteric movements of the individual components. Such machines are now known to perform many of the most important reactions in cells.

Proteins are subjected to many reversible post-translational modifications, such as the covalent addition of a phosphate or an acetyl group to a specific amino acid side chain. The addition of these modifying groups is used to regulate the activity of a protein, changing its conformation, its binding to other proteins and its location inside the cell. A typical protein in a cell will interact with more than five different partners. Using the new technologies of proteomics, biologists can analyze thousands of proteins in one set of experiments. One important result is the production of detailed protein interaction maps, which aim at describing all of the binding interactions between the thousands of distinct proteins in a cell.

PROBLEMS

Which statements are true? Explain why or why not.

3–1 Each strand in a β sheet is a helix with two amino acids per turn.

3–2 Loops of polypeptide that protrude from the surface of a protein often form the binding sites for other molecules.

3–3 An enzyme reaches a maximum rate at high substrate concentration because it has a fixed number of active sites where substrate binds.

3–4 Higher concentrations of enzyme give rise to a higher turnover number.

3–5 Enzymes such as aspartate transcarbamoylase that undergo cooperative allosteric transitions invariably contain multiple identical subunits.

3–6 Continual addition and removal of phosphates by protein kinases and protein phosphatases is wasteful of energy—since their combined action consumes ATP—but it is a necessary consequence of effective regulation by phosphorylation.

Discuss the following problems.

3–7 Consider the following statement. "To produce one molecule of each possible kind of polypeptide chain, 300 amino acids in length, would require more atoms than exist in the universe." Given the size of the universe, do you suppose this statement could possibly be correct? Since counting atoms is a tricky business, consider the problem from the standpoint of mass. The mass of the observable universe is estimated to be about 10^{80} grams, give or take an order of magnitude or so. Assuming that the average mass of an amino acid is 110 daltons, what would be the mass of one molecule of each possible kind of polypeptide chain 300 amino acids in length? Is this greater than the mass of the universe?

3–8 A common strategy for identifying distantly related proteins is to search the database using a short signature sequence indicative of the particular protein function. Why is it better to search with a short sequence than with a long sequence? Do you not have more chances for a 'hit' in the database with a long sequence?

3–9 The so-called kelch motif consists of a four-stranded β sheet, which forms what is known as a β propeller. It is usually found to be repeated four to seven times, forming a kelch repeat domain in a multidomain protein. One such kelch repeat domain is shown in **Figure Q3–1**. Would you classify this domain as an 'in-line' or 'plug-in' type domain?

3–10 Titin, which has a molecular weight of 3×10^6 daltons, is the largest polypeptide yet described. Titin molecules extend from muscle thick filaments to the Z disc; they are thought to act as springs to keep the thick filaments centered in the sarcomere. Titin is composed of a large number of repeated immunoglobulin (Ig) sequences of 89 amino acids, each of which is folded into a domain about 4 nm in length (**Figure Q3–2**A).

You suspect that the springlike behavior of titin is caused by the sequential unfolding (and refolding) of individual Ig domains. You test this hypothesis using the atomic force microscope, which allows you to pick up one end of a protein molecule and pull with an accurately measured force. For a fragment of titin containing seven repeats of the Ig domain, this experiment gives the sawtooth force-versus-extension curve shown in Figure Q3–2B. When the experiment is repeated in a solution of 8 M urea (a protein denaturant), the peaks disappear and the measured extension becomes much longer for a given force. If the experiment is repeated after the protein has been cross-linked by treatment with glutaraldehyde, once again the peaks disappear but the extension becomes much smaller for a given force.

A. Are the data consistent with your hypothesis that titin's springlike behavior is due to the sequential unfolding of individual Ig domains? Explain your reasoning.

B. Is the extension for each putative domain-unfolding event the magnitude you would expect? (In an extended polypeptide chain, amino acids are spaced at intervals of 0.34 nm.)

C. Why is each successive peak in Figure Q3–2B a little higher than the one before?

D. Why does the force collapse so abruptly after each peak?

3–11 It is often said that protein complexes are made from subunits (that is, individually synthesized proteins) rather than as one long protein because the former is more likely to give a correct final structure.

A. Assuming that the protein synthesis machinery incorporates one incorrect amino acid for each 10,000 it inserts,

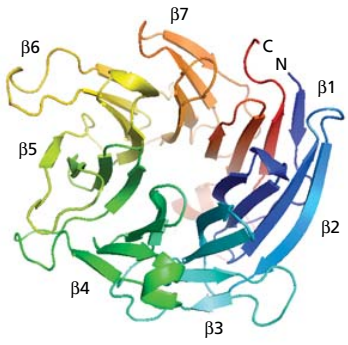


Figure Q3–1 The kelch repeat domain of galactose oxidase from *D. dendroides* (Problem 3–9). The seven individual β propellers are indicated. The N- and C-termini are indicated by N and C.

