Лекция 2.
Структура и функция белка

PART

Structure and Catalysis

In Part I we contrasted the complex structure and function of living cells with the relative simplicity of the monomeric units from which the enzymes, supramolecular complexes, and organelles of the cells are constructed. Part II is devoted to the structure and function of the major classes of cellular constituents: amino acids and proteins (Chapters 5 through 8), fatty acids, lipids, and membranes (Chapters 9 and 10), sugars and polysaccharides (Chapter 11), and nucleotides and nucleic acids (Chapter 12). We begin in each case by considering the covalent structure of the simple subunits (amino acids, fatty acids, monosaccharides, and nucleotides). These subunits are a major part of the language of biochemistry; familiarity with them is a prerequisite for understanding more advanced topics covered in this book, as well as the rapidly growing and exciting literature of biochemistry.

After describing the covalent chemistry of the monomeric units, we consider the structure of the macromolecules and supramolecular complexes derived from them. An overriding theme is that the polymeric macromolecules in living systems, though large, are highly ordered chemical entities, with specific sequences of monomeric subunits giving rise to discrete structures and functions. This fundamental theme can be broken down into three interrelated principles: (1) the unique structure of each macromolecule determines its function; (2) noncovalent interactions play a critical role in the structure and function of macromolecules; and (3) the specific sequences of monomeric subunits in polymeric macromolecules contain the information upon which the ordered living state depends. Each of these principles deserves further comment.

The relationship between structure and function is especially evident in proteins, which exhibit an extraordinary diversity of functions. One particular polymeric sequence of amino acids produces a strong, fibrous structure found in hair and wool; another produces a protein that transports oxygen in the blood. Similarly, the special functions of lipids, polysaccharides, and nucleic acids can be understood as a direct manifestation of their chemical structure, with their characteristic monomeric subunits linked in precise functional groups or polymers. Lipids aggregate to form membranes; sugars linked together become energy stores and structural fibers; nucleotides in a polymer become the blueprint for an entire organism.

As we move from monomeric units to larger and larger polymers, the chemical focus shifts from covalent bonds to noncovalent interactions. The covalent nature of monomeric units, and of the bonds that connect them in polymers, places strong constraints upon the shapes

Facing page: End-on view of the triple-stranded collagen superhelix. Collagen, a component of connective tissue, provides tensile strength and resiliency. Its strength is derived in part from the three tightly wrapped identical helical strands (shown in gray, purple, and blue), much the way a length of rope is stronger than its constituent fibers. The tight wrapping is made possible by the presence of glycine, shown in red, at every third position along each strand, where the strands are in contact. Glycine's small size allows for very close contact.

assumed by large molecules. It is the numerous noncovalent interactions, however, that dictate the stable native conformation and provide the flexibility necessary for the biological function of these large molecules. We will see that noncovalent interactions are essential to the catalytic power of enzymes, the arrangement and properties of lipids in a membrane, and the critical interaction of complementary base pairs in nucleic acids.

The principle that sequences of monomeric subunits are information-rich emerges fully in the discussion of nucleic acids in Chapter 12. However, proteins and some polysaccharides are also information-rich molecules. The amino acid sequence is a form of information that directs the folding of the protein into its unique three-dimensional structure, and ultimately determines the function of the protein. Some polysaccharides also have unique sequences and three-dimensional structures that can be recognized by other macromolecules.

For each class of molecules we find a similar structural hierarchy, in which subunits of fixed structure are connected by bonds of limited flexibility, to form macromolecules with three-dimensional structures determined by noncovalent interactions. Together, the molecules described in Part II are the "stuff" of life. We begin with the amino acids.

CHAPTER

Amino Acids and Peptides

Proteins are the most abundant macromolecules in living cells, occurring in all cells and all parts of cells. Proteins also occur in great variety; thousands of different kinds may be found in a single cell. Moreover, proteins exhibit great diversity in their biological function. Their central role is made evident by the fact that proteins are the most important final products of the information pathways discussed in Part IV of this book. In a sense, they are the molecular instruments through which genetic information is expressed. It is appropriate to begin the study of biological macromolecules with the proteins, whose name derives from the Greek *prōtos,* meaning "first" or "foremost."

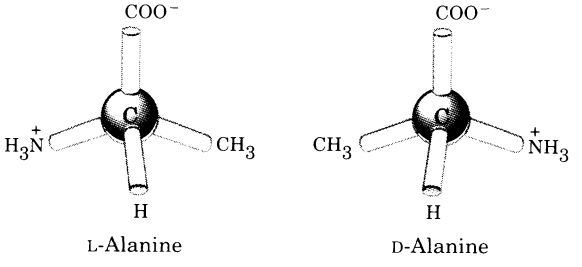
Relatively simple monomeric subunits provide the key to the structure of the thousands of different proteins. All proteins, whether from the most ancient lines of bacteria or from the most complex forms of life, are constructed from the same ubiquitous set of 20 amino acids, covalently linked in characteristic linear sequences. Because each of these amino acids has a distinctive side chain that determines its chemical properties, this group of 20 precursor molecules may be regarded as the alphabet in which the language of protein structure is written.

Proteins are chains of amino acids, each joined to its neighbor by a specific type of covalent bond. What is most remarkable is that cells can produce proteins that have strikingly different properties and activities by joining the same 20 amino acids in many different combinations and sequences. From these building blocks different organisms can make such widely diverse products as enzymes, hormones, antibodies, the lens protein of the eye, feathers, spider webs, rhinoceros horns (Fig. 5–1), milk proteins, antibiotics, mushroom poisons, and a myriad of other substances having distinct biological activities.

Protein structure and function is the topic for the next four chapters. In this chapter we begin with a description of amino acids and the covalent bonds that link them together in peptides and proteins.

Figure 5–1 The protein keratin is formed by all vertebrates. It is the chief structural component of hair, scales, horn, wool, nails, and feathers. The black rhinoceros is nearing extinction in the wild because of the myths prevalent in some parts of the world that a powder derived from its horn has aphrodisiac properties. In reality, the chemical properties are no different from those of powdered bovine hooves or human fingernails.

Figure 5–2 General structure of the amino acids found in proteins. With the exception of the nature of the R group, this structure is common to all the α-amino acids. (Proline, because it is an imino acid, is an exceptional component of proteins.) The α carbon is shown in blue. R (in red) represents the R group or side chain, which is different in each amino acid. In all amino acids except glycine (shown for comparison) the α-carbon atom has four different substituent groups.



(a)

(b)

(c)

Amino Acids

Proteins can be reduced to their constituent amino acids by a variety of methods, and the earliest studies of proteins naturally focused on the free amino acids derived from them. The first amino acid to be discovered in proteins was asparagine, in 1806. The last of the 20 to be found, threonine, was not identified until 1938. All the amino acids have trivial or common names, in some cases derived from the source from which they were first isolated. Asparagine was first found in asparagus, as one might guess; glutamate was found in wheat gluten; tyrosine was first isolated from cheese (thus its name is derived from the Greek *tyros,* "cheese"); and glycine (Greek *glykos,* "sweet") was so named because of its sweet taste.

Amino Acids Have Common Structural Features

All of the 20 amino acids found in proteins have a carboxyl group and an amino group bonded to the same carbon atom (the α carbon) (Fig. 5–2). They differ from each other in their side chains, or R groups, which vary in structure, size, and electric charge, and influence the solubility of amino acids in water. When the R group contains additional carbons in a chain, they are designated β, γ, δ, ϵ, etc., proceeding out from the α carbon. The 20 amino acids of proteins are often referred to as the standard, primary, or normal amino acids, to distinguish them from amino acids within proteins that are modified after the proteins are synthesized, and from many other kinds of amino acids present in living organisms but not in proteins. The standard amino acids have been assigned three-letter abbreviations and one-letter symbols (Table 5–1), which are used as shorthand to indicate the composition and sequence of amino acids in proteins.

We note in Figure 5–2 that for all the standard amino acids except one (glycine) the α carbon is asymmetric, bonded to four different substituent groups: a carboxyl group, an amino group, an R group, and a hydrogen atom. The α-carbon atom is thus a **chiral center** (see Fig. 3–9). Because of the tetrahedral arrangement of the bonding orbitals around the α-carbon atom of amino acids, the four different substituent groups can occupy two different arrangements in space, which are nonsuperimposable mirror images of each other (Fig. 5–3). These two forms are called **enantiomers** or **stereoisomers** (see Fig. 3–9). All molecules with a chiral center are also **optically active**—i.e., they can rotate plane-polarized light, with the direction of the rotation differing for different stereoisomers.

Figure 5–3 (a) The two stereoisomers of alanine. L- and D-alanine are nonsuperimposable mirror images of each other. **(b, c)** Two different conventions for showing the configurations in space of stereoisomers. In perspective formulas **(b)** the wedge-shaped bonds project out of the plane of the paper, the dashed bonds behind it. In projection formulas **(c)** the horizontal bonds are assumed to project out of the plane of the paper, the vertical bonds behind. However, projection formulas are often used casually without reference to stereochemical configuration.

Table 5–1 Properties and conventions associated with the standard amino acids

Amino acid	Abbreviated names		M_r	pK_1 (—COOH)	pK_2 (—NH_3^+)	pK_R (R group)	pI	Hydropathy index*	Occurrence in Proteins (%)†
Nonpolar, aliphatic R groups									
Glycine	**Gly**	**G**	75	2.34	9.60		5.97	−0.4	7.5
Alanine	**Ala**	**A**	89	2.34	9.69		6.01	1.8	9.0
Valine	**Val**	**V**	117	2.32	9.62		5.97	4.2	6.9
Leucine	**Leu**	**L**	131	2.36	9.60		5.98	3.8	7.5
Isoleucine	**Ile**	**I**	131	2.36	9.68		6.02	4.5	4.6
Proline	**Pro**	**P**	115	1.99	10.96		6.48	−1.6	4.6
Aromatic R groups									
Phenylalanine	**Phe**	**F**	165	1.83	9.13		5.48	2.8	3.5
Tyrosine	**Tyr**	**Y**	181	2.20	9.11	10.07	5.66	−1.3	3.5
Tryptophan	**Trp**	**W**	204	2.38	9.39		5.89	−0.9	1.1
Polar, uncharged R groups									
Serine	**Ser**	**S**	105	2.21	9.15	13.60	5.68	−0.8	7.1
Threonine	**Thr**	**T**	119	2.11	9.62	13.60	5.87	−0.7	6.0
Cysteine	**Cys**	**C**	121	1.96	8.18	10.28	5.07	2.5	2.8
Methionine	**Met**	**M**	149	2.28	9.21		5.74	1.9	1.7
Asparagine	**Asn**	**N**	132	2.02	8.80		5.41	−3.5	4.4
Glutamine	**Gln**	**Q**	146	2.17	9.13		5.65	−3.5	3.9
Negatively charged R groups									
Aspartate	**Asp**	**D**	133	1.88	9.60	3.65	2.77	−3.5	5.5
Glutamate	**Glu**	**E**	147	2.19	9.67	4.25	3.22	−3.5	6.2
Positively charged R groups									
Lysine	**Lys**	**K**	146	2.18	8.95	10.53	9.74	−3.9	7.0
Arginine	**Arg**	**R**	174	2.17	9.04	12.48	10.76	−4.5	4.7
Histidine	**His**	**H**	155	1.82	9.17	6.00	7.59	−3.2	2.1

* A scale combining hydrophobicity and hydrophilicity; can be used to predict which amino acids will be found in an aqueous environment (− values) and which will be found in a hydrophobic environment (+ values). See Box 10-2. From Kyte, J. & Doolittle, R.F. (1982) *J. Mol. Biol.* **157,** 105–132.

† Average occurrence in over 200 proteins. From Klapper, M.H. (1977) *Biochem. Biophys. Res. Commun.* **78,** 1018–1024.

The classification and naming of stereoisomers is based on the **absolute configuration** of the four substituents of the asymmetric carbon atom. For this purpose a reference compound has been chosen, to which all other optically active compounds are compared. This reference compound is the 3-carbon sugar glyceraldehyde (Fig. 5–4), the smallest sugar to have an asymmetric carbon atom. The naming of configurations of both simple sugars and amino acids is based on the absolute configuration of glyceraldehyde, as established by x-ray diffraction analysis. The stereoisomers of all chiral compounds having a configuration related to that of L-glyceraldehyde are designated L (for levorotatory, derived from *levo,* meaning "left"), and the stereoisomers related to D-glyceraldehyde are designated D (for dextrorotatory, derived from *dextro,* meaning "right"). The symbols L and D thus refer to the absolute configuration of the four substituents around the chiral carbon.

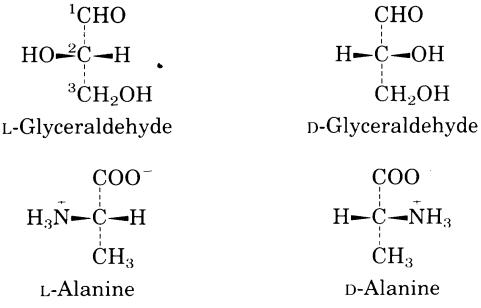


Figure 5–4 Steric relationship of the stereoisomers of alanine to the absolute configuration of L- and D-glyceraldehyde. In these perspective formulas, the carbons are lined up vertically, with the chiral atom in the center. The carbons in these molecules are numbered beginning with the aldehyde or carboxyl carbons on the end, or 1 to 3 from top to bottom as shown. When presented in this way, the R group of the amino acid (in this case the methyl group of alanine) is always below the α carbon. L-Amino acids are those with the α-amino group on the left, and D-amino acids have the α-amino group on the right.

Proteins Contain L-Amino Acids

Nearly all biological compounds with a chiral center occur naturally in only one stereoisomeric form, either D or L. The amino acids in protein molecules are the L stereoisomers. D-Amino acids have been found only in small peptides of bacterial cell walls and in some peptide antibiotics (see Fig. 5–19).

It is remarkable that the amino acids of proteins are all L stereoisomers. As we noted in Chapter 3, when chiral compounds are formed by ordinary chemical reactions, a racemic mixture of D and L isomers results. Whereas the L and D forms of chiral molecules are difficult for a chemist to distinguish and isolate, they are as different as night and day to a living system. The ability of cells to specifically synthesize the L isomer of amino acids reflects one of many extraordinary properties of enzymes (Chapter 8). The stereospecificity of the reactions catalyzed by some enzymes is made possible by the asymmetry of their active sites. The characteristic three-dimensional structures of proteins (Chapter 7), which dictate their diverse biological activities, require that *all* their constituent amino acids be of one stereochemical series.

Amino Acids Are Ionized in Aqueous Solutions

Amino acids in aqueous solution are ionized and can act as acids or bases. Knowledge of the acid–base properties of amino acids is extremely important in understanding the physical and biological properties of proteins. Moreover, the technology of separating, identifying, and quantifying the different amino acids, which are necessary steps in determining the amino acid composition and sequence of protein molecules, is based largely on their characteristic acid–base behavior.

Those α-amino acids having a single amino group and a single carboxyl group crystallize from neutral aqueous solutions as fully ionized species known as **zwitterions** (German for "hybrid ions"), each having both a positive and a negative charge (Fig. 5–5). These ions are electrically neutral and remain stationary in an electric field. The dipolar nature of amino acids was first suggested by the observation that crystalline amino acids have melting points much higher than those of other organic molecules of similar size. The crystal lattice of amino acids is held together by strong electrostatic forces between positively and negatively charged functional groups of neighboring molecules, resembling the stable ionic crystal lattice of NaCl (see Fig. 4–6).

Figure 5–5 Nonionic and zwitterionic forms of amino acids. Note the separation of the + and − charges in the zwitterion, which makes it an electric dipole. The nonionic form does not occur in significant amounts in aqueous solutions. The zwitterion predominates at neutral pH.

Amino Acids Can Be Classified by R Group

An understanding of the chemical properties of the standard amino acids is central to an understanding of much of biochemistry. The topic can be simplified by grouping the amino acids into classes based on the properties of their R groups (Table 5–1), in particular, their **polarity** or tendency to interact with water at biological pH (near pH 7.0). The polarity of the R groups varies widely, from totally nonpolar or hydrophobic (water-insoluble) to highly polar or hydrophilic (water-soluble).

The structures of the 20 standard amino acids are shown in Figure 5–6, and many of their properties are listed in Table 5–1. There are five main classes of amino acids, those whose R groups are: nonpolar and aliphatic; aromatic (generally nonpolar); polar but uncharged; negatively charged; and positively charged. Within each class there are gradations of polarity, size, and shape of the R groups.

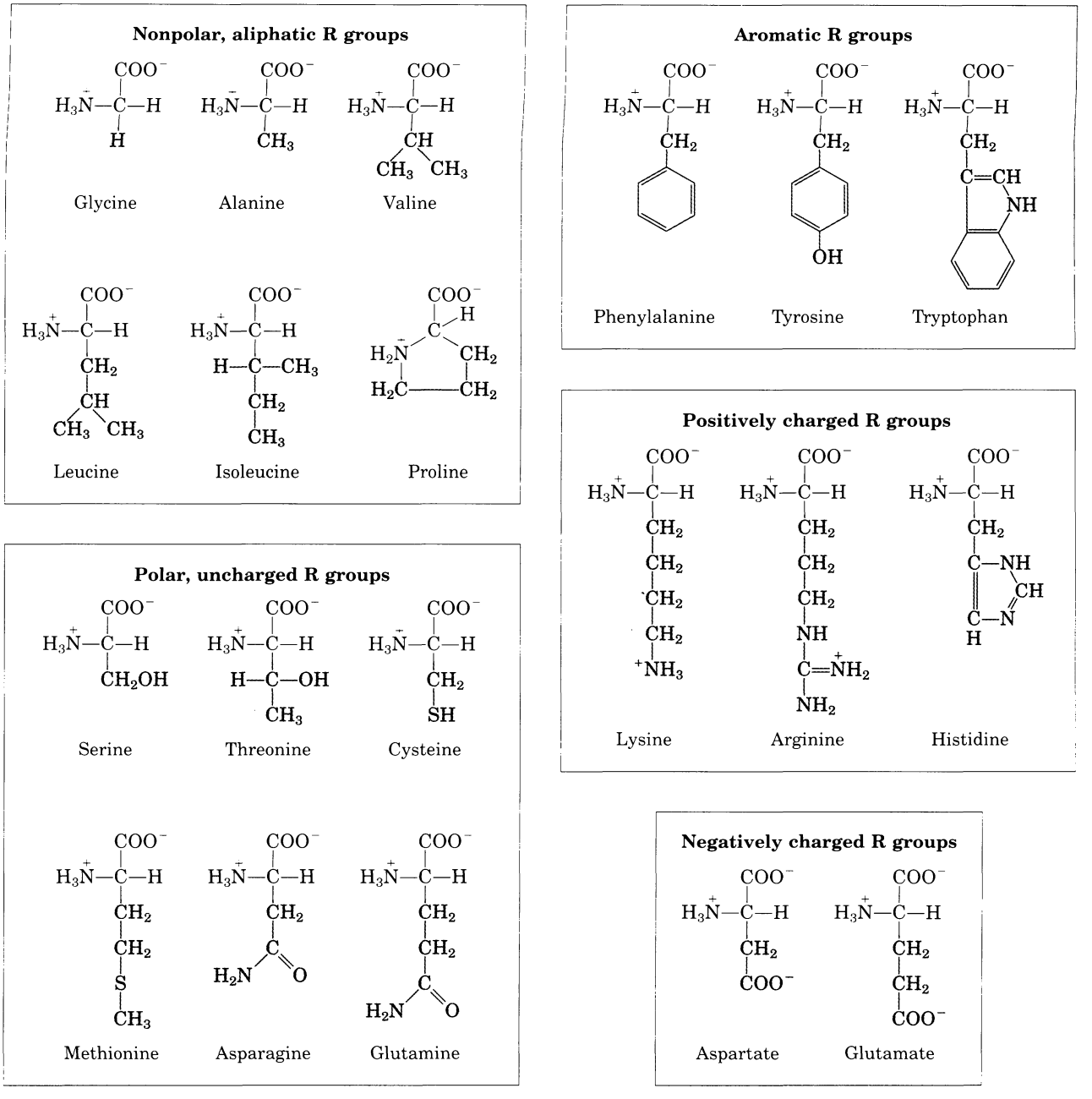


Figure 5–6 The 20 standard amino acids of proteins. They are shown with their amino and carboxyl groups ionized, as they would occur at pH 7.0. The portions in black are those common to all the amino acids; the portions shaded in red are the R groups.

Nonpolar, Aliphatic R Groups The hydrocarbon R groups in this class of amino acids are nonpolar and hydrophobic (Fig. 5–6). The bulky side chains of **alanine, valine, leucine,** and **isoleucine,** with their distinctive shapes, are important in promoting hydrophobic interactions within protein structures. **Glycine** has the simplest amino acid structure. Where it is present in a protein, the minimal steric hindrance of the glycine side chain allows much more structural flexibility than the other amino acids. **Proline** represents the opposite structural extreme. The secondary amino (imino) group is held in a rigid conformation that reduces the structural flexibility of the protein at that point.

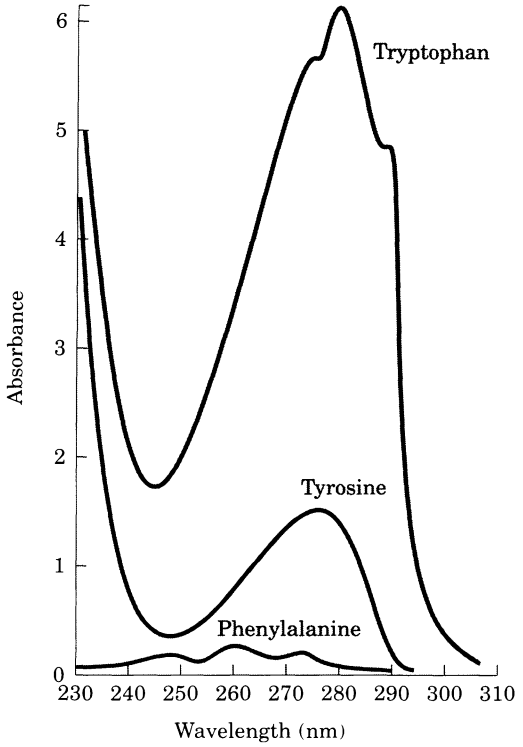


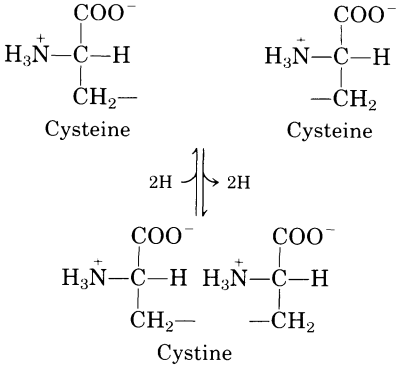
Figure 5–7 Comparison of the light absorbance spectra of the aromatic amino acids at pH 6.0. The amino acids are present in equimolar amounts (10^{-3} M) under identical conditions. The light absorbance of tryptophan is as much as fourfold higher than that of tyrosine. Phenylalanine absorbs less light than either tryptophan or tyrosine. Note that the absorbance maximum for tryptophan and tyrosine occurs near a wavelength of 280 nm.

Aromatic R Groups **Phenylalanine, tyrosine,** and **tryptophan,** with their aromatic side chains (Fig. 5–6), are relatively nonpolar (hydrophobic). All can participate in hydrophobic interactions, which are particularly strong when the aromatic groups are stacked on one another. The hydroxyl group of tyrosine can form hydrogen bonds, and it acts as an important functional group in the activity of some enzymes. Tyrosine and tryptophan are significantly more polar than phenylalanine because of the tyrosine hydroxyl group and the nitrogen of the tryptophan indole ring.

Tryptophan and tyrosine, and to a lesser extent phenylalanine, absorb ultraviolet light (Fig. 5–7 and Box 5–1). This accounts for the characteristic strong absorbance of light by proteins at a wavelength of 280 nm, and is a property exploited by researchers in the characterization of proteins.

Polar, Uncharged R Groups The R groups of these amino acids (Fig. 5–6) are more soluble in water, or hydrophilic, than those of the nonpolar amino acids, because they contain functional groups that form hydrogen bonds with water. This class of amino acids includes **serine, threonine, cysteine, methionine, asparagine,** and **glutamine.** The polarity of serine and threonine is contributed by their hydroxyl groups; that of cysteine and methionine by their sulfur atom; and that of asparagine and glutamine by their amide groups.

Asparagine and glutamine are the amides of two other amino acids also found in proteins, aspartate and glutamate, respectively, to which asparagine and glutamine are easily hydrolyzed by acid or base. Cysteine has an R group (a thiol group) that is approximately as acidic as the hydroxyl group of tyrosine. Cysteine requires special mention for another reason. It is readily oxidized to form a covalently linked dimeric amino acid called **cystine,** in which two cysteine molecules are joined by a disulfide bridge. Disulfide bridges of this kind occur in many proteins, stabilizing their structures.



Negatively Charged (Acidic) R Groups The two amino acids having R groups with a net negative charge at pH 7.0 are **aspartate** and **glutamate,** each with a second carboxyl group (Fig. 5–6). These amino acids are the parent compounds of asparagine and glutamine, respectively.

Positively Charged (Basic) R Groups The amino acids in which the R groups have a net positive charge at pH 7.0 are **lysine,** which has a second amino group at the ϵ position on its aliphatic chain; **arginine,** which has a positively charged guanidino group; and **histidine,** containing an imidazole group (Fig. 5–6). Histidine is the only standard amino acid having a side chain with a pK_a near neutrality.

BOX 5–1 Absorption of Light by Molecules: The Lambert–Beer Law

Measurement of light absorption is an important tool for analysis of many biological molecules. The fraction of the incident light absorbed by a solution at a given wavelength is related to the thickness of the absorbing layer (path length) and the concentration of the absorbing species. These two relationships are combined into the Lambert–Beer law, given in integrated form as

$$\log \frac{I_0}{I} = \epsilon c l$$

where I_0 is the intensity of the incident light, I is the intensity of the transmitted light, ϵ is the molar absorption coefficient (in units of liters per mole-centimeter), c the concentration of the absorbing species (in moles per liter), and l the path length of the light-absorbing sample (in centimeters). The Lambert–Beer law assumes that the incident light is parallel and monochromatic and that the solvent and solute molecules are randomly oriented. The expression $\log (I_0/I)$ is called the absorbance, designated A.

It is important to note that each millimeter path length of absorbing solution in a 1.0 cm cell absorbs not a constant amount but a constant fraction of the incident light. However, with an absorbing layer of fixed path length, *the absorbance A is directly proportional to the concentration of the absorbing solute.*

The molar absorption coefficient varies with the nature of the absorbing compound, the solvent, the wavelength, and also with pH if the light-absorbing species is in equilibrium with another species having a different spectrum through gain or loss of protons.

In practice, absorbance measurements are usually made on a set of standard solutions of known concentration at a fixed wavelength. A sample of unknown concentration can then be compared with the resulting standard curve, as shown in Figure 1.

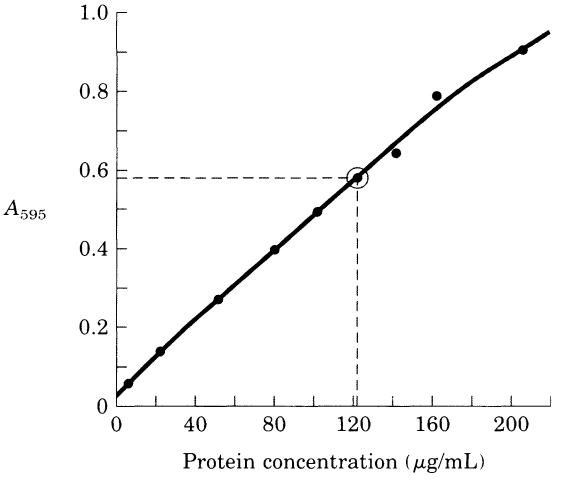
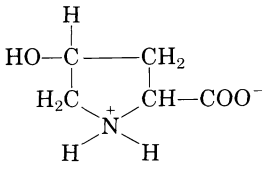


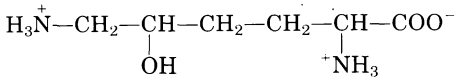
Figure 1 Eight standard solutions containing known amounts of protein and one sample containing an unknown amount of protein were reacted with the Bradford reagent. This reagent contains a dye that shifts its absorption maximum to 595 nm when it binds amino acid residues. The A_{595} (absorbance at 595 nm) of the standard samples was plotted against the protein concentration to create the standard curve, shown here. The A_{595} of the unknown sample, 0.58, corresponds to a protein concentration of 122 μg/mL.

Cells Also Contain Nonstandard Amino Acids

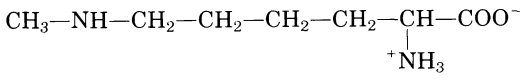
In addition to the 20 standard amino acids that are common in all proteins, other amino acids have been found as components of only certain types of proteins (Fig. 5–8a). Each of these is derived from one of the 20 standard amino acids, in a modification reaction that occurs after the standard amino acid has been inserted into a protein. Among the nonstandard amino acids are **4-hydroxyproline,** a derivative of proline, and **5-hydroxylysine;** the former is found in plant cell-wall proteins, and both are found in the fibrous protein collagen of connective tissues. ***N*-Methyllysine** is found in myosin, a contractile protein of muscle. Another important nonstandard amino acid is



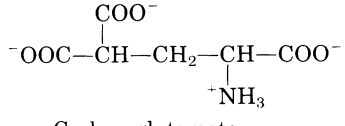
4-Hydroxyproline



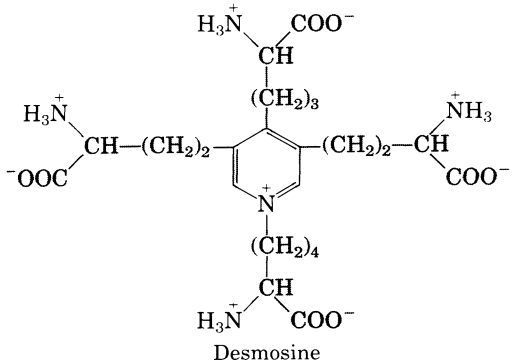
5-Hydroxylysine



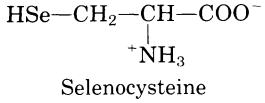
6-*N*-Methyllysine



γ-Carboxyglutamate

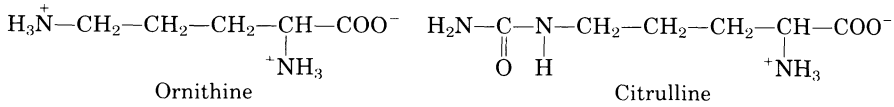


Desmosine



Selenocysteine

(a)



Ornithine Citrulline

(b)

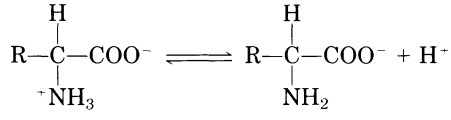
Figure 5–8 (a) Some nonstandard amino acids found in proteins; all are derived from standard amino acids. The extra functional groups are shown in red. Desmosine is formed from four residues of lysine, whose carbon backbones are shaded in gray. Selenocysteine is derived from serine. **(b)** Ornithine and citrulline are intermediates in the biosynthesis of arginine and in the urea cycle. Note that two systems are used to number carbons in the naming of these amino acids. The α, β, γ system used for γ-carboxyglutamate begins at the α carbon (see Fig. 5–2) and extends into the R group. The α-carboxyl group is not included. In contrast, the numbering system used to identify the modified carbon in 4-hydroxyproline, 5-hydroxylysine, and 6-*N*-methyllysine includes the α-carboxyl carbon, which is designated carbon 1 (*or C-1*).

γ-carboxyglutamate, found in the blood-clotting protein prothrombin as well as in certain other proteins that bind Ca^{2+} in their biological function. More complicated is the nonstandard amino acid **desmosine,** a derivative of four separate lysine residues, found in the fibrous protein elastin. **Selenocysteine** contains selenium rather than the oxygen of serine, and is found in glutathione peroxidase and a few other proteins.

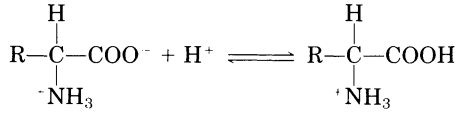
Some 300 additional amino acids have been found in cells and have a variety of functions but are not substituents of proteins. **Ornithine** and **citrulline** (Fig. 5–8b) deserve special note because they are key intermediates in the biosynthesis of arginine and in the urea cycle. These pathways are described in Chapters 21 and 17, respectively.

Amino Acids Can Act as Acids and as Bases

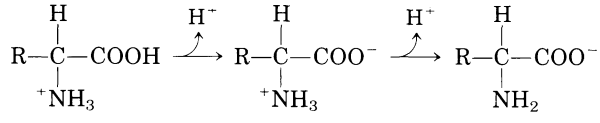
When a crystalline amino acid, such as alanine, is dissolved in water, it exists in solution as the dipolar ion, or zwitterion, which can act either as an acid (proton donor):



or as a base (proton acceptor):



Substances having this dual nature are **amphoteric** and are often called **ampholytes,** from "amphoteric electrolytes." A simple monoamino monocarboxylic α-amino acid, such as alanine, is actually a diprotic acid when it is fully protonated, that is, when both its carboxyl group and amino group have accepted protons. In this form it has two groups that can ionize to yield protons, as indicated in the following equation:



Amino Acids Have Characteristic Titration Curves

Titration involves the gradual addition or removal of protons. Figure 5–9 shows the titration curve of the diprotic form of glycine. Each molecule of added base results in the net removal of one proton from

one molecule of amino acid. The plot has two distinct stages, each corresponding to the removal of one proton from glycine. Each of the two stages resembles in shape the titration curve of a monoprotic acid, such as acetic acid (see Fig. 4–10), and can be analyzed in the same way. At very low pH, the predominant ionic species of glycine is $^{+}H_3N—CH_2—COOH$, the fully protonated form. At the midpoint in the first stage of the titration, in which the —COOH group of glycine loses its proton, equimolar concentrations of proton-donor ($^{+}H_3N—CH_2—COOH$) and proton-acceptor ($^{+}H_3N—CH_2—COO^{-}$) species are present. At the midpoint of a titration (see Fig. 4–11), the pH is equal to the pK_a of the protonated group being titrated. For glycine, the pH at the midpoint is 2.34, thus its —COOH group has a pK_a of 2.34. [Recall that pH and pK_a are simply convenient notations for proton concentration and the equilibrium constant for ionization, respectively (Chapter 4). The pK_a is a measure of the tendency of a group to give up a proton, with that tendency decreasing tenfold as the pK_a increases by one unit.] As the titration proceeds, another important point is reached at pH 5.97. Here there is a point of inflection, at which removal of the first proton is essentially complete, and removal of the second has just begun. At this pH the glycine is present largely as the dipolar ion $^{+}H_3N—CH_2—COO^{-}$. We shall return to the significance of this inflection point in the titration curve shortly.

The second stage of the titration corresponds to the removal of a proton from the $—NH_3^+$ group of glycine. The pH at the midpoint of this stage is 9.60, equal to the pK_a for the $—NH_3^+$ group. The titration is complete at a pH of about 12, at which point the predominant form of glycine is $H_2N—CH_2—COO^{-}$.

From the titration curve of glycine we can derive several important pieces of information. First, it gives a quantitative measure of the pK_a of each of the two ionizing groups, 2.34 for the —COOH group and 9.60 for the $—NH_3^+$ group. Note that the carboxyl group of glycine is over 100 times more acidic (more easily ionized) than the carboxyl group of acetic acid, which has a pK_a of 4.76. This effect is caused by the nearby positively charged amino group on the α-carbon atom, as described in Figure 5–10.

The second piece of information given by the titration curve of glycine (Fig. 5–9) is that this amino acid has *two* regions of buffering power (see Fig. 4–12). One of these is the relatively flat portion of the curve centered about the first pK_a of 2.34, indicating that glycine is a good buffer near this pH. The other buffering zone extends for ~1.2 pH units centered around pH 9.60. Note also that glycine is not a good buffer at the pH of intracellular fluid or blood, about 7.4.

The Henderson–Hasselbalch equation (Chapter 4) can be used to calculate the proportions of proton-donor and proton-acceptor species of glycine required to make a buffer at a given pH within the buffering ranges of glycine; it also makes it possible to solve other kinds of buffer problems involving amino acids (see Box 4–2).

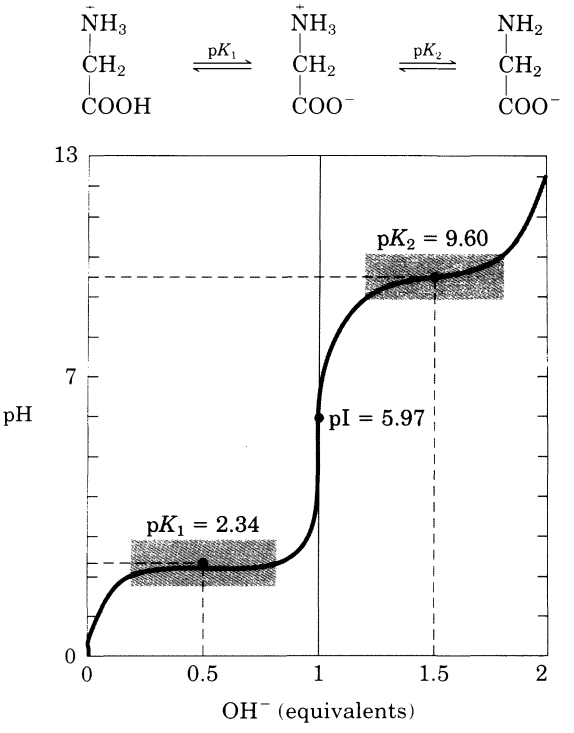
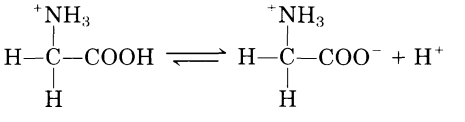


Figure 5–9 The titration curve of 0.1 M glycine at 25 °C. The ionic species predominating at key points in the titration are shown above the graph. The shaded boxes, centered about $pK_1 = 2.34$ and $pK_2 = 9.60$, indicate the regions of greatest buffering power.

Figure 5–10 **(a)** Interactions between the α-amino and α-carboxyl groups in an α-amino acid. The nearby positive charge of the $—NH_3^+$ group makes ionization of the carboxyl group more likely (i.e., lowers the pK_a for —COOH). This is due to a stabilizing interaction between opposite charges on the zwitterion and a repulsive interaction between the positive charges of the amino group and the departing proton. **(b)** The normal pK_a for a carboxyl group is approximately 4.76, as for acetic acid.

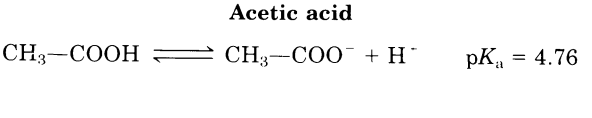
α-Amino acid (glycine)



$pK_a = 2.34$

(a)

Acetic acid



$pK_a = 4.76$

(b)

The Titration Curve Predicts the Electric Charge of Amino Acids

Another important piece of information derived from the titration curve of an amino acid is the relationship between its net electric charge and the pH of the solution. At pH 5.97, the point of inflection between the two stages in its titration curve, glycine is present as its dipolar form, fully ionized but with no *net* electric charge (Fig. 5–9). This characteristic pH is called the **isoelectric point** or **isoelectric pH,** designated **pI** or pH_I. For an amino acid such as glycine, which has no ionizable group in the side chain, the isoelectric point is the arithmetic mean of the two pK_a values:

$$pI = \frac{1}{2}(pK_1 + pK_2)$$

which in the case of glycine is

$$pI = \frac{1}{2}(2.34 + 9.60) = 5.97$$

As is evident in Figure 5–9, glycine has a net negative charge at any pH above its pI and will thus move toward the positive electrode (the anode) when placed in an electric field. At any pH below its pI, glycine has a net positive charge and will move toward the negative electrode, the cathode. The farther the pH of a glycine solution is from its isoelectric point, the greater the net electric charge of the population of glycine molecules. At pH 1.0, for example, glycine exists entirely as the form $^+H_3N—CH_2—COOH$, with a net positive charge of 1.0. At pH 2.34, where there is an equal mixture of $^+H_3N—CH_2—COOH$ and $^+H_3N—CH_2—COO^-$, the average or net positive charge is 0.5. The sign and the magnitude of the net charge of any amino acid at any pH can be predicted in the same way.