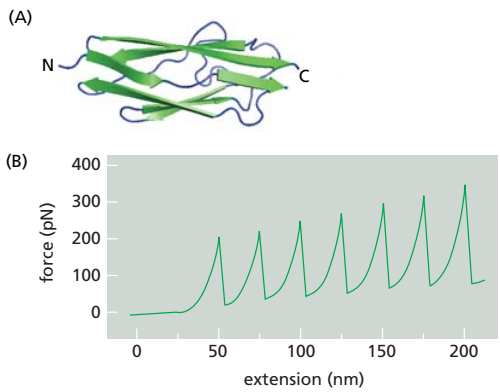


Figure Q3–2 Springlike behavior of titin (Problem 3–10). (A) The structure of an individual Ig domain. (B) Force in piconewtons versus extension in nanometers obtained by atomic force microscopy.

calculate the fraction of bacterial ribosomes that would be assembled correctly if the proteins were synthesized as one large protein versus built from individual proteins? For the sake of calculation assume that the ribosome is composed of 50 proteins, each 200 amino acids in length, and that the subunits—correct and incorrect—are assembled with equal likelihood into the complete ribosome. [The probability that a polypeptide will be made correctly, P_C, equals the fraction correct for each operation, f_C, raised to a power equal to the number of operations, n: $P_C = (f_C)^n$. For an error rate of 1/10,000, $f_C = 0.9999$.]

B. Is the assumption that correct and incorrect subunits assemble equally well likely to be true? Why or why not? How would a change in that assumption affect the calculation in part A?

3–12 Rous sarcoma virus (RSV) carries an oncogene called *Src*, which encodes a continuously active protein tyrosine kinase that leads to unchecked cell proliferation. Normally, Src carries an attached fatty acid (myristoylate) group that allows it to bind to the cytoplasmic side of the plasma membrane. A mutant version of Src that does not allow attachment of myristoylate does not bind to the membrane. Infection of cells with RSV encoding either the normal or the mutant form of Src leads to the same high level of protein tyrosine kinase activity, but the mutant Src does not cause cell proliferation.

A. Assuming that the normal Src is all bound to the plasma membrane and that the mutant Src is distributed throughout the cytoplasm, calculate their relative concentrations in the neighborhood of the plasma membrane. For the purposes of this calculation, assume that the cell is a sphere with a radius of 10 μm and that the mutant Src is distributed throughout, whereas the normal Src is confined to a 4-nm-thick layer immediately beneath the membrane. [For this problem, assume that the membrane has no thickness. The volume of a sphere is $(4/3)\pi r^3$.]

B. The target (X) for phosphorylation by *Src* resides in the membrane. Explain why the mutant Src does not cause cell proliferation.

3–13 An antibody binds to another protein with an equilibrium constant, K, of 5×10^9 M^{-1}. When it binds to a second, related protein, it forms three fewer hydrogen bonds, reducing its binding affinity by 2.8 kcal/mole. What is the K for its binding to the second protein? (Free-energy change is related to the equilibrium constant by the equation $\Delta G° = -2.3\,RT \log K$, where R is 1.98×10^{-3} kcal/(mole K) and T is 310 K.)

3–14 The protein SmpB binds to a special species of tRNA, tmRNA, to eliminate the incomplete proteins made from truncated mRNAs in bacteria. If the binding of SmpB to tmRNA is plotted as fraction tmRNA bound versus SmpB concentration, one obtains a symmetrical S-shaped curve as shown in **Figure Q3–3**. This curve is a visual display of a very useful relationship between K_d and concentration, which has broad applicability. The general expression for fraction of ligand bound is derived from the equation for K_d (K_d = [Pr][L]/[Pr–L]) by substituting ($[L]_{TOT} - [L]$) for [Pr–L] and rearranging. Because the total concentration of ligand ($[L]_{TOT}$) is equal to the free ligand ([L]) plus bound ligand ([Pr–L]),

$$\text{fraction bound} = [L]/[L]_{TOT} = [Pr]/([Pr] + K_d)$$

For SmpB and tmRNA, the fraction bound = $[tmRNA]/[tmRNA]_{TOT} = [SmpB]/([SmpB] + K_d)$. Using this relationship, calculate the fraction of tmRNA bound for SmpB concentrations equal to 10^4K_d, 10^3K_d, 10^2K_d, 10^1K_d, K_d, $10^{-1}K_d$, $10^{-2}K_d$, $10^{-3}K_d$, and $10^{-4}K_d$.

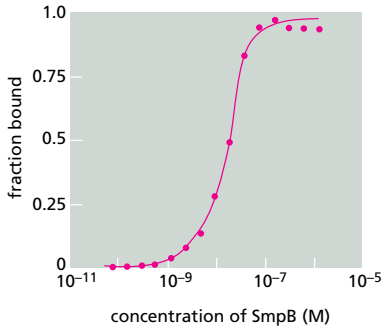


Figure Q3–3 Fraction of tmRNA bound versus SmpB concentration (Problem 3–14).

3–15 Many enzymes obey simple Michaelis–Menten kinetics, which are summarized by the equation

$$\text{rate} = V_{max}[S]/([S] + K_m)$$

where V_{max} = maximum velocity, [S] = concentration of substrate, and K_m = the Michaelis constant.

It is instructive to plug a few values of [S] into the equation to see how rate is affected. What are the rates for [S] equal to zero, equal to K_m, and equal to infinite concentration?

3–16 The enzyme hexokinase adds a phosphate to D-glucose but ignores its mirror image, L-glucose. Suppose that you were able to synthesize hexokinase entirely from D-amino acids, which are the mirror image of the normal L-amino acids.

A. Assuming that the 'D' enzyme would fold to a stable conformation, what relationship would you expect it to bear to the normal 'L' enzyme?

B. Do you suppose the 'D' enzyme would add a phosphate to L-glucose, and ignore D-glucose?

3–17 How do you suppose that a molecule of hemoglobin is able to bind oxygen efficiently in the lungs, and yet release it efficiently in the tissues?

3–18 Synthesis of the purine nucleotides AMP and GMP proceeds by a branched pathway starting with ribose 5-phosphate (R5P), as shown schematically in **Figure Q3–4**. Using the principles of feedback inhibition, propose a regulatory strategy for this pathway that ensures an adequate supply of both AMP and GMP and minimizes the buildup of the intermediates (*A*–*I*) when supplies of AMP and GMP are adequate.

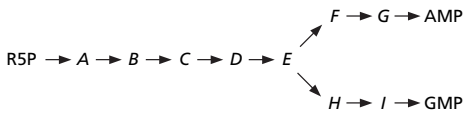


Figure Q3–4 Schematic diagram of the metabolic pathway for synthesis of AMP and GMP from R5P (Problem 3–18).

REFERENCES

General

Berg JM, Tymoczko JL & Stryer L (2006) Biochemistry, 6th ed. New York: WH Freeman.

Branden C & Tooze J (1999) Introduction to Protein Structure, 2nd ed. New York: Garland Science.

Dickerson, RE (2005) Present at the Flood: How Structural Molecular Biology Came About. Sunderland, MA: Sinauer

Kyte J (2006) Structure in Protein Chemistry. New York: Routledge.

Petsko GA & Ringe D (2004) Protein Structure and Function. London: New Science Press.

Perutz M (1992) Protein Structure: New Approaches to Disease and Therapy. New York: WH Freeman.

The Shape and Structure of Proteins

Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181:223–230.

Bray D (2005) Flexible peptides and cytoplasmic gels. *Genome Biol* 6:106–109.

Burkhard P, Stetefeld J & Strelkov SV (2001) Coiled coils: a highly versatile protein folding motif. *Trends Cell Biol* 11:82–88.

Caspar DLD & Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27:1–24.

Doolittle RF (1995) The multiplicity of domains in proteins. *Annu Rev Biochem* 64:287–314.

Eisenberg D (2003) The discovery of the alpha-helix and beta-sheet, the principle structural features of proteins. *Proc Natl Acad Sci USA* 100:11207–11210.

Fraenkel-Conrat H & Williams RC (1955) Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc Natl Acad Sci USA* 41:690–698.

Goodsell DS & Olson AJ (2000) Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct* 29:105–153.

Harrison SC (1992) Viruses. *Curr Opin Struct Biol* 2:293–299.

Harrison SC (2004) Whither structural biology? *Nature Struct Mol Biol* 11:12–15.

Hudder A, Nathanson L & Deutscher MP (2003) Organization of mammalian cytoplasm. *Mol Cell Biol* 23:9318–9326.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Meiler J & Baker D (2003) Coupled prediction of protein secondary and tertiary structure. *Proc Natl Acad Sci USA* 100:12105–12110.

Nomura M (1973) Assembly of bacterial ribosomes. *Science* 179: 864–873.

Orengo CA & Thornton JM (2005) Protein families and their evolution—a structural perspective. *Annu Rev Biochem* 74:867–900.

Pauling L & Corey RB (1951) Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc Natl Acad Sci USA* 37:729–740.

Pauling L, Corey RB & Branson HR (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA* 37:205–211.