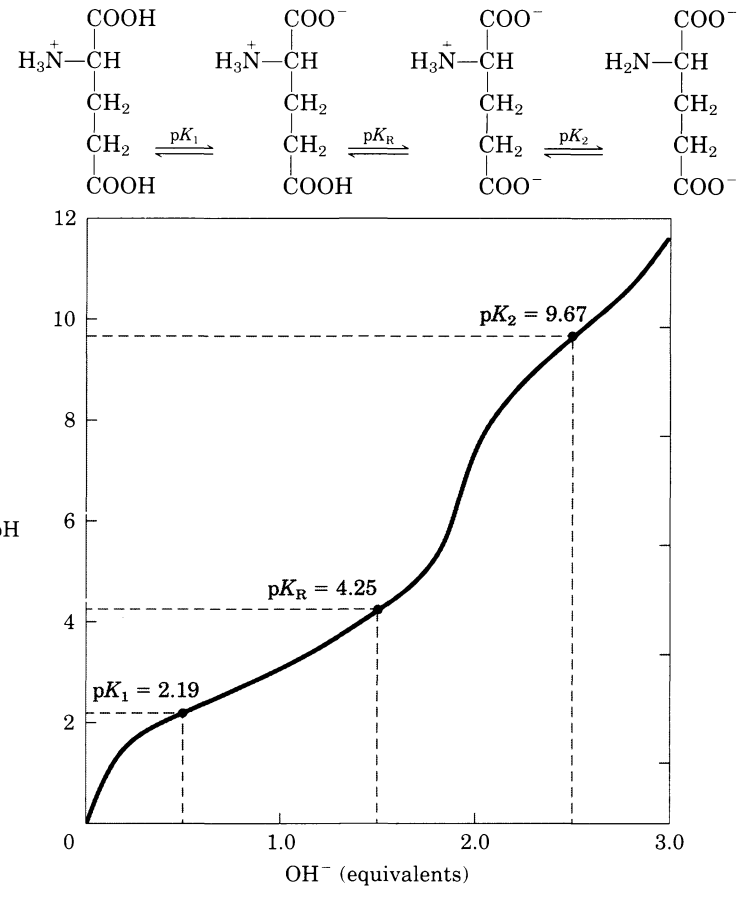
This information has practical importance. For a solution containing a mixture of amino acids, the different amino acids can be separated on the basis of the direction and relative rate of their migration when placed in an electric field at a known pH.

Amino Acids Differ in Their Acid–Base Properties

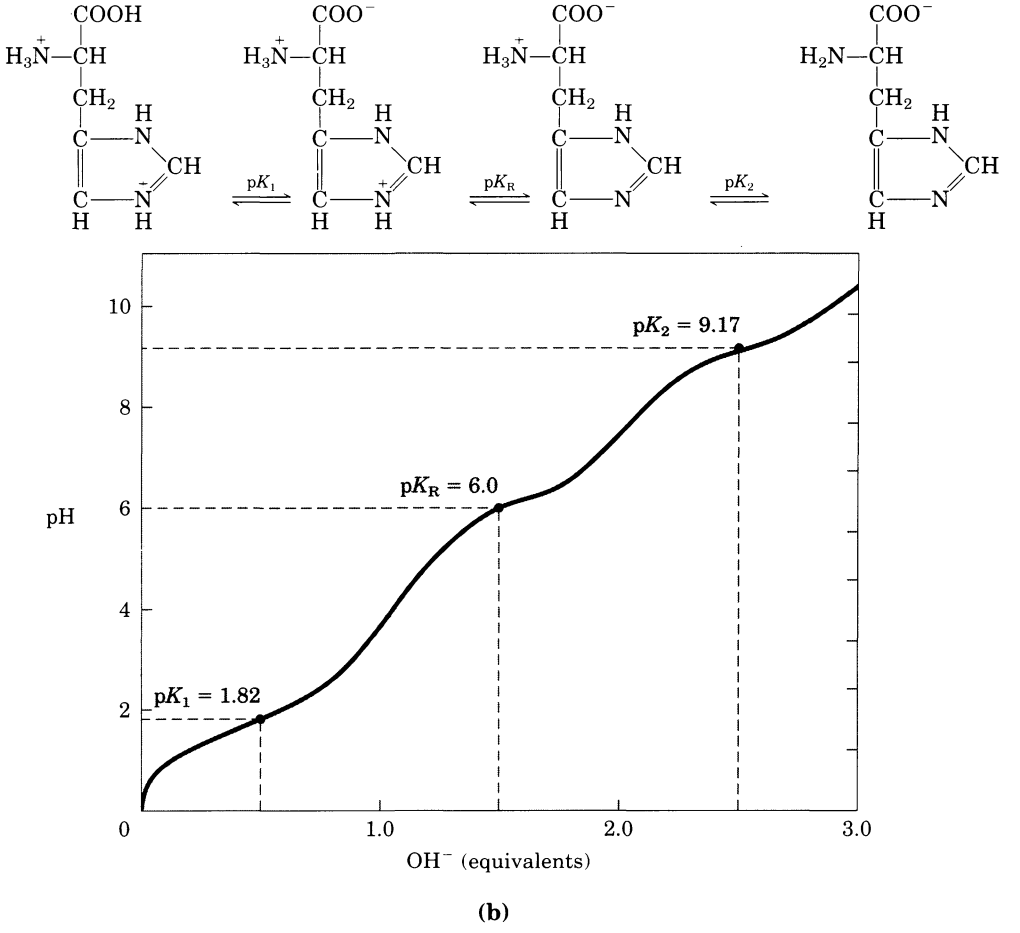
The shared properties of many amino acids permit some simplifying generalizations about the acid–base behavior of different classes of amino acids.

All amino acids with a single α-amino group, a single α-carboxyl group, and an R group that does not ionize have titration curves resembling that of glycine (Fig. 5–9). This group of amino acids is characterized by having very similar, although not identical, values for pK_1 (the pK of the —COOH group) in the range of 1.8 to 2.4 and for pK_2 (of the $—NH_3^+$ group) in the range of 8.8 to 11.0 (Table 5–1).

Amino acids with an ionizable R group (Table 5–1) have more complex titration curves with *three* stages corresponding to the three possible ionization steps; thus they have three pK_a values. The third stage for the titration of the ionizable R group merges to some extent with the others. The titration curves of two representatives of this group, glutamate and histidine, are shown in Figure 5–11. The isoelectric points of amino acids in this class reflect the type of ionizing R groups present. For example, glutamate has a pI of 3.22, considerably lower than that of glycine. This is a result of the presence of two carboxyl



(a)



(b)

Figure 5–11 The titration curves of **(a)** glutamate and **(b)** histidine. The pK_a of the R group is designated pK_R.

groups which, at the average of their pK_a values (3.22), contribute a net negative charge of −1 that balances the +1 contributed by the amino group. Similarly, the pI of histidine, with two groups that are positively charged when protonated, is 7.59 (the average of the pK_a values of the amino and imidazole groups), much higher than that of glycine.

Another important generalization can be made about the acid–base behavior of the 20 standard amino acids. Under the general condition of free and open exposure to the aqueous environment, only histidine has an R group (pK_a = 6.0) providing significant buffering power near the neutral pH usually found in the intracellular and intercellular fluids of most animals and bacteria. All the other amino acids have pK_a values too far away from pH 7 to be effective physiological buffers (Table 5–1), although in the interior of proteins the pK_a values of amino acid side chains are often altered.

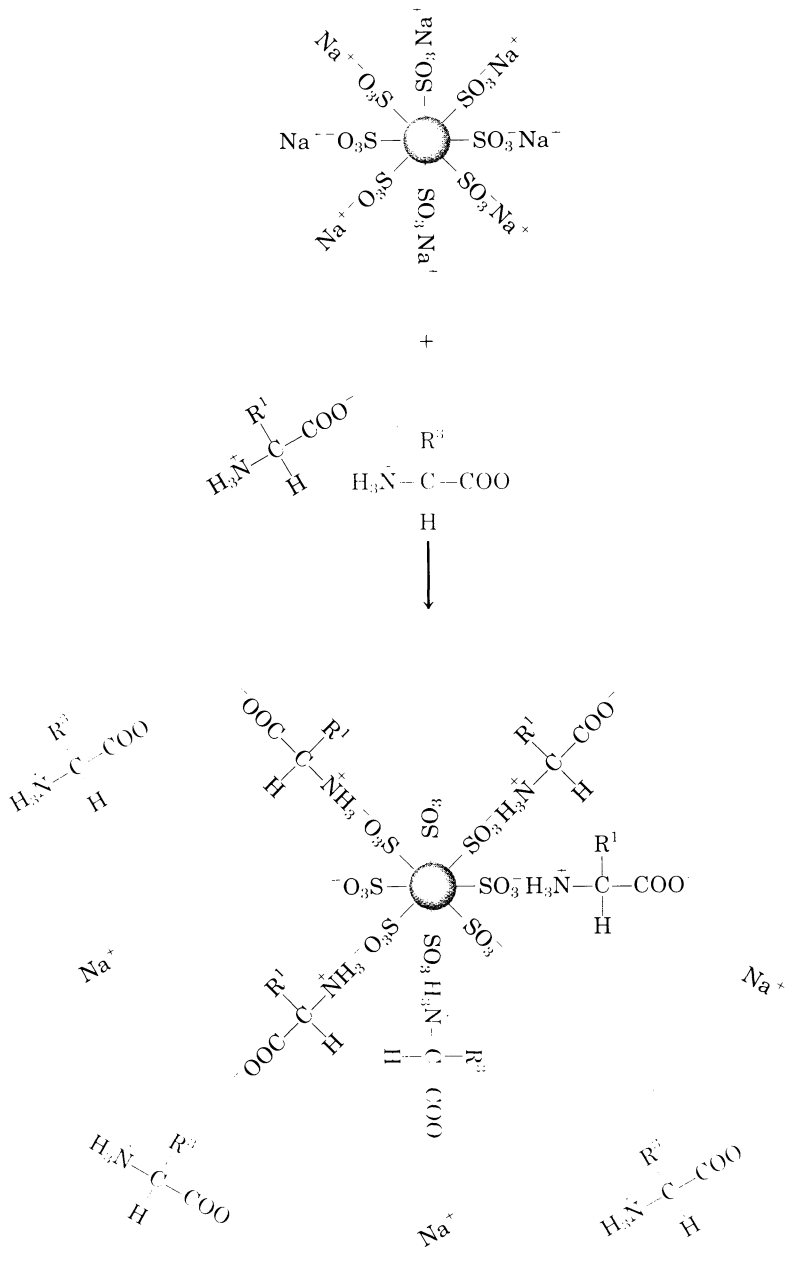
Ion-Exchange Chromatography Separates Amino Acids by Electric Charge

Ion-exchange chromatography is the most widely used method for separating, identifying, and quantifying the amounts of each amino acid in a mixture. This technique primarily exploits differences in the sign and magnitude of the net electric charges of amino acids at a given pH, which are predictable from their pK_a values or their titration curves.

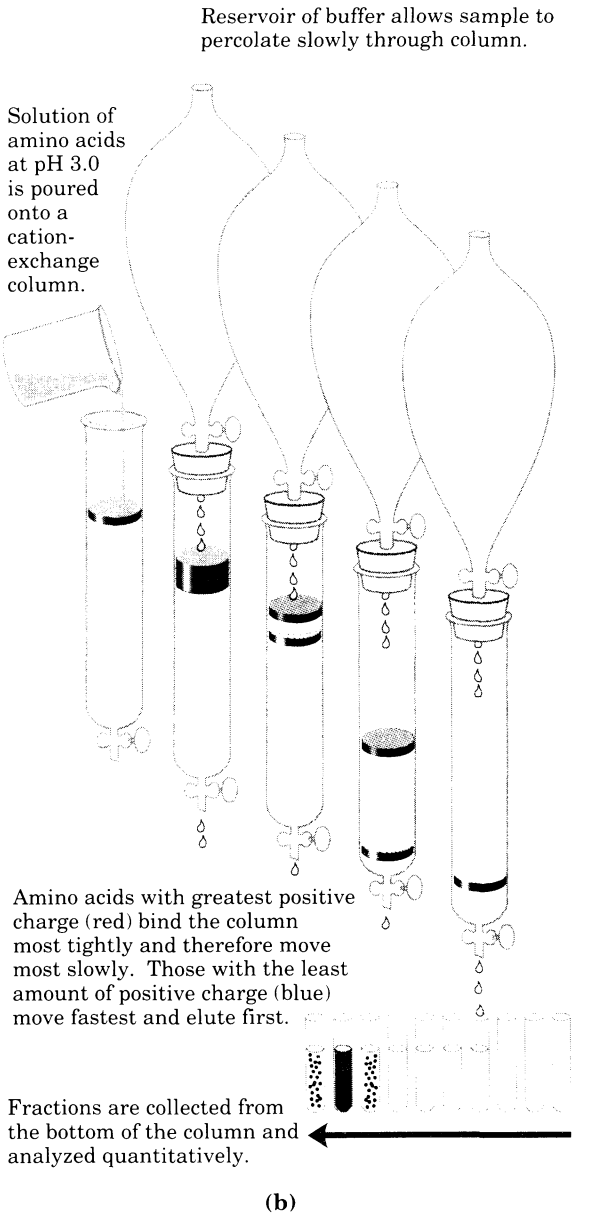
The chromatographic column consists of a long tube filled with particles of a synthetic resin containing fixed charged groups; those with fixed anionic groups are called **cation-exchange** resins and those with fixed cationic groups, **anion-exchange** resins. A simple form of ion-exchange chromatography on a cation-exchange resin is described in Figure 5–12. The affinity of each amino acid for the resin is affected by pH (which determines the ionization state of the molecule) and the concentration of other salt ions that may compete with the resin by associating with the amino acid. Separation of amino acids can therefore be optimized by gradually changing the pH and/or salt concentration of the solution being passed through the column so as to create a pH or salt gradient. A modern enhancement of this and other chromatographic techniques is called **high-performance liquid chromatography** (HPLC). This takes advantage of stronger resin material and improved apparatus designed to permit chromatography at high pressures, allowing better separations in a much shorter time. For amino acids, the entire procedure has been automated, so that elution, collection of fractions, analysis of each fraction, and recording of data are performed automatically in an **amino-acid analyzer.** Figure 5–13 shows a chromatogram of an amino acid mixture analyzed in this way.

Amino Acids Undergo Characteristic Chemical Reactions

As for all organic compounds, the chemical reactions of amino acids are those characteristic of their functional groups. Because all amino acids contain amino and carboxyl groups, all will undergo chemical reactions characteristic for these groups. For example, their amino groups can be acetylated or formylated, and their carboxyl groups can be esterified. We will not examine all such organic reactions of amino acids, but several widely used reactions are noteworthy because they greatly simplify the detection, measurement, and identification of amino acids.



(a)



(b)

Figure 5–12 Ion-exchange chromatography. An example of a cation-exchange resin is presented. **(a)** Negatively charged sulfonate groups ($-SO_3^-$) on the resin surface attract and bind cations, such as H^+, Na^+, or cationic forms of amino acids. **(b)** An acidic solution (pH 3.0) of the amino acid mixture is poured on a column packed with resin and allowed to percolate through slowly. At pH 3.0 the amino acids are largely cations with net positive charges, but they differ in the pK_a values of their R groups, and hence in the extent to which they are ionized and in their tendency to bind to the anionic resin. As a result, they move through the column at different rates.

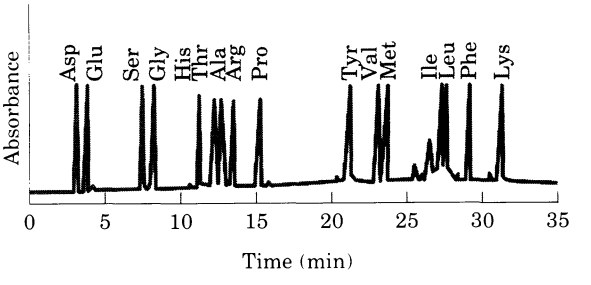
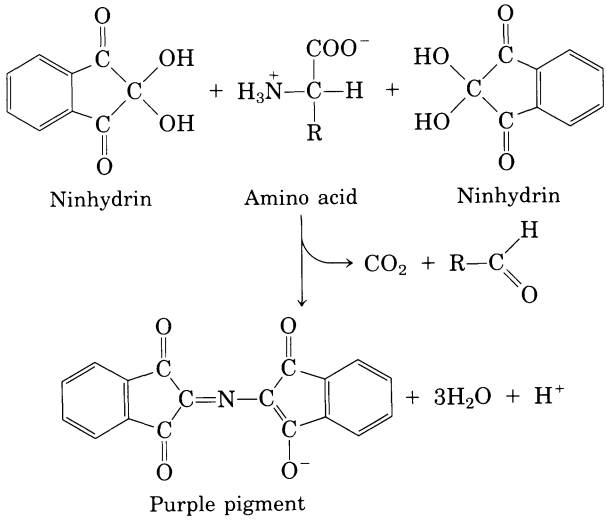


Figure 5–13 Automatically recorded high-performance liquid chromatographic analysis of amino acids on a cation-exchange resin. The area under each peak on the chromatogram is proportional to the amount of each amino acid in the mixture.



One of the most important, technically and historically, is the ninhydrin reaction, which has been used for many years to detect and quantify microgram amounts of amino acids. When amino acids are heated with excess ninhydrin, all those having a free α-amino group yield a purple product. Proline, in which the α-amino group is substituted (forming an imino group), yields a yellow product. Under appropriate conditions the intensity of color produced (optical absorbance of the solution; see Box 5–1) is proportional to the amino acid concentration. Comparing the absorbance to that of appropriate standard solutions is an accurate and technically simple method for measuring amino acid concentration.

Several other convenient reagents are available that react with the α-amino group to form colored or fluorescent derivatives. Unlike ninhydrin, these have the advantage that the intact R group of the amino acid remains part of the product, so that derivatives of different amino acids can be distinguished. Fluorescamine reacts rapidly with amino acids and provides great sensitivity, yielding a highly fluorescent derivative that permits the detection of nanogram quantities of amino acids (Fig. 5–14). Dabsyl chloride, dansyl chloride, and 1-fluoro-2,4-dinitrobenzene (Fig. 5–14) yield derivatives that are stable under harsh conditions such as those used in the hydrolysis of proteins.

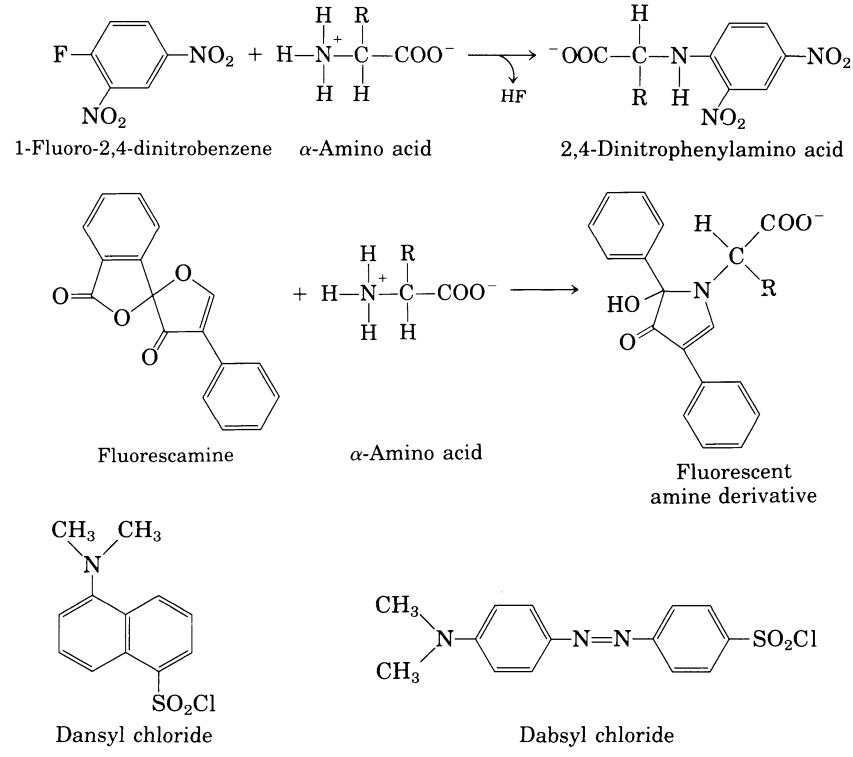


Figure 5–14 Reagents that react with the α-amino group of amino acids. The reactions producing 2,4-dinitrophenyl and fluorescamine derivatives are illustrated. The reactions of dansyl chloride and dabsyl chloride are similar to that of 1-fluoro-2,4-dinitrobenzene (Sanger's reagent). Because the derivatives of these reagents absorb light, they greatly facilitate the detection and quantification of the amino acids.

Peptides

We now turn to polymers of amino acids, the **peptides.** Biologically occurring peptides range in size from small molecules containing only two or three amino acids to macromolecules containing thousands of amino acids. The focus here is on the structure and chemical properties of the smaller peptides, providing a prelude to the discussion of the large peptides called proteins in the next two chapters.

Peptides Are Chains of Amino Acids

Two amino acid molecules can be covalently joined through a substituted amide linkage, termed a **peptide bond,** to yield a dipeptide. Such a linkage is formed by removal of the elements of water from the α-carboxyl group of one amino acid and the α-amino group of another (Fig. 5–15). Peptide-bond formation is an example of a condensation reaction, a common class of reaction in living cells. Note that as shown in Figure 5–15, this reaction has an equilibrium that favors reactants rather than products. To make the reaction thermodynamically more favorable, the carboxyl group must be chemically modified or activated so that the hydroxyl group can be more readily eliminated. A chemical approach to this problem is outlined at the end of this chapter (see Box 5–2). The biological approach to peptide bond formation is a major topic of Chapter 26.

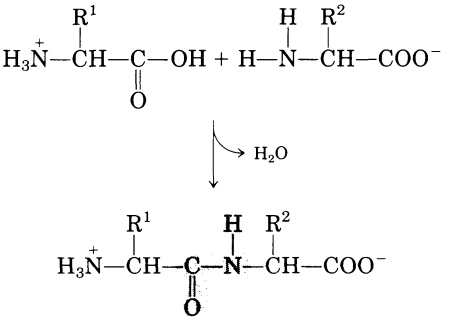


Figure 5–15 Formation of a peptide bond (shaded in gray) in a dipeptide. This is a condensation reaction. The α-amino group of amino acid 2 acts as a nucleophile (see Table 3–6) to displace the hydroxyl group of amino acid 1 (red). Amino groups are good nucleophiles, but the hydroxyl group is a poor leaving group and is not readily displaced. At physiological pH the reaction as shown does not occur to any appreciable extent. Peptide bond formation is endergonic, with a free-energy change of about +21 kJ/mol.

Three amino acids can be joined by two peptide bonds to form a tripeptide; similarly, amino acids can be linked to form tetrapeptides and pentapeptides. When a few amino acids are joined in this fashion, the structure is called an **oligopeptide.** When many amino acids are joined, the product is called a **polypeptide.** Proteins may have thousands of amino acid units. Although the terms "protein" and "polypeptide" are sometimes used interchangeably, molecules referred to as polypeptides generally have molecular weights below 10,000.

Figure 5–16 shows the structure of a pentapeptide. The amino acid units in a peptide are often called **residues** (each has lost a hydrogen atom from its amino group and a hydroxyl moiety from its carboxyl group). The amino acid residue at that end of a peptide having a free α-amino group is the **amino-terminal** (or N-terminal) residue; the residue at the other end, which has a free carboxyl group, is the **carboxyl-terminal** (C-terminal) residue. By convention, short peptides are named from the sequence of their constituent amino acids, beginning at the left with the amino-terminal residue and proceeding toward the carboxyl terminus at the right (Fig. 5–16).

Although hydrolysis of peptide bonds is an exergonic reaction, it occurs slowly because of its high activation energy. As a result, the peptide bonds in proteins are quite stable under most intracellular conditions.

The peptide bond is the single most important covalent bond linking amino acids in peptides and proteins. The only other type of covalent bond that occurs frequently enough to deserve special mention here is the disulfide bond sometimes formed between two cysteine residues. Disulfide bonds play a special role in the structure of many proteins, particularly those that function extracellularly, such as the hormone insulin and the immunoglobulins or antibodies.

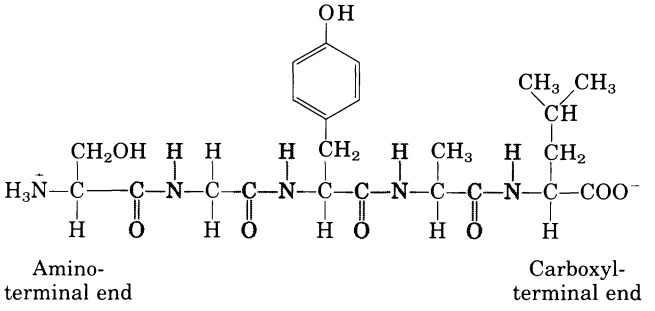
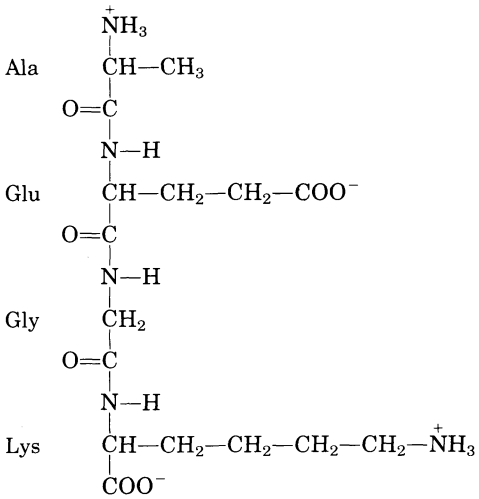


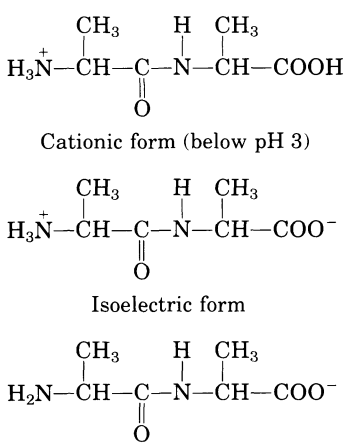
Figure 5–16 Structure of the pentapeptide seryl-glycyltyrosinylalanylleucine, or Ser–Gly–Tyr–Ala–Leu. Peptides are named beginning with the amino-terminal residue, which by convention is placed at the left. The peptide bonds are shown shaded in gray, the R groups in red.

Alanylglutamylglycyllysine



(a)

Alanylalanine



(b)

Figure 5–17 Ionization and electric charge of peptides. The groups ionized at pH 7.0 are in red. **(a)** A tetrapeptide with two ionizable R groups. **(b)** The cationic, isoelectric, and anionic forms of a dipeptide lacking ionizable R groups.

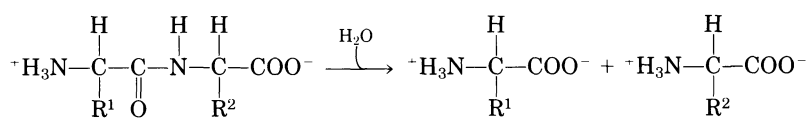
Peptides Can Be Distinguished by Their Ionization Behavior

Peptides contain only one free α-amino group and one free α-carboxyl group (Fig. 5–17). These groups ionize as they do in simple amino acids, although the ionization constants are different because the oppositely charged group is absent from the α carbon. The α-amino and α-carboxyl groups of all other constituent amino acids are covalently joined in the form of peptide bonds, which do not ionize and thus do not contribute to the total acid–base behavior of peptides. However, the R groups of some amino acids can ionize (Table 5–1), and in a peptide these contribute to the overall acid–base properties (Fig. 5–17). Thus the acid–base behavior of a peptide can be predicted from its single free α-amino and α-carboxyl groups and the nature and number of its ionizable R groups. Like free amino acids, peptides have characteristic titration curves and a characteristic isoelectric pH at which they do not move in an electric field. These properties are exploited in some of the techniques used to separate peptides and proteins (Chapter 6).

Peptides Undergo Characteristic Chemical Reactions

Like other organic molecules, peptides undergo chemical reactions that are characteristic of their functional groups: the free amino and carboxyl groups and the R groups.

Peptide bonds can be hydrolyzed by boiling with either strong acid (typically 6 M HCl) or base to yield the constituent amino acids.



Hydrolysis of peptide bonds in this manner is a necessary step in determining the amino acid composition of proteins. The reagents shown in Figure 5–14 label only free amino groups: those of the amino-terminal residue and the R groups of any lysines present. If dabsyl chloride, dansyl chloride, or 1-fluoro-2,4-dinitrobenzene is used before acid hydrolysis of the peptide, the amino-terminal residue can be separated and identified (Fig. 5–18).

Peptide bonds can also be hydrolyzed by certain enzymes called **proteases.** Proteolytic (protein-cleaving) enzymes are found in all cells and tissues, where they degrade unneeded or damaged proteins or aid in the digestion of food.

Some Small Polypeptides Have Biological Activity

Much of the material in the chapters to follow will revolve around the activities of proteins with molecular weights measured in the tens and even hundreds of thousands. Not all polypeptides are so large, however. There are many naturally occurring small polypeptides and oligopeptides, some of which have important biological activities and exert their effects at very low concentrations. For example, a number of vertebrate hormones (intercellular chemical messengers) (Chapter 22) are small polypeptides. The hormone insulin contains two polypeptide chains, one having 30 amino acid residues and the other 21. Other polypeptide hormones include glucagon, a pancreatic hormone of 29 residues that opposes the action of insulin, and corticotropin, a 39-

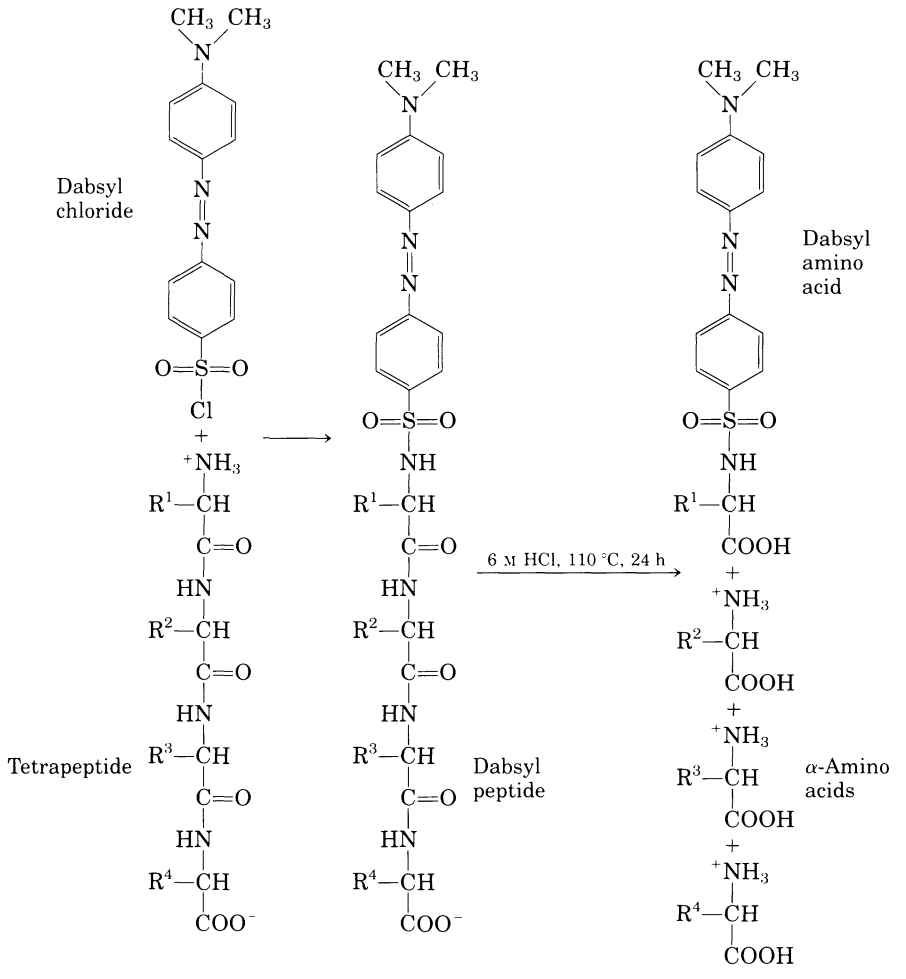
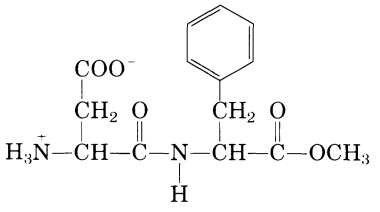


Figure 5–18 The amino-terminal residue of a tetrapeptide can be identified by labeling it with dabsyl chloride, then hydrolyzing the peptide bonds in strong acid. The result is a mixture of amino acids of which only the amino-terminal amino acid (and lysine) is labeled.

residue hormone of the anterior pituitary gland that stimulates the adrenal cortex.

Some biologically important peptides have only a few amino acid residues. That small peptides can have large biological effects is readily illustrated by the activity of the commercially synthesized dipeptide, L-aspartylphenylalanyl methyl ester. This compound is an artificial sweetener better known as aspartame or NutraSweet®:



L-Aspartyl-L-phenylalanyl methyl ester (aspartame)

Among naturally occurring small peptides are hormones such as oxytocin (nine amino acid residues), which is secreted by the posterior pituitary and stimulates uterine contractions; bradykinin (nine residues), which inhibits inflammation of tissues; and thyrotropin-releasing factor (three residues), which is formed in the hypothalamus and stimulates the release of another hormone, thyrotropin, from the anterior pituitary gland (Fig. 5–19). Also noteworthy among short peptides are the enkephalins, compounds formed in the central nervous system

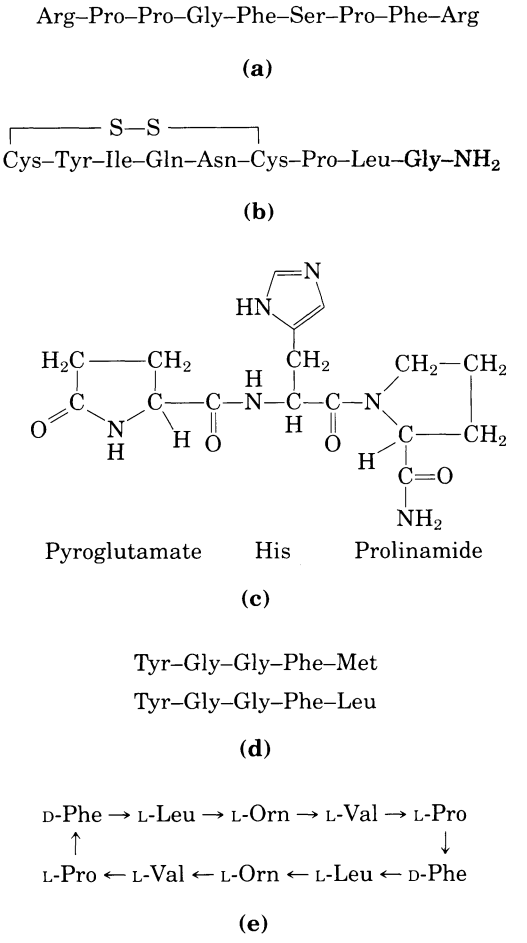


Figure 5–19 Some naturally occurring peptides with intense biological activity. The amino-terminal residues are at the left end. **(a)** Bradykinin, a hormonelike peptide that inhibits inflammatory reactions. **(b)** Oxytocin, formed by the posterior pituitary gland. The shaded portion is a residue of glycinamide ($H_2N—CH_2—CONH_2$). **(c)** Thyrotropin-releasing factor, formed by the hypothalamus. **(d)** Two enkephalins, brain peptides that affect the perception of pain. **(e)** Gramicidin S, an antibiotic produced by the bacterium *Bacillus brevis*. The arrows indicate the direction from the amino toward the carboxyl end of each residue. The peptide has no termini because it is circular. Orn is the symbol for ornithine, an amino acid that generally does not occur in proteins. Note that gramicidin S contains two residues of a D-amino acid (D-phenylalanine).

BOX 5–2 Chemical Synthesis of Peptides and Small Proteins

Many peptides are potentially useful as pharmacological reagents, and their synthesis is of considerable commercial importance. There are three ways to obtain a peptide: (1) purification from tissue, a task often made difficult by the vanishingly low concentrations of some peptides; (2) genetic engineering; or (3) direct chemical synthesis. Powerful techniques now make direct chemical synthesis an attractive option in many cases. In addition to commercial applications, the synthesis of specific peptide portions of larger proteins is an increasingly important tool for the study of protein structure and function.

The complexity of proteins makes the traditional synthetic approaches of organic chemistry impractical for peptides with more than four or five amino acids. One problem is the difficulty of purifying the product after each step, because the chemical properties of the peptide change each time a new amino acid is added.

The major breakthrough in this technology was provided by R. Bruce Merrifield. His innovation involved synthesizing a peptide while keeping it attached at one end to a solid support. The support is an insoluble polymer (resin) contained within a column, similar to that used for chromatographic procedures. The peptide is built up on this support one amino acid at a time using a standard set of reactions in a repeating cycle (Fig. 1).

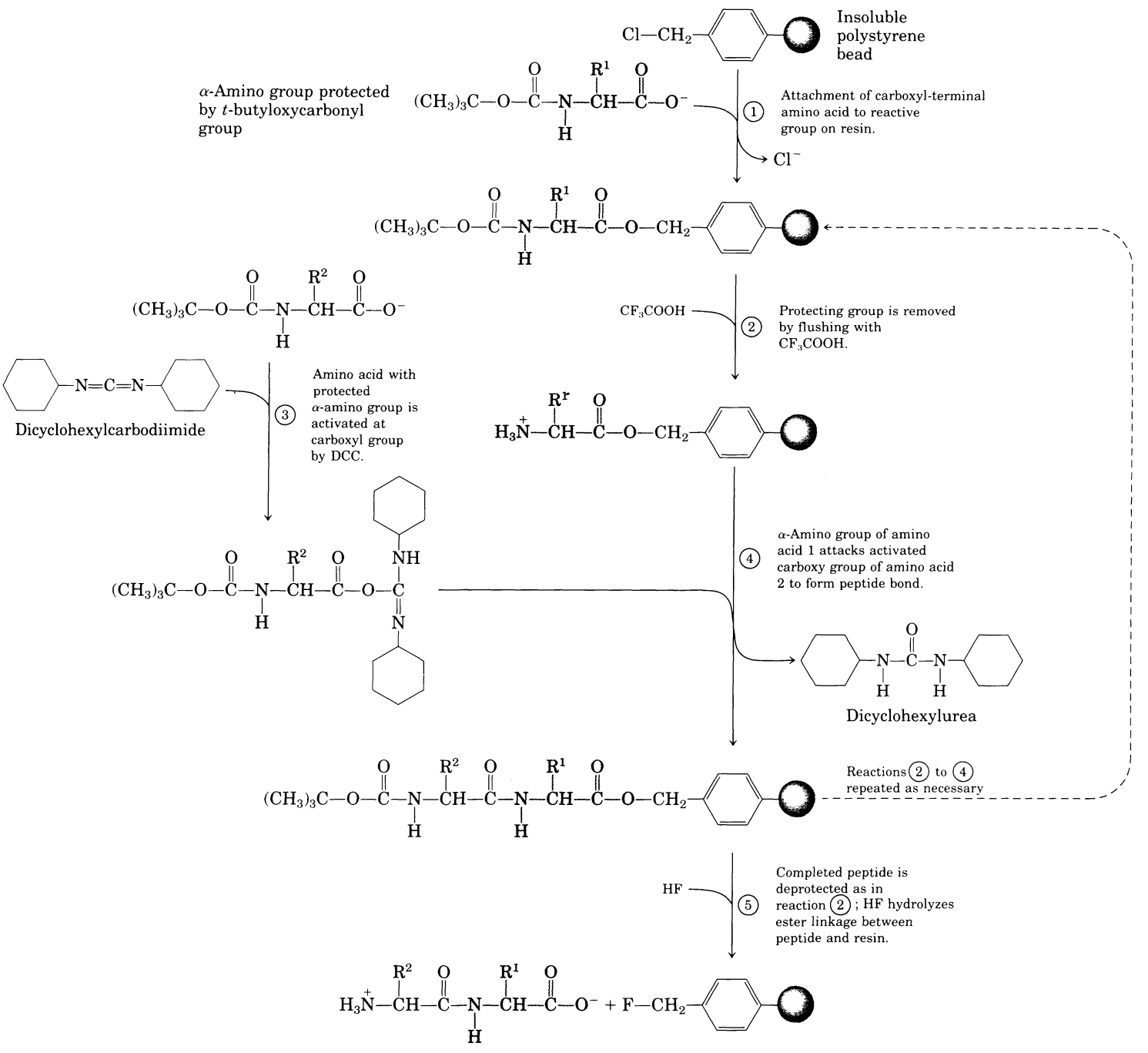
The technology for chemical peptide synthesis has been automated, and several commercial instruments are now available. The most important limitation of the process involves the efficiency of each amino acid addition, as can be seen by calculating the overall yields of peptides of various lengths when the yield for addition of each new amino acid is 96.0 versus 99.8% (Table 1). The chemistry has been optimized to permit the synthesis of proteins 100 amino acids long in about 4 days in reasonable yield. A very similar approach is used to synthesize nucleic acids (Fig. 12–38). It is worth noting that this technology, impressive as it is, still pales when compared with biological processes. The same 100 amino-acid protein would be synthesized with exquisite fidelity in about 5 seconds in a bacterial cell.

Table 1 Effect of stepwise yield on overall yields in peptide synthesis

Number of residues in the final polypeptide	Overall yields of final peptide (%) when the yield of each step is: 96.0%	99.8%
11	66	98
21	44	96
31	29	94
51	13	90
100	1.7	82

Figure 1 Chemical synthesis of a peptide on a solid support. Reactions ② through ④ are necessary for the formation of each peptide bond.

that bind to receptors in certain cells of the brain and induce analgesia (deadening of pain sensations). Enkephalins represent one of the body's own mechanisms for control of pain. The enkephalin receptors also bind morphine, heroin, and other addicting opiate drugs (although these are not peptides). Some extremely toxic mushroom poisons, such as amanitin, are also peptides, as are many antibiotics.



A growing number of small peptides are proving to be important commercially as pharmaceutical reagents. Unfortunately, they are often present in exceedingly small amounts and hence are hard to purify. For these and other reasons, the chemical synthesis of peptides has become one of the major technologies associated with biochemistry (Box 5–2).

Summary

The 20 amino acids commonly found as hydrolysis products of proteins contain an α-carboxyl group, an α-amino group, and a distinctive R group substituted on the α-carbon atom. The α-carbon atom of the amino acids (except glycine) is asymmetric, and thus amino acids can exist in at least two stereoisomeric forms. Only the L stereoisomers, which are related to the absolute configuration of L-glyceraldehyde, are found in proteins. The amino acids are classified on the basis of the polarity of their R groups. The nonpolar, aliphatic class includes alanine, glycine, isoleucine, leucine, proline, and valine. Phenylalanine, tryptophan, and tyrosine have aromatic side chains and are also relatively hydrophobic. The polar, uncharged class includes asparagine, cysteine, glutamine, methionine, serine, and threonine. The negatively charged (acidic) amino acids are aspartate and glutamate; the positively charge (basic) ones are arginine, histidine, and lysine. There are also a large number of nonstandard amino acids that occur in some proteins (as a result of the modification of standard amino acids) or as free metabolites in cells.

Monoamino monocarboxylic amino acids are diprotic acids ($^+H_3NCH(R)COOH$) at low pH. As the pH is raised to about 6, near the isoelectric point, the proton is lost from the carboxyl group to form the dipolar or zwitterionic species $^+H_3NCH(R)COO^-$, which is electrically neutral. Further increase in pH causes loss of the second proton, to yield the ionic species $H_2NCH(R)COO^-$. Amino acids with ionizable R groups may exist in additional ionic species, depending on the pH and the pK_a of the R group. Thus amino acids vary in their acid–base properties. Amino acids form colored derivatives with ninhydrin. Other colored or fluorescent derivatives are formed in reactions of the α-amino group of amino acids with fluorescamine, dansyl chloride, dabsyl chloride, and 1-fluoro-2,4-dinitrobenzene. Complex mixtures of amino acids can be separated and identified by ion-exchange chromatography or HPLC.

Amino acids can be joined covalently through peptide bonds to form peptides, which can also be formed by incomplete hydrolysis of polypeptides. The acid–base behavior and chemical reactions of a peptide are functions of its amino-terminal amino group, its carboxyl-terminal carboxyl group, and its R groups. Peptides can be hydrolyzed to yield free amino acids. Some peptides occur free in cells and tissues and have specific biological functions. These include some hormones and antibiotics, as well as other peptides with powerful biological activity.

Further Reading

General

Cantor, C.R. & Schimmel, P.R. (1980) *Biophysical Chemistry,* Part I: *The Conformation of Biological Macromolecules,* W.H. Freeman and Company, San Francisco.
Excellent textbook outlining the properties of biological macromolecules and their monomeric subunits.

Creighton, T.E. (1984) *Proteins: Structures and Molecular Properties,* W.H. Freeman and Company, New York.
Very useful general source.

Dickerson, R.E. & Geis, I. (1983) *Proteins: Structure, Function, and Evolution,* 2nd edn, The Benjamin/Cummings Publishing Company, Menlo Park, CA.
Beautifully illustrated and interesting account.

Amino Acids

Corrigan, J.J. (1969) D-Amino acids in animals. *Science* **169,** 142–148.

Meister, A. (1965) *Biochemistry of the Amino Acids,* 2nd edn, Vols. 1 and 2, Academic Press, Inc., New York.
Encyclopedic treatment of the properties, occurrence, and metabolism of amino acids.

Montgomery, R. & Swenson, C.A. (1976) *Quantitative Problems in the Biochemical Sciences,* 2nd edn, W.H. Freeman and Company, New York.

Segel, I.H. (1976) *Biochemical Calculations,* 2nd edn, John Wiley & Sons, New York.

Peptides

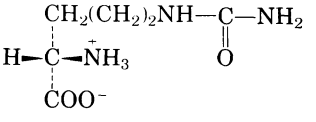
Haschemeyer, R.H. & Haschemeyer, A.E.V. (1973) *Proteins: A Guide to Study by Physical and Chemical Methods,* John Wiley & Sons, New York.

Merrifield, B. (1986) Solid phase synthesis. *Science* **232,** 341–347.

Smith, L.M. (1988) Automated synthesis and sequence analysis of biological macromolecules. *Analyt. Chem.* **60,** 381A–390A.

Problems

1. *Absolute Configuration of Citrulline* Is citrulline isolated from watermelons (shown below) a D- or L-amino acid? Explain.



2. *Relation between the Structures and Chemical Properties of the Amino Acids* The structures and chemical properties of the amino acids are crucial to understanding how proteins carry out their biological functions. The structures of the side chains of 16 amino acids are given below. Name the amino acid that contains each structure and match the R group with the most appropriate description of its properties, (a) to (m). Some of the descriptions may be used more than once.

(a) Small polar R group containing a hydroxyl group; this amino acid is important in the active site of some enzymes.

(b) Provides the least amount of steric hindrance.

(c) R group has $pK_a \approx 10.5$, making it positively charged at physiological pH.

(d) Sulfur-containing R group; neutral at any pH.

(e) Aromatic R group, hydrophobic in nature and neutral at any pH.

(f) Saturated hydrocarbon, important in hydrophobic interactions.

(g) The only amino acid having an ionizing R group with a pK_a near 7; it is an important group in the active site of some enzymes.

(h) The only amino acid having a substituted α-amino group; it influences protein folding by forcing a bend in the chain.

(i) R group has a pK_a near 4 and thus is negatively charged at pH 7.

(j) An aromatic R group capable of forming hydrogen bonds; it has a pK_a near 10.

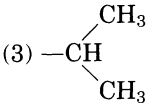
(k) Forms disulfide cross-links between polypeptide chains; the pK_a of its functional group is about 10.

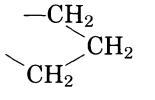
(l) R group with $pK_a \approx 12$, making it positively charged at physiological pH.

(m) When this polar but uncharged R group is hydrolyzed, the amino acid is converted into another amino acid having a negatively charged R group at pH near 7.

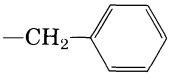
(1) —H

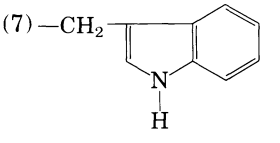
(2) —CH_3

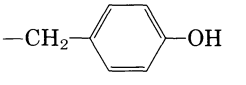
(3) 

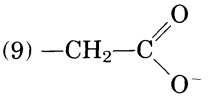
(4) 

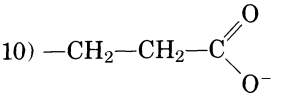
(5) —CH_2OH

(6) 

(7) 

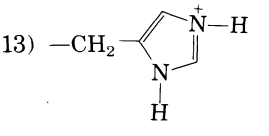
(8) 

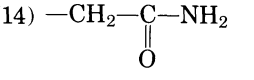
(9) 

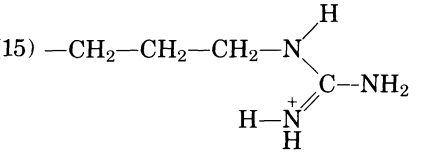
(10) 

(11) —CH_2—CH_2—S—CH_3

(12) —CH_2—SH

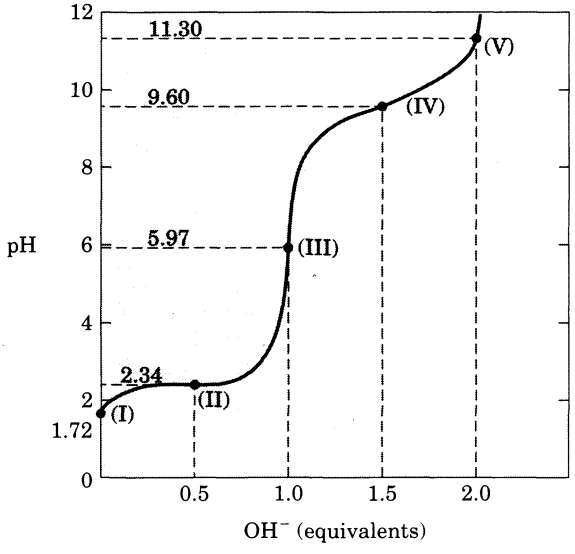
(13) 

(14) 

(15) 

(16) —CH_2—CH_2—CH_2—CH_2—$\overset{+}{N}H_3$

3. *Relationship between the Titration Curve and the Acid–Base Properties of Glycine* A 100 mL solution of 0.1 M glycine at pH 1.72 was titrated with 2 M NaOH solution. During the titration, the pH was monitored and the results were plotted in the graph shown. The key points in the titration are designated I to V on the graph. For each of the statements below, *identify* the appropriate key point in the titration and *justify* your choice.



(a) At what point will glycine be present predominantly as the species $^{+}H_3N—CH_2—COOH$?

(b) At what point is the *average* net charge of glycine $+\frac{1}{2}$?

(c) At what point is the amino group of half of the molecules ionized?

(d) At what point is the pH equal to the pK_a of the carboxyl group?

(e) At what point is the pH equal to the pK_a of the protonated amino group?

(f) At what points does glycine have its maximum buffering capacity?

(g) At what point is the *average* net charge zero?

(h) At what point has the carboxyl group been completely titrated (first equivalence point)?

(i) At what point are half of the carboxyl groups ionized?

(j) At what point is glycine completely titrated (second equivalence point)?

(k) At what point is the structure of the predominant species $^{+}H_3N—CH_2—COO^{-}$?

(l) At what point do the structures of the predominant species correspond to a 50:50 mixture of $^{+}H_3N—CH_2—COO^{-}$ and $H_2N—CH_2—COO^{-}$?

(m) At what point is the *average* net charge of glycine −1?

(n) At what point do the structures of the predominant species consist of a 50:50 mixture of $^{+}H_3N—CH_2—COOH$ and $^{+}H_3N—CH_2—COO^{-}$?

(o) What point corresponds to the isoelectric point?

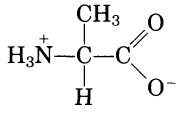
(p) At what point is the *average* net charge on glycine $-\frac{1}{2}$?

(q) What point represents the end of the titration?

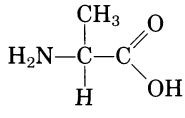
(r) If one wanted to use glycine as an efficient buffer, which points would represent the *worst* pH regions for buffering power?

(s) At what point in the titration is the predominant species $H_2N—CH_2—COO^{-}$?

4. *How Much Alanine Is Present as the Completely Uncharged Species?* At a pH equal to the isoelectric point, the *net* charge on alanine is zero. Two structures can be drawn that have a net charge of zero (zwitterionic and uncharged forms), but the predominant form of alanine at its pI is zwitterionic.



Zwitterionic



Uncharged

(a) Explain why the form of alanine at its pI is zwitterionic rather than completely uncharged.

(b) Estimate the fraction of alanine present at its pI as the completely uncharged form. Justify your assumptions.

5. *Ionization State of Amino Acids* Each ionizable group of an amino acid can exist in one of two states, charged or neutral. The electric charge on the functional group is determined by the relationship between its pK_a and the pH of the solution. This relationship is described by the Henderson–Hasselbalch equation.

(a) Histidine has three ionizable functional groups. Write the relevant equilibrium equations for its three ionizations and assign the proper pK_a for each ionization. Draw the structure of histidine in each ionization state. What is the net charge on the histidine molecule in each ionization state?

(b) Draw the structures of the predominant ionization state of histidine at pH 1, 4, 8, and 12. Note that the ionization state can be approximated by treating each ionizable group independently.

(c) What is the net charge of histidine at pH 1, 4, 8, and 12? For each pH, will histidine migrate toward the anode (+) or cathode (−) when placed in an electric field?

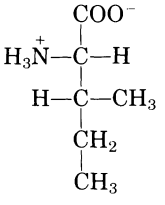
6. *Preparation of a Glycine Buffer* Glycine is commonly used as a buffer. Preparation of a 0.1 M glycine buffer starts with 0.1 M solutions of glycine hydrochloride ($HOOC—CH_2—NH_3^{+}Cl^{-}$) and gly-

cine ($^-OOC—CH_2—NH_3^+$), two commercially available forms of glycine. What volumes of these two solutions must be mixed to prepare 1 L of 0.1 M glycine buffer having a pH of 3.2? (Hint: See Box 4–2)

7. *Separation of Amino Acids by Ion-Exchange Chromatography* Mixtures of amino acids are analyzed by first separating the mixture into its components through ion-exchange chromatography. On a cation-exchange resin containing sulfonate groups (see Fig. 5–12), the amino acids flow down the column at different rates because of two factors that retard their movement: (1) ionic attraction between the $—SO_3^-$ residues on the column and positively charged functional groups on the amino acids and (2) hydrophobic interaction between amino acid side chains and the strongly hydrophobic backbone of the polystyrene resin. For each pair of amino acids listed, determine which member will be eluted first from an ion-exchange column by a pH 7.0 buffer.

(a) Asp and Lys
(b) Arg and Met
(c) Glu and Val
(d) Gly and Leu
(e) Ser and Ala

8. *Naming the Stereoisomers of Isoleucine* The structure of the amino acid isoleucine is:



(a) How many chiral centers does it have?
(b) How many optical isomers?
(c) Draw perspective formulas for all the optical isomers of isoleucine.

9. *Comparison of the pK_a Values of an Amino Acid and Its Peptides* The titration curve of the amino acid alanine shows the ionization of two functional groups with pK_a values of 2.34 and 9.69, corresponding to the ionization of the carboxyl and the protonated amino groups, respectively. The titration of di-, tri-, and larger oligopeptides of alanine also shows the ionization of only two functional groups, although the experimental pK_a values are different. The trend in pK_a values is summarized in the table.

Amino acid or peptide	pK_1	pK_2
Ala	2.34	9.69
Ala–Ala	3.12	8.30
Ala–Ala–Ala	3.39	8.03
Ala–$(Ala)_n$–Ala, $n \geq 4$	3.42	7.94

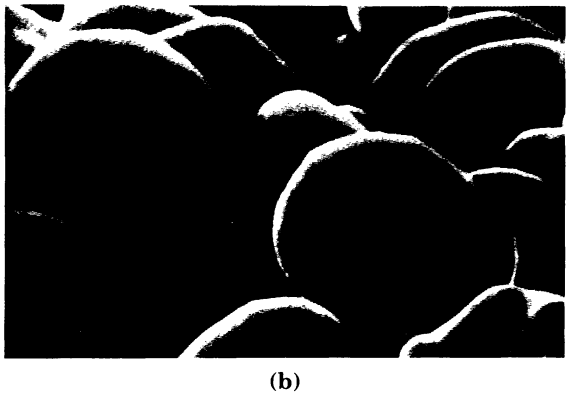
(a) Draw the structure of Ala–Ala–Ala. Identify the functional groups associated with pK_1 and pK_2.
(b) The value of pK_1 *increases* in going from Ala to an Ala oligopeptide. Provide an explanation for this trend.
(c) The value of pK_2 *decreases* in going from Ala to an Ala oligopeptide. Provide an explanation for this trend.

10. *Peptide Synthesis* In the synthesis of polypeptides on solid supports, the α-amino group of each new amino acid is "protected" by a *t*-butyloxycarbonyl group (see Box 5–2). What would happen if this protecting group were not present?

CHAPTER

An Introduction to Proteins

(a)



(b)

Figure 6–1 Functions of proteins. **(a)** The light produced by fireflies is the result of a light-producing reaction involving luciferin and ATP that is catalyzed by the enzyme luciferase (see Box 13–3). **(b)** Erythrocytes contain large amounts of the oxygen-transporting protein hemoglobin. **(c)** The white color of milk is derived primarily from the protein casein. **(d)** The movement of cilia in protozoans depends on the action of the protein dynein. **(e)** The protein fibroin is the major structural component of spider webs. **(f)** Castor beans contain a highly toxic protein called ricin. **(g)** Cancerous tumors are often made up of cells that have defects involving one or more of the proteins that regulate cell division.

Almost everything that occurs in the cell involves one or more proteins. Proteins provide structure, catalyze cellular reactions, and carry out a myriad of other tasks. Their central place in the cell is reflected in the fact that genetic information is ultimately expressed as protein. For each protein there is a segment of DNA (a gene; see Chapters 12 and 23) that encodes information specifying its sequence of amino acids. There are thousands of different kinds of proteins in a typical cell, each encoded by a gene and each performing a specific function. Proteins are among the most abundant biological macromolecules and are also extremely versatile in their functions.

The chapter begins with a discussion of some of the general properties of proteins. This is followed by a short summary of some common techniques used to purify and study proteins. Finally, we will examine the **primary structure** of protein molecules: the covalent backbone structure and the sequence of amino acid residues. One goal is to discover the relationships between amino acid sequence and biological function.

Properties of Proteins

An understanding of these important macromolecules must begin with the fundamentals. What do proteins do? How big are they? What forms or shapes do they take? What are their chemical properties? The answers serve as an orientation to much that follows.

Proteins Have Many Different Biological Functions

We can classify proteins according to their biological roles.

Enzymes The most varied and most highly specialized proteins are those with catalytic activity—the enzymes. Virtually all the chemical reactions of organic biomolecules in cells are catalyzed by enzymes. Many thousands of different enzymes, each capable of catalyzing a different kind of chemical reaction, have been discovered in different organisms (Fig. 6–1a).

Transport Proteins Transport proteins in blood plasma bind and carry specific molecules or ions from one organ to another. Hemoglobin of erythrocytes (Fig. 6–1b) binds oxygen as the blood passes through the lungs, carries it to the peripheral tissues, and there releases it to participate in the energy-yielding oxidation of nutrients. The blood

plasma contains lipoproteins, which carry lipids from the liver to other organs. Other kinds of transport proteins are present in the plasma membranes and intracellular membranes of all organisms; these are adapted to bind glucose, amino acids, or other substances and transport them across the membrane.

Nutrient and Storage Proteins The seeds of many plants store nutrient proteins required for the growth of the germinating seedling. Particularly well-studied examples are the seed proteins of wheat, corn, and rice. Ovalbumin, the major protein of egg white, and casein, the major protein of milk, are other examples of nutrient proteins (Fig. 6–1c). The ferritin found in some bacteria and in plant and animal tissues stores iron.

Contractile or Motile Proteins Some proteins endow cells and organisms with the ability to contract, to change shape, or to move about. Actin and myosin function in the contractile system of skeletal muscle and also in many nonmuscle cells. Tubulin is the protein from which microtubules are built. Microtubules act in concert with the protein dynein in flagella and cilia (Fig. 6–1d) to propel cells.

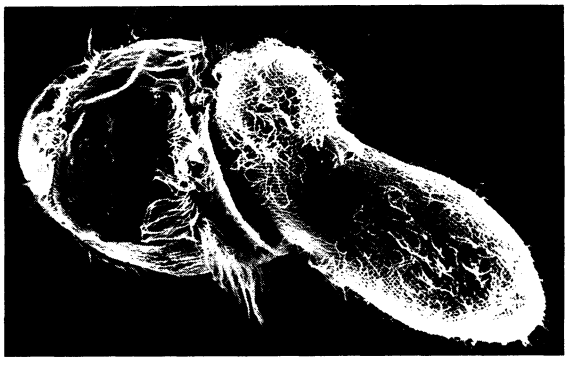
Structural Proteins Many proteins serve as supporting filaments, cables, or sheets, to give biological structures strength or protection. The major component of tendons and cartilage is the fibrous protein collagen, which has very high tensile strength. Leather is almost pure collagen. Ligaments contain elastin, a structural protein capable of stretching in two dimensions. Hair, fingernails, and feathers consist largely of the tough, insoluble protein keratin. The major component of silk fibers and spider webs is fibroin (Fig. 6–1e). The wing hinges of some insects are made of resilin, which has nearly perfect elastic properties.

Defense Proteins Many proteins defend organisms against invasion by other species or protect them from injury. The immunoglobulins or antibodies, specialized proteins made by the lymphocytes of vertebrates, can recognize and precipitate or neutralize invading bacteria, viruses, or foreign proteins from another species. Fibrinogen and thrombin are blood-clotting proteins that prevent loss of blood when the vascular system is injured. Snake venoms, bacterial toxins, and toxic plant proteins, such as ricin, also appear to have defensive functions (Fig. 6–1f). Some of these, including fibrinogen, thrombin, and some venoms, are also enzymes.

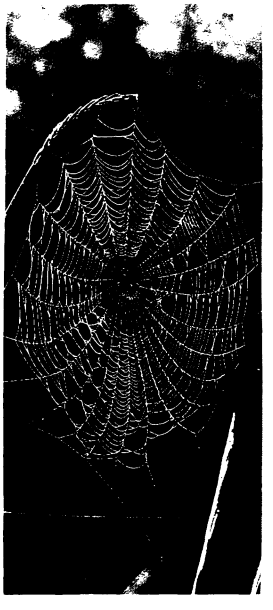
Regulatory Proteins Some proteins help regulate cellular or physiological activity. Among them are many hormones. Examples include insulin, which regulates sugar metabolism, and the growth hormone of the pituitary. The cellular response to many hormonal signals is often mediated by a class of GTP-binding proteins called G proteins (GTP is closely related to ATP, with guanine replacing the adenine portion of the molecule; see Figs. 1–12 and 3–16b.) Other regulatory proteins bind to DNA and regulate the biosynthesis of enzymes and RNA molecules involved in cell division in both prokaryotes and eukaryotes (Fig. 6–1g).

Other Proteins There are numerous other proteins whose functions are rather exotic and not easily classified. Monellin, a protein of an African plant, has an intensely sweet taste. It is being studied as a

(c)

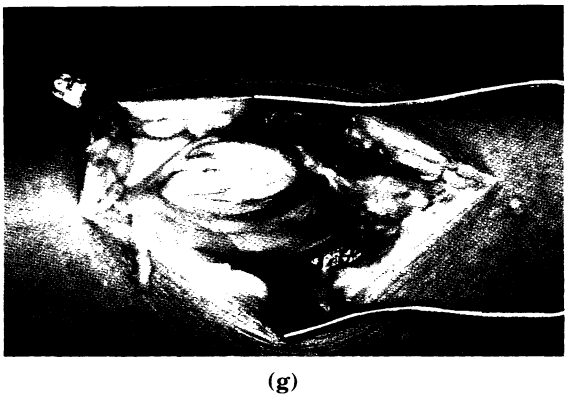


(d)



(e)

(f)



(g)

nonfattening, nontoxic food sweetener for human use. The blood plasma of some Antarctic fish contains antifreeze proteins, which protect their blood from freezing.

It is extraordinary that all these proteins, with their very different properties and functions, are made from the same group of 20 amino acids.

Proteins Are Very Large Molecules

How long are the polypeptide chains in proteins? Table 6–1 shows that human cytochrome *c* has 104 amino acid residues linked in a single chain; bovine chymotrypsinogen has 245 amino acid residues. Probably near the upper limit of size is the protein apolipoprotein B, a cholesterol-transport protein with 4,536 amino acid residues in a single polypeptide chain of molecular weight 513,000. Most naturally occurring polypeptides contain less than 2,000 amino acid residues.

Table 6–1 Molecular data on some proteins

	Molecular weight	Number of residues	Number of polypeptide chains
Insulin (bovine)	5,733	51	2
Cytochrome *c* (human)	13,000	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (egg white)	13,930	129	1
Myoglobin (equine heart)	16,890	153	1
Chymotrypsin (bovine pancreas)	21,600	241	3
Chymotrypsinogen (bovine)	22,000	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	68,500	~550	1
Hexokinase (yeast)	102,000	~800	2
Immunoglobulin G (human)	145,000	~1,320	4
RNA polymerase (*E. coli*)	450,000	~4,100	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamate dehydrogenase (bovine liver)	1,000,000	~8,300	~40

Some proteins consist of a single polypeptide chain, but others, called **multisubunit** proteins, have two or more (Table 6–1). The individual polypeptide chains in a multisubunit protein may be identical or different. If at least some are identical, the protein is sometimes called an **oligomeric** protein and the subunits themselves are referred to as **protomers.** The enzyme ribonuclease has one polypeptide chain. Hemoglobin has four: two identical α chains and two identical β chains, all four held together by noncovalent interactions.

The molecular weights of proteins, which can be determined by various physicochemical methods, may range from little more than 10,000 for small proteins such as cytochrome *c* (104 residues), to more than 10^6 for proteins with very long polypeptide chains or those with several subunits. The molecular weights of some typical proteins are given in Table 6–1. No simple generalizations can be made about the molecular weights of proteins in relation to their function.

One can calculate the approximate number of amino acid residues in a simple protein containing no other chemical group by dividing its molecular weight by 110. Although the average molecular weight of the

20 standard amino acids is about 138, the smaller amino acids predominate in most proteins; when weighted for the proportions in which the various amino acids occur in proteins (see Table 5–1), the average molecular weight is nearer to 128. Because a molecule of water (M_r 18) is removed to create each peptide bond, the average molecular weight of an amino acid residue in a protein is about 128 − 18 = 110. Table 6–1 shows the number of amino acid residues in several proteins.

Proteins Have Characteristic Amino Acid Compositions

As is true for simple peptides, hydrolysis of proteins with acid or base yields a mixture of free α-amino acids. When completely hydrolyzed, each type of protein yields a characteristic proportion or mixture of the different amino acids. Table 6–2 shows the composition of the amino acid mixtures obtained on complete hydrolysis of human cytochrome *c* and of bovine chymotrypsinogen, the inactive precursor of the digestive enzyme chymotrypsin. These two proteins, with very different functions, also differ significantly in the relative numbers of each kind of amino acid they contain. The 20 amino acids almost never occur in equal amounts in proteins. Some amino acids may occur only once per molecule or not at all in a given type of protein; others may occur in large numbers.

Table 6–2 Amino acid composition of two proteins

Amino acid	Number of residues per molecule of protein	
	Human cytochrome *c*	Bovine chymotrypsinogen
Ala	6	22
Arg	2	4
Asn	5	15
Asp	3	8
Cys	2	10
Gln	2	10
Glu	8	5
Gly	13	23
His	3	2
Ile	8	10
Leu	6	19
Lys	18	14
Met	3	2
Phe	3	6
Pro	4	9
Ser	2	28
Thr	7	23
Trp	1	8
Tyr	5	4
Val	3	23
Total	104	245

Some Proteins Contain Chemical Groups Other Than Amino Acids

Many proteins, such as the enzymes ribonuclease and chymotrypsinogen, contain only amino acids and no other chemical groups; these are considered simple proteins. However, some proteins contain chemical components in addition to amino acids; these are called **conjugated proteins.** The non–amino acid part of a conjugated protein is usually called its **prosthetic group.** Conjugated proteins are classified on the basis of the chemical nature of their prosthetic groups (Table 6–3); for example, **lipoproteins** contain lipids, **glycoproteins** contain sugar groups, and **metalloproteins** contain a specific metal. A number of proteins contain more than one prosthetic group. Usually the prosthetic group plays an important role in the protein's biological function.

Table 6–3 Conjugated proteins

Class	Prosthetic group	Example
Lipoproteins	Lipids	β_1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

Working with Proteins

The aggregate biochemical picture of protein structure and function is derived from the study of many individual proteins. To study a protein in any detail it must be separated from all other proteins in a cell, and techniques must be available to determine its properties. The necessary methods come from protein chemistry, a discipline as old as biochemistry itself and one that retains a central position in biochemical research. Modern techniques are providing ever newer experimental insights into the critical relationship between the structure of a protein and its function.

Proteins Can Be Separated and Purified

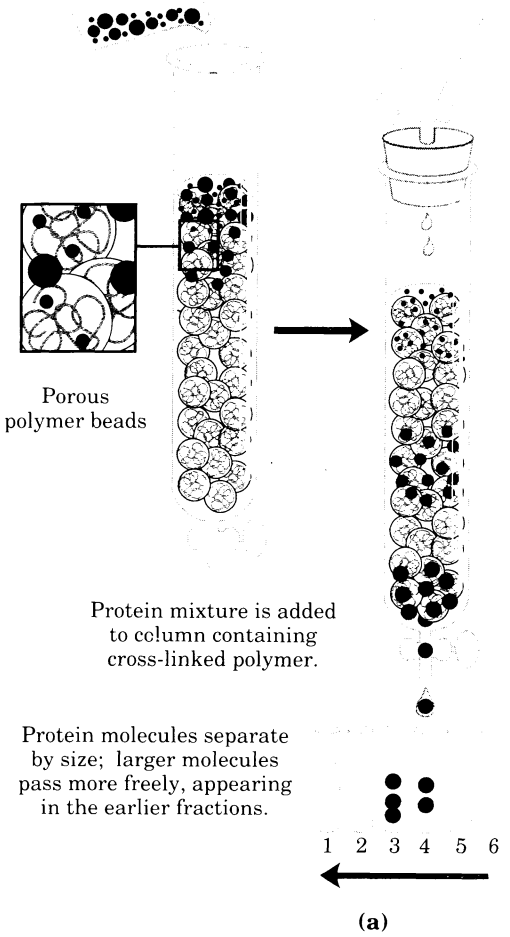
Cells contain thousands of different kinds of proteins. A pure preparation of a given protein is essential before its properties, amino acid composition, and sequence can be determined. How, then, can one protein be purified?

Methods for separating proteins take advantage of properties such as charge, size, and solubility, which vary from one protein to the next. Because many proteins bind to other biomolecules, proteins can also be separated on the basis of their binding properties. The source of a protein is generally tissue or microbial cells. The cells must be broken open and the protein released into a solution called a **crude extract.** If necessary, differential centrifugation can be used to prepare subcellular fractions or to isolate organelles (see Fig. 2–24). Once the extract or organelle preparation is ready, a variety of methods are available for separation of proteins. Ion-exchange chromatography (see Fig. 5–12) can be used to separate proteins with different charges in much the same way that it separates amino acids. Other chromatographic methods take advantage of differences in size, binding affinity, and solubility (Fig. 6–2). Nonchromatographic methods include the selective precipitation of proteins with salt, acid, or high temperatures.

The approach to the purification of a "new" protein, one not previously isolated, is guided both by established precedents and common sense. In most cases, several different methods must be used sequentially to completely purify a protein. The choice of method is somewhat empirical, and many protocols may be tried before the most effective is determined. Trial and error can often be minimized by using purification procedures developed for similar proteins as a guide. Published purification protocols are available for many thousands of proteins. Common sense dictates that inexpensive procedures be used first, when the total volume and number of contaminants is greatest. Chromatographic methods are often impractical at early stages because the amount of chromatographic medium needed increases with sample size. As each purification step is completed, the sample size generally becomes smaller (Table 6–4) and more sophisticated (and expensive) chromatographic procedures can be applied.

Individual Proteins Can Be Quantified

In order to purify a protein, it is essential to have an assay to detect and quantify that protein in the presence of many other proteins. Often, purification must proceed in the absence of any information about the size and physical properties of the protein, or the fraction of the total protein mass it represents in the extract.



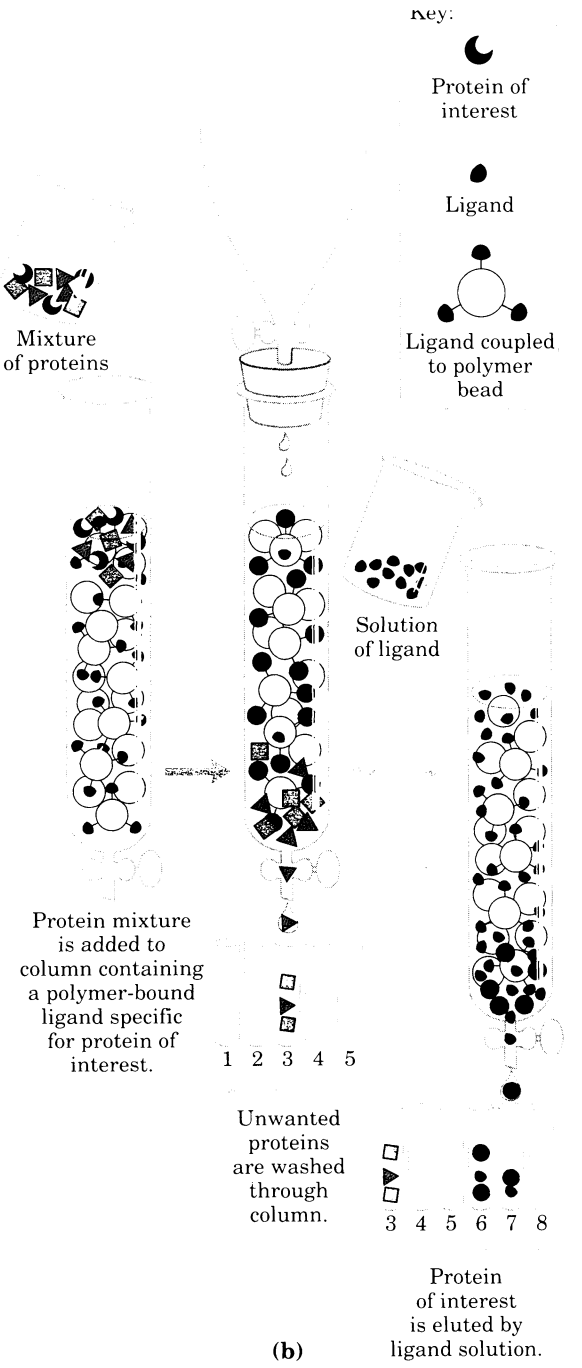


Figure 6–2 Two types of chromatographic methods used in protein purification. **(a)** Size-exclusion chromatography; also called gel filtration. This method separates proteins according to size. The column contains a cross-linked polymer with pores of selected size. Larger proteins migrate faster than smaller ones, because they are too large to enter the pores in the beads and hence take a more direct route through the column. The smaller proteins enter the pores and are slowed by the more labyrinthian path they take through the column. **(b)** Affinity chromatography separates proteins by their binding specificities. The proteins retained on the column are those that bind specifically to a ligand cross-linked to the beads. (In biochemistry, the term "ligand" is used to refer to a group or molecule that is bound.) After nonspecific proteins are washed through the column, the bound protein of particular interest is eluted by a solution containing free ligand.

Table 6–4 A purification table for a hypothetical enzyme*

Procedure or step	Fraction volume (ml)	Total protein (mg)	Activity (units)	Specific activity (units/mg)
1. Crude cellular extract	1,400	10,000	100,000	10
2. Precipitation	280	3,000	96,000	32
3. Ion-exchange chromatography	90	400	80,000	200
4. Size-exclusion chromatography	80	100	60,000	600
5. Affinity chromatography	6	3	45,000	15,000

* All data represent the status of the sample *after* the procedure indicated in the first column has been carried out.

The amount of an enzyme in a given solution or tissue extract can be assayed in terms of the catalytic effect it produces, that is, the *increase* in the rate at which its substrate is converted to reaction products when the enzyme is present. For this purpose one must know (1) the overall equation of the reaction catalyzed, (2) an analytical procedure for determining the disappearance of the substrate or the appearance of the reaction products, (3) whether the enzyme requires cofactors such as metal ions or coenzymes, (4) the dependence of the enzyme activity on substrate concentration, (5) the optimum pH, and (6) a temperature zone in which the enzyme is stable and has high activity. Enzymes are usually assayed at their optimum pH and at some convenient temperature within the range 25 to 38 °C. Also, very high substrate concentrations are generally required so that the initial reaction rate, which is measured experimentally, is proportional to enzyme concentration (Chapter 8).

By international agreement, 1.0 unit of enzyme activity is defined as the amount of enzyme causing transformation of 1.0 μmol of substrate per minute at 25 °C under optimal conditions of measurement. The term **activity** refers to the total units of enzyme in the solution. The **specific activity** is the number of enzyme units per milligram of protein (Fig. 6–3). The specific activity is a measure of enzyme purity: it increases during purification of an enzyme and becomes maximal and constant when the enzyme is pure (Table 6–4).

After each purification step, the activity of the preparation (in units) is assayed, the total amount of protein is determined independently, and their ratio gives the specific activity. Activity and total protein generally decrease with each step. Activity decreases because some loss always occurs due to inactivation or nonideal interactions with chromatographic materials or other molecules in the solution. Total protein decreases because the objective is to remove as much nonspecific protein as possible. In a successful step, the loss of nonspecific protein is much greater than the loss of activity; therefore, specific activity increases even as total activity falls. The data are then assembled in a purification table (Table 6–4). A protein is generally considered pure when further purification steps fail to increase specific activity, and when only a single protein species can be detected (by methods to be described later).

For proteins that are not enzymes, other quantification methods are required. Transport proteins can be assayed by their binding to the molecule they transport, and hormones and toxins by the biological effect they produce; for example, growth hormones will stimulate the growth of certain cultured cells. Some structural proteins represent such a large fraction of a tissue mass that they can be readily extracted and purified without an assay. The approaches are as varied as the proteins themselves.

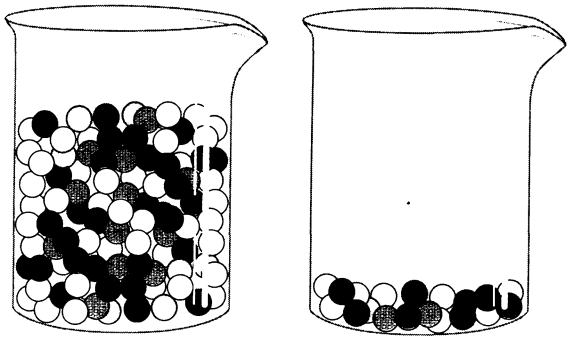
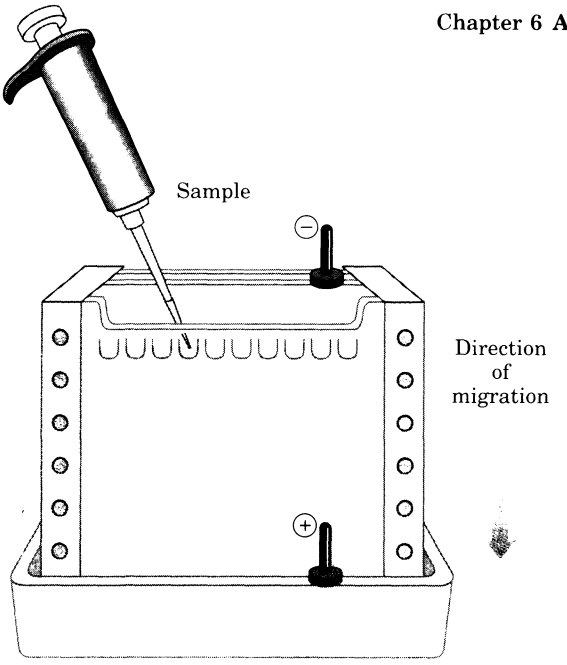
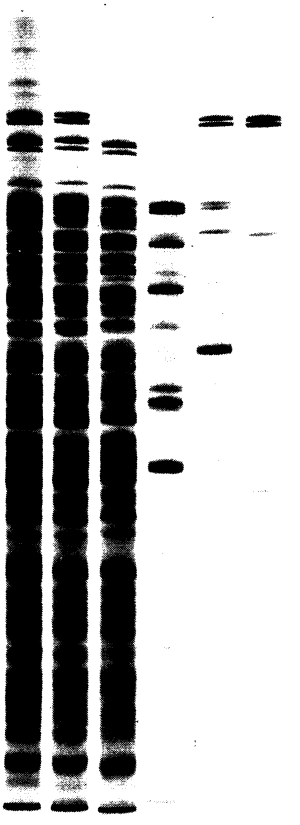


Figure 6–3 Activity versus specific activity. The difference between these two terms can be illustrated by considering two jars of marbles. The jars contain the same number of red marbles (representing an unknown protein), but different amounts of marbles of other colors. If the marbles are taken to represent proteins, both jars contain the same *activity* of the protein represented by the red marbles. The second jar, however, has the higher *specific activity* because here the red marbles represent a much higher fraction of the total.



(a)



(b)

Figure 6–4 Electrophoresis. **(a)** Different samples are loaded in wells or depressions at the top of the polyacrylamide gel. The proteins move into the gel when an electric field is applied. The gel minimizes convection currents caused by small temperature gradients, and it minimizes protein movements other than those induced by the electric field. **(b)** Proteins can be visualized after electrophoresis by treating the gel with a stain such as Coomassie blue, which binds to the proteins but not to the gel itself. Each band on the gel represents a different protein (or protein subunit); smaller proteins are found nearer the bottom of the gel. This gel illustrates the purification of the enzyme RNA polymerase from the bacterium *E. coli*. The first lane shows the proteins present in the crude cellular extract. Successive lanes show the proteins present after each purification step. The purified protein contains four subunits, as seen in the last lane on the right.

Proteins Can Be Characterized by Electrophoresis

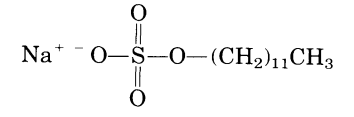
In addition to chromatography, another important set of methods is available for the separation of proteins, based on the migration of charged proteins in an electric field, a process called **electrophoresis.** These procedures are not often used to purify proteins in large amounts because simpler alternative methods are usually available and electrophoretic methods often inactivate proteins. Electrophoresis is, however, especially useful as an analytical method. Its advantage is that proteins can be visualized as well as separated, permitting a researcher to estimate quickly the number of proteins in a mixture or the degree of purity of a particular protein preparation. Also, electrophoresis allows determination of crucial properties of a protein such as its isoelectric point and approximate molecular weight.

In electrophoresis, the force moving the macromolecule (nucleic acids as well as proteins are separated this way) is the electrical potential, E. The electrophoretic mobility of the molecule, μ, is the ratio of the velocity of the particle, V, to the electrical potential. Electrophoretic mobility is also equal to the net charge of the molecule, Z, divided by the frictional coefficient, f. Thus:

$$\mu = \frac{V}{E} = \frac{Z}{f}$$

Electrophoresis of proteins is generally carried out in gels made up of the cross-linked polymer polyacrylamide (Fig. 6–4). The polyacrylamide gel acts as a molecular sieve, slowing the migration of proteins approximately in proportion to their mass, or molecular weight.