Ponting CP, Schultz J, Copley RR et al. (2000) Evolution of domain families. *Adv Protein Chem* 54:185–244.

Trinick J (1992) Understanding the functions of titin and nebulin. *FEBS Lett* 307:44–48.

Vogel C, Bashton M, Kerrison ND et al (2004) Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* 14:208–216.

Zhang C & Kim SH (2003) Overview of structural genomics: from structure to function. *Curr Opin Chem Biol* 7:28–32.

Protein Function

Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92:291–294.

Benkovic SJ (1992) Catalytic antibodies. *Annu Rev Biochem* 61:29–54.

Berg OG & von Hippel PH (1985) Diffusion-controlled macromolecular interactions. *Annu Rev Biophys Biophys Chem* 14:131–160.

Bhattacharyya RP, Remenyi A, Yeh BJ & Lim WA (2006) Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu Rev Biochem* 75:655–680.

Bourne HR (1995) GTPases: a family of molecular switches and clocks. *Philos Trans R Soc Lond B* 349:283–289.

Braden BC & Poljak RJ (1995) Structural features of the reactions between antibodies and protein antigens. *FASEB J* 9:9–16.

Dickerson RE & Geis I (1983) Hemoglobin: Structure, Function, Evolution and Pathology. Menlo Park, CA: Benjamin Cummings.

Dressler D & Potter H (1991) Discovering Enzymes. New York: Scientific American Library.

Eisenberg D, Marcotte, EM, Xenarios, I & Yeates TO (2000) Protein function in the post-genomic era. *Nature* 405:823–826.

Fersht AR (1999) Structure and Mechanisms in Protein Science: A Guide to Enzyme Catalysis. New York: WH Freeman.

Johnson, LN & Lewis RJ (2001) Structural basis for control by phosphorylation. *Chem Rev* 101:2209–2242.

Kantrowitz ER & Lipscomb WN (1988) *Escherichia coli* aspartate transcarbamoylase: the relation between structure and function. *Science* 241:669–674.

Khosla C & Harbury PB (2001) Modular enzymes. *Nature* 409:247–252.

Kim E & Sheng M (2004) PDZ domain proteins of synapses. *Nature Rev Neurosci* 5:771–781.

Koshland DE, Jr (1984) Control of enzyme activity and metabolic pathways. *Trends Biochem Sci* 9:155–159.

Kraut DA, Carroll KS & Herschlag D (2003) Challenges in enzyme mechanism and energetics. *Annu Rev Biochem* 72:517–571.

Krogan NJ, Cagney G, Yu H et al (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisisae*. *Nature* 440:637–643.

Lichtarge O, Bourne HR & Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* 257:342–358.

Marcotte EM, Pellegrini M, Ng HL et al (1999) Detecting protein function and protein–protein interactions from genome sequences. *Science* 285:751–753.

Monod J, Changeux JP & Jacob F (1963) Allosteric proteins and cellular control systems. *J Mol Biol* 6:306–329.

Pawson T & Nash P (2003) Assembly of regulatory systems through protein interaction domains. *Science* 300:445–452.

Pavletich NP (1999) Mechanisms of cyclin-dependent kinase regulation: structures of Cdks, their cyclin activators, and Cip and INK4 inhibitors. *J Mol Biol* 287:821–828.

Pellicena P & Kuriyan J (2006) Protein-protein interactions in the allosteric regulation of protein kinases. *Curr Opin Struct Biol* 16:702–709.

Perutz M (1990) Mechanisms of Cooperativity and Allosteric Regulation in Proteins. Cambridge: Cambridge University Press.

Raushel F, Thoden, JB & Holden HM (2003) Enzymes with molecular tunnels. *Acc Chem Res* 36:539–548.

Radzicka A & Wolfenden R (1995) A proficient enzyme. *Science* 267:90–93.

Sato TK, Overduin M & Emr S (2001) Location, location, location: Membrane targeting directed by PX domains. *Science* 294:1881–1885.

Schramm VL (1998) Enzymatic transition states and transition state analog design. *Annu Rev Biochem* 67:693–720.

Schultz PG & Lerner RA (1995) From molecular diversity to catalysis: lessons from the immune system. *Science* 269:1835–1842.

Vale RD & Milligan RA (2000) The way things move: looking under the hood of molecular motor proteins. *Science* 288:88–95.

Vocadlo DJ, Davies GJ, Laine R & Withers SG (2001) Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* 412:835–838.

Walsh C (2001) Enabling the chemistry of life. *Nature* 409:226–231.

Yang XJ (2005) Multisite protein modification and intramolecular signaling. *Oncogene* 24:1653–1662.

Zhu H, Bilgin M & Snyder M (2003) Proteomics. *Annu Rev Biochem* 72:783–812.