An electrophoretic method commonly used for estimation of purity and molecular weight makes use of the detergent **sodium dodecyl sulfate** (SDS). SDS binds to most proteins (probably by hydrophobic interactions; see Chapter 4) in amounts roughly proportional to the molecular weight of the protein, about one molecule of SDS for every two amino acid residues. The bound SDS contributes a large net negative charge, rendering the intrinsic charge of the protein insignificant.



Sodium dodecyl sulfate (SDS)

In addition, the native conformation of a protein is altered when SDS is bound, and most proteins assume a similar shape, and thus a similar ratio of charge to mass. Electrophoresis in the presence of SDS therefore separates proteins almost exclusively on the basis of mass (molecular weight), with smaller polypeptides migrating more rapidly. After electrophoresis, the proteins are visualized by adding a dye such as Coomassie blue (Fig. 6–4b) which binds to proteins but not to the gel itself. This type of gel provides one method to monitor progress in isolating a protein, because the number of protein bands should decrease as the purification proceeds. When compared with the positions to which proteins of known molecular weight migrate in the gel, the position of an unknown protein can provide an excellent measure of its molecular weight (Fig. 6–5). If the protein has two or more different subunits, each subunit will generally be separated by the SDS treatment, and a separate band will appear for each.

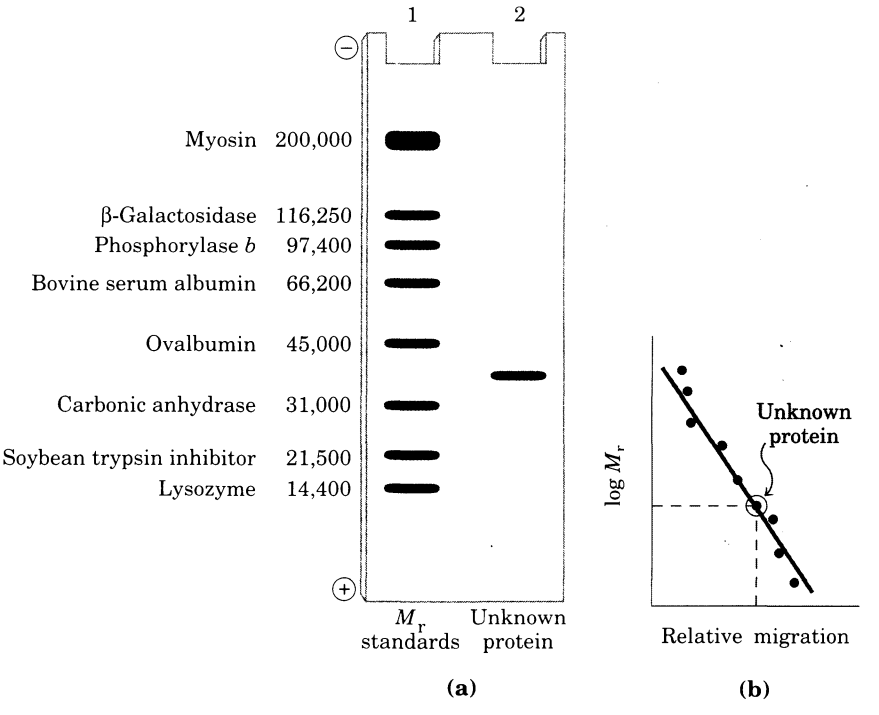


Figure 6–5 Estimating the molecular weight of a protein. The electrophoretic mobility of a protein on an SDS polyacrylamide gel is related to its molecular weight, M_r. **(a)** Standard proteins of known molecular weight are subjected to electrophoresis (lane 1). These marker proteins can be used to estimate the M_r of an unknown protein (lane 2). **(b)** A plot of log M_r of the marker proteins versus relative migration during electrophoresis allows the M_r of the unknown protein to be read from the graph.

Table 6–5 The isoelectric points of some proteins

	pI
Pepsin	~1.0
Egg albumin	4.6
Serum albumin	4.9
Urease	5.0
β-Lactoglobulin	5.2
Hemoglobin	6.8
Myoglobin	7.0
Chymotrypsinogen	9.5
Cytochrome *c*	10.7
Lysozyme	11.0

Isoelectric focusing is a procedure used to determine the isoelectric point (pI) of a protein (Fig. 6–6). A pH gradient is established by allowing a mixture of low molecular weight organic acids and bases (ampholytes; see p. 118) to distribute themselves in an electric field generated across the gel. When a protein mixture is applied, each protein migrates until it reaches the pH that matches its pI. Proteins with different isoelectric points are thus distributed differently throughout the gel (Table 6–5).

Combining these two electrophoretic methods in two-dimensional gels permits the resolution of complex mixtures of proteins (Fig. 6–7). This is a more sensitive analytical method than either isoelectric focusing or SDS electrophoresis alone. Two-dimensional electrophoresis separates proteins of identical molecular weight that differ in pI, or proteins with similar pI values but different molecular weights.

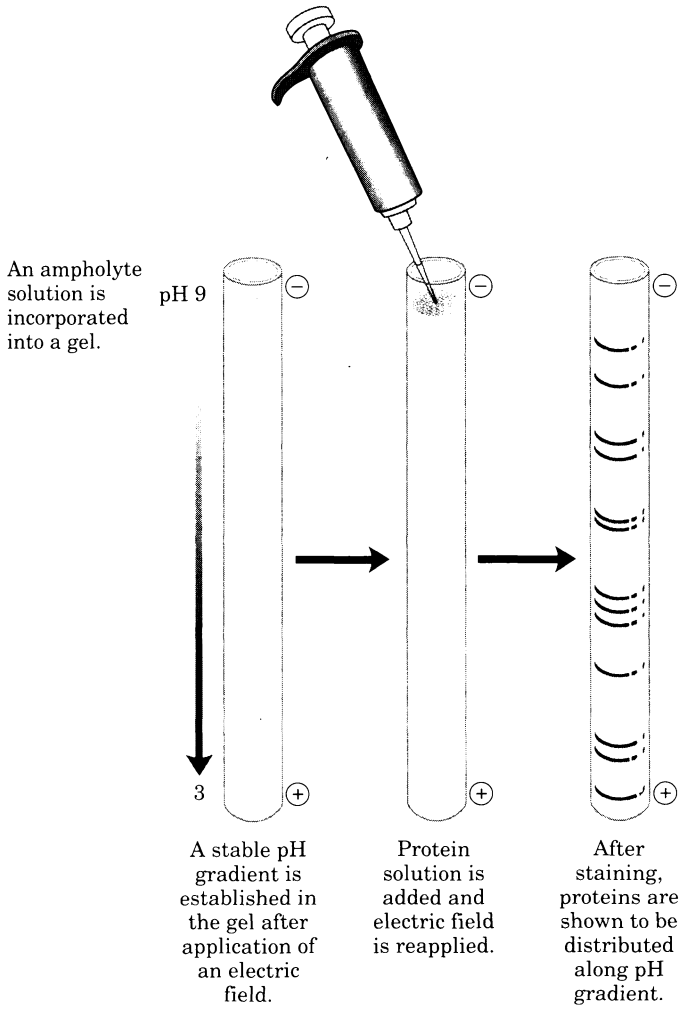


Figure 6–6 Isoelectric focusing. This technique separates proteins according to their isoelectric points. A stable pH gradient is established in the gel by the addition of appropriate ampholytes. A protein mixture is placed in a well on the gel. With an applied electric field, proteins enter the gel and migrate until each reaches a pH equivalent to its pI. Remember that the net charge of a protein is zero when pH = pI.

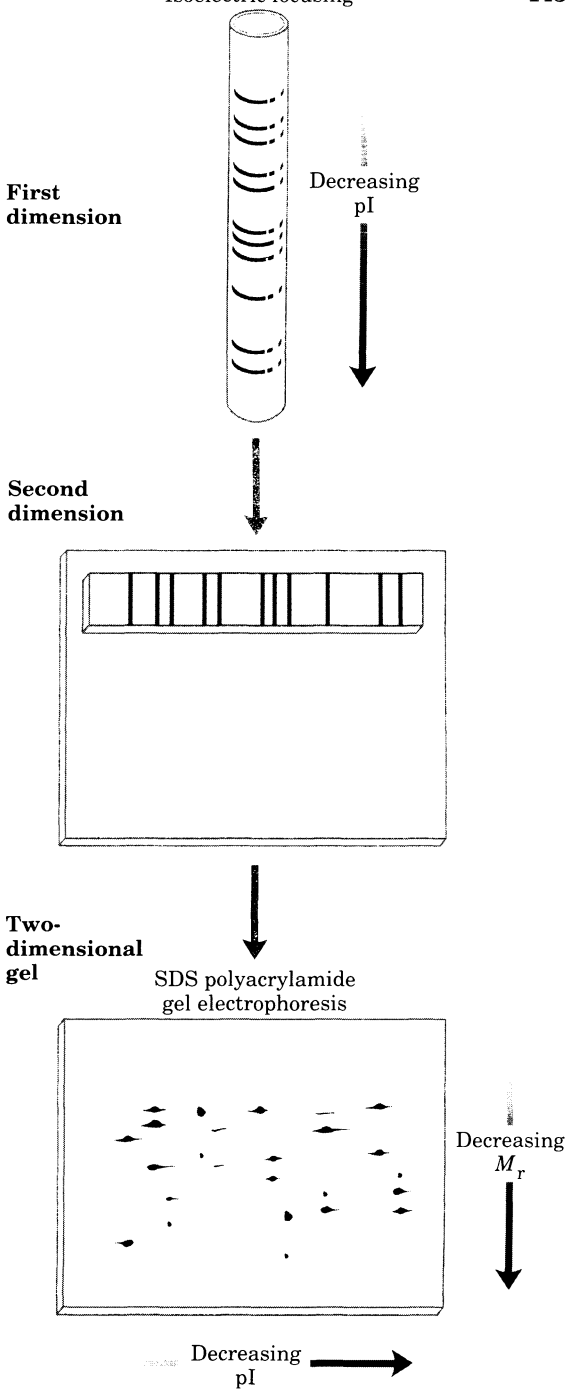
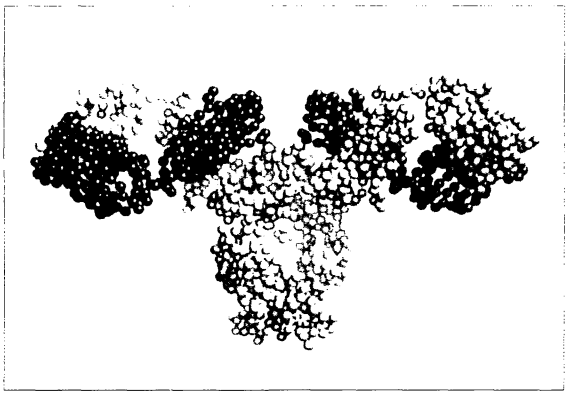
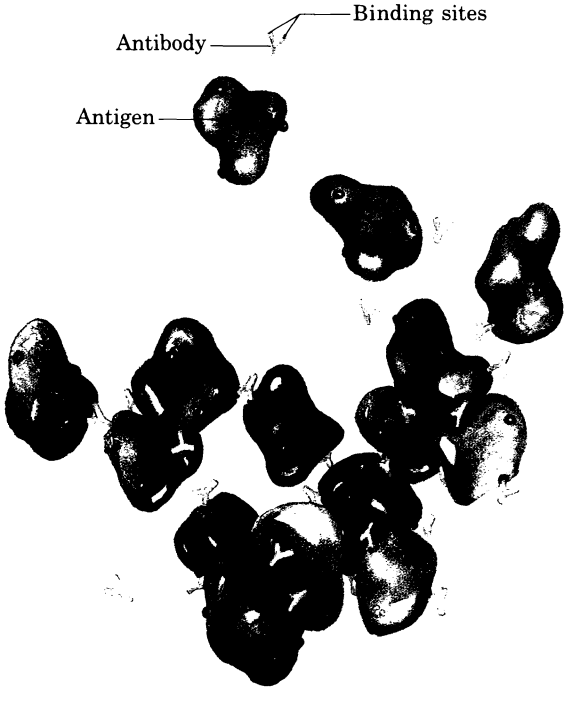


Figure 6–7 Two-dimensional electrophoresis. **(a)** Proteins are first separated by isoelectric focusing. The gel is then laid horizontally on a second gel, and the proteins are separated by SDS polyacrylamide gel electrophoresis. In this two-dimensional gel, horizontal separation reflects differences in pI; vertical separation reflects differences in molecular weight. **(b)** More than 1,000 different proteins from *E. coli* can be resolved using this technique.



(a)



(b)

Figure 6–8 The immune response and the action of antibodies. **(a)** A molecule of immunoglobulin G (IgG) consists of two polypeptides known as heavy chains (white and light blue) and two known as light chains (purple and dark blue). Immunoglobulins are glycoproteins and contain bound carbohydrate (yellow). **(b)** Each antigen evokes a specific set of antibodies, which will recognize and combine only with that antigen or closely related molecules. (Antibody-binding sites are shown as red areas on the antigen.) The Y-shaped antibodies each have two binding sites for the antigen, and can precipitate the antigen by forming an insoluble, latticelike aggregate.

The Antibody–Antigen Interaction Is Used to Quantify and Localize Proteins

Several sensitive analytical procedures have been developed from the study of a class of proteins called **antibodies** or **immunoglobulins.** Antibody molecules appear in the blood serum and certain tissues of a vertebrate animal in response to injection of an **antigen,** a protein or other macromolecule foreign to that individual. Each foreign protein elicits the formation of a set of different antibodies, which can combine with the antigen to form an antigen–antibody complex. The production of antibodies is part of a general defense mechanism in vertebrates called the **immune response.**

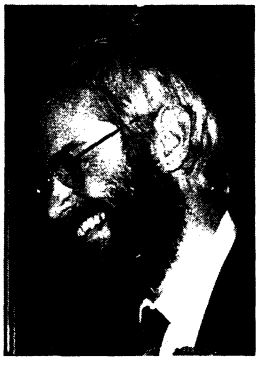
Antibodies are Y-shaped proteins consisting of four polypeptide chains. They have two binding sites that are complementary to specific structural features of the antigen molecule, making possible the formation of a three-dimensional lattice of alternating antigen and antibody molecules (Fig. 6–8). If sufficient antigen is present in a sample, the addition of antibodies or blood serum from an immunized animal will result in the formation of a quantifiable precipitate. No such precipitate is formed when serum of an unimmunized animal is mixed with the antigen.

Antibodies are highly specific for the foreign proteins or other macromolecules that evoke their formation. It is this specificity that makes them valuable analytical reagents. A rabbit antibody formed to horse serum albumin, for example, will combine with the latter but will not usually combine with other horse proteins, such as horse hemoglobin.

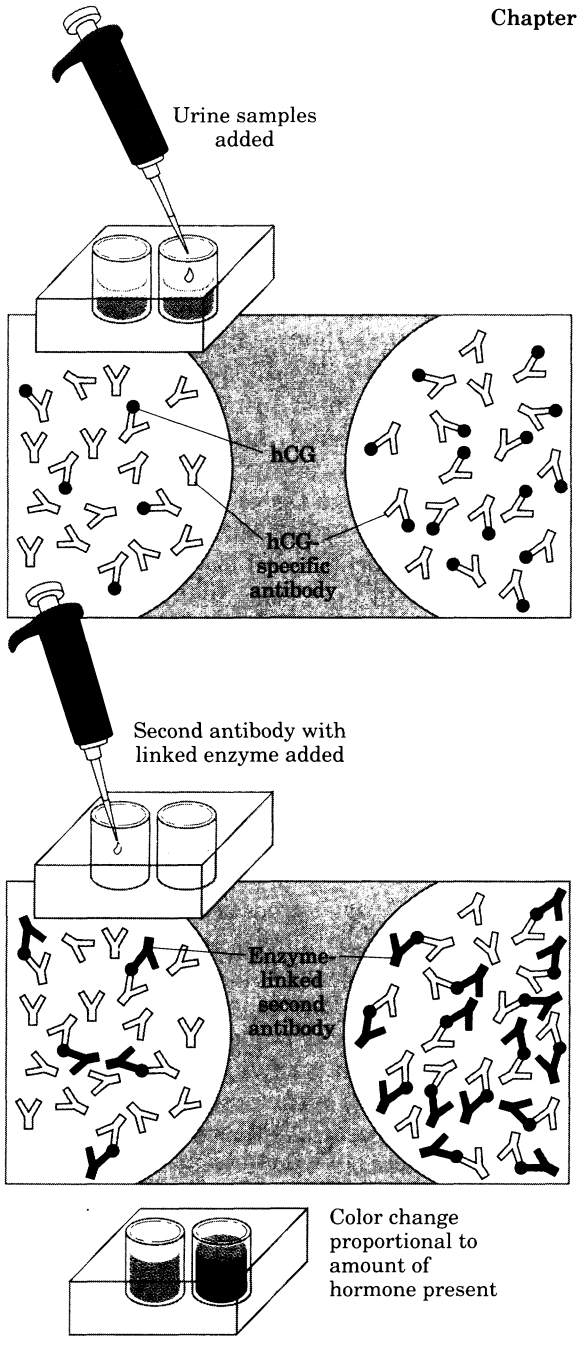
Two types of antibody preparations are in use: **polyclonal** and **monoclonal.** Polyclonal antibodies are those produced by many different types (or populations) of antibody-producing cells in an animal immunized with an antigen (in this case a protein). Each type of cell produces an antibody that binds only to a specific, small part of the antigen protein. Consequently, polyclonal preparations contain a mixture of antibodies that recognize different parts of the protein. Monoclonal antibodies, in contrast, are synthesized by a population of identical cells (a **clone**) grown in cell culture. These antibodies are homogeneous, all recognizing the same specific part of the protein. The techniques for producing monoclonal antibodies were worked out by Georges Köhler and Cesar Milstein.

Antibodies are so exquisitely specific that they can in some cases distinguish between two proteins differing by only a single amino acid.

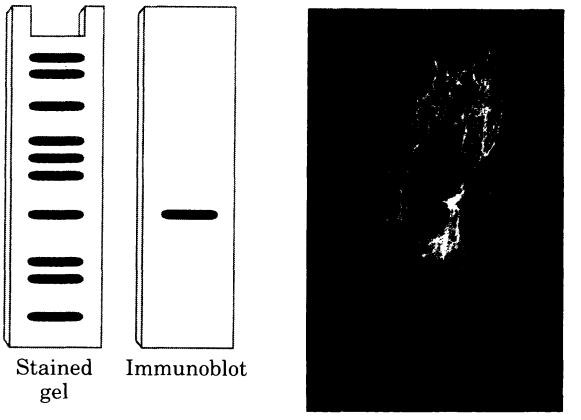
Cesar Milstein



Georges Köhler



(a)



(b) (c)

Figure 6–9 Analytical methods based on the interaction of antibodies with antigen. **(a)** An enzyme-linked immunosorbent assay (ELISA) used in testing for human pregnancy. Human chorionic gonadotropin (hCG), a hormone produced by the placenta, is detectable in maternal urine a few days after conception. In the ELISA, an antibody specific for hCG is attached to the bottom of a well in a plastic tray, to which a few drops of urine are added. If any hCG is present, it will bind to the antibodies. The tube is then washed, and a second antibody (also specific for hCG) is added. This second antibody is linked to an enzyme that catalyzes the conversion of a colorless compound to a colored one; the amount of colored compound produced provides a sensitive measure of the amount of hormone present. The ELISA has been adapted for use in determining the amount of specific proteins in tissue samples, in blood, or in urine.

(b) Immunoblot (or Western blot) technique. Proteins are separated by electrophoresis, then antibodies are used to determine the presence and size of the proteins. After separation, the proteins are transferred electrophoretically from an SDS polyacrylamide gel to a special paper (which makes them more accessible). Specific, labeled antibody is added, then the paper is washed to remove unbound antibody. The label can be a radioactive element, a fluorescent compound, or an enzyme as in the ELISA. The position of the labeled antibody defines the M_r of the detected protein. All of the proteins are seen in the stained gel; only the protein bound to the antibody is seen in the immunoblot.

(c) In immunocytochemistry, labeled antibodies are introduced into cells to reveal the subcellular location of a specific protein. Here, fluorescently labeled antibodies and a fluorescence microscope have been used to locate tubulin filaments in a human fibroblast.

When a mixture of proteins is added to a chromatography column in which the antibody is covalently attached to a resin, the antibody will specifically bind its target protein and retain it on the column while other proteins are washed through. The target protein can then be eluted from the resin by a salt solution or some other agent. This can be a powerful tool for protein purification.

A variety of other analytical techniques rely on antibodies. In each case the antibody is attached to a radioactive label or some other reagent to make it easy to detect. The antibody binds the target protein, and the label reveals its presence in a solution or its location in a gel or even a living cell. Several variations of this procedure are illustrated in Figure 6–9. We shall examine some other aspects of antibodies in chapters to follow; they are of extreme importance in medicine and also tell much about the structure of proteins and the action of genes.

The Covalent Structure of Proteins

All proteins in all species, regardless of their function or biological activity, are built from the same set of 20 amino acids (Chapter 5). What is it, then, that makes one protein an enzyme, another a hormone, another a structural protein, and still another an antibody? How do they differ chemically? Quite simply, proteins differ from each other because each has a distinctive number and *sequence* of amino acid residues. The amino acids are the alphabet of protein structure; they can be arranged in an almost infinite number of sequences to make an almost infinite number of different proteins. A specific sequence of amino acids folds up into a unique three-dimensional structure, and this structure in turn determines the function of the protein.

The amino acid sequence of a protein, or its **primary structure,** can be very informative to a biochemist. No other property so clearly distinguishes one protein from another. This now becomes the focus of the remainder of the chapter. We first consider empirical clues that amino acid sequence and protein function are closely linked, then describe how amino acid sequence is determined, and finally outline the many uses to which this information can be put.

The Function of a Protein Depends on Its Amino Acid Sequence

The bacterium *E. coli* produces about 3,000 different proteins. A human being produces 50,000 to 100,000 different proteins. In both cases, each separate type of protein has a unique structure and this structure confers a unique function. Each separate type of protein also has a unique amino acid sequence. Intuition suggests that the amino acid sequence must play a fundamental role in determining the three-dimensional structure of the protein, and ultimately its function, but is this expectation correct? A quick survey of proteins and how they vary in amino acid sequence provides a number of empirical clues that help substantiate the important relationship between amino acid sequence and biological function. First, as we have already noted, proteins with different functions always have different amino acid sequences. Second, more than 1,400 human genetic diseases have been traced to the production of defective proteins (Table 6–6). Perhaps a third of these proteins are defective because of a single change in the amino acid sequence; hence, if the primary structure is altered, the function of the protein may also be changed. Finally, on comparing proteins with similar functions from different species, we find that these proteins often have similar amino acid sequences. An extreme case is ubiquitin, a 76 amino acid protein involved in regulating the degradation of other proteins. The amino acid sequence of ubiquitin is identical in species as disparate as fruit flies and humans.

Is the amino acid sequence absolutely fixed, or invariant, for a particular protein? No; some flexibility is possible. An estimated 20 to 30% of the proteins in humans are **polymorphic,** having amino acid sequence variants in the human population. Many of these variations in sequence have little or no effect on the function of the protein. Furthermore, proteins that carry out a broadly similar function in distantly related species often differ greatly in overall size and amino acid

sequence. An example is DNA polymerase, the primary enzyme involved in DNA synthesis. The DNA polymerase of a bacterium is very different in much of its sequence from that of a mouse cell.

Table 6–6 A sampling of genetic diseases linked to loss or defect of a single enzyme or protein

Disease	Physiological effects	Affected enzyme or protein
Cystic fibrosis	Abnormal secretion in lungs, pancreas, sweat glands; chronic pulmonary disease generally leading to death in children or young adults	Chloride channel
Lesch–Nyhan syndrome	Neurological defects, self-mutilation, mental retardation	Hypoxanthine-guanine phosphoribosyl transferase
Immunodeficiency disease	Severe loss of immune response	Purine nucleoside phosphorylase
Immunodeficiency disease	Severe loss of immune response (children must live in a sterile bubble)	Adenosine deaminase
Gaucher's disease	Erosion of bones, hip joints; sometimes brain damage	Glucocerebrosidase
Gout, primary	Overproduction of uric acid resulting in recurring attacks of acute arthritis	Phosphoribosyl pyrophosphate synthetase
Rickets, vitamin D-dependent	Short stature, convulsions	25-Hydroxycholecalciferol-1-hydroxylase
Familial hypercholesterolemia	Atherosclerosis resulting from elevated cholesterol levels in blood; sometimes early death from heart failure	Low-density lipoprotein receptor
Tay-Sachs disease	Motor weakness, mental deterioration, death by age 3 yr	Hexosaminidase-A
Sickle-cell anemia	Pain, swelling in hands and feet; can lead to sudden severe pain in bones or joints and death	Hemoglobin

The amino acid sequence of a protein is inextricably linked to its function. Proteins often contain crucial substructures within their amino acid sequence that are essential to their biological functions. The amino acid sequence in other regions might vary considerably without affecting these functions. The fraction of the sequence that is critical varies from protein to protein, complicating the task of relating sequence to structure, and structure to function. Before we can consider this problem further, however, we must examine how sequence information is obtained.

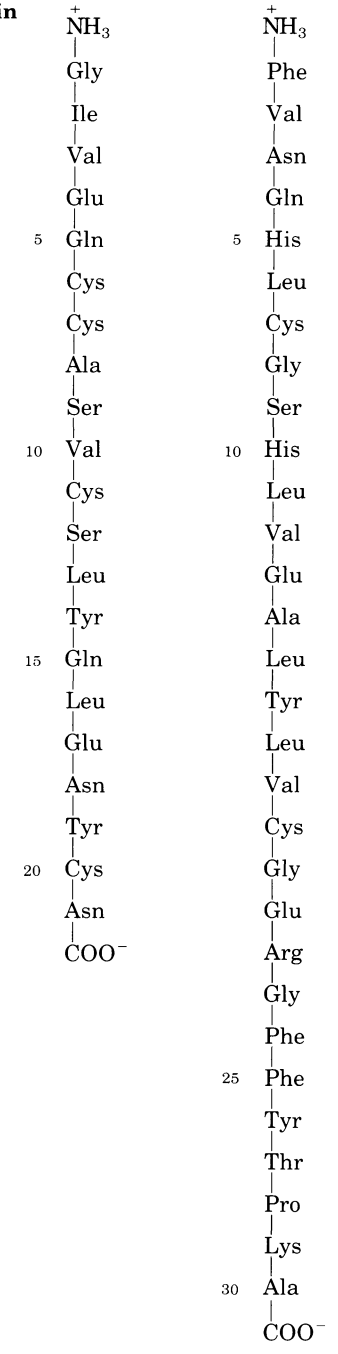


Figure 6–10 The amino acid sequence of the two chains of bovine insulin, which are joined by disulfide cross-linkages. The A chain is identical in human, pig, dog, rabbit, and sperm whale insulins. The B chains of the cow, pig, dog, goat, and horse are identical. Such identities between similar proteins of different species are discussed later in this chapter.

The Amino Acid Sequence of Polypeptide Chains Can Be Determined

Two major discoveries in 1953 ushered in the modern era of biochemistry. In that year James D. Watson and Francis Crick deduced the double-helical structure of DNA and proposed a structural basis for the precise replication of DNA (Chapter 12). Implicit in their proposal was the idea that the sequence of nucleotide units in DNA bears encoded genetic information. In that same year, Frederick Sanger worked out the sequence of amino acids in the polypeptide chains of the hormone insulin (Fig. 6–10), surprising many researchers who had long thought that elucidation of the amino acid sequence of a polypeptide would be a hopelessly difficult task. These achievements together suggested that the nucleotide sequence of DNA and the amino acid sequence of proteins were somehow related. Within just over a decade, the nucleotide code that determines the amino acid sequence of protein molecules had been revealed (Chapter 26).

Today the amino acid sequences of thousands of different proteins from many species are known, determined using principles first developed by Sanger. These methods are still in use, although with many variations and improvements in detail.

Short Polypeptides Are Sequenced Using Automated Procedures

Three procedures are used in the determination of the sequence of a polypeptide chain (Fig. 6–11). The first is to hydrolyze it and determine its amino acid composition (Fig. 6–11a). This information is often valuable in later steps, and can also be useful in itself. Because amino acid composition differs from one protein to the next, it can serve as a kind of fingerprint. It can be used, for example, to help determine whether proteins isolated by different laboratories are the same or different.

Often, the next step is to identify the amino-terminal amino acid residue (Fig. 6–11b). For this purpose Sanger developed the reagent 1-fluoro-2,4-dinitrobenzene (FDNB; see Fig. 5–14). Other reagents used to label the amino-terminal residue are dansyl chloride and dabsyl chloride (see Figs. 5–14 and 5–18). The dansyl derivative is highly fluorescent and can be detected and measured in much lower concentrations than dinitrophenyl derivatives. The dabsyl derivative is intensely colored and also provides greater sensitivity than the dinitrophenyl compounds. These methods destroy the polypeptide and their utility is therefore limited to identification of the amino-terminal residue.

Frederick Sanger

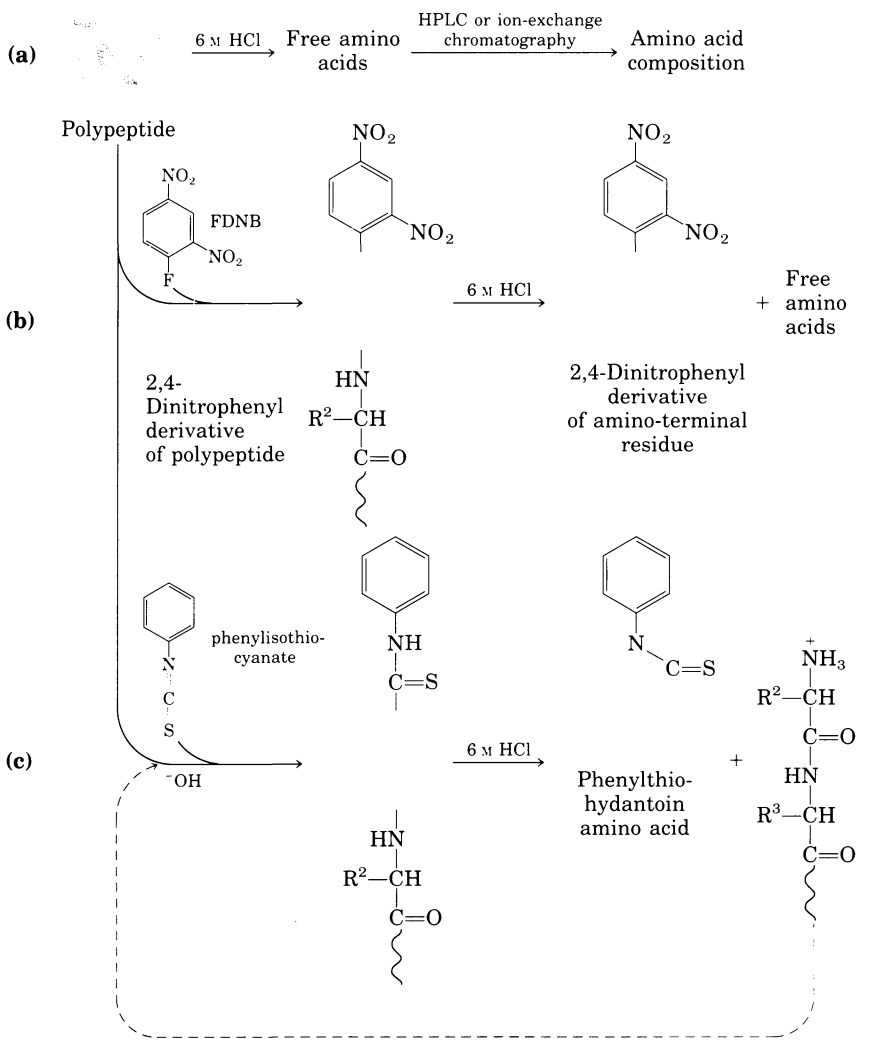


Figure 6–11 Steps in sequencing a polypeptide. **(a)** Determination of amino acid composition and **(b)** identification of the amino-terminal residue are the first steps for many polypeptides. Sanger's method for identifying the amino-terminal residue is shown here. The Edman degradation procedure **(c)** reveals the entire sequence of a peptide. For shorter peptides, this method alone readily yields the entire sequence, and steps **(a)** and **(b)** are often omitted. The latter procedures are useful in the case of larger polypeptides, which are often fragmented into smaller peptides for sequencing (see Fig. 6–13).

To sequence the entire polypeptide, a chemical method devised by Pehr Edman is usually employed. The **Edman degradation** procedure labels and removes only the amino-terminal residue from a peptide, leaving all other peptide bonds intact (Fig. 6–11c). The peptide is reacted with phenylisothiocyanate, and the amino-terminal residue is ultimately removed as a phenylthiohydantoin derivative. After removal and identification of the amino-terminal residue, the *new* amino-terminal residue so exposed can be labeled, removed, and identified by repeating the same series of reactions. This procedure is repeated until the entire sequence is determined. Refinements of each step permit the sequencing of up to 50 amino acid residues in a large peptide.

The many individual steps and the careful bookkeeping required in the determination of the amino acid sequence of long polypeptide chains are usually carried out by programmed and automated analyzers. The Edman degradation is carried out on a programmed machine, called a **sequenator,** which mixes reagents in the proper proportions, separates the products, identifies them, and records the results. Such instruments have greatly reduced the time and labor required to determine the amino acid sequence of polypeptides. These methods are extremely sensitive. Often, less than a microgram of protein is sufficient to determine its complete amino acid sequence.

Large Proteins Must Be Sequenced in Smaller Segments

The overall accuracy for determination of an amino acid sequence generally declines as the length of the polypeptide increases, especially for polypeptides longer than 50 amino acids. The very large polypeptides found in proteins must usually be broken down into pieces small enough to be sequenced efficiently. There are several steps in this process. First, any disulfide bonds are broken, and the protein is cleaved into a set of specific fragments by chemical or enzymatic methods. Each fragment is then purified, and sequenced by the Edman procedure. Finally, the order in which the fragments appear in the original protein is determined and disulfide bonds (if any) are located.

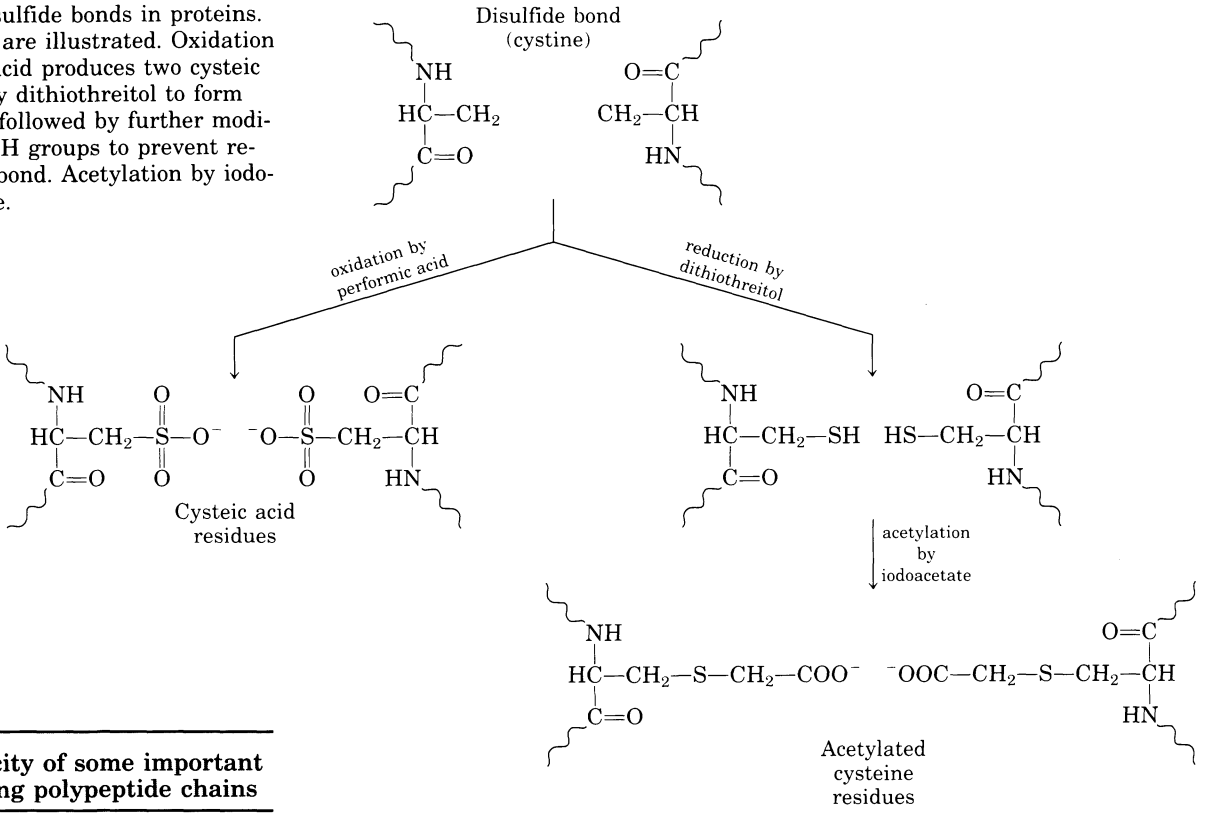


Figure 6–12 Breaking disulfide bonds in proteins. The two common methods are illustrated. Oxidation of cystine with performic acid produces two cysteic acid residues. Reduction by dithiothreitol to form cysteine residues must be followed by further modification of the reactive —SH groups to prevent reformation of the disulfide bond. Acetylation by iodoacetate serves this purpose.

Breaking Disulfide Bonds Disulfide bonds interfere with the sequencing procedure. A cystine residue (p. 116) that has one of its peptide bonds cleaved by the Edman procedure will remain attached to the polypeptide. Disulfide bonds also interfere with the enzymatic or chemical cleavage of the polypeptide (described below). Two approaches to irreversible breakage of disulfide bonds are outlined in Figure 6–12.

Cleaving the Polypeptide Chain Several methods can be used for fragmenting the polypeptide chain. These involve a set of enzymes (proteases) and chemical reagents that cleave peptide chains adjacent to specific amino acid residues (Table 6–7). The digestive enzyme trypsin, for example, catalyzes the hydrolysis of only those peptide bonds in

Table 6–7 The specificity of some important methods for fragmenting polypeptide chains

Treatment*	Cleavage points†
Trypsin	Lys, Arg (C)
Submaxillarus protease	Arg (C)
Chymotrypsin	Phe, Trp, Tyr (C)
Staphylococcus aureus V8 protease	Asp, Glu (C)
Asp-*N*-protease	Asp, Glu (N)
Pepsin	Phe, Trp, Tyr (N)
Cyanogen bromide	Met (C)

* All of the enzymes or reagents listed are available from commercial sources.

† Residues furnishing the primary recognition point for the protease; peptide bond cleavage occurs either on the carbonyl (C) or amino (N) side of the indicated group of amino acids.

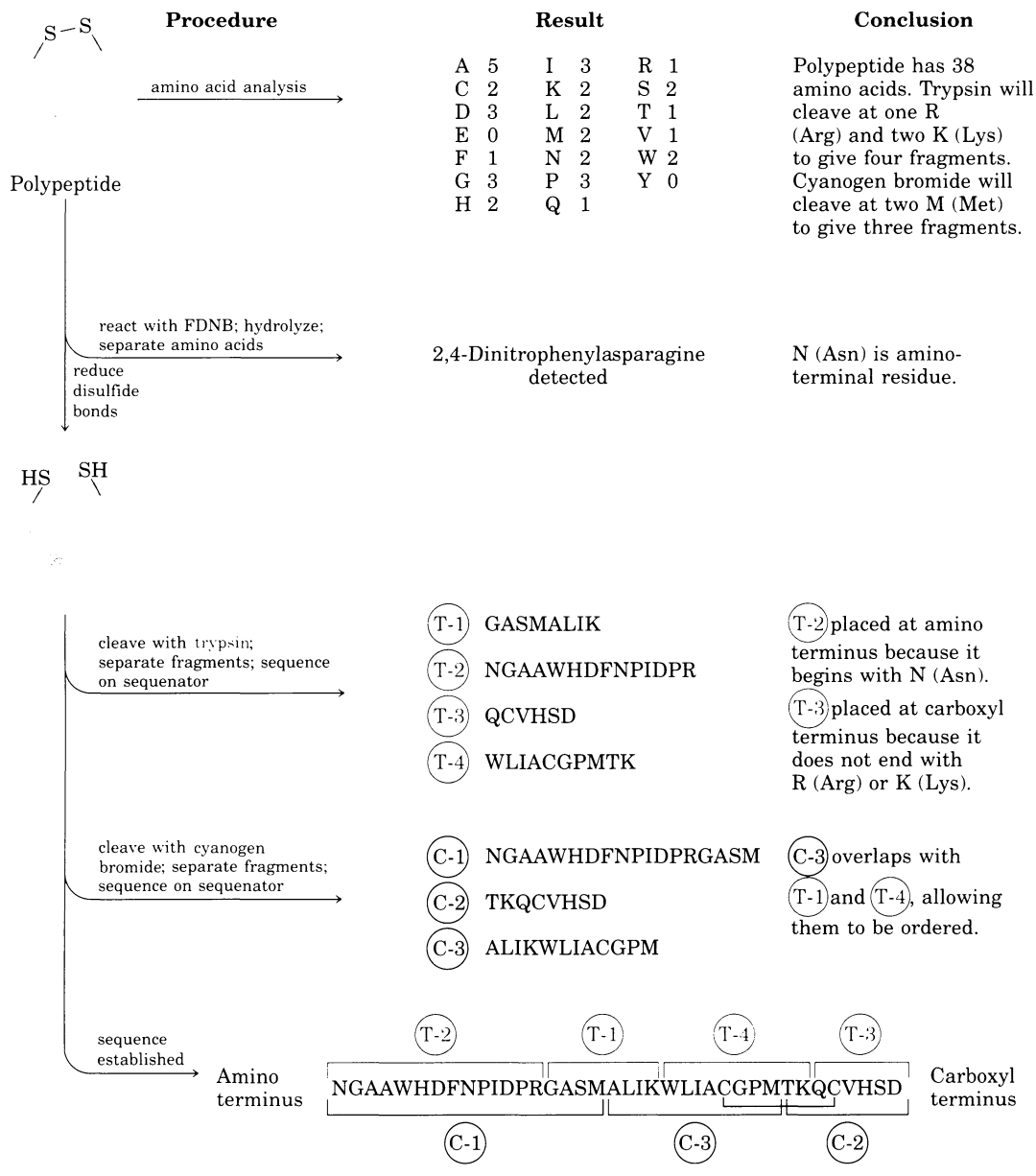


Figure 6–13 Fragmenting proteins prior to sequencing, and placing peptide fragments in their proper order with overlaps. The one-letter abbreviations for amino acids are given in Table 5–1. In this example, there are only two Cys residues, thus one possibility for location of the disulfide bridge (black bracket). In polypeptides with three or more Cys residues, disulfide bridges can be located as described in the text.

which the carbonyl group is contributed by either a Lys or an Arg residue, regardless of the length or amino acid sequence of the chain. The number of smaller peptides produced by trypsin cleavage can thus be predicted from the total number of Lys or Arg residues in the original polypeptide (Fig. 6–13). A polypeptide with five Lys and/or Arg residues will usually yield six smaller peptides on cleavage with trypsin. Moreover, all except one of these will have a carboxyl-terminal Lys or Arg. The fragments produced by trypsin action are separated by chromatographic or electrophoretic methods.

Sequencing of Peptides All the peptide fragments resulting from the action of trypsin are sequenced separately by the Edman procedure.

Ordering Peptide Fragments The order of these trypsin fragments in the original polypeptide chain must now be determined. Another sample of the intact polypeptide is cleaved into small fragments using a different enzyme or reagent, one that cleaves peptide bonds at points other than those cleaved by trypsin. For example, the reagent cyanogen bromide cleaves only those peptide bonds in which the carbonyl group is contributed by Met (Table 6–7). The fragments resulting from this new procedure are then separated and sequenced as before.

The amino acid sequences of each fragment obtained by the two cleavage procedures are examined, with the objective of finding peptides from the second procedure whose sequences establish continuity, because of overlaps, between the fragments obtained by the first cleavage procedure (Fig. 6–13). Overlapping peptides obtained from the second fragmentation yield the correct order of the peptide fragments produced in the first. Moreover, the two sets of fragments can be compared for possible errors in determining the amino acid sequence of each fragment. If the amino-terminal amino acid has been identified before the original cleavage of the protein, this information can be used to establish which fragment is derived from the amino terminus.

If the second cleavage procedure fails to establish continuity between all peptides from the first cleavage, a third or even a fourth cleavage method must be used to obtain a set of peptides that can provide the necessary overlap(s). A variety of proteolytic enzymes with different specificities are available (Table 6–7).

Locating Disulfide Bonds After sequencing is completed, locating the disulfide bonds requires an additional step. A sample of the protein is again cleaved with a reagent such as trypsin, this time without first breaking the disulfide bonds. When the resulting peptides are separated by electrophoresis and compared with the original set of peptides generated by trypsin, two of the original peptides will be missing and a new, larger peptide will appear. The two missing peptides represent the regions of the intact polypeptide that are linked by a disulfide bond.

Amino Acid Sequences Can Be Deduced from DNA Sequences

The approach outlined above is not the only way to obtain amino acid sequences. The development of rapid DNA sequencing methods (Chapter 12), the elucidation of the genetic code (Chapter 26), and the development of techniques for the isolation of genes (Chapter 28) make it possible to deduce the sequence of a polypeptide by determining the sequence of nucleotides in its gene (Fig. 6–14). The two techniques are complementary. When the gene is available, sequencing the DNA can be faster and more accurate than sequencing the protein. If the gene has not been isolated, direct sequencing of peptides is necessary, and this can provide information (e.g., the location of disulfide bonds) not available in a DNA sequence. In addition, a knowledge of the amino acid sequence can greatly facilitate the isolation of the corresponding gene (Chapter 28).

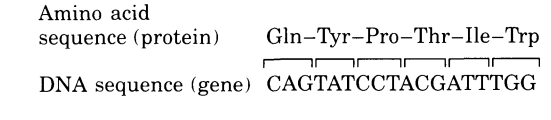


Figure 6–14 Correspondence of DNA and amino acid sequences. Each amino acid is encoded by a specific sequence of three nucleotides (triplet) in DNA. The genetic code is described in detail in Chapter 26.

Amino Acid Sequences Provide Important Biochemical Information

The sequence of amino acids in a protein can offer insights into its three-dimensional structure and its function, cellular location, and evolution. Most of these insights are derived by searching for similarities with other known sequences. Thousands of sequences are known and available in computerized data bases. The comparison of a newly obtained sequence with this large bank of stored sequences often reveals relationships both surprising and enlightening.

The relationship between amino acid sequence and three-dimensional structure, and between structure and function, is not understood in detail. However, a growing number of protein families are being revealed that have at least some shared structural and functional features that can be readily identified on the basis of amino acid sequence similarities alone. For example, there are four major families of proteases, several families of naturally occurring protease inhibitors, a large number of closely related protein kinases, and a similar large number of related protein phosphatases. Individual proteins are generally assigned to families by the degree of similarity in amino acid sequence (identical to other members of the family across 30% or more of the sequence), and proteins in these families generally share at least some structural and functional characteristics. Some families are defined, however, by identities involving only a few amino acids that are critical to a certain function. Many membrane-bound protein receptors share important structural features and have similar amino acid sequences, even though the extracellular molecules they bind are quite different. Even the immunoglobulin family includes a host of extracellular and cell-surface proteins in addition to antibodies.

The similarities may involve the entire protein or may be confined to relatively small segments of it. A number of similar substructures (domains) occur in many functionally unrelated proteins. An example is a 40 to 45 amino acid sequence called the EGF (epidermal growth factor) domain that makes up part of the structure of urokinase, the low-density lipoprotein receptor, several proteins involved in blood clotting, and many others. These domains often fold up into structural configurations that have an unusual degree of stability or that are specialized for a certain environment. Evolutionary relationships can also be inferred from the structural and functional similarities within protein families.

Certain amino acid sequences often serve as signals that determine the cellular location, chemical modification, and half-life of a protein. Special signal sequences, usually at the amino terminus, are used to target certain proteins for export from the cell, while other proteins are distributed to the nucleus, the cell surface, the cytosol, and other cellular locations. Other sequences act as attachment sites for prosthetic groups, such as glycosyl groups in glycoproteins and lipids in lipoproteins. Some of these signals are well characterized, and are easily recognized if they occur in the sequence of a newly discovered protein.

The probability that information about a new protein can be deduced from its primary structure improves constantly with the almost daily addition to the number of published amino acid sequences stored in shared databanks.

Homologous Proteins from Different Species Have Homologous Sequences

Several important conclusions have come from study of the amino acid sequences of homologous proteins from different species. **Homologous proteins** are those that are evolutionarily related. They usually perform the same function in different species; an example is hemoglobin, which has the same oxygen-transport function in different vertebrates. Homologous proteins from different species often have polypeptide chains that are identical or nearly identical in length. Many positions in the amino acid sequence are occupied by the same amino acid in all species and are thus called **invariant residues.** But in other positions there may be considerable variation in the amino acid from one species to another; these are called **variable residues.**

The functional significance of sequence homology can be illustrated by **cytochrome *c*,** an iron-containing mitochondrial protein that transfers electrons during biological oxidations in eukaryotic cells. The polypeptide chain of this protein has a molecular weight of about 13,000 and has about 100 amino acid residues in most species. The amino acid sequences of cytochrome *c* from over 60 different species have been determined, and 27 positions in the chain of amino acid residues are invariant in all species tested (Fig. 6–15), suggesting that they are the most important residues specifying the biological activity of cytochrome *c*. The residues in other positions in the chain exhibit some interspecies variation. There are clear gradations in the number of changes observed in the variable residues. In some positions, all substitutions involve similar amino acid residues (e.g., Arg will replace Lys, both of which are positively charged); these are called **conservative substitutions.** At other positions the substitutions are more random. As we will show in the next chapter, the polypeptide chains of proteins are folded into characteristic and specific conformations and these conformations depend on amino acid sequence. Clearly, the invariant residues are more critical to the structure and function of a protein than the variable ones. Recognizing which amino acids fall into each category is an important step in deciphering the complicated question of how amino acid sequence is translated into a specific three-dimensional structure.

The variable amino acids provide information of another sort. Evolution is sometimes regarded as a theory that is accepted but difficult to test, yet the phylogenetic trees established by taxonomy have been tested and experimentally confirmed through biochemistry. The exam-

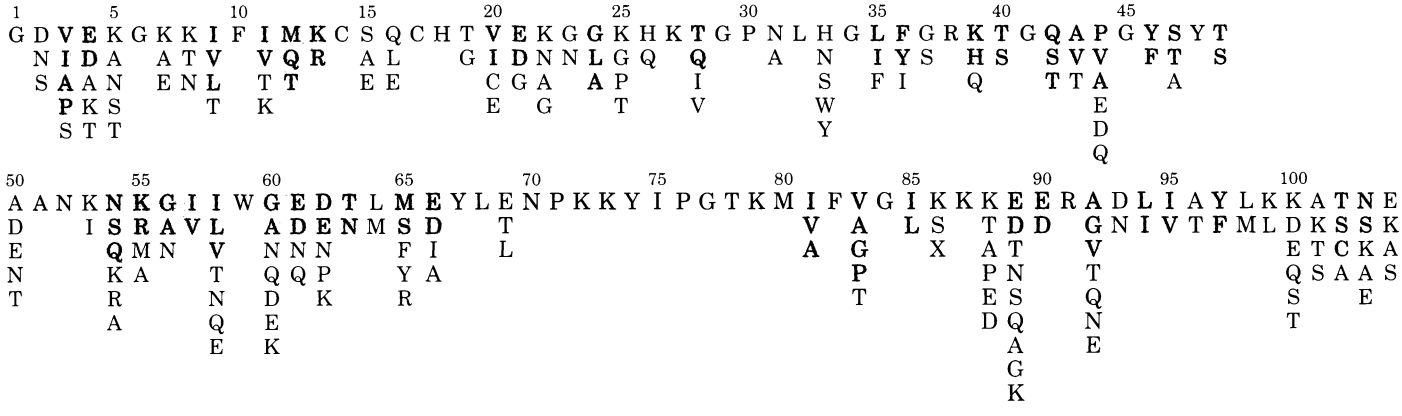


Figure 6–15 The amino acid sequence of human cytochrome *c*. Amino acid substitutions found at different positions in the cytochrome *c* of other species are listed below the sequence of the human protein. The amino acids are color-coded to help distinguish conservative and nonconservative substitutions: invariant amino acids are shaded in yellow, conservative amino acid substitutions are shaded in blue, and nonconservative substitutions are unshaded. X is an unusual amino acid, trimethyllysine. The one-letter abbreviations for amino acids are used here (see Table 5–1).

ination of sequences of cytochrome *c* and other homologous proteins has led to an important conclusion: the number of residues that differ in homologous proteins from any two species is in proportion to the phylogenetic difference between those species. For example, 48 amino acid residues differ in the cytochrome *c* molecules of the horse and of yeast, which are very widely separated species, whereas only two residues differ in the cytochrome *c* of the much more closely related duck and chicken. In fact, the cytochrome *c* molecule has identical amino acid sequences in the chicken and the turkey, and in the pig, cow, and sheep. Information on the number of residue differences between homologous proteins of different species allows the construction of evolutionary maps that show the origin and sequence of development of different animals and plants during the evolution of species (Fig. 6–16). The relationships established by taxonomy and biochemistry agree well.

Figure 6–16 Main branches of the evolutionary tree constructed from the number of amino acid differences between cytochrome *c* molecules of different species. The numbers represent the number of residues by which the cytochrome *c* of a given line of organism differs from its ancestors.

Summary

Cells generally contain thousands of different proteins, each with a different function or biological activity. These functions include enzymatic catalysis, molecular transport, nutrition, cell or organismal motility, structural roles, organismal defense, regulation, and many others. Proteins consist of very long polypeptide chains having from 100 to over 2,000 amino acid residues joined by peptide linkages. Some proteins have several polypeptide chains, which are then referred to as subunits. Simple proteins yield only amino acids on hydrolysis; conjugated proteins contain in addition some other component, such as a metal ion or organic prosthetic group.

Proteins are purified by taking advantage of properties in which they differ, such as size, shape, binding affinities, charge, etc. Purification also requires a method for quantifying or assaying a particular protein in the presence of others. Proteins can be both separated and visualized by electrophoretic methods. Antibodies that specifically bind a certain protein can be used to detect and locate that protein in a solution, a gel, or even in the interior of a cell.

All proteins are made from the same set of 20 amino acids. Their differences in function result from differences in the composition and sequence of their amino acids. The amino acid sequences of polypeptide chains can be established by fragmenting them into smaller pieces using several specific reagents, and determining the amino acid sequence of each fragment by the Edman degradation procedure. The sequencing of suitably sized peptide fragments has been automated. The peptide fragments are then placed in the correct order by finding sequence overlaps between fragments generated by different methods. Protein sequences can also be deduced from the nucleotide sequence of the corresponding gene in the DNA. The amino acid sequence can be compared with the thousands of known sequences, often revealing insights into the structure, function, cellular location, and evolution of the protein.

Homologous proteins from different species show sequence homology: certain positions in the polypeptide chains contain the same amino acids, regardless of the species. In other positions the amino acids may differ. The invariant residues are evidently essential to the function of the protein. The degree of similarity between amino acid sequences of homologous proteins from different species correlates with the evolutionary relationship of the species.

Further Reading

See Chapter 5 for additional useful references.

Properties of Proteins

Creighton, T.E. (1984) *Proteins: Structures and Molecular Properties,* W.H. Freeman and Company, New York.

Dickerson, R.E. & Geis, I. (1983) *Proteins: Structure, Function, and Evolution,* 2nd edn, The Benjamin/Cummings Publishing Company, Menlo Park, CA.

A beautifully illustrated introduction to proteins.

Doolittle, R.F. (1985) Proteins. *Sci. Am.* **253** (October), 88–99.

An overview that highlights evolutionary relationships.

Srinavasan, P.R., Fruton, J.S., & Edsall, J.T. (eds) (1979) The Origins of Modern Biochemistry: A Retrospect on Proteins. *Ann. N.Y. Acad. Sci.* **325.**

A collection of very interesting articles on the history of protein research.

Structure and Function of Proteins. (1989) *Trends Biochem. Sci.* **14** (July).

A special issue devoted to reviews on protein chemistry and protein structure.

Working with Proteins

Hirs, C.H.W. & Timasheff, S.N. (eds) (1983) *Methods in Enzymology,* Vol. 91, Part I: *Enzyme Structure,* Academic Press, Inc., New York.

An excellent collection of authoritative articles on techniques in protein chemistry. Includes information on sequencing.

Kornberg, A. (1990) Why purify enzymes? In *Methods in Enzymology,* Vol. 182: *Guide to Protein Purification* (Deutscher, M.P., ed), pp. 1–5, Academic Press, Inc., New York.

The critical role of classical biochemical methods in a new age.

O'Farrell, P.H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250,** 4007–4021.

An interesting attempt to count all the proteins in the E. coli *cell.*

Plummer, David T. (1987) *An Introduction to Practical Biochemistry,* 3rd edn, McGraw-Hill, London.

Good descriptions of many techniques for beginning students.

Scopes, R.K. (1987) *Protein Purification: Principles and Practice,* 2nd edn, Springer-Verlag, New York.

Tonegawa, S. (1985) The molecules of the immune system. *Sci. Am.* **253** (October), 122–131.

The Covalent Structure of Proteins

Dickerson, R.E. (1972) The structure and history of an ancient protein. *Sci. Am.* **226** (April), 58–72.

A nice summary of information gleaned from interspecies comparisons of cytochrome c *sequences.*

Doolittle, R. (1981) Similar amino acid sequences: chance or common ancestry? *Science* **214,** 149–159.

A good discussion of what can be learned by comparing amino acid sequences.

Hunkapiller, M.W., Strickler, J.E., & Wilson, K.J. (1984) Contemporary methodology for protein structure determination. *Science* **226,** 304–311.

Reidhaar-Olson, J.F. & Sauer, R.T. (1988) Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* **241,** 53–57.

A systematic study of possible amino acid substitutions in a short segment of one protein.

Wilson, A.C. (1985) The molecular basis of evolution. *Sci. Am.* **253** (October), 164–173.

Problems

1. *How Many β-Galactosidase Molecules Are Present in an* E. coli *Cell? E. coli* is a rod-shaped bacterium 2 μm long and 1 μm in diameter. When grown on lactose (a sugar found in milk), the bacterium synthesizes the enzyme β-galactosidase (M_r 450,000), which catalyzes the breakdown of lactose. The average density of the bacterial cell is 1.2 g/mL, and 14% of its total mass is soluble protein, of which 1.0% is β-galactosidase. Calculate the number of β-galactosidase molecules in an *E. coli* cell grown on lactose.

2. *The Number of Tryptophan Residues in Bovine Serum Albumin* A quantitative amino acid analysis reveals that bovine serum albumin contains 0.58% by weight of tryptophan, which has a molecular weight of 204.

(a) Calculate the minimum molecular weight of bovine serum albumin (i.e., assuming there is only one tryptophan residue per protein molecule).

(b) Gel filtration of bovine serum albumin gives a molecular weight estimate of about 70,000. How many tryptophan residues are present in a molecule of serum albumin?

3. *The Molecular Weight of Ribonuclease* Lysine makes up 10.5% of the weight of ribonuclease. Calculate the minimum molecular weight of ribonuclease. The ribonuclease molecule contains ten lysine residues. Calculate the molecular weight of ribonuclease.

4. *The Size of Proteins* What is the approximate molecular weight of a protein containing 682 amino acids in a single polypeptide chain?

5. *Net Electric Charge of Peptides* A peptide isolated from the brain has the sequence

Glu–His–Trp–Ser–Tyr–Gly–Leu–Arg–Pro–Gly

Determine the net charge on the molecule at pH 3. What is the net charge at pH 5.5? At pH 8? At pH 11? Estimate the pI for this peptide. (Use pK_a values for side chains and terminal amino and carboxyl groups as given in Table 5–1.)

6. *The Isoelectric Point of Pepsin* Pepsin of gastric juice (pH ≈ 1.5) has a pI of about 1, much lower than that of other proteins (see Table 6–5). What functional groups must be present in relatively large numbers to give pepsin such a low pI? What amino acids can contribute such groups?

7. *The Isoelectric Point of Histones* Histones are proteins of eukaryotic cell nuclei. They are tightly

bound to deoxyribonucleic acid (DNA), which has many phosphate groups. The pI of histones is very high, about 10.8. What amino acids must be present in relatively large numbers in histones? In what way do these residues contribute to the strong binding of histones to DNA?

8. *Solubility of Polypeptides* One method for separating polypeptides makes use of their differential solubilities. The solubility of large polypeptides in water depends upon the relative polarity of their R groups, particularly on the number of ionized groups: the more ionized groups there are, the more soluble the polypeptide. Which of each pair of polypeptides below is more soluble at the indicated pH?

(a) $(Gly)_{20}$ or $(Glu)_{20}$ at pH 7.0

(b) $(Lys–Ala)_3$ or $(Phe–Met)_3$ at pH 7.0

(c) $(Ala–Ser–Gly)_5$ or $(Asn–Ser–His)_5$ at pH 6.0

(d) $(Ala–Asp–Gly)_5$ or $(Asn–Ser–His)_5$ at pH 3.0

9. *Purification of an Enzyme* A biochemist discovers and purifies a new enzyme, generating the purification table below:

Procedure	Total protein (mg)	Activity (units)
1. Crude extract	20,000	4,000,000
2. Precipitation (salt)	5,000	3,000,000
3. Precipitation (pH)	4,000	1,000,000
4. Ion-exchange chromatography	200	800,000
5. Affinity chromatography	50	750,000
6. Size-exclusion chromatography	45	675,000

(a) From the information given in the table, calculate the specific activity of the enzyme solution after each purification procedure.

(b) Which of the purification procedures used for this enzyme is most effective (i.e., gives the greatest increase in purity)?

(c) Which of the purification procedures is least effective?

(d) Is there any indication in this table that the enzyme is now pure? What else could be done to estimate the purity of the enzyme preparation?

10. *Fragmentation of a Polypeptide Chain by Proteolytic Enzymes* Trypsin and chymotrypsin are specific enzymes that catalyze the hydrolysis of polypeptides at specific locations (Table 6–7). The sequence of the B chain of insulin is shown below. Note that the cystine cross-linkage between the A and B chains has been cleaved through the action of performic acid (see Fig. 6–12).

Phe–Val–Asn–Gln–His–Leu–$CysSO_3^-$–Gly–Ser–His–Leu–Val–Glu–Ala–Leu–Tyr–Leu–Val–$CysSO_3^-$–Gly–Glu–Arg–Gly–Phe–Phe–Tyr–Thr–Pro–Lys–Ala

Indicate the points in the B chain that are cleaved by (a) trypsin and (b) chymotrypsin. Note that these proteases will not remove single amino acids from either end of a polypeptide chain.

11. *Sequence Determination of the Brain Peptide Leucine Enkephalin* A group of peptides that influence nerve transmission in certain parts of the brain has been isolated from normal brain tissue. These peptides are known as opioids, because they bind to specific receptors that bind opiate drugs, such as morphine and naloxone. Opioids thus mimic some of the properties of opiates. Some researchers consider these peptides to be the brain's own pain killers. Using the information below, determine the amino acid sequence of the opioid leucine enkephalin. Explain how your structure is consistent with each piece of information.

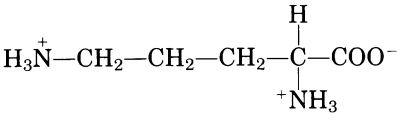
(a) Complete hydrolysis by 1 M HCl at 110 °C followed by amino acid analysis indicated the presence of Gly, Leu, Phe, and Tyr, in a 2:1:1:1 molar ratio.

(b) Treatment of the peptide with 1-fluoro-2,4-dinitrobenzene followed by complete hydrolysis and chromatography indicated the presence of the 2,4-dinitrophenyl derivative of tyrosine. No free tyrosine could be found.

(c) Complete digestion of the peptide with pepsin followed by chromatography yielded a dipeptide containing Phe and Leu, plus a tripeptide containing Tyr and Gly in a 1:2 ratio.

12. *Structure of a Peptide Antibiotic from* Bacillus brevis Extracts from the bacterium *Bacillus brevis* contain a peptide with antibiotic properties. Such peptide antibiotics form complexes with metal ions and apparently disrupt ion transport across the cell membrane, killing certain bacterial species. The structure of the peptide has been determined from the following observations.

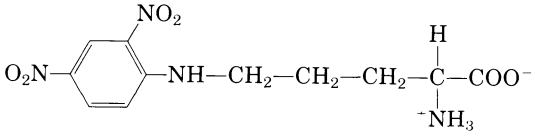
(a) Complete acid hydrolysis of the peptide followed by amino acid analysis yielded equimolar amounts of Leu, Orn, Phe, Pro, and Val. Orn is ornithine, an amino acid not present in proteins but present in some peptides. It has the structure



(b) The molecular weight of the peptide was estimated as about 1,200.

(c) When treated with the enzyme carboxypeptidase, the peptide failed to undergo hydrolysis.

(d) Treatment of the intact peptide with 1-fluoro-2,4-dinitrobenzene, followed by complete hydrolysis and chromatography, yielded only free amino acids and the following derivative:



(Hint: Note that the 2,4-dinitrophenyl derivative involves the amino group of a side chain rather than the α-amino group.)

(e) Partial hydrolysis of the peptide followed by chromatographic separation and sequence analysis yielded the di- and tripeptides below (the amino-terminal amino acid is always at the left):

Leu–Phe Phe–Pro Orn–Leu Val–Orn
Val–Orn–Leu Phe–Pro–Val Pro–Val–Orn

Given the above information, deduce the amino acid sequence of the peptide antibiotic. Show your reasoning. When you have arrived at a structure, go back and demonstrate that it is consistent with *each* experimental observation.

CHAPTER

The Three-Dimensional Structure of Proteins

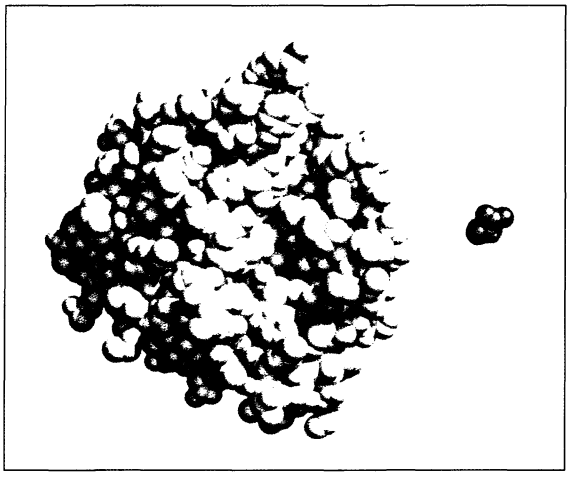


Figure 7–1 The structure of the enzyme chymotrypsin, a globular protein. A molecule of glycine (blue) is shown for size comparison.

The covalent backbone of proteins is made up of hundreds of individual bonds. If free rotation were possible around even a fraction of these bonds, proteins could assume an almost infinite number of three-dimensional structures. Each protein has a specific chemical or structural function, however, strongly suggesting that each protein has a unique three-dimensional structure (Fig. 7–1). The simple fact that proteins can be crystallized provides strong evidence that this is the case. The ordered arrays of molecules in a crystal can generally form only if the molecular units making up the crystal are identical. The enzyme urease (M_r 483,000) was among the first proteins crystallized, by James Sumner in 1926. This accomplishment demonstrated dramatically that even very large proteins are discrete chemical entities with unique structures, and it revolutionized thinking about proteins.

In this chapter, we will explore the three-dimensional structure of proteins, emphasizing several principles. First, the three-dimensional structure of a protein is determined by its amino acid sequence. Second, the function of a protein depends upon its three-dimensional structure. Third, the three-dimensional structure of a protein is unique, or nearly so. Fourth, the most important forces stabilizing the specific three-dimensional structure maintained by a given protein are noncovalent interactions. Finally, even though the structure of proteins is complicated, several common patterns can be recognized.

The relationship between the amino acid sequence and the three-dimensional structure of a protein is an intricate puzzle that has yet to be solved in detail. Polypeptides with very different amino acid sequences sometimes assume similar structures, and similar amino acid sequences sometimes yield very different structures. To find and understand patterns in this biochemical labyrinth requires a renewed appreciation for fundamental principles of chemistry and physics.

Overview of Protein Structure

The spatial arrangement of atoms in a protein is called a **conformation.** The term conformation refers to a structural state that can, without breaking any covalent bonds, interconvert with other structural states. A change in conformation could occur, for example, by rotation about single bonds. Of the innumerable conformations that are theoretically possible in a protein containing hundreds of single bonds, one generally predominates. This is usually the conformation that is ther-

modynamically the most stable, having the lowest Gibbs' free energy (G). Proteins in their functional conformation are called **native** proteins.

What principles determine the most stable conformation of a protein? Although protein structures can seem hopelessly complex, close inspection reveals recurring structural patterns. The patterns involve different levels of structural complexity, and we now turn to a biochemical convention that serves as a framework for much of what follows in this chapter.

There Are Four Levels of Architecture in Proteins

Conceptually, protein structure can be considered at four levels (Fig. 7–2). **Primary structure** includes all the covalent bonds between amino acids and is normally defined by the sequence of peptide-bonded amino acids and locations of disulfide bonds. The relative spatial arrangement of the linked amino acids is unspecified.

Polypeptide chains are not free to take up any three-dimensional structure at random. Steric constraints and many weak interactions stipulate that some arrangements will be more stable than others. **Secondary structure** refers to regular, recurring arrangements in space of adjacent amino acid residues in a polypeptide chain. There are a few common types of secondary structure, the most prominent being the α helix and the β conformation. **Tertiary structure** refers to the spatial relationship among all amino acids in a polypeptide; it is the complete three-dimensional structure of the polypeptide. The boundary between secondary and tertiary structure is not always clear. Several different types of secondary structure are often found within the three-dimensional structure of a large protein. Proteins with several polypeptide chains have one more level of structure: **quaternary structure,** which refers to the spatial relationship of the polypeptides, or subunits, within the protein.

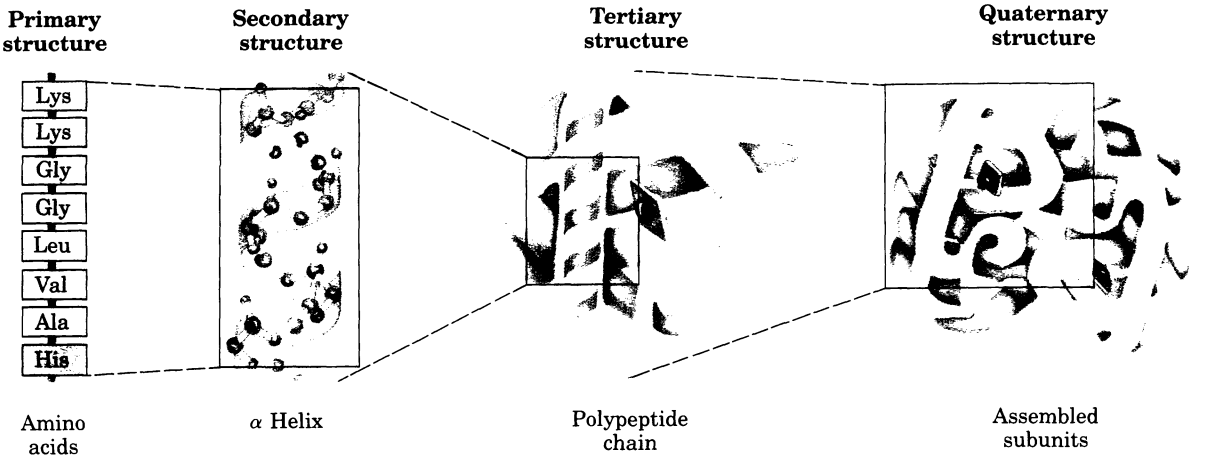


Figure 7–2 Levels of structure in proteins. The *primary structure* consists of a sequence of amino acids linked together by covalent peptide bonds, and includes any disulfide bonds. The resulting polypeptide can be coiled into an α helix, one form of *secondary structure*. The helix is a part of the *tertiary structure* of the folded polypeptide, which is itself one of the subunits that make up the *quaternary structure* of the multimeric protein, in this case hemoglobin.

Continued advances in the understanding of protein structure, folding, and evolution have made it necessary to define two additional structural levels intermediate between secondary and tertiary structure. A stable clustering of several elements of secondary structure is sometimes referred to as **supersecondary structure.** The term is used to describe particularly stable arrangements that occur in many

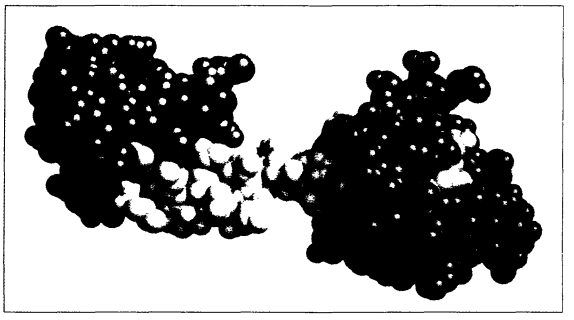


Figure 7–3 The different structural domains in the polypeptide troponin C, a calcium-binding protein associated with muscle. The separate calcium-binding domains, indicated in blue and purple, are connected by a long α helix, shown in white.

different proteins and sometimes many times in a single protein. A somewhat higher level of structure is the **domain.** This refers to a compact region, including perhaps 40 to 400 amino acids, that is a distinct structural unit within a larger polypeptide chain. A polypeptide that is folded into a dumbbell-like shape might be considered to have two domains, one at either end. Many domains fold independently into thermodynamically stable structures. A large polypeptide chain can contain several domains that often are readily distinguishable within the overall structure (Fig. 7–3). In some cases the individual domains have separate functions. As we will see, important patterns exist at each of these levels of structure that provide clues to understanding the overall structure of large proteins.

A Protein's Conformation Is Stabilized Largely by Weak Interactions

The native conformation of a protein is only marginally stable; the difference in free energy between the folded and unfolded states in typical proteins under physiological conditions is in the range of only 20 to 65 kJ/mol. A given polypeptide chain can theoretically assume countless different conformations, and as a result the unfolded state of a protein is characterized by a high degree of conformational entropy. This entropy, and the hydrogen-bonding interactions of many groups in the polypeptide chain with solvent (water), tend to maintain the unfolded state. The chemical interactions that counteract these effects and stabilize the native conformation include disulfide bonds and the weak (noncovalent) interactions described in Chapter 4: hydrogen bonds, and hydrophobic, ionic, and van der Waals interactions. An appreciation of the role of these weak interactions is especially important to understanding how polypeptide chains fold into specific secondary, tertiary, and quaternary structures.

Every time a bond is formed between two atoms, some free energy is released in the form of heat or entropy. In other words, the formation of bonds is accompanied by a favorable (negative) change in free energy. The ΔG for covalent bond formation is generally in the range of -200 to -460 kJ/mol. For weak interactions, $\Delta G = -4$ to -30 kJ/mol. Although covalent bonds are clearly much stronger, weak interactions predominate as a stabilizing force in protein structure because of their number. In general, the protein conformation with the lowest free energy (i.e., the most stable) is the one with the maximum number of weak interactions.

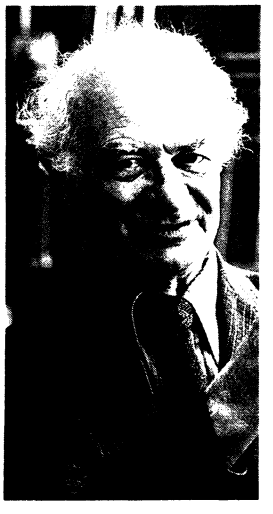
The stability of a protein is not simply the sum of the free energies of formation of the many weak interactions within it, however. We have already noted that the stability of proteins is marginal. Every hydrogen-bonding group in a polypeptide chain was hydrogen bonded to water prior to folding. For every hydrogen bond formed in a protein, hydrogen bonds (of similar strength) between the same groups and water were broken. The net stability contributed by a given weak interaction, or the *difference* in free energies of the folded and unfolded state, is close to zero. We must therefore explain why the native conformation of a protein is favored. The contribution of weak interactions to protein stability can be understood in terms of the properties of water (Chapter 4). Pure water contains a network of hydrogen-bonded water molecules. No other molecule has the hydrogen-bonding potential of water, and other molecules present in an aqueous solution will disrupt

the hydrogen bonding of water to some extent. Optimizing the hydrogen bonding of water around a hydrophobic molecule results in the formation of a highly structured shell or solvation layer of water in the immediate vicinity, resulting in an unfavorable decrease in the entropy of water. The association among hydrophobic or nonpolar groups results in a decrease in this structured solvation layer, or a favorable increase in entropy. As described in Chapter 4, this entropy term is the major thermodynamic driving force for the association of hydrophobic groups in aqueous solution, and hydrophobic amino acid side chains therefore tend to be clustered in a protein's interior, away from water.

The formation of hydrogen bonds and ionic interactions in a protein is also driven largely by this same entropic effect. Polar groups can generally form hydrogen bonds with water and hence are soluble in water. However, the number of hydrogen bonds per unit mass is generally greater for pure water than for any other liquid or solution, and there are limits to the solubility of even the most polar molecules because of the net decrease in hydrogen bonding that occurs when they are present. Therefore, a solvation shell of structured water will also form to some extent around polar molecules. Even though the energy of formation of an intramolecular hydrogen bond or ionic interaction between two polar groups in a macromolecule is largely canceled out by the elimination of such interactions between the same groups and water, the release of structured water when the intramolecular interaction is formed provides an entropic driving force for folding. Most of the net change in free energy that occurs when weak interactions are formed within a protein is therefore derived from the increase in entropy in the surrounding aqueous solution.

Of the different types of weak interactions, hydrophobic interactions are particularly important in stabilizing a protein conformation; the interior of a protein is generally a densely packed core of hydrophobic amino acid side chains. It is also important that any polar or charged groups in the protein interior have suitable partners for hydrogen bonding or ionic interactions. One hydrogen bond makes only a small apparent contribution to the stability of a native structure, but the presence of a single hydrogen-bonding group without a partner in the hydrophobic core of a protein can be so *destabilizing* that conformations containing such a group are often thermodynamically untenable.

Most of the structural patterns outlined in this chapter reflect these two simple rules: (1) hydrophobic residues must be buried in the protein interior and away from water, and (2) the number of hydrogen bonds must be maximized. Insoluble proteins and proteins within membranes (Chapter 10) follow somewhat different rules because of their function or their environment, but weak interactions are still critical structural elements.



Linus Pauling

Robert Corey
1897–1971

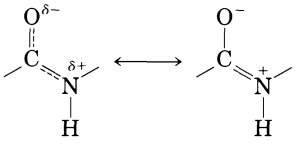
Protein Secondary Structure

Several types of secondary structure are particularly stable and occur widely in proteins. The most prominent are the α helix and β conformations described below. Using fundamental chemical principles and a few experimental observations, Linus Pauling and Robert Corey predicted the existence of these secondary structures in 1951, several years before the first complete protein structure was elucidated.

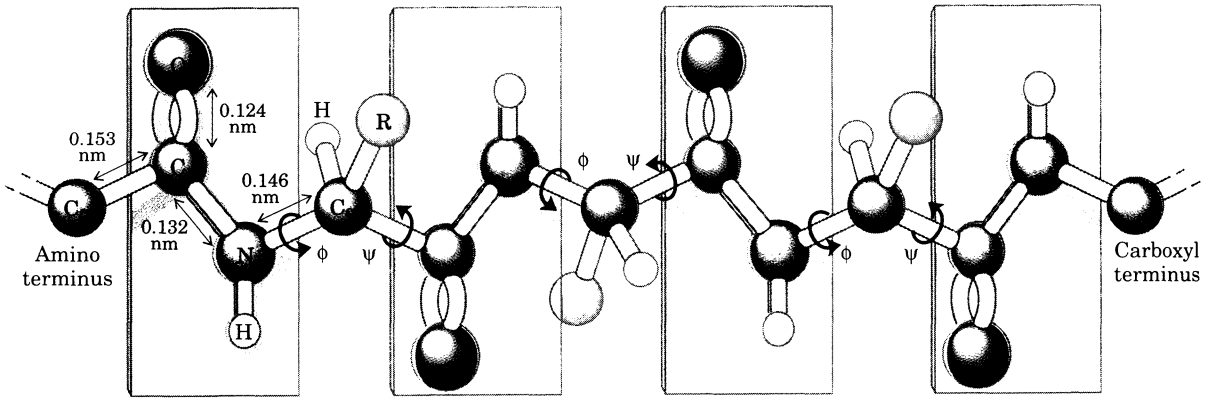
In considering secondary structure, it is useful to classify proteins into two major groups: fibrous proteins, having polypeptide chains arranged in long strands or sheets, and globular proteins, with polypeptide chains folded into a spherical or globular shape. Fibrous proteins play important structural roles in the anatomy and physiology of vertebrates, providing external protection, support, shape, and form. They may constitute one-half or more of the total body protein in larger animals. Most enzymes and peptide hormones are globular proteins. Globular proteins tend to be structurally complex, often containing several types of secondary structure; fibrous proteins usually consist largely of a single type of secondary structure. Because of this structural simplicity, certain fibrous proteins played a key role in the development of the modern understanding of protein structure and provide particularly clear examples of the relationship between structure and function; they are considered in some detail after the general discussion of secondary structure.

The Peptide Bond Is Rigid and Planar

Pauling and Corey began their work on protein structure in the late 1930s by first focusing on the structure of the peptide bond. The α carbons of adjacent amino acids are separated by three covalent bonds, arranged C_α—C—N—C_α. X-ray diffraction studies of crystals of amino acids and of simple dipeptides and tripeptides demonstrated that the amide C—N bond in a peptide is somewhat shorter than the C—N bond in a simple amine and that the atoms associated with the bond are coplanar. This indicated a resonance or partial sharing of two pairs of electrons between the carbonyl oxygen and the amide nitrogen (Fig.



(a)



(c)

Figure 7–4 **(a)** The planar peptide group. Each peptide bond has some double-bond character due to resonance and cannot rotate. The carbonyl oxygen has a partial negative charge and the amide nitrogen a partial positive charge, setting up a small electric dipole. Note that the oxygen and hydrogen atoms in the plane are on opposite sides of the C—N bond. This is the trans configuration. Virtually all peptide bonds in proteins occur in this configuration, although an exception is noted in Fig. 7–10. **(b)** Three bonds separate sequential C_α carbons in a polypeptide chain. The N—C_α and C_α—C bonds can rotate, with bond angles designated ϕ and ψ, respectively. **(c)** Limited rotation can occur around two of the three types of bonds in a polypeptide chain. The C—N bonds in the planar peptide groups (shaded in blue), which make up one-third of all the backbone bonds, are not free to

7–4a). The oxygen has a partial negative charge and the nitrogen a partial positive charge, setting up a small electric dipole. The four atoms of the peptide group lie in a single plane, in such a way that the oxygen atom of the carbonyl group and the hydrogen atom of the amide nitrogen are trans to each other. From these studies Pauling and Corey concluded that the amide C—N bonds are unable to rotate freely because of their partial double-bond character. The backbone of a polypeptide chain can thus be pictured as a series of rigid planes separated by substituted methylene groups, —CH(R)— (Fig. 7–4c). The rigid peptide bonds limit the number of conformations that can be assumed by a polypeptide chain.

Rotation is permitted about the N—C_α and the C_α—C bonds. By convention the bond angles resulting from rotations are labeled ϕ (phi) for the N—C_α bond and ψ (psi) for the C_α—C bond. Again by convention, both ϕ and ψ are defined as 0° in the conformation in which the two peptide bonds connected to a single α carbon are in the same plane, as shown in Figure 7–4d. In principle, ϕ and ψ can have any value between −180° and +180°, but many values of ϕ and ψ are prohibited by steric interference between atoms in the polypeptide backbone and amino acid side chains. The conformation in which ϕ and ψ are both 0° is prohibited for this reason; this is used merely as a reference point for describing the angles of rotation.

Every possible secondary structure is described completely by the two bond angles ϕ and ψ that are repeated at each residue. Allowed values for ϕ and ψ can be shown graphically by simply plotting ψ versus ϕ, an arrangement known as a **Ramachandran plot.** The Ramachandran plot in Figure 7–5 shows the conformations permitted for most amino acid residues.

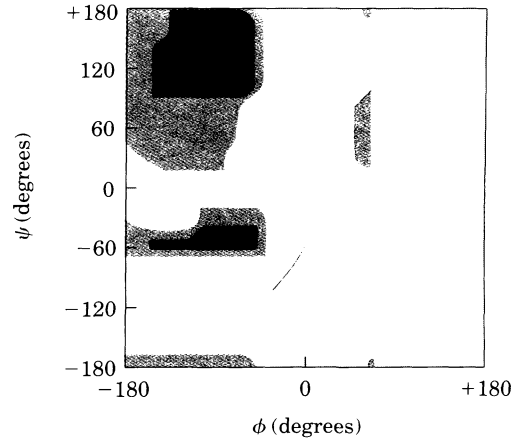
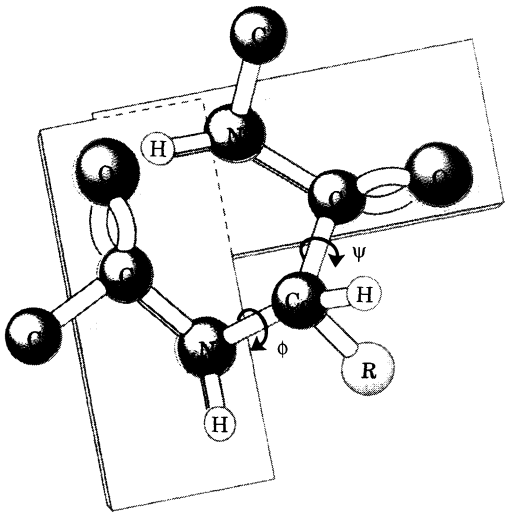


Figure 7–5 A Ramachandran plot. The theoretically allowed conformations of peptides are shown, defined by the values of ϕ and ψ. The shaded areas reflect conformations that can be take up by all amino acids (dark shading) or all except valine and isoleucine (medium shading); the lightest shading reflects conformations that are somewhat unstable but are found in some protein structures.



(d)

rotate. Other single bonds in the backbone may also be rotationally hindered, depending on the size and charge of the R groups. **(d)** By convention, ϕ and ψ are both defined as 0° when the two peptide bonds flanking an α carbon are in the same plane. In a protein, this conformation is prohibited by steric overlap between a carbonyl oxygen and an α-amino hydrogen atom.

The α Helix Is a Common Protein Secondary Structure

Pauling and Corey were aware of the importance of hydrogen bonds in orienting polar chemical groups such as the —C=O and —N—H groups of the peptide bond. They also had the experimental results of William Astbury, who in the 1930s had conducted pioneering x-ray studies of proteins. Astbury demonstrated that the protein that makes up hair and wool (the fibrous protein α-keratin) has a regular structure that repeats every 0.54 nm. With this information and their data on the peptide bond, and with the help of precisely constructed models, Pauling and Corey set out to determine the likely conformations of protein molecules.

The simplest arrangement the polypeptide chain could assume with its rigid peptide bonds (but with the other single bonds free to rotate) is a helical structure, which Pauling and Corey called the **α helix** (Fig. 7–6). In this structure the polypeptide backbone is tightly wound around the long axis of the molecule, and the R groups of the amino acid residues protrude outward from the helical backbone. The repeating unit is a single turn of the helix, which extends about 0.56 nm along the long axis, corresponding closely to the periodicity

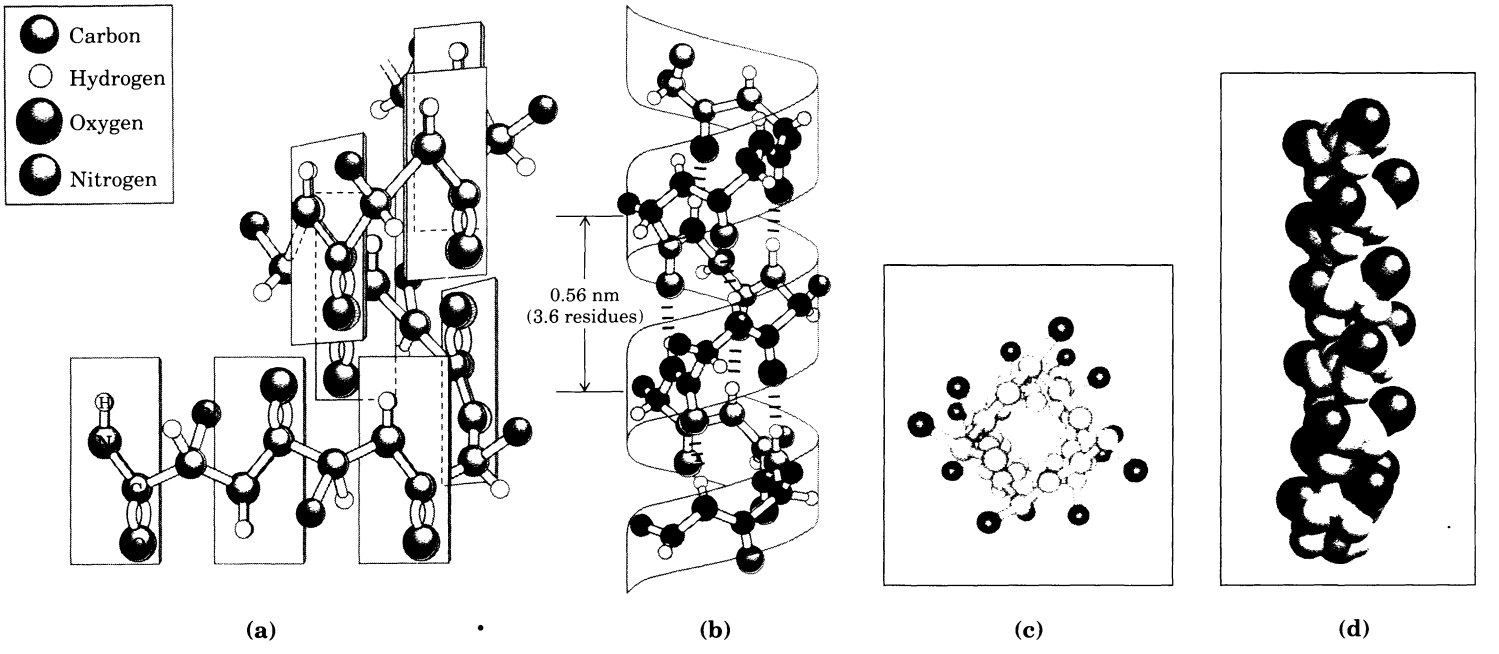
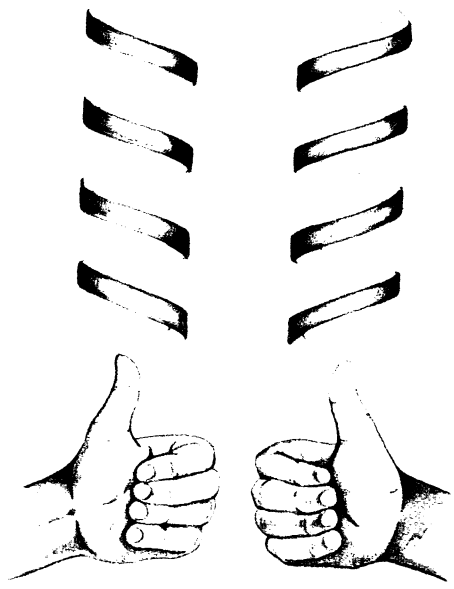


Figure 7–6 Four models of the α helix, showing different aspects of its structure. **(a)** Formation of a right-handed α helix. The planes of the rigid peptide bonds are parallel to the long axis of the helix. **(b)** Ball-and-stick model of a right-handed α helix, showing the intrachain hydrogen bonds. The repeat unit is a single turn of the helix, 3.6 residues. **(c)** The α helix as viewed from one end, looking down the longitudinal axis. Note the positions of the R groups, represented by red spheres. **(d)** A space-filling model of the α helix.

BOX 7–1

Knowing the Right Hand from the Left

There is a simple method for determining the handedness of a helical structure, whether right-handed or left-handed. Make fists of your two hands with thumbs outstretched and pointing away from you. Looking at your right hand, think of a helix spiraling away in the direction indicated by your right thumb, and the spiral occurring in the direction in which the other four fingers are curled as shown (clockwise). The resulting helix is right-handed. Repeating the process with your left hand will produce an image of a left-handed helix, which rotates in the counterclockwise direction as it spirals away from you.



Astbury observed on x-ray analysis of hair keratin. The amino acid residues in an α helix have conformations with $\psi = -45°$ to $-50°$ and $\phi = -60°$, and each helical turn includes 3.6 amino acids. The twisting of the helix has a right-handed sense (Box 7–1) in the most common form of the α helix, although a very few left-handed variants have been observed.

The α helix is one of two prominent types of secondary structure in proteins. It is the predominant structure in α-keratins. In globular proteins, about one-fourth of all amino acid residues are found in α helices, the fraction varying greatly from one protein to the next.

Why does such a helix form more readily than many other possible conformations? The answer is, in part, that it makes optimal use of internal hydrogen bonds. The structure is stabilized by a hydrogen bond between the hydrogen atom attached to the electronegative nitrogen atom of each peptide linkage and the electronegative carbonyl oxygen atom of the fourth amino acid on the amino-terminal side of it in the helix (Fig. 7–6b). Every peptide bond of the chain participates in such hydrogen bonding. Each successive coil of the α helix is held to the adjacent coils by several hydrogen bonds, which in summation give the entire structure considerable stability.

Further model-building experiments have shown that an α helix can form with either L- or D-amino acids. However, all residues must be of one stereoisomeric series; a D-amino acid will disrupt a regular structure consisting of L-amino acids, and vice versa. Naturally occurring L-amino acids can form either right- or left-handed helices, but, with rare exceptions, only right-handed helices are found in proteins.

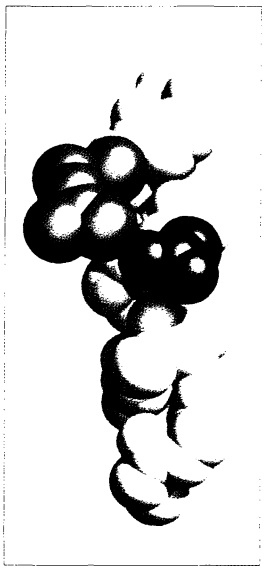


Figure 7–7 Interactions between R groups of amino acids three residues apart in an α helix. An ionic interaction between Asp^{100} and Arg^{103} in an α-helical region of the protein troponin C is shown in this space-filling model. The polypeptide backbone (carbons, α-amino nitrogens, and α-carbonyl oxygens) is shown in white for a helix segment about 12 amino acids long. The only side chains shown are the interacting Asp and Arg residues, with the aspartate in red and the arginine in blue. The side chain interaction illustrated occurs within the white connecting helix in Fig. 7–3.

Amino Acid Sequence Affects α Helix Stability

Not all polypeptides can form a stable α helix. Additional interactions occur between amino acid side chains that can stabilize or destabilize this structure. For example, if a polypeptide chain has many Glu residues in a long block, this segment of the chain will not form an α helix at pH 7.0. The negatively charged carboxyl groups of adjacent Glu residues repel each other so strongly that they overcome the stabilizing influence of hydrogen bonds on the α helix. For the same reason, if there are many adjacent Lys and/or Arg residues, with positively charged R groups at pH 7.0, they will also repel each other and prevent formation of the α helix. The bulk and shape of certain R groups can also destabilize the α helix or prevent its formation. For example, Asn, Ser, Thr, and Leu residues tend to prevent formation of the α helix if they occur close together in the chain.

The twist of an α helix ensures that critical interactions occur between an amino acid side chain and the side chain three (and sometimes four) residues away on either side of it (Fig. 7–7). Positively charged amino acids are often found three residues away from negatively charged amino acids, permitting the formation of an ionic interaction. Two aromatic amino acids are often similarly spaced, resulting in a hydrophobic interaction.

A minor constraint on the formation of the α helix is the presence of Pro residues. In proline the nitrogen atom is part of a rigid ring (Fig. 5–6), and rotation about the N—C_α bond is not possible. In addition, the nitrogen atom of a Pro residue in peptide linkage has no substituent hydrogen-to-hydrogen bond with other residues. For these reasons, proline is only rarely found within an α helix.

A final factor affecting the stability of an α helix is the identity of the amino acids located near the ends of the α-helical segment of a polypeptide. A small electric dipole exists in each peptide bond (see Fig. 7–4). These dipoles add across the hydrogen bonds in the helix so that the net dipole increases as helix length increases (Fig. 7–8). The four amino acids at either end of the helix do not participate fully in the helix hydrogen bonds. The partial positive and negative charges of the helix dipole actually reside on the peptide amino and carbonyl groups near the amino-terminal and carboxyl-terminal ends of the helix, respectively. For this reason, negatively charged amino acids are often found near the amino terminus of the helical segment, where they have a stabilizing interaction with the positive charge of the helix dipole; a positively charged amino acid at the amino-terminal end is destabilizing. The opposite is true at the carboxyl-terminal end of the helical segment.

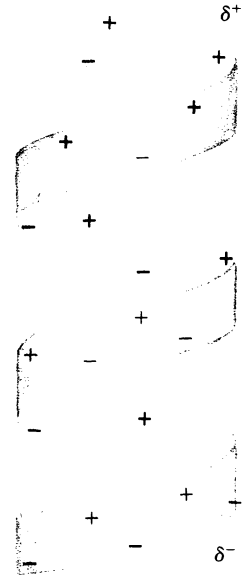


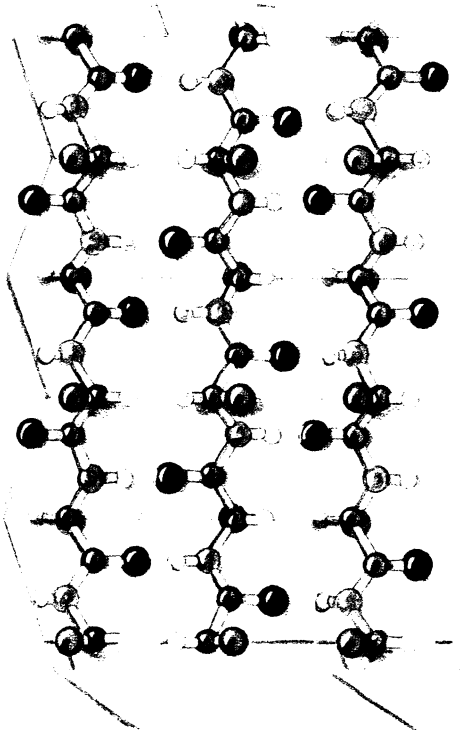
Figure 7–8 The electric dipole of a peptide bond (Fig. 7–4a) is transmitted along an α-helical segment through the intrachain hydrogen bonds, resulting in an overall helix dipole. In this illustration, the amino and carbonyl constituents of each peptide bond are indicated by + and − symbols, respectively. Unbonded amino and carbonyl constituents in the peptide bonds near either end of the α-helical region are shown in red.

Thus there are five different kinds of constraints that affect the stability of an α helix: (1) the electrostatic repulsion (or attraction) between amino acid residues with charged R groups, (2) the bulkiness of adjacent R groups, (3) the interactions between amino acid side chains spaced three (or four) residues apart, (4) the occurrence of Pro residues, and (5) the interaction between amino acids at the ends of the helix and the electric dipole inherent to this structure.

The β Conformation Organizes Polypeptide Chains into Sheets

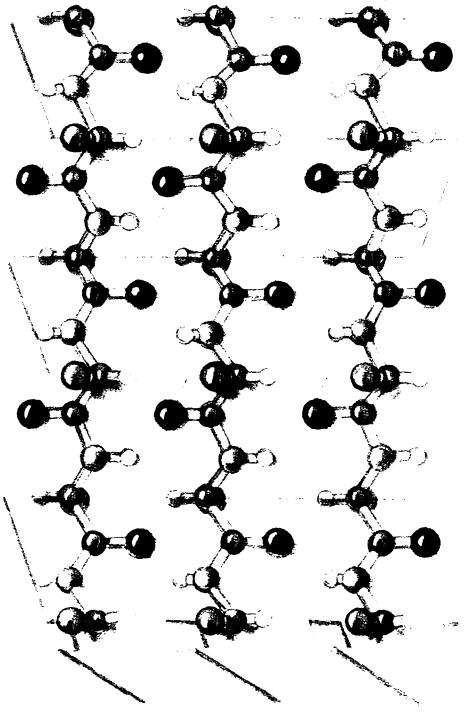
Pauling and Corey predicted a second type of repetitive structure, the **β conformation.** This is the more extended conformation of the polypeptide chains, as seen in the silk protein fibroin (a member of a class of fibrous proteins called β-keratins), and its structure has been confirmed by x-ray analysis. In the β conformation, which like the α helix is common in proteins, the backbone of the polypeptide chain is extended into a zigzag rather than helical structure (Fig. 7–9). In fibroin the zigzag polypeptide chains are arranged side by side to form a structure resembling a series of pleats; such a structure is called a β pleated sheet. In the β conformation the hydrogen bonds can be either intrachain, or interchain between the peptide linkages of adjacent polypeptide chains. All the peptide linkages of β-keratin participate in interchain hydrogen bonding. The R groups of adjacent amino acids protrude in opposite directions from the zigzag structure, creating an alternating pattern as seen in the side view (Fig. 7–9c).

Figure 7–9 The β conformation of polypeptide chains. Views show the R groups extending out from the β pleated sheet and emphasize the pleated sheet described by the planes of the peptide bonds. Hydrogen-bond cross-links between adjacent chains are also shown. **(a)** Antiparallel β sheets, in which the amino-terminal to carboxyl-terminal orientation of adjacent chains (arrows) is inverse. **(b)** Parallel β sheets. **(c)** Silk fibers are made up of the protein fibroin. Its structure consists of layers of antiparallel β sheets rich in Ala (purple) and Gly (yellow) residues. The small side chains interdigitate and allow close packing of each layered sheet, as shown in this side view.



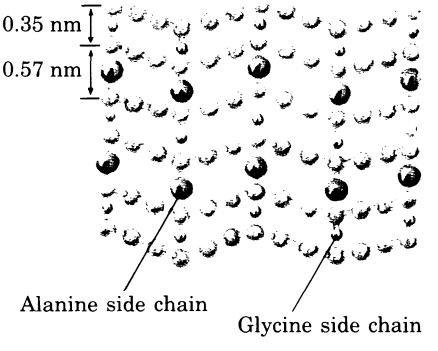
Antiparallel

(a)



Parallel

(b)



(c)

The adjacent polypeptide chains in a β pleated sheet can be either parallel (having the same amino-to-carboxyl polypeptide orientation) or antiparallel (having the opposite amino-to-carboxyl orientation). The structures are similar, although the repeat period is shorter for the parallel conformation (0.65 nm, as opposed to 0.7 nm for antiparallel).

In some structural situations there are limitations to the kinds of amino acids that can occur in the β structure. When two or more pleated sheets are layered closely together within a protein, the R groups of the amino acid residues on the contact surfaces must be relatively small. β-Keratins such as silk fibroin and the protein of spider webs have a very high content of Gly and Ala residues, those with the smallest R groups. Indeed, in silk fibroin Gly and Ala alternate over large parts of the sequence (Fig. 7–9c).

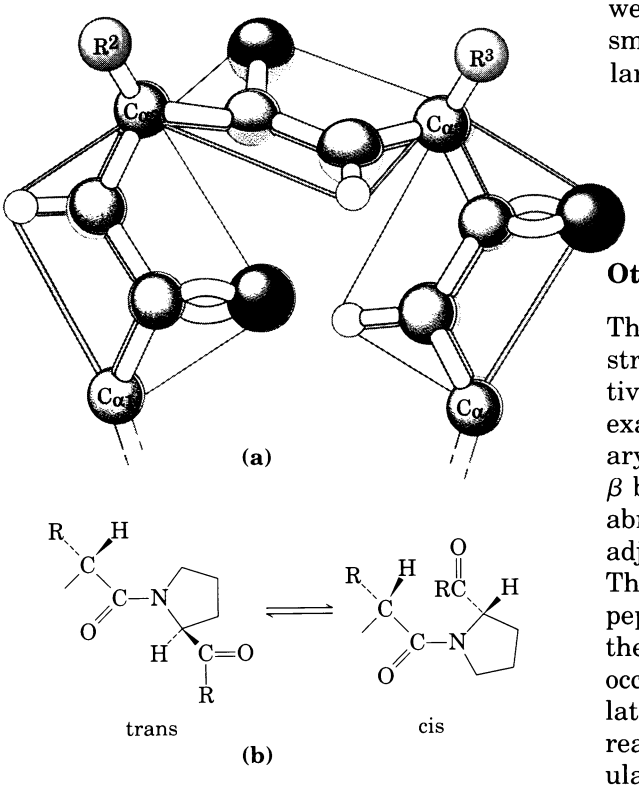


Figure 7–10 Structure of a β turn or β bend. **(a)** Note the hydrogen bond between the peptide groups of the first and fourth residues involved in the bend. **(b)** The trans and cis isomers of a peptide bond involving the imino nitrogen of proline. Over 99.95% of the peptide bonds between amino acid residues other than Pro are in the trans configuration. About 6% of the peptide bonds involving the imino nitrogen of proline, however, are in the cis configuration, and many of these occur at β turns.

Other Secondary Structures Occur in Some Proteins

The α helix and the β conformation are the major repetitive secondary structures easily recognized in a wide variety of proteins. Other repetitive structures exist, often in only one or a few specialized proteins. An example is the collagen helix (see Fig. 7–14). One other type of secondary structure is common enough to deserve special mention. This is a β bend or β turn (Fig. 7–10), often found where a polypeptide chain abruptly reverses direction. (These turns often connect the ends of two adjacent segments of an antiparallel β pleated sheet, hence the name.) The structure is a tight turn (~180°) involving four amino acids. The peptide groups flanking the first amino acid are hydrogen bonded to the peptide groups flanking the fourth. Gly and Pro residues often occur in β turns, the former because it is small and flexible; and the latter because peptide bonds involving the imino nitrogen of proline readily assume the cis configuration (Fig. 7–10b), a form that is particularly amenable to a tight turn. β Turns are often found near the surface of a protein.

Secondary Structure Is Affected by Several Factors

The α helix and β conformation are stable because steric repulsion is minimized and hydrogen bonding is maximized. As shown by a Ramachandran plot, these structures fall within a range of sterically allowed structures that is relatively restricted. Values of ϕ and ψ for common secondary structures are shown in Figure 7–11. Most values of ϕ and ψ for amino acid residues, taken from known protein structures, fall into the expected regions, with high concentrations near the α helix and β conformation values as expected. The only amino acid often found in a conformation outside these regions is glycine. Because its hydrogen side chain is small, a Gly residue can take up many conformations that are sterically forbidden for other amino acids.

Some amino acids are accommodated in the different types of secondary structures better than others. An overall summary is presented in Figure 7–12. Some biases, such as the presence of Pro and Gly residues in β turns, can be explained readily; other evident biases are not understood.

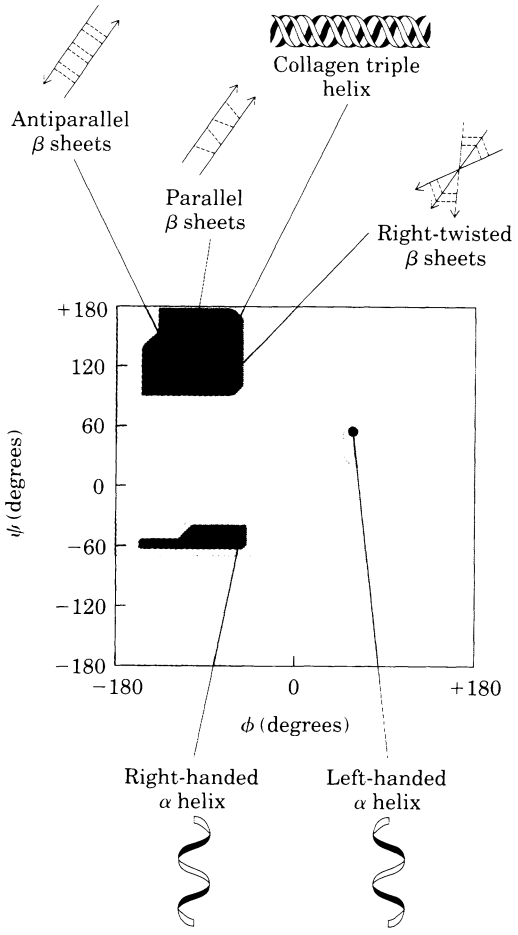


Figure 7–11 A Ramachandran plot. The values of ϕ and ψ for the various secondary structures are overlaid on the plot from Fig. 7–5.

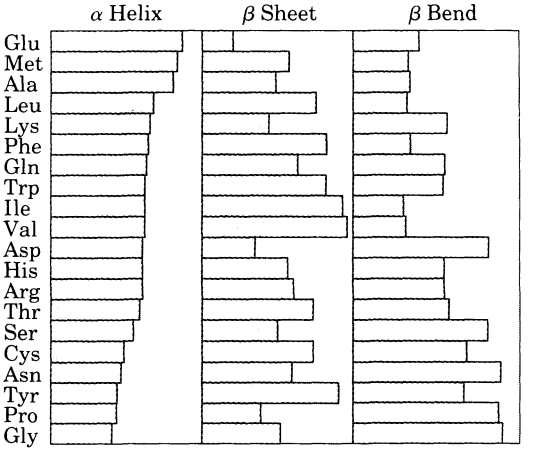


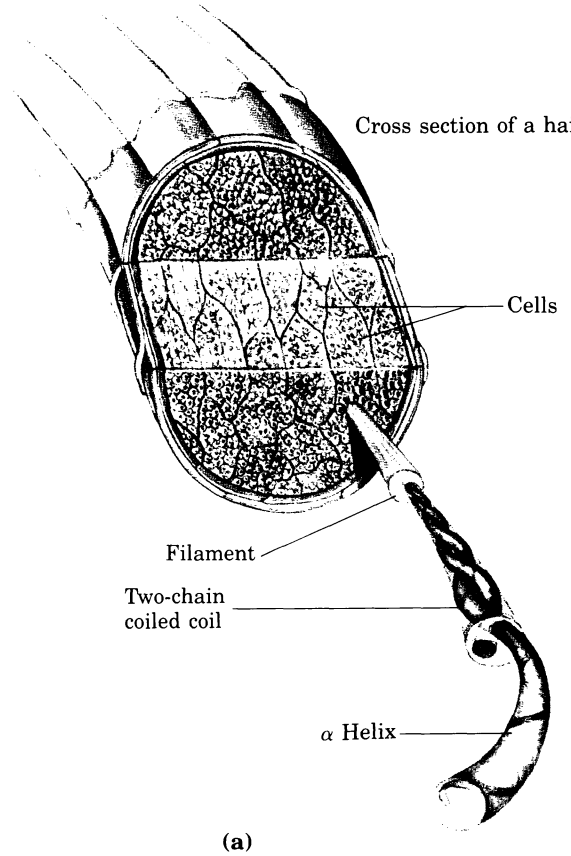
Figure 7–12 Relative probabilities that a given amino acid will occur in the three common types of secondary structure.

Fibrous Proteins Are Adapted for a Structural Function

α-Keratin, collagen, and elastin provide clear examples of the relationship between protein structure and biological function (Table 7–1). These proteins share properties that give strength and/or elasticity to structures in which they occur. They have relatively simple structures, and all are insoluble in water, a property conferred by a high concentration of hydrophobic amino acids both in the interior of the protein and on the surface. These proteins represent an exception to the rule that hydrophobic groups must be buried. The hydrophobic core of the molecule therefore contributes less to structural stability, and covalent bonds assume an especially important role.

Table 7–1 Secondary structures and properties of fibrous proteins

Structure	Characteristics	Examples of occurrence
α Helix, cross-linked by disulfide bonds	Tough, insoluble protective structures of varying hardness and flexibility	α-Keratin of hair, feathers, and nails
β Conformation	Soft, flexible filaments	Fibroin of silk
Collagen triple helix	High tensile strength, without stretch	Collagen of tendons, bone matrix
Elastin chains cross-linked by desmosine and lysinonorleucine	Two-way stretch with elasticity	Elastin of ligaments



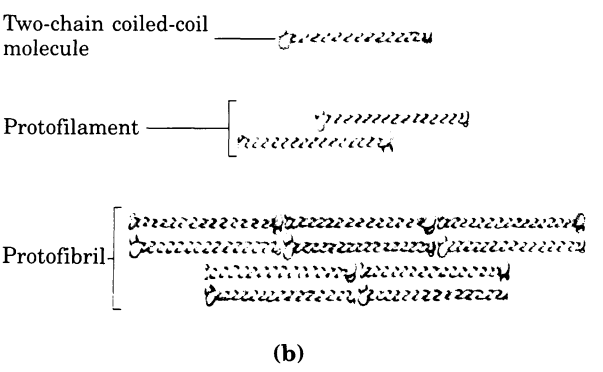


Figure 7–13 **(a)** Hair α-keratin is an elongated α helix with somewhat thicker domains near the amino and carboxy termini. Pairs of these helices are interwound, probably in a left-handed sense, to form two-chain coiled coils. These then combine in higher-order structures called protofilaments and protofibrils, as shown in **(b)**. (About four protofibrils combine to form a filament.) The individual two-chain coiled coils in the various substructures also appear to be interwound, but the handedness of the interwinding and other structural details are unknown.

***α*-Keratin** and **collagen** have evolved for strength. In vertebrates, α-keratins constitute almost the entire dry weight of hair, wool, feathers, nails, claws, quills, scales, horns, hooves, tortoise shell, and much of the outer layer of skin. Collagen is found in connective tissue such as tendons, cartilage, the organic matrix of bones, and the cornea of the eye. The polypeptide chains of both proteins have simple helical structures. The α-keratin helix is the right-handed α helix found in many other proteins (Fig. 7–13). However, the collagen helix is unique. It is left-handed (see Box 7–1) and has three amino acid residues per turn (Fig. 7–14). In both α-keratin and collagen, a few amino acids predominate. α-Keratin is rich in the hydrophobic residues Phe, Ile, Val, Met, and Ala. Collagen is 35% Gly, 11% Ala, and 21% Pro and Hyp (hydroxyproline; see Fig. 5–8). The unusual amino acid content of collagen is imposed by structural constraints unique to the collagen helix. The amino acid sequence in collagen is generally a repeating tripeptide unit, Gly–X–Pro or Gly–X–Hyp, where X can be any amino acid. The food product gelatin is derived from collagen. Although it is protein, it has little nutritional value because collagen lacks significant amounts of many amino acids that are essential in the human diet.

In both α-keratin and collagen, strength is amplified by wrapping multiple helical strands together in a superhelix, much the way strings are twisted to make a strong rope (Figs. 7–13, 7–14). In both proteins the helical path of the supertwists is opposite in sense to the twisting of the individual polypeptide helices, a conformation that permits the closest possible packing of the multiple polypeptide chains. The super-

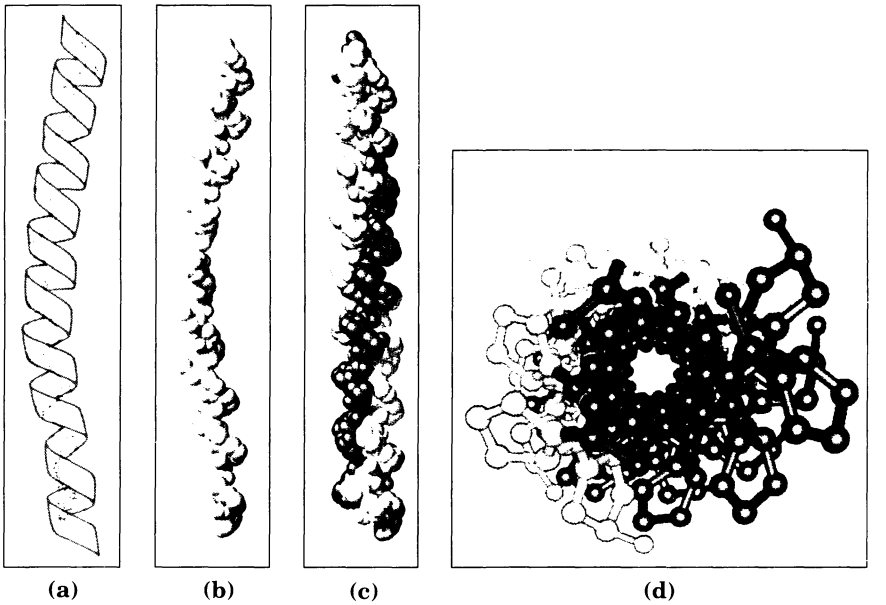
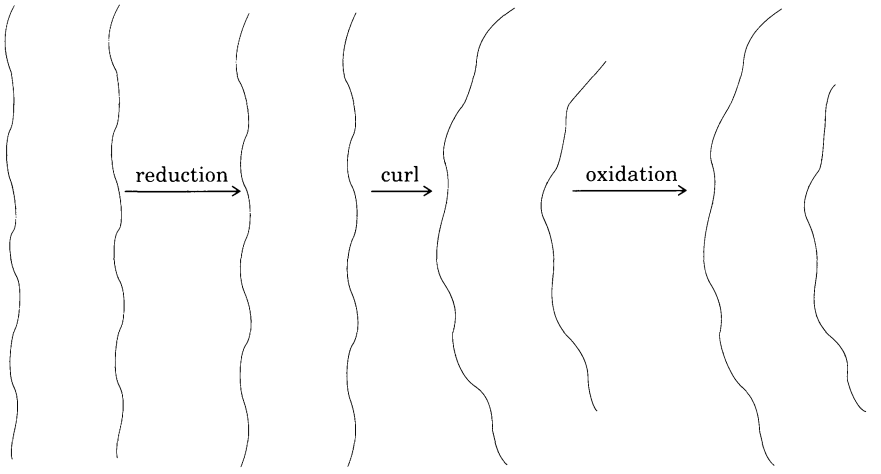


Figure 7–14 Structure of collagen. The collagen helix is a repeating secondary structure unique to this protein. **(a)** The repeating tripeptide sequence Gly–X–Pro or Gly–X–Hyp adopts a left-handed helical structure with three residues per turn. The repeating sequence used to generate this model is Gly–Pro–Hyp. **(b)** Space-filling model of the collagen helix shown in **(a)**. **(c)** Three of these helices wrap around one another with a right-handed twist. The resulting three-stranded molecule is referred to as tropocollagen (see Fig. 7–15). **(d)** The three-stranded collagen superhelix shown from one end, in a ball-and-stick representation. Glycine residues are shown in red. Glycine, because of its small size, is required at the tight junction where the three chains are in contact.

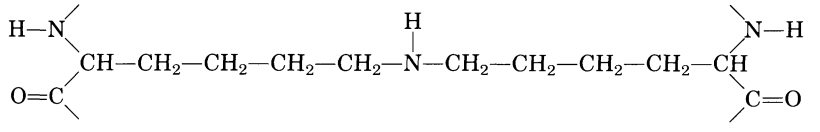
BOX 7–2 Permanent Waving Is Biochemical Engineering



α-Keratins exposed to moist heat can be stretched into the β conformation, but on cooling revert to the α-helical conformation spontaneously. This is because the R groups of α-keratins are larger on average than those of β-keratins and thus are not compatible with a stable β conformation. This characteristic of α-keratins, as well as their content of disulfide cross-linkages, is the basis of permanent waving. The hair to be waved is first bent around a form of appropriate shape. A solution of a reducing agent, usually a compound containing a thiol or sulfhydryl group (—SH), is then applied with heat. The reducing agent cleaves the disulfide cross-linkages by reducing each cystine to two cysteine residues, one in each adjacent chain. The moist heat breaks hydrogen bonds and causes the α-helical structure of the polypeptide chains to uncoil and stretch. After a time the reducing solution is removed, and an oxidizing agent is added to establish *new* disulfide bonds between pairs of Cys residues of adjacent polypeptide chains, but not the same pairs that existed before the treatment. On washing and cooling the hair, the polypeptide chains revert to their α-helical conformation. The hair fibers now curl in the desired fashion because new disulfide cross-linkages have been formed where they will exert some torsion or twist on the bundles of α-helical coils in the hair fibers.

helical twisting is probably left-handed in α-keratin (Fig. 7–13) and right-handed in collagen (Fig. 7–14). The tight wrapping of the collagen triple helix provides great tensile strength with no capacity to stretch. Collagen fibers can support up to 10,000 times their own weight and are said to have greater tensile strength than a steel wire of equal cross section.

The strength of these structures is also enhanced by covalent cross-links between polypeptide chains within the multi-helical "ropes" and between adjacent ones. In α-keratin, the cross-links are contributed by disulfide bonds (Box 7–2). In the hardest and toughest α-keratins, such as those of tortoise shells and rhinoceros horns, up to 18% of the residues are cysteines involved in disulfide bonds. The arrangement of α-keratin to form a hair fiber is shown in Figure 7–13. In collagen, the cross-links are contributed by an unusual type of covalent link between two Lys residues that creates a nonstandard amino acid residue called lysinonorleucine, found only in certain fibrous proteins.



Polypeptide chain | Lys residue minus ϵ-amino group (norleucine) | Lys residue | Polypeptide chain

Lysinonorleucine

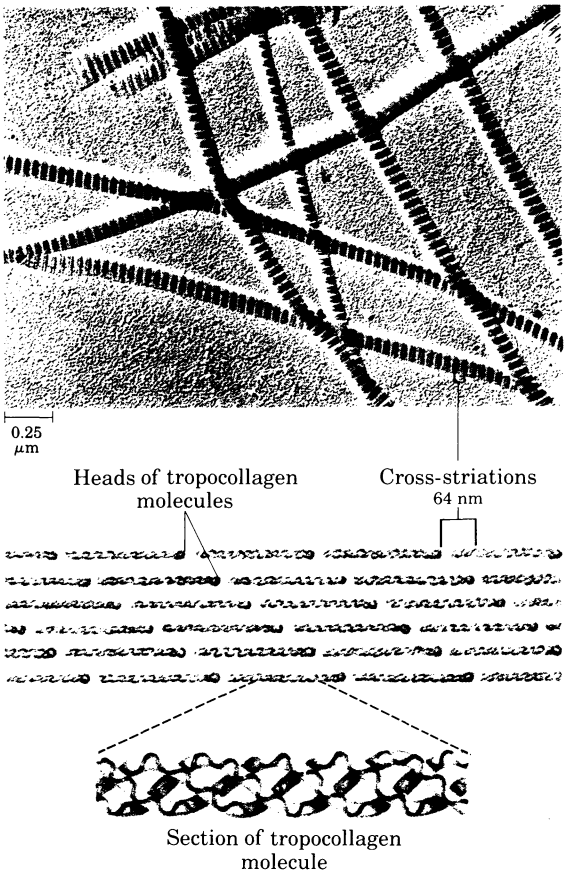


Figure 7–15 The structure of collagen fibers. Tropocollagen (M_r 300,000) is a rod-shaped molecule, about 300 nm long and only 1.5 nm thick. The three helically intertwined polypeptides are of equal length, each having about 1,000 amino acid residues. In some collagens all three chains are identical in amino acid sequence, but in others two chains are identical and the third differs. The heads of adjacent molecules are staggered, and the alignment of the head groups of every fourth molecule produces characteristic cross-striations 64 nm apart that are evident in an electron micrograph.

Collagen fibrils consist of recurring three-stranded polypeptide units called tropocollagen, arranged head to tail in parallel bundles (Fig. 7–15). The rigid, brittle character of the connective tissue in older people is the result of an accumulation of covalent cross-links in collagen as we age.

Human genetic defects involving collagen illustrate the close relationship between amino acid sequence and three-dimensional structure in this protein. Osteogenesis imperfecta results in abnormal bone formation in human babies. Ehlers-Danlos syndrome is characterized by loose joints. Both can be lethal and both result from the substitution of a Cys or Ser residue, respectively, for a Gly (a different Gly residue in each case) in the amino acid sequence of collagen. These seemingly small substitutions have a catastrophic effect on collagen function because they disrupt the Gly–X–Pro repeat that gives collagen its unique helical structure.

Elastic connective tissue contains the fibrous protein **elastin,** which resembles collagen in some of its properties but is very different in others. The polypeptide subunit of elastin fibrils is tropoelastin (M_r 72,000), containing about 800 amino acid residues. Like collagen, it is rich in Gly and Ala residues. Tropoelastin differs from tropocollagen in having many Lys but few Pro residues; it forms a special type of helix, different from the α helix and the collagen helix. Tropoelastin consists of lengths of helix rich in Gly residues separated by short regions containing Lys and Ala residues. The helical portions stretch on applying tension but revert to their original length when tension is released.

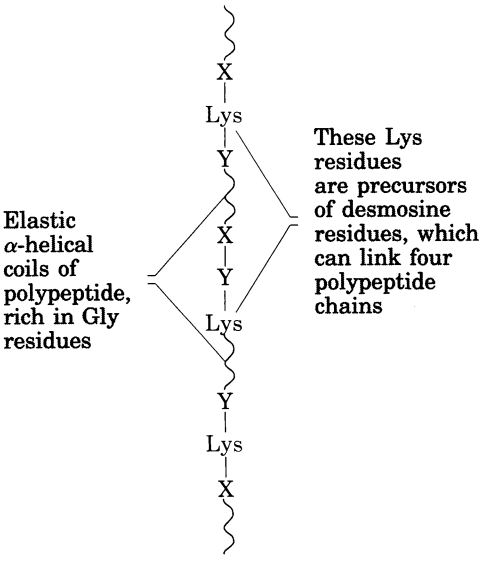


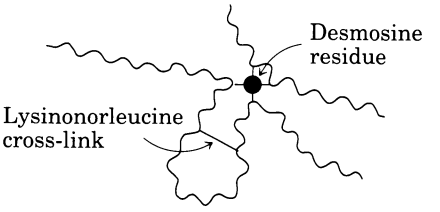
Figure 7–16 Tropoelastin molecules and their linkage to form a network of polypeptide chains in elastin. Elastin consists of tropoelastin molecules cross-linked to give two-dimensional or three-dimensional elasticity. In addition to desmosine residues (in red), which can link two, three, or four tropoelastin molecules, as shown, elastin contains other kinds of cross-linkages, such as lysinonorleucine, also designated in red.

The regions containing Lys residues form covalent cross-links. Four Lys side chains come together and are enzymatically converted into desmosine (see Fig. 5–8) and a related compound, isodesmosine; these amino acids are found only in elastin. Lysinonorleucine (p. 173) also occurs in elastin. These nonstandard amino acids are capable of joining tropoelastin chains into arrays that can be stretched reversibly in all directions (Fig. 7–16).

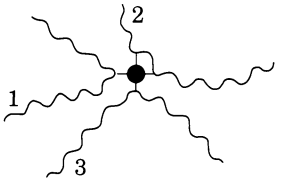
Protein Tertiary Structure

Although fibrous proteins generally have only one type of secondary structure, globular proteins can incorporate several types of secondary structure in the same molecule. Globular proteins—including enzymes, transport proteins, some peptide hormones, and immunoglobulins—are folded structures much more compact than α or β conformations (as shown for serum albumin in Figure 7–17).

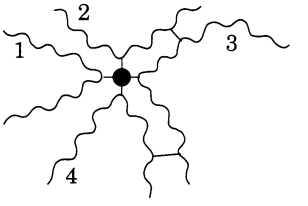
The three-dimensional arrangement of all atoms in a protein is referred to as the tertiary structure, and this now becomes our focus. Whereas the secondary structure of polypeptide chains is determined by the *short-range* structural relationship of amino acid residues, tertiary structure is conferred by *longer-range* aspects of amino acid sequence. Amino acids that are far apart in the polypeptide sequence and reside in different types of secondary structure may interact when the



Cross-linking two chains



Linking three chains



Linking four chains

β Conformation
200 x 0.5 nm

α Helix
90 x 1.1 nm

Native globular form
13 x 3 nm

Figure 7–17 Bovine serum albumin (M_r 64,500) has 584 residues in a single chain. Shown above are the approximate dimensions its single polypeptide chain would have if it occurred entirely in extended β conformation or as an α helix. Also shown (left) is the actual size of native serum albumin in its native globular form, as determined by physicochemical measurements; the polypeptide chain must be very compactly folded to fit into these dimensions.

protein is folded. The formation of bends in the polypeptide chain during folding and the direction and angle of these bends are determined by the number and location of specific bend-producing amino acids, such as Pro, Thr, Ser, and Gly residues. Moreover, loops of the highly folded polypeptide chain are held in their characteristic tertiary positions by different kinds of weak-bonding interactions (and sometimes by covalent bonds such as disulfide cross-links) between R groups of adjacent loops.

We will now consider how secondary structures contribute to the tertiary folding of a polypeptide chain in a globular protein, and how this structure is stabilized by weak interactions, in particular by hydrophobic interactions involving nonpolar amino acid side chains in the tightly packed core of the protein.

X-Ray Analysis of Myoglobin Revealed Its Tertiary Structure

The breakthrough in understanding globular protein structure came from x-ray diffraction studies of the protein myoglobin carried out by John Kendrew and his colleagues in the 1950s (Box 7–3). Myoglobin is a relatively small (M_r 16,700), oxygen-binding protein of muscle cells that functions in the storage and transport of oxygen for mitochondrial oxidation of cell nutrients. Myoglobin contains a single polypeptide chain of 153 amino acid residues of known sequence and a single iron-porphyrin, or **heme,** group (Fig. 7–18), identical to that of hemoglobin, the oxygen-binding protein of erythrocytes. The heme group is responsible for the deep red-brown color of both myoglobin and hemoglobin. Myoglobin is particularly abundant in the muscles of diving mammals such as the whale, seal, and porpoise, whose muscles are so rich in this protein that they are brown. Storage of oxygen by muscle myoglobin permits these animals to remain submerged for long periods of time.

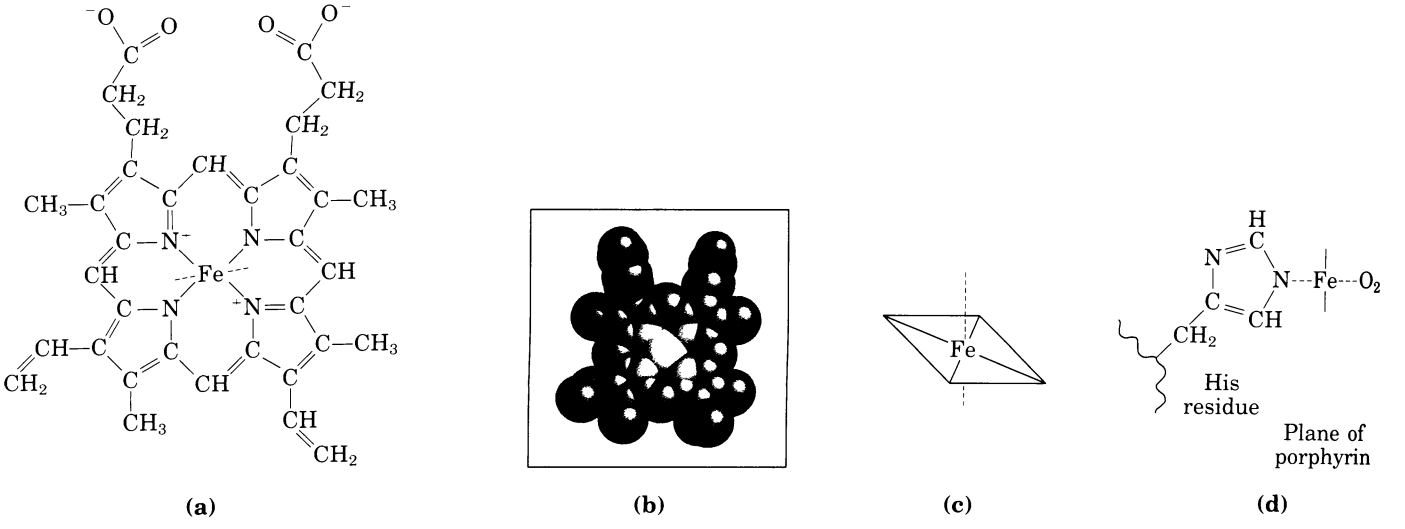


Figure 7–18 The heme group, present in myoglobin, hemoglobin, cytochrome *b,* and many other heme proteins, consists of a complex organic ring structure, protoporphyrin, to which is bound an iron atom in its ferrous (Fe^{2+}) state. Two representations are shown in **(a)** and **(b). (c)** The iron atom has six coordination bonds, four in the plane of, and bonded to, the flat porphyrin molecule and two perpendicular to it. **(d)** In myoglobin and hemoglobin, one of the perpendicular coordination bonds is bound to a nitrogen atom of a His residue. The other is "open" and serves as the binding site for an O_2 molecule, as shown here in the edge view.

BOX 7–3 X-Ray Diffraction

The spacing of atoms in a crystal lattice can be determined by measuring the angles and the intensities at which a beam of x rays of a given wavelength is diffracted by the electron shells around the atoms. For example, x-ray analysis of sodium chloride crystals shows that Na^+ and Cl^- ions are arranged in a simple cubic lattice. The spacing of the different kinds of atoms in complex organic molecules, even very large ones such as proteins, can also be analyzed by x-ray diffraction methods. However, this is far more difficult than for simple salt crystals because the very large number of atoms in a protein molecule yields thousands of diffraction spots that must be analyzed by computer.

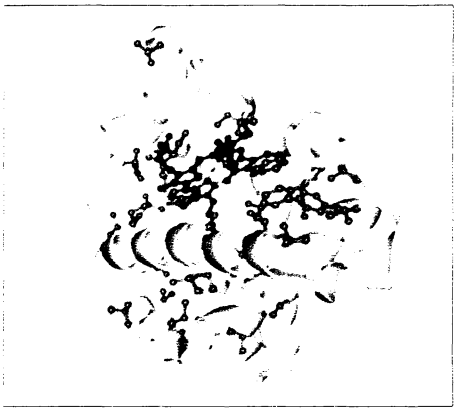
The process may be understood at an elementary level by considering how images are generated in a light microscope. Light from a point source is focused on an object. The light waves are scattered by the object, and these scattered waves are recombined by a series of lenses to generate an enlarged image of the object. The limit to the size of an object whose structure can be determined by such a system (i.e., its resolving power) is determined by the wavelength of the light. Objects smaller than half the wavelength of the incident light cannot be resolved. This is why x rays, with wavelengths in the range of a few tenths of a nanometer (often measured in angstroms, Å; 1 Å=0.1 nm), must be used for proteins. There are no lenses that can recombine x rays to form an image; the pattern of diffracted light is collected directly and converted into an image by computer analysis.

Operationally, there are several steps in x-ray structural analysis. The amount of information obtained depends on the degree of structural order in the sample. Some important structural parameters were obtained from early studies of the diffraction patterns of the fibrous proteins that occur in fairly regular arrays in hair and wool. More detailed three-dimensional structural information, however, requires a highly ordered crystal of a protein. Protein crystallization is something of an empirical science, and the structures of many important proteins are not yet known simply because they have proven difficult to crystallize. Once a crystal is obtained, it is placed in an x-ray beam between the x-ray source and a detector. A regular array of spots called reflections (Fig. 1) is generated by precessional motion of the crystal. The spots represent reflections of the x-ray beam, and each atom in a molecule makes a contribution to each spot. The overall pattern of spots is related to the structure of the protein through a mathematical device called a Fourier transform. The intensity of each spot is measured from the positions and intensities of the spots in several of these diffraction patterns, and the precise three-dimensional structure of the protein is calculated.

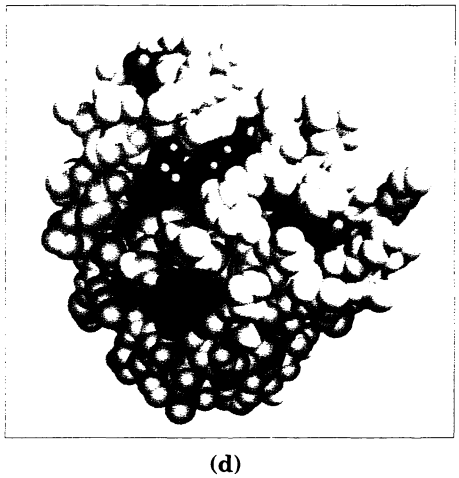
John Kendrew found that the x-ray diffraction pattern of crystalline myoglobin from muscles of the sperm whale is very complex, with nearly 25,000 reflections. Computer analysis of these reflections took place in stages. The resolution improved at each stage, until in 1959 the positions of virtually all the atoms in the protein could be determined. The amino acid sequence deduced from the structure agreed with that obtained by chemical analysis. The structures of hundreds of proteins have since been determined to a similar level of resolution, many of them much more complex than myoglobin.

Figure 1 Photograph of the x-ray diffraction pattern of crystalline sperm whale myoglobin.

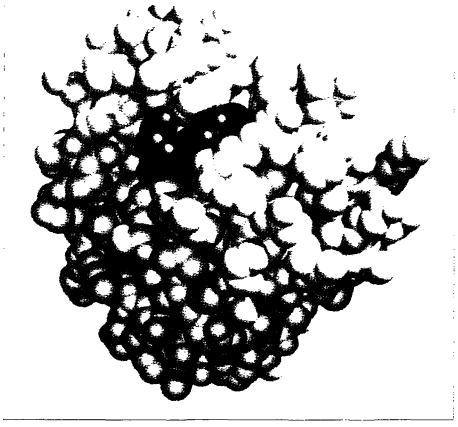
(a)



(c)



(d)



(b)

Figure 7–19 Tertiary structure of sperm whale myoglobin. The orientation of the protein is the same in all panels; the heme group is shown in red. **(a)** The polypeptide backbone, shown in a ribbon representation of a type introduced by Jane Richardson; this highlights regions of secondary structure. The α-helical regions in myoglobin are evident. Amino acid side chains are not shown. **(b)** A space-filling model, showing that the heme group is largely buried. All amino acid side chains are included. **(c)** A ribbon representation, including side chains (purple) for the hydrophobic residues Leu, Ile, Val, and Phe. **(d)** A space-filling model with all amino acid side chains. The hydrophobic residues are again shown in purple; most are not visible because they are buried in the interior of the protein.

Figure 7–19 shows several structural representations of myoglobin, illustrating how the polypeptide chain is folded in three dimensions—its tertiary structure. The backbone of the myoglobin molecule is made up of eight relatively straight segments of α helix interrupted by bends. The longest α helix has 23 amino acid residues and the shortest only seven; all are right-handed. More than 70% of the amino acids in the myoglobin molecule are in these α-helical regions. X-ray analysis also revealed the precise position of each of the R groups, which occupy nearly all the open space between the folded loops.

Other important conclusions were drawn from the structure of myoglobin. The positioning of amino acid side chains reflects a structure that derives much of its stability from hydrophobic interactions. Most of the hydrophobic R groups are in the interior of the myoglobin molecule, hidden from exposure to water. All but two of the polar R groups are located on the outer surface of the molecule, and all of them are hydrated. The myoglobin molecule is so compact that in its interior there is room for only four molecules of water. This dense hydrophobic core is typical of globular proteins. The fraction of space occupied by atoms in an organic liquid is 0.25 to 0.35; in a typical solid the fraction is 0.75. In a protein the fraction is 0.72 to 0.76, very comparable to that in a solid. In this closely packed environment weak interactions strengthen and reinforce each other. For example, the nonpolar side chains in the core are so close together that short-range van der Waals interactions make a significant contribution to stabilizing hydrophobic interactions. By contrast, in an oil droplet suspended in water, the van der Waals interactions are minimal and the cohesiveness of the droplet is based almost exclusively on entropy.

The structure of myoglobin both confirmed some expectations and introduced some new elements of secondary structure. As predicted by Pauling and Corey, all the peptide bonds are in the planar trans configuration. The α helices in myoglobin provided the first direct experimental evidence for the existence of this type of secondary structure. Each of the four Pro residues of myoglobin occurs at a bend (recall that the rigid R group of proline is largely incompatible with α-helical structure). Other bends contain Ser, Thr, and Asn residues, which are among the amino acids that tend to be incompatible with α-helical structure if they are in close proximity (p. 168).

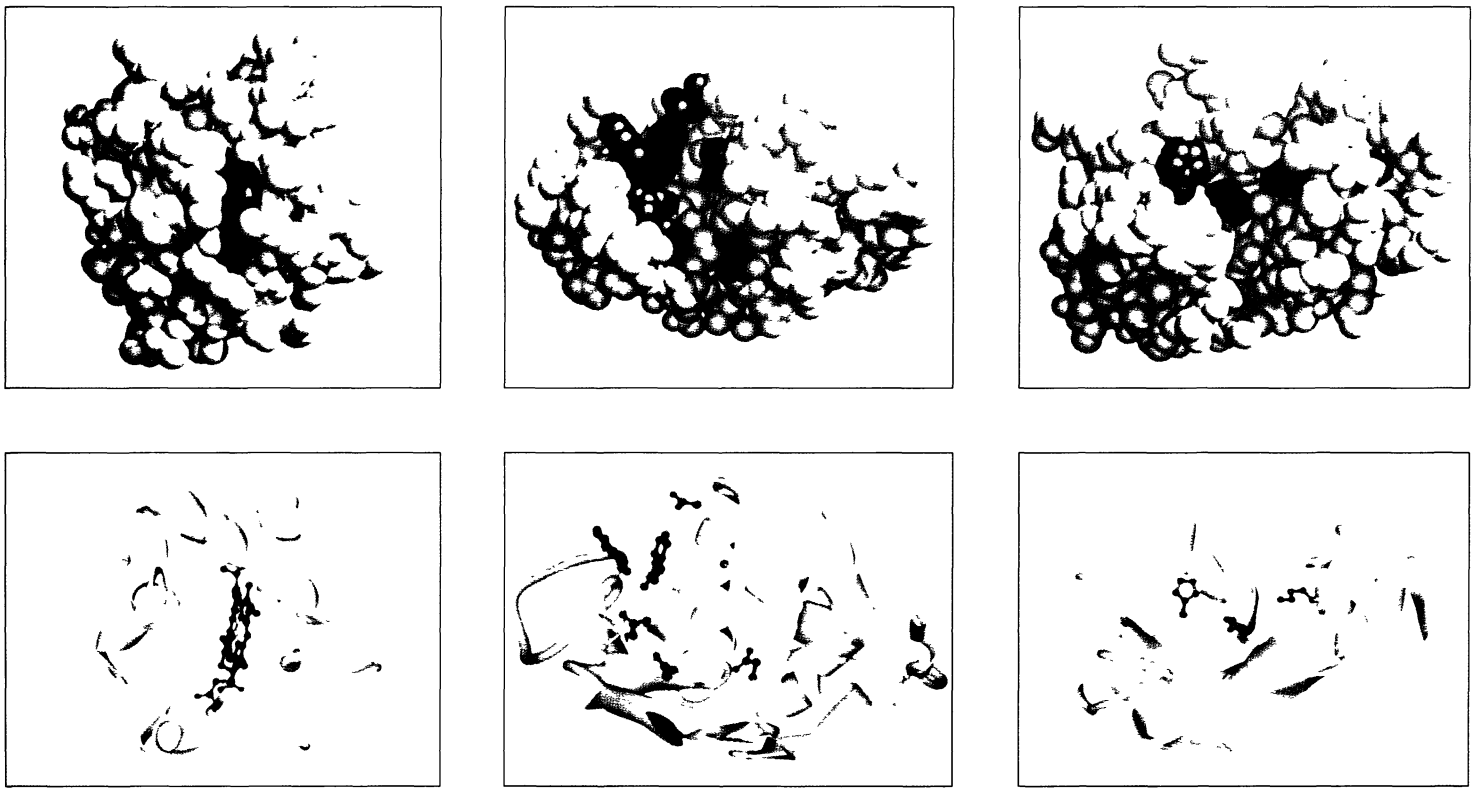
The flat heme group rests in a crevice, or pocket, in the myoglobin molecule. The iron atom in the center of the heme group has two bonding (coordination) positions perpendicular to the plane of the heme.

One of these is bound to the R group of the His residue at position 93; the other is the site to which an O_2 molecule is bound. Within this pocket, the accessibility of the heme group to solvent is highly restricted. This is important for function because free heme groups in an oxygenated solution are rapidly oxidized from the ferrous (Fe^{2+}) form, which is active in the reversible binding of O_2, to the ferric (Fe^{3+}) form, which does not bind O_2.

Proteins Differ in Tertiary Structure

With the elucidation of the tertiary structures of hundreds of other globular proteins by x-ray analysis, it is clear that myoglobin represents only one of many ways in which a polypeptide chain can be folded. In Figure 7–20 the structures of cytochrome *c*, lysozyme, and ribonuclease are compared. All have different amino acid sequences and different tertiary structures, reflecting differences in function. Like myoglobin, cytochrome *c* is a small heme protein (M_r 12,400) containing a single polypeptide chain of about 100 residues and a single heme group, which in this case is covalently attached to the polypeptide. It functions as a component of the respiratory chain of mitochondria (Chapter 18). X-ray analysis of cytochrome *c* (Fig. 7–20) shows that only about 40% of the polypeptide is in α-helical segments, compared with almost 80% of the myoglobin chain. The rest of the cytochrome *c* chain contains bends, turns, and irregularly coiled and extended segments. Thus, cytochrome *c* and myoglobin differ markedly in structure, even though both are small heme proteins.

Figure 7–20 The three-dimensional structures of three small proteins: cytochrome *c*, lysozyme, and ribonuclease. For lysozyme and ribonuclease the active site of the enzyme faces the viewer. Key functional groups (the heme in cytochrome *c*, and amino acid side chains in the active site of lysozyme and ribonuclease) are shown in red; disulfide bonds are shown in yellow. Two representations of each protein are shown: a space-filling model and a ribbon representation. In the ribbon depictions, the β structures are represented by flat arrows and the α helices by spiral ribbons; the orientation in each case is the same as that of the space-filling model, to facilitate comparison.



Cytochrome *c* **Lysozyme** **Ribonuclease**

Table 7–2 Approximate amounts of α helix and β conformation in some single-chain proteins*

Protein (total residues)	Residues (%)	
	α Helix	β Conformation
Myoglobin (153)	78	0
Cytochrome *c* (104)	39	0
Lysozyme (129)	40	12
Ribonuclease (124)	26	35
Chymotrypsin (247)	14	45
Carboxypeptidase (307)	38	17

Source: Data from Cantor, C.R. & Schimmel, P.R. (1980) *Biophysical Chemistry*, Part I: *The Conformation of Biological Macromolecules*, p. 100, W.H. Freeman and Company, New York.

* Portions of the polypeptide chains that are not accounted for by α helix or β conformation consist of bends and irregularly coiled or extended stretches. Segments of α helix and β conformation sometimes deviate slightly from their normal dimensions and geometry.

Lysozyme (M_r 14,600) is an enzyme in egg white and human tears that catalyzes the hydrolytic cleavage of polysaccharides in the protective cell walls of some families of bacteria. Lysozyme is so named because it can lyse, or degrade, bacterial cell walls and thus serve as a bactericidal agent. Like cytochrome *c*, about 40% of its 129 amino acid residues are in α-helical segments, but the arrangement is different and some β structure is also present. Four disulfide bonds contribute stability to this structure. The α helices line a long crevice in the side of the molecule (Fig. 7–20), called the active site, which is the site of substrate binding and action. The bacterial polysaccharide that is the substrate for lysozyme fits into this crevice.

Ribonuclease, another small globular protein (M_r 13,700), is an enzyme secreted by the pancreas into the small intestine, where it catalyzes the hydrolysis of certain bonds in the ribonucleic acids present in ingested food. Its tertiary structure, determined by x-ray analysis, shows that little of its 124 amino acid polypeptide chain is in α-helical conformation, but it contains many segments in the β conformation. Like lysozyme, ribonuclease has four disulfide bonds between loops of the polypeptide chain (Fig. 7–20).

Table 7–2 shows the relative percentages of α helix and β conformation among several small, single-chain, globular proteins. Each of these proteins has a distinct structure, adapted for its particular biological function. These proteins do share several important properties, however. Each is folded compactly, and in each case the hydrophobic amino acid side chains are oriented toward the interior (away from water) and the hydrophilic side chains are on the surface. These specific structures are also stabilized by a multitude of hydrogen bonds and some ionic interactions.

Proteins Lose Structure and Function on Denaturation

The way to demonstrate the importance of a specific protein structure for biological function is to alter the structure and determine the effect on function. One extreme alteration is the total loss or randomization of three-dimensional structure, a process called **denaturation.** This is the familiar process that occurs when an egg is cooked. The white of the egg, which contains the soluble protein egg albumin, coagulates to a white solid on heating. It will not redissolve on cooling to yield a clear solution of protein as in the original unheated egg white. Heating of egg albumin has therefore changed it, seemingly in an irreversible manner. This effect of heat occurs with virtually all globular proteins, regardless of their size or biological function, although the precise temperature at which it occurs may vary and it is not always irreversible. The change in structure brought about by denaturation is almost invariably associated with loss of function. This is an expected consequence of the principle that the specific three-dimensional structure of a protein is critical to its function.

Proteins can be denatured not only by heat, but also by extremes of pH, by certain miscible organic solvents such as alcohol or acetone, by certain solutes such as urea, or by exposure of the protein to detergents. Each of these denaturing agents represents a relatively mild treatment in the sense that no covalent bonds in the polypeptide chain are broken. Boiling a protein solution disrupts a variety of weak interactions. Organic solvents, urea, and detergents act primarily by disrupting the hydrophobic interactions that make up the stable core of globular proteins; extremes of pH alter the net charge on the protein,

causing electrostatic repulsion and disruption of some hydrogen bonding. Remember that the native structure of most proteins is only marginally stable. It is not necessary to disrupt *all* of the stabilizing weak interactions to reduce the thermodynamic stability to a level that is insufficient to keep the protein conformation intact.

Amino Acid Sequence Determines Tertiary Structure

The most important proof that the tertiary structure of a globular protein is determined by its amino acid sequence came from experiments showing that denaturation of some proteins is reversible. Some globular proteins denatured by heat, extremes of pH, or denaturing reagents will regain their native structure and their biological activity, a process called **renaturation,** if they are returned to conditions in which the native conformation is stable.

A classic example is the denaturation and renaturation of ribonuclease. Purified ribonuclease can be completely denatured by exposure to a concentrated urea solution in the presence of a reducing agent. The reducing agent cleaves the four disulfide bonds to yield eight Cys residues, and the urea disrupts the stabilizing hydrophobic interactions, thus freeing the entire polypeptide from its folded conformation. Under these conditions the enzyme loses its catalytic activity and undergoes complete unfolding to a randomly coiled form (Fig. 7–21). When the urea and the reducing agent are removed, the randomly coiled, denatured ribonuclease spontaneously refolds into its correct tertiary structure, with full restoration of its catalytic activity (Fig. 7–21). The refolding of ribonuclease is so accurate that the four intrachain disulfide bonds are reformed in the same positions in the renatured molecule as in the native ribonuclease. In theory, the eight Cys residues could have recombined at random to form up to four disulfide bonds in 105 different ways. This classic experiment, carried out by Christian Anfinsen in the 1950s, proves that the amino acid sequence of the polypeptide chain of proteins contains all the information required to fold the chain into its native, three-dimensional structure.

The study of homologous proteins has strengthened this conclusion. We have seen that in a series of homologous proteins, such as cytochrome *c,* from different species, the amino acid residues at certain positions in the sequence are invariant, whereas at other positions the amino acids may vary (see Fig. 6–15). This is also true for myoglobins isolated from different species of whales, from the seal, and from some terrestrial vertebrates. The similarity of the tertiary structures and amino acid sequences of myoglobins from different sources led to the conclusion that the amino acid sequence of myoglobin somehow must determine its three-dimensional folding pattern, an idea substantiated by the similar structures found by x-ray analysis of myoglobins from different species. Other sets of homologous proteins also show this relationship; in each case there are sequence homologies as well as similar tertiary structures.

Many of the invariant amino acid residues of homologous proteins appear to occur at critical points along the polypeptide chain. Some are found at or near bends in the chain, others at cross-linking points between loops in the tertiary structure, such as Cys residues involved in disulfide bonds. Still others occur at the catalytic sites of enzymes or at the binding sites for prosthetic groups, such as the heme group of cytochrome *c.*

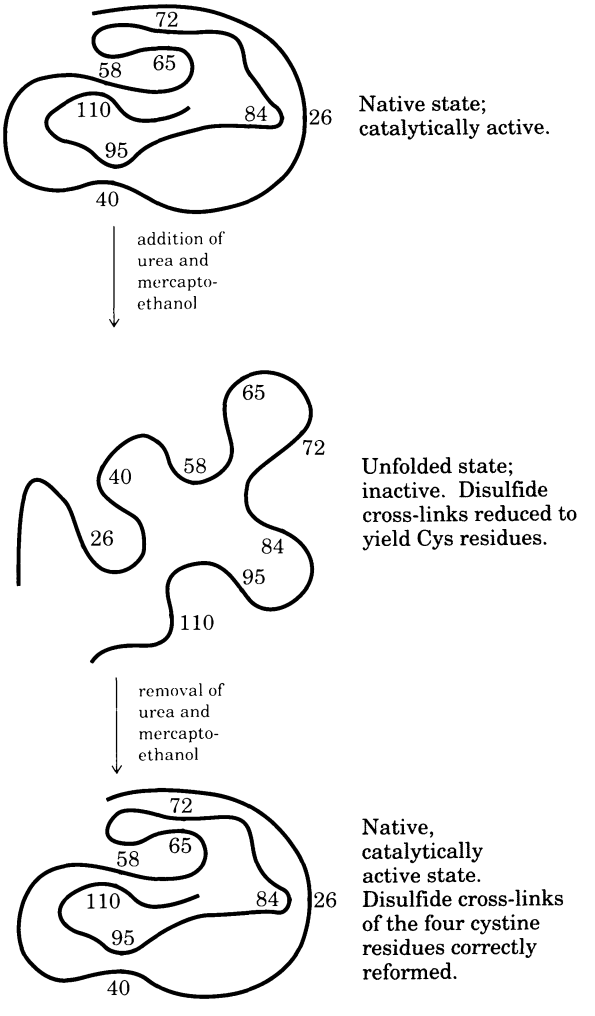


Figure 7–21 Renaturation of unfolded, denatured ribonuclease, with reestablishment of correct disulfide cross-links. Urea is added to denature ribonuclease, and mercaptoethanol ($HOCH_2CH_2SH$) to reduce and thus cleave the disulfide bonds of the four cystine residues to yield eight cysteine residues.

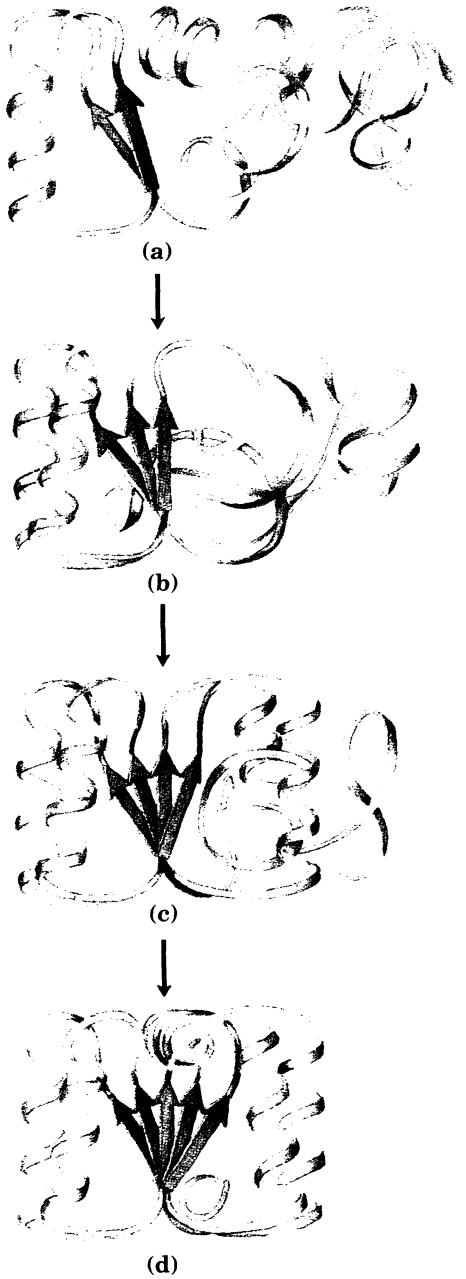


Figure 7–22 A possible protein-folding pathway. **(a)** Protein folding often begins with spontaneous formation of a structural nucleus consisting of a few particularly stable regions of secondary structure. **(b)** As other regions adopt secondary structure, they are stabilized by long-range interactions with the structural nucleus. **(c)** The folding process continues until most of the polypeptide has assumed regular secondary structure. **(d)** The final structure generally represents the most thermodynamically stable conformation.

Looking at naturally occurring amino acid substitutions has an important limitation. Any change that abolishes the function of an essential protein (e.g., a change in an invariant residue) usually results in death of the organism very early in development. This severe form of natural selection eliminates many potentially informative changes from study. Fortunately, biochemists have devised methods to specifically alter amino acid sequences in the laboratory and examine the effects of these changes on protein structure and function. These methods are derived from recombinant DNA technology (Chapter 28) and rely on altering the genetic material encoding the protein. By this process, called **site-directed mutagenesis,** specific amino acid sequences can be changed by deleting, adding, rearranging, or substituting amino acid residues. The catalytic roles of certain amino acids lining the active sites of enzymes such as triose phosphate isomerase and chymotrypsin have been elucidated by substituting different amino acids in their place. The importance of certain amino acids in protein folding and structure is being addressed in the same way.

Tertiary Structures Are Not Rigid

Although the native tertiary conformation of a globular protein is the thermodynamically most stable form its polypeptide chain can assume, this conformation must not be regarded as absolutely rigid. Globular proteins have a certain amount of flexibility in their backbones and undergo short-range internal fluctuations. Many globular proteins also undergo small conformational changes in the course of their biological function. In many instances, these changes are associated with the binding of a ligand. The term **ligand** in this context refers to a specific molecule that is bound by a protein (from Latin, *ligare,* "to tie" or "bind"). For example, the hemoglobin molecule, which we shall examine later in this chapter, has one conformation when oxygen is bound, and another when the oxygen is released. Many enzyme molecules also undergo a conformational change on binding their substrates, a process that is part of their catalytic action (Chapter 8).

Polypeptides Fold Rapidly by a Stepwise Process

In living cells, proteins are made from amino acids at a very high rate. For example, *Escherichia coli* cells can make a complete, biologically active protein molecule containing 100 amino acid residues in about 5 s at 37 °C. Yet calculations show that at least 10^{50} yr would be required for a polypeptide chain of 100 amino acid residues to fold itself spontaneously by a random process in which it tries out all possible conformations around every single bond in its backbone until it finds its native, biologically active form. Thus protein folding cannot be a completely random, trial-and-error process. There simply must be shortcuts.

The folding pathway of a large polypeptide chain is unquestionably complicated, and the principles that guide this process have not yet been worked out in detail. For several proteins, however, there is evidence that folding proceeds through several discrete intermediates, and that some of the earliest steps involve local folding of regions of secondary structure. In one model (Fig. 7–22), the process is envisioned as hierarchical, following the levels of structure outlined at the beginning of this chapter. Local secondary structures would form first, followed by longer-range interactions between, say, two α helices with compatible amino acid side chains, a process continuing until folding

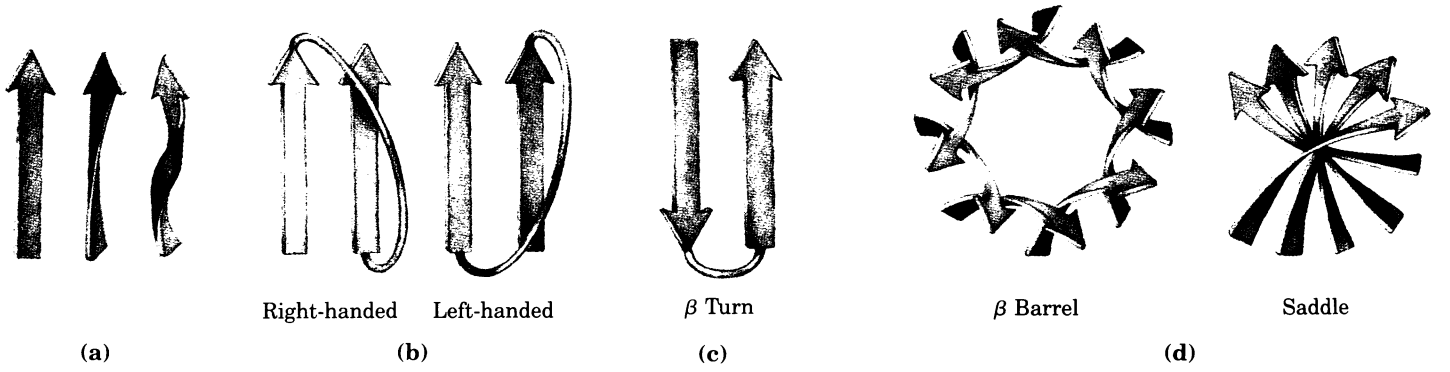


Figure 7–23 Extended β chains of amino acids tend to twist in a right-handed sense because the slightly twisted conformation is more stable than the linear conformation **(a).** This influences the conformation of the polypeptide segments that connect two β strands, and also the stable conformations assumed by several adjacent β strands. **(b)** Connections between parallel β chains are right-handed. **(c)** The β turn is a common connector between antiparallel β chains. **(d)** The tendency for right-handed twisting is seen in two particularly stable arrangements of adjacent β chains: the β barrel and the saddle; these structures form the stable core of many proteins.

was complete. In an alternative model, folding is initiated by a spontaneous collapse of the polypeptide into a compact state mediated by hydrophobic interactions among nonpolar residues. The state resulting from this "hydrophobic collapse" may have a high content of secondary structure, but many amino acid side chains are not entirely fixed. Either or both models (and perhaps others) may apply to a given protein.

A number of structural constraints help to guide the interaction of regions of secondary structure. The most common patterns are sometimes referred to as supersecondary structures. A prominent one is a tendency for extended β conformations to twist in a right-handed sense (Fig. 7–23a). This influences both the arrangement of β sheets relative to one another and the path of the polypeptide segment connecting two β strands. Two parallel β strands, for example, must be connected by a crossover strand (Fig. 7–23b). In principle, this crossover could have a right- or left-handed conformation, but only the right-handed form is found in proteins. The twisting of β sheets also leads to a characteristic twisting of the structure formed when many sheets are put together. Two examples of resulting structures are the β barrel and saddle shapes (Fig. 7–23d), which form the core of many larger structures.

Weak-bonding interactions represent the ultimate thermodynamic constraint on the interaction of different regions of secondary structure. The R groups of amino acids project outward from α-helical and β structures, and thus the need to bury hydrophobic residues means that water-soluble proteins must have more than one layer of secondary structure. One simple structural method for burying hydrophobic residues is a supersecondary structural unit called a βαβ loop (Fig. 7–24), a structure often repeated multiple times in larger proteins. More elaborate structures are domains made up of facing β sheets (with hydrophobic residues sandwiched between), and β sheets covered on one side with several α helices, as described later.

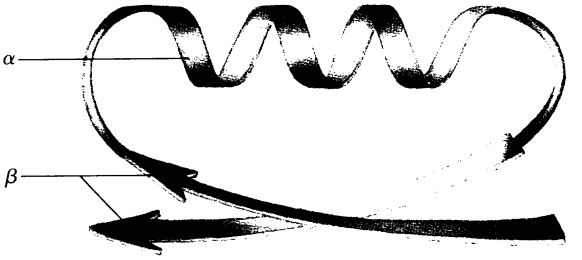


Figure 7–24 The βαβ loop. The shaded region denotes the area where stabilizing hydrophobic interactions occur.

It becomes more difficult to bury hydrophobic residues in smaller structures, and the number of potential weak interactions available for stabilization decreases. For this reason, smaller proteins are often held together with a number of covalent bonds, principally disulfide linkages. Recall the multiple disulfide bonds in the small proteins insulin (see Fig. 6–10) and ribonuclease (Fig. 7–21). Other types of covalent bonds also occur. The heme group in cytochrome *c*, for example, is covalently linked to the protein on two sides, providing a significant stabilization of the entire protein structure.

Not all proteins fold spontaneously as they are synthesized in the cell. Proteins that facilitate the folding of other proteins have been found in a wide variety of cells. These are called **polypeptide chain binding proteins** or molecular chaperones. Several of these proteins can bind to polypeptide chains, preventing nonspecific aggregation of weak-bonding side chains. They guide the folding of some polypeptides, as well as the assembly of multiple polypeptides into larger structures. Dissociation of polypeptide chain binding proteins from polypeptides is often coupled to ATP hydrolysis. One family of such proteins has structures that are highly conserved in organisms ranging from bacteria to mammals. These proteins (M_r 70,000), as well as several other families of polypeptide chain binding proteins, were originally identified as "heat shock" proteins because they are induced in many cells when heat stress is applied, and apparently help stabilize other proteins.

Some proteins have also been found that promote polypeptide folding by catalyzing processes that otherwise would limit the rate of folding, such as the reversible formation of disulfide bonds or proline isomerization (the interconversion of the cis and trans isomers of peptide bonds involving the imino nitrogen of proline; see Fig. 7–10).

There Are a Few Common Tertiary Structural Patterns

Following the folding patterns outlined above and others yet to be discovered, a newly synthesized polypeptide chain quickly assumes its most stable tertiary structure. Although each protein has a unique structure, several patterns of tertiary structure seem to occur repeatedly in proteins that differ greatly in biological function and amino acid sequence (Fig. 7–25). This may reflect an unusual degree of stability and/or functional flexibility conferred by these particular tertiary structures. It also demonstrates that biological function is determined *not only by the overall three-dimensional shape of the protein, but also* by the arrangement of amino acids within that shape.

One structural motif is made up of eight β strands arranged in a circle with each β strand connected to its neighbor by an α helix. The β regions are arranged in the barrel structure described in Figure 7–23, and they influence the overall tertiary structure, giving rise to the name α/β barrel (Fig. 7–25a). This structure is found in many enzymes; a binding site for a cofactor or substrate is often found in a pocket formed near an end of the barrel.

Another structural motif is the four-helix bundle (Fig. 7–25b), in which four α helices are connected by three peptide loops. The helices are slightly tilted to form a pocket in the middle, which often contains a binding site for a metal or other cofactors essential for biological function. A somewhat similar structure in which seven helices are ar-

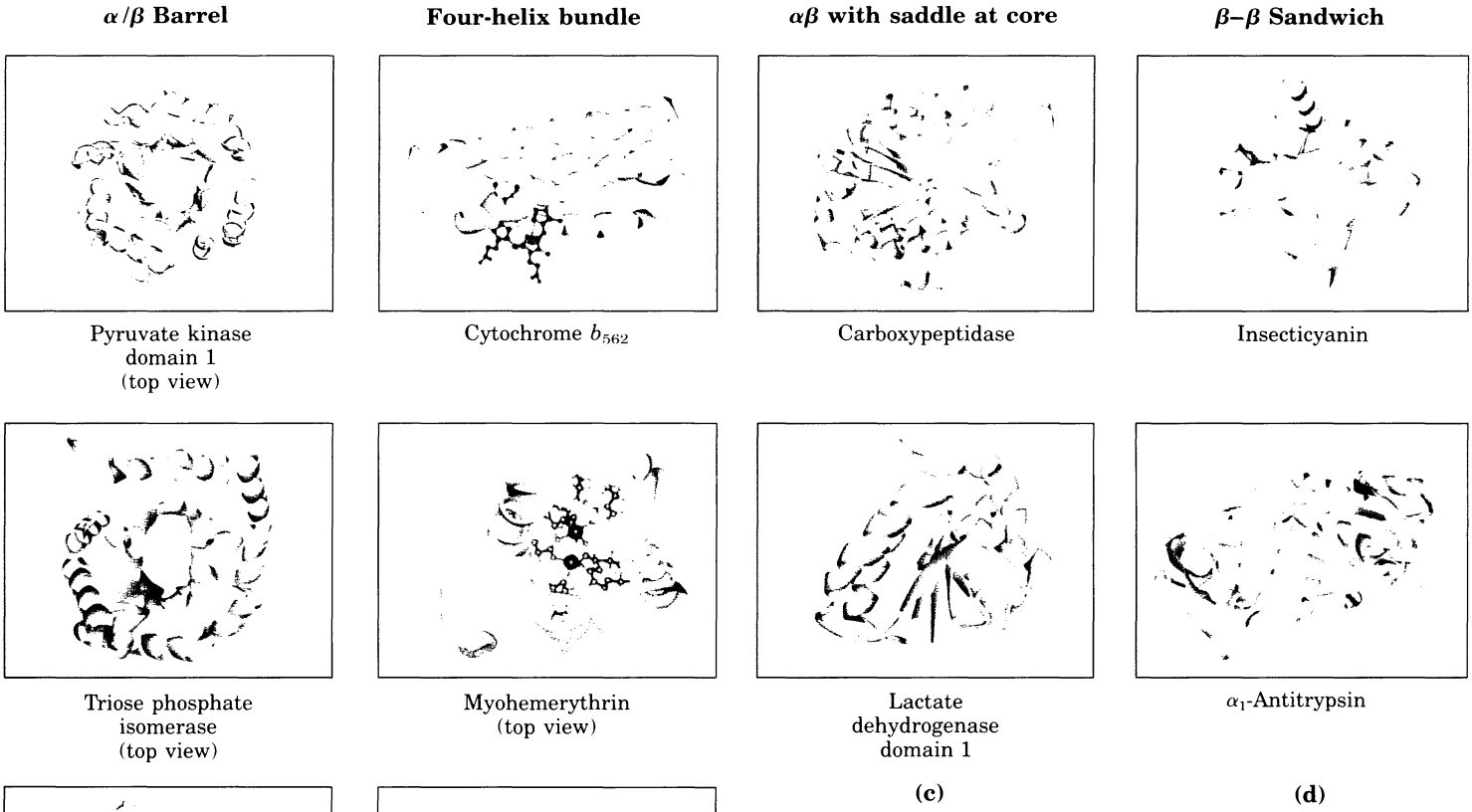


Figure 7–25 Examples of some common structural motifs in proteins. **(a)** The α/β barrel, found in pyruvate kinase and triose phosphate isomerase, enzymes of the glycolytic pathway. This structure also occurs in the larger domain of ribulose-1,5-bisphosphate carboxylase/oxygenase (known also as rubisco), an enzyme essential to the fixation of CO_2 by plants; in glycolate oxidase, an enzyme in photorespiration; and in a number of other unrelated proteins. **(b)** The four-helix bundle, shown here in cytochrome b_{562} and myohemerythrin. A dinuclear iron center and coordinating amino acids in myohemerythrin are shown in orange. Myohemerythrin is a nonheme oxygen-transporting protein found in certain worms and mollusks. The four-helix bundle is also found in apoferritin and the tobacco mosaic virus coat-protein. Apoferritin is a widespread protein involved in iron transport and storage. **(c)** $\alpha\beta$ with saddle at core, in carboxypeptidase, a protein-hydrolyzing (proteolytic) enzyme, and lactate dehydrogenase, a glycolytic enzyme. **(d)** β–β Sandwich. In the protein insecticyanin of moths, the hydrophobic pocket binds biliverdin, a colored substance that plays a role in camouflage. α_1-Antitrypsin is a naturally occurring inhibitor of the proteolytic enzyme trypsin.

ranged in a barrel-like motif is found in some membrane proteins (see Fig. 10–10). The seven helices often surround a channel that spans the membrane.

A third motif has a β sheet in the "saddle" conformation forming a stable core, often surrounded by a number of α-helical regions (Fig. 7–25c). Structures of this kind are found in many enzymes. The location of the substrate binding site varies, determined by the placement of the α helices and other variable structural elements.

One final motif makes use of a sandwich of β sheets, layered so that the strands of the sheets form a quiltlike cross-hatching when viewed from above (Fig. 7–25d). This creates a hydrophobic pocket between the β sheets that is often a binding site for a planar hydrophobic molecule.

Protein Quaternary Structure

Some proteins contain two or more separate polypeptide chains or subunits, which may be identical or different in structure. One of the best-known examples of a multisubunit protein is hemoglobin, the oxygen-carrying protein of erythrocytes. Among the larger, more complex multisubunit proteins are the enzyme RNA polymerase of *E. coli,* responsible for initiation and synthesis of RNA chains; the enzyme aspartate transcarbamoylase (12 chains; see Fig. 8–26), important in the synthesis of nucleotides; and, as an extreme case, the enormous pyruvate dehydrogenase complex of mitochondria, which is a cluster of three enzymes containing a total of 102 polypeptide chains.

The arrangement of proteins and protein subunits in three-dimensional complexes constitutes quaternary structure. The interactions between subunits are stabilized and guided by the same forces that stabilize tertiary structure: multiple noncovalent interactions. The association of polypeptide chains can serve a variety of functions. Many multisubunit proteins serve regulatory functions; their activities are altered by the binding of certain small molecules. Interactions between subunits can permit very large changes in enzyme activity in response to small changes in the concentration of substrate or regulatory molecules (Chapter 8). In other cases, separate subunits can take on separate but related functions. Entire metabolic pathways are often organized by the association of a supramolecular complex of enzymes, permitting an efficient channeling of pathway intermediates from one enzyme to the next. Other associations, such as the histones in a nucleosome or the coat proteins of a virus, serve primarily structural roles. Large assemblies sometimes reflect complex functions. One obvious example is the complicated structure of ribosomes (see Fig. 26–12), which carry out protein synthesis.

X-ray and other analytical methods for structure determination become more difficult as the size and number of subunits in a protein increases. Nevertheless, sufficient data are already available to yield some very important information about the structure and function of multisubunit proteins.

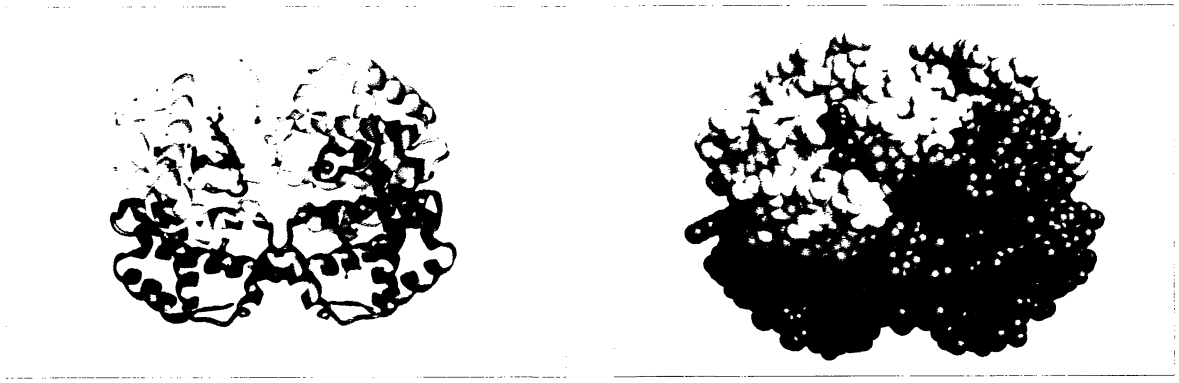
Max Perutz

John Kendrew

X-Ray Analysis Revealed the Complete Structure of Hemoglobin

The first oligomeric protein to be subjected to x-ray analysis was hemoglobin (M_r 64,500), which contains four polypeptide chains and four heme prosthetic groups, in which the iron atoms are in the ferrous (Fe^{2+}) state. The protein portion, called globin, consists of two α chains (141 residues each) and two β chains (146 residues each). Note that α and β do not refer to secondary structures in this case. Because hemoglobin is four times as large as myoglobin, much more time and effort were required to solve its three-dimensional structure, finally achieved by Max Perutz, John Kendrew, and their colleagues in 1959.

The hemoglobin molecule is roughly spherical, with a diameter of about 5.5 nm. The α and β chains contain several segments of α helix separated by bends, with a tertiary structure very similar to that of the single polypeptide of myoglobin. In fact, there are 27 invariant amino acid residues in these three polypeptide chains, and closely related amino acids at 40 additional positions, indicating that these polypeptides (myoglobin and the α and β chains of hemoglobin) are evolutionarily related. The four polypeptide chains in hemoglobin fit together in an approximately tetrahederal arrangement (Fig. 7–26).



(a) (b)

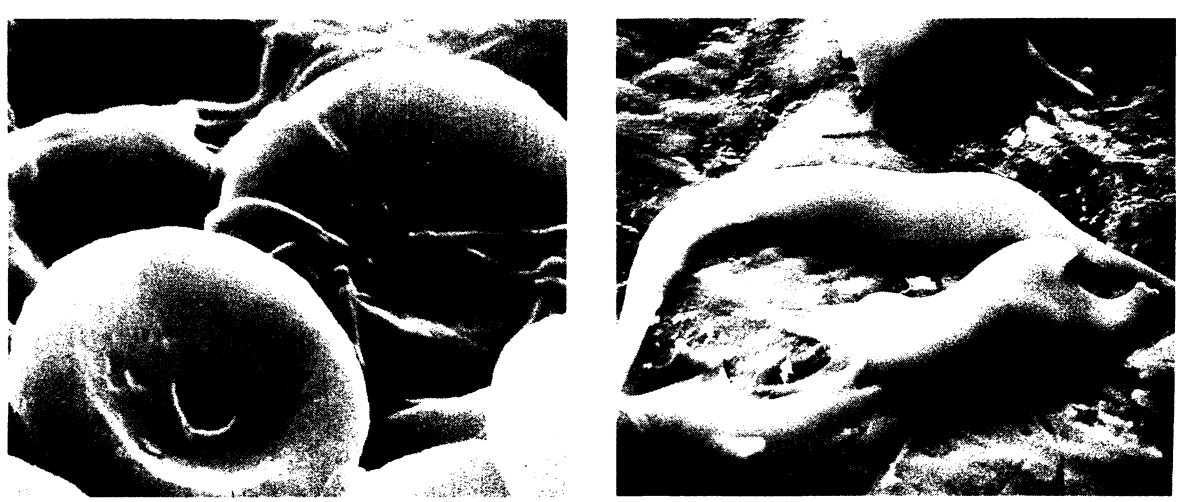
Figure 7–26 The three-dimensional (quaternary) structure of deoxyhemoglobin, revealed by x-ray diffraction analysis, showing how the four subunits are packed together. **(a)** A ribbon representation. **(b)** A space-filling model. The α subunits are shown in white and light blue; the β subunits are shown in pink and purple. Note that the heme groups, shown in red, are relatively far apart.

One heme is bound to each polypeptide chain of hemoglobin. The oxygen-binding sites are rather far apart given the size of the molecule, about 2.5 nm from one another. Each heme is partially buried in a pocket lined with hydrophobic amino acid side chains. It is bound to its polypeptide chain through a coordination bond of the iron atom to the R group of a His residue (see Fig. 7–18). The sixth coordination bond of the iron atom of each heme is available to bind O_2.

Closer examination of the quaternary structure of hemoglobin, with the help of molecular models, shows that although there are few contacts between the two α chains or between the two β chains, there are many contact points between the α and β chains. These contact points consist largely of hydrophobic side chains of amino acid residues, but also include ionic interactions involving the carboxyl-terminal residues of the four subunits.

Naturally occurring changes in the amino acid sequence of hemoglobin provide some useful insights into the relationship between structure and function in proteins. More than 300 genetic variants of hemoglobin are known to occur in the human population. Most of these variations are single amino acid changes that have only minor structural or functional effects. An exception is a substitution of valine for glutamate at position 6 of the β chain. This residue is on the outer surface of the molecule, and the change produces a "sticky" hydrophobic spot on the surface that results in abnormal quaternary association of hemoglobin. When oxygen concentrations are below a critical level, the subunits polymerize into linear arrays of fibers that distort cell shape. The result is a sickling of erythrocytes (Fig. 7–27), the cause of sickle-cell anemia.

Figure 7–27 Scanning electron micrographs of **(a)** normal and **(b)** sickled human erythrocytes. The sickled cells are fragile, and their breakdown causes anemia.



(a) (b)

Conformational Changes in Hemoglobin Alter Its Oxygen-Binding Capacity

Hemoglobin is an instructive model for studying the function of many regulatory oligomeric proteins. The blood in a human being must carry about 600 L of oxygen from the lungs to the tissues every day, but very little of this is carried by the blood plasma because oxygen is only sparingly soluble in aqueous solutions. Nearly all the oxygen carried by whole blood is bound and transported by the hemoglobin of the erythrocytes. Normal human erythrocytes are small (6 to 9 μm), biconcave disks (Fig. 7–27a). They have no nucleus, mitochondria, endoplasmic reticulum, or other organelles. The hemoglobin of the erythrocytes in arterial blood passing from the lungs to the peripheral tissues is about 96% saturated with oxygen. In the venous blood returning to the heart, the hemoglobin is only about 64% saturated. Thus blood passing through a tissue releases about one-third of the oxygen it carries.

The special properties of the hemoglobin molecule that make it such an effective oxygen carrier are best understood by comparing the O_2-binding or O_2-saturation curves of myoglobin and hemoglobin (Fig. 7–28). These show the percentage of O_2-binding sites of hemoglobin or myoglobin that are occupied by O_2 molecules when solutions of these proteins are in equilibrium with different partial pressures of oxygen in the gas phase. (The partial pressure of oxygen, abbreviated pO_2, is the pressure contributed by oxygen to the overall pressure of a mixture of gases, and is directly related to the concentration of oxygen in the mixture.)

From its saturation curve, it is clear that myoglobin has a very high affinity for oxygen (Fig. 7–28). Furthermore, the O_2-saturation curve of myoglobin is a simple hyperbolic curve, as might be expected from the mass action of oxygen on the equilibrium myoglobin + $O_2 \rightleftharpoons$ oxymyoglobin. In contrast, the oxygen affinity of each of the four O_2-binding sites of deoxyhemoglobin is much lower, and the O_2-saturation curve of hemoglobin is sigmoid (S-shaped) (Fig. 7–28). This shape indicates that whereas the affinity of hemoglobin for binding the first O_2 molecule (to any of the four sites) is relatively low, the second, third, and fourth O_2 molecules are bound with a very much higher affinity. This accounts for the steeply rising portion of the sigmoid curve. The increase in the affinity of hemoglobin for oxygen after the first O_2 molecule is bound is almost 500-fold. Thus the oxygen affinity of each heme–polypeptide subunit of hemoglobin depends on whether O_2 is bound to neighboring subunits. The conversion of deoxyhemoglobin to oxyhemoglobin requires the disruption of ionic interactions involving the carboxyl-terminal residues of the four subunits, interactions that con-

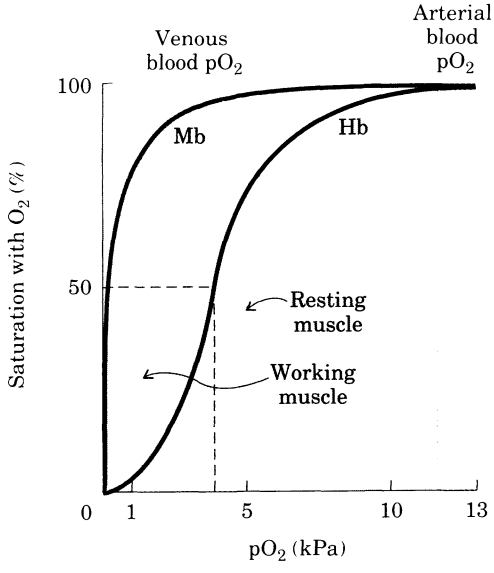


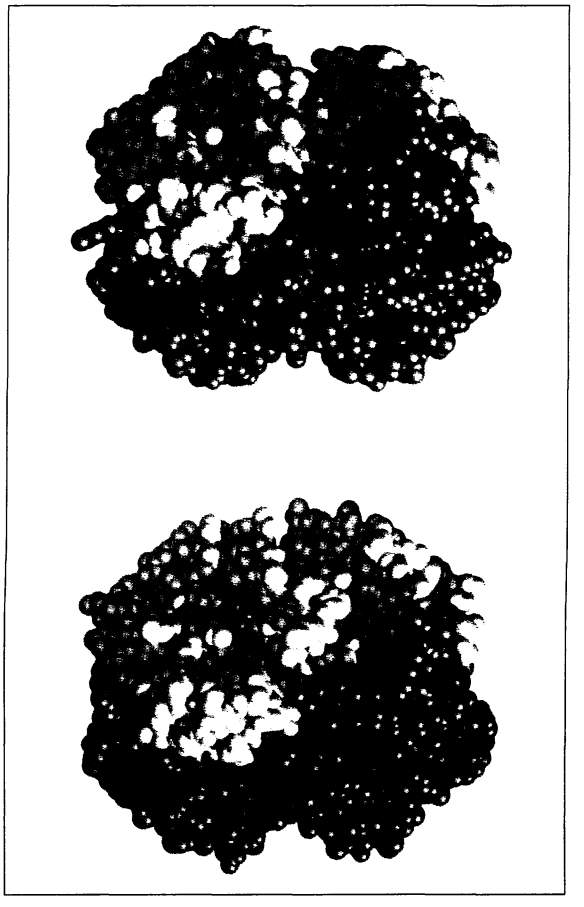
Figure 7–28 The oxygen-binding curves of myoglobin (Mb) and hemoglobin (Hb). Myoglobin has a much greater affinity for oxygen than does hemoglobin. It is 50% saturated at oxygen partial pressures (pO_2) of only 0.15 to 0.30 kPa, whereas hemoglobin requires a pO_2 of about 3.5 kPa for 50% saturation. Note that although both hemoglobin and myoglobin are more than 95% saturated at the pO_2 in arterial blood leaving the lungs (~13 kPa), hemoglobin is only about 75% saturated in resting muscle, where the pO_2 is about 5 kPa, and only 10% saturated in working muscle, where the pO_2 is only about 1.5 kPa. Thus hemoglobin can release its oxygen very effectively in muscle and other peripheral tissues. Myoglobin, on the other hand, is still about 80% saturated at a pO_2 of 1.5 kPa, and therefore unloads very little oxygen even at very low pO_2. Thus the sigmoid O_2-saturation curve of hemoglobin is a molecular adaptation for its transport function in erythrocytes, assuring the binding and release of oxygen in the appropriate tissues.

strain the overall structure in a low-affinity state. The increase in affinity for successive O_2 molecules reflects the fact that more of these ionic interactions must be broken for binding the first O_2 than for binding later ones.

Once the first heme–polypeptide subunit binds an O_2 molecule, it communicates this information to the remaining subunits through interactions at the subunit interfaces. The subunits respond by greatly increasing their oxygen affinity. This involves a change in the conformation of hemoglobin that occurs when oxygen binds (Fig. 7–29). Such communication among the four heme–polypeptide subunits of hemoglobin is the result of cooperative interactions among the subunits. Because binding of one O_2 molecule increases the probability that further O_2 molecules will be bound by the remaining subunits, hemoglobin is said to have **positive cooperativity.** Sigmoid binding curves, like that of hemoglobin for oxygen, are characteristic of positive cooperative binding. Cooperative oxygen binding does not occur with myoglobin, which has only one heme group within a single polypeptide chain and thus can bind only one O_2 molecule; its saturation curve is therefore hyperbolic. The multiple subunits of hemoglobin and the interactions between these subunits result in a fundamental difference between the O_2-binding actions of myoglobin and hemoglobin.

Positive cooperativity is not the only result of subunit interactions in oligomeric proteins. Some oligomeric proteins show **negative cooperativity:** binding of one ligand molecule *decreases* the probability that further ligand molecules will be bound. These and additional regulatory mechanisms used by these proteins are considered in Chapter 8.

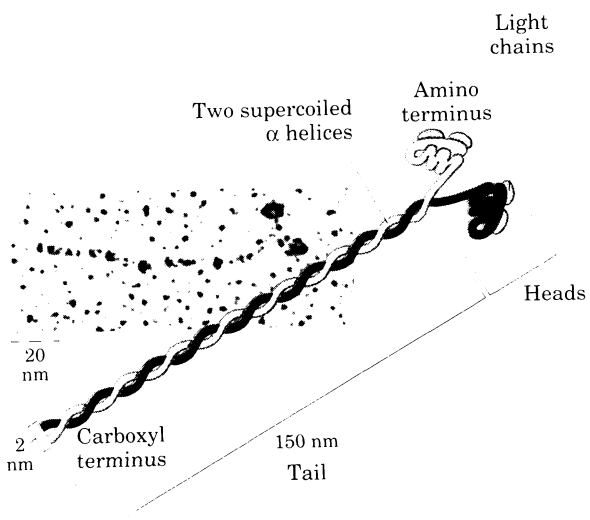
Figure 7–29 Conformational changes induced in hemoglobin when oxygen binds. (The oxygen-bound form is shown at bottom.) There are multiple structural changes, some not visible here; most of the changes are subtle. The α and β subunits are colored as in Fig. 7–26.



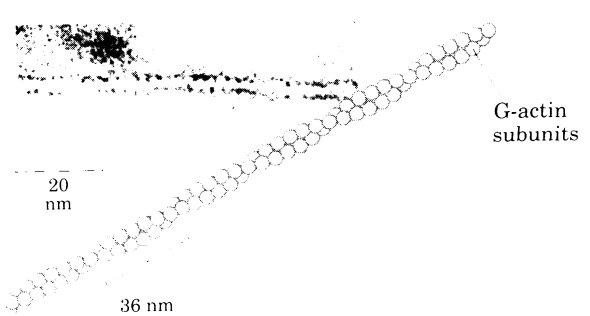
Hemoglobin Binds Oxygen in the Lungs and Releases It in Peripheral Tissues

In the lungs the pO_2 in the air spaces is about 13 kPa; at this pressure hemoglobin is about 96% saturated with oxygen. However, in the cells of a working muscle the pO_2 is only about 1.5 kPa because muscle cells use oxygen at a high rate and thus lower its local concentration. As the blood passes through the muscle capillaries, oxygen is released from the nearly saturated hemoglobin in the erythrocytes into the blood plasma and thence into the muscle cells. As is evident from the O_2-saturation curve in Figure 7–28, hemoglobin releases about a third of its bound oxygen as it passes through the muscle capillaries, so that when it leaves the muscle, it is only about 64% saturated. When the blood returns to the lungs, where the pO_2 is much higher (13 kPa), the hemoglobin quickly binds more oxygen until it is 96% saturated again.

Now suppose that the hemoglobin in the erythrocyte were replaced by myoglobin. We see from the hyperbolic O_2-saturation curve of myoglobin (Fig. 7–28) that only 1 or 2% of the bound oxygen can be released from myoglobin as the pO_2 decreases from 13 kPa in the lungs to 3 kPa in the muscle. Myoglobin therefore is not very well adapted for carrying oxygen from the lungs to the tissues, because it has a much higher affinity for oxygen and releases very little of it at the pO_2 in muscles and other peripheral tissues. However, in its true biological function *within* muscle cells, which is to store oxygen and make it available to the mitochondria, myoglobin is in fact much better suited than hemoglobin, because its very high affinity for oxygen at low pO_2 enables it to bind and store oxygen effectively. Thus hemoglobin and myoglobin are specialized and adapted for different kinds of O_2-binding functions.



(a)



(b)

Figure 7–30 Myosin and actin, the two filamentous proteins of contractile systems. **(a)** The myosin molecule has a long tail consisting of two supercoiled α-helical polypeptide chains (heavy chains). The head of each heavy chain is associated with two light chains and is an enzyme capable of hydrolyzing ATP. **(b)** A representation of an F-actin fiber, which consists of two chains of G-actin subunits coiled about each other to form a filament.

There Are Limits to the Size of Proteins

The relatively large size of proteins reflects their functions. The function of an enzyme, for example, requires a protein large enough to form a specifically structured pocket to bind its substrate. The size of proteins has limits, however, imposed by the genetic coding capacity of nucleic acids and the accuracy of the protein biosynthetic process. The use of many copies of one or a few proteins to make a large enclosing structure is important for viruses because this strategy conserves genetic material. Remember that there is a linear correspondence between the sequence of a gene in nucleic acid and the amino acid sequence of the protein for which it codes (see Fig. 6–14). The nucleic acids of viruses are much too small to encode the information required for a protein shell made of a single polypeptide. By using many copies of much smaller proteins for the virus coat, a much shorter nucleic acid is needed for the protein subunits, and this nucleic acid can be efficiently used over and over again. Cells also use large protein complexes in muscle, cilia, the cytoskeleton, and other structures. It is simply more efficient to make many copies of a small protein than one copy of a very large one. The second factor limiting the size of proteins is the error frequency during protein biosynthesis. This error frequency is low but can become significant for very large proteins. Simply put, the potential for incorporating a "wrong" amino acid in a protein is greater for a large protein than a small one.

Some Proteins Form Supramolecular Complexes

The same principles that govern the stability of secondary, tertiary, and quaternary structure in proteins guide the formation of very large protein complexes. These function, for example, as biological engines (muscle and cilia), large structural enclosures (virus coats), cellular skeletons (actin and tubulin filaments), DNA-packaging complexes (chromatin), and machines for protein synthesis (ribosomes). In many cases the complex consists of a small number of distinct proteins, specialized so that they spontaneously polymerize to form large structures.

Muscle provides an example of a supramolecular complex of multiple copies of a limited number of proteins. The contractile force of muscle is generated by the interaction of two proteins, actin and myosin (Chapter 2). Myosin is a long, rodlike molecule (M_r 540,000) consisting of six polypeptide chains, two so-called heavy chains (M_r ~230,000) and four light chains (M_r ~20,000) (Fig. 7–30a). The two heavy chains have long α-helical tails that twist around each other in a left-handed fashion. The large head domain, at one end of each heavy chain, interacts with actin and contains a catalytic site for ATP hydrolysis. Many myosin molecules assemble together to form the **thick filaments** of skeletal muscle (Fig. 7–31).

The other protein, actin, is a polymer of the globular protein G-actin (M_r 42,000); two such polymers coil around each other in a right-handed helix to form a **thin filament** (Fig. 7–30b). The interaction between actin and myosin is dynamic; contacts consist of multiple weak interactions that are strong enough to provide a stable association but weak enough to allow dissociation when needed. Hydrolysis of ATP in the myosin head is coupled to a series of conformational changes that bring about muscle contraction (Fig. 7–32). A similar engine involving an interaction between tubulin and dynein brings about the motion of cilia (Chapter 2).

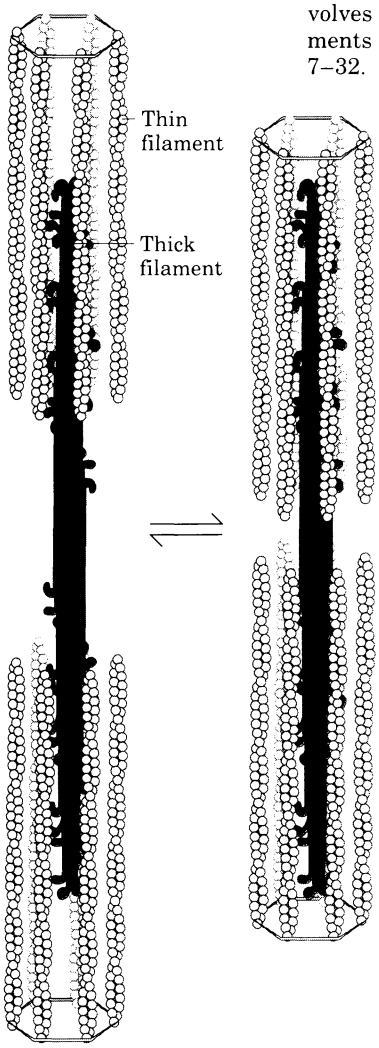


Figure 7–31 The thick and thin filaments of muscle. Many myosin molecules assemble in a bundle to form a thick filament. Muscle contraction involves the sliding of thick filaments past thin filaments of actin, by a mechanism described in Fig. 7–32.

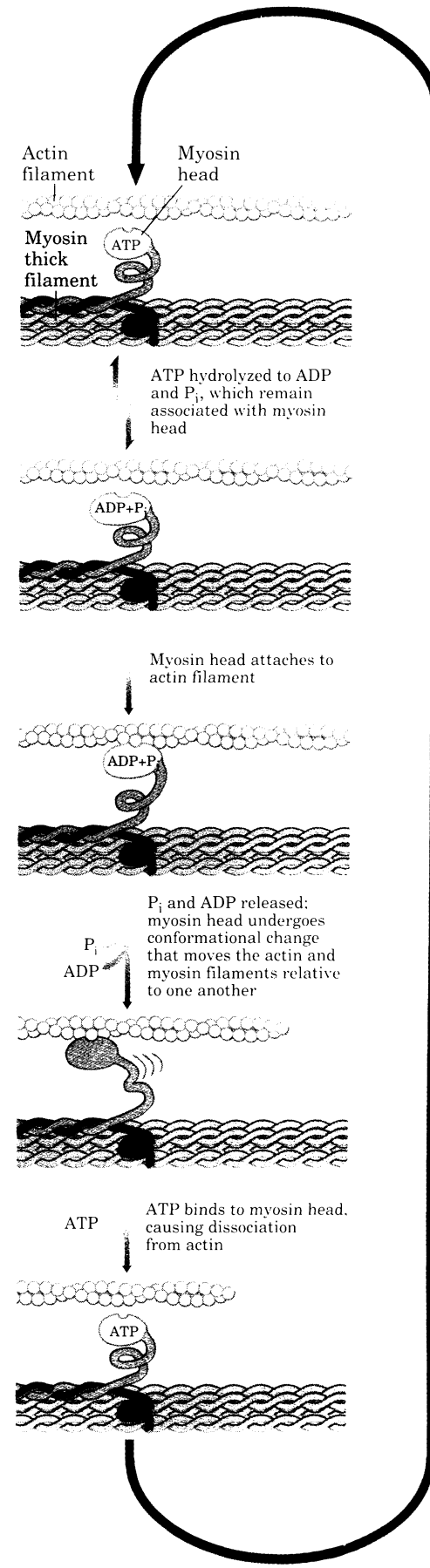


Figure 7–32 The sliding of the thick and thin filaments in muscle involves an interaction between actin and myosin mediated by ATP hydrolysis ($ATP \rightleftharpoons ADP + P_i$, where P_i is inorganic phosphate, or PO_4^{3-}). Conformational changes in the myosin head that are coupled to stages in the ATP hydrolytic cycle cause myosin to successively dissociate from one actin subunit and then associate with another farther along the actin filament. In this way the myosin heads "walk" along the thin filaments, and draw the thin filament array into the thick filament array.

The protein structures in virus coats (called capsids) generally function simply as enclosures. In many cases capsids are made up of one or a few proteins that assemble spontaneously around a viral DNA or RNA molecule. Two types of viral structures are shown in Figure 7–33. The tobacco mosaic virus is a right-handed helical filament with 2,130 copies of a single protein that interact to form a cylinder enclosing the RNA genome. Another common structure for virus coats is the icosahedron, a regular 12-cornered polyhedron having 20 equilateral triangular faces. Two examples are poliovirus and human rhinovirus 14 (a common cold virus), each made up of 60 protein units (Fig. 7–33). Each protein unit consists of single copies of four different polypeptide chains, three of which are accessible at the outer surface. The resulting shell encloses the genetic material (RNA) of the virus.

The primary forces guiding the assembly of even these very large structures are the weak noncovalent interactions that have dominated this discussion. Each protein has several surfaces that are complementary to surfaces in adjacent protein subunits. Each protein is most stable only when it is part of the larger structure.

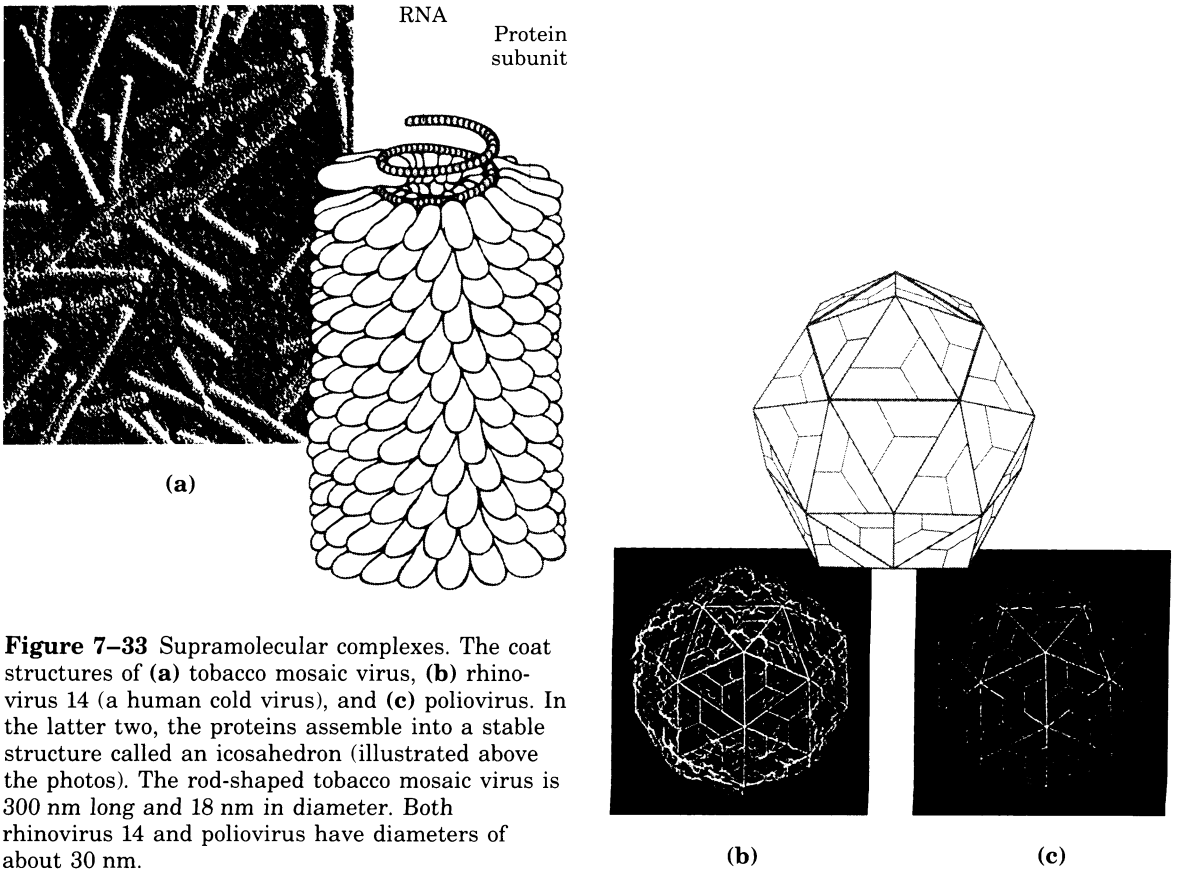


Figure 7–33 Supramolecular complexes. The coat structures of **(a)** tobacco mosaic virus, **(b)** rhinovirus 14 (a human cold virus), and **(c)** poliovirus. In the latter two, the proteins assemble into a stable structure called an icosahedron (illustrated above the photos). The rod-shaped tobacco mosaic virus is 300 nm long and 18 nm in diameter. Both rhinovirus 14 and poliovirus have diameters of about 30 nm.

Every protein has a unique three-dimensional structure that reflects its function, a structure stabilized by multiple weak interactions. Hydrophobic interactions provide the major contribution to stabilizing the globular form of most soluble proteins; hydrogen bonds and ionic interactions are optimized in the specific structure that is thermodynamically most stable.

There are four generally recognized levels of protein structure. Primary structure refers to the amino acid sequence and the location of disulfide bonds. Secondary structure refers to the spatial relationship of adjacent amino acids. Tertiary structure is the three-dimensional conformation of an entire polypeptide chain. Quaternary structure involves the spatial relationship of multiple polypeptide chains (e.g., enzyme subunits) that are tightly associated.

The nature of the bonds in the polypeptide chain places constraints on structure. The peptide bond is characterized by a partial double-bond character that keeps the entire amide group in a rigid planar configuration. The N—C_α and C_α—C bonds can rotate with bond angles ϕ and ψ, respectively. Secondary structure can be defined completely by these two bond angles.

There are two general classes of proteins: fibrous and globular. Fibrous proteins, which serve mainly structural roles, have simple repeating structures and provided excellent models for the early studies of protein structure. Two major types of secondary structure were predicted by model building based on information obtained from fibrous proteins: the α helix and the β conformation. Both are characterized by optimal hydrogen bonding between amide nitrogens and carbonyl oxygens in the peptide backbone. The stability of these structures within a protein is influenced by their amino acid content and by the relative placement of amino acids in the sequence. Another nonrepeating type of secondary structure common in proteins is the β bend.

In fibrous proteins such as keratin and collagen, a single type of secondary structure predominates. The polypeptide chains are supertwisted into ropes and then combined in larger bundles to provide strength. The structure of elastin permits stretching.

Globular proteins have more complicated tertiary structures, often containing several types of secondary structure in the same polypeptide chain. The first globular protein structure to be determined, using x-ray diffraction methods, was that of myoglobin. This structure confirmed that a predicted secondary structure (α helix) occurs in proteins; that hydrophobic amino acids are located in the protein interior; and that globular proteins are compact. Subsequent research on protein structure has reinforced these conclusions while demonstrating that different proteins often differ in tertiary structure.

The three-dimensional structure of proteins can be destroyed by treatments that disrupt weak interactions, a process called denaturation. Denaturation destroys protein function, demonstrating a relationship between structure and function. Some denatured proteins (e.g., ribonuclease) can renature spontaneously to give active protein, showing that the tertiary structure of a protein is determined by its amino acid sequence.

The folding of globular proteins is believed to begin with local formation of regions of secondary structure, followed by interactions of these regions and adjustments to reach the final tertiary structure. Sometimes regions of a polypeptide chain, called domains, fold up separately and can have separate functions. The final structure and the steps taken to reach it are influenced by the need to bury hydrophobic amino acid side chains in the protein interior away from water, the tendency of a polypeptide chain to twist in a right-handed sense, and the need to maximize hydrogen bonds and ionic interactions. These constraints give rise to structural patterns such as the $\beta\alpha\beta$ fold and twisted β pleated sheets. Even at the level of tertiary structure, some common patterns are found in proteins that have no known functional relationship.

Quaternary structure refers to the interaction between the subunits of oligomeric proteins or large protein assemblies. The best-studied oligomeric protein is hemoglobin. The four subunits of hemoglobin exhibit cooperative interactions on oxygen binding. Binding of oxygen to one subunit facilitates oxygen binding to the next, giving rise to a sigmoid binding curve. These effects are mediated by subunit–subunit interactions and subunit conformational changes. Very large protein structures consisting of many copies of one or a few different proteins are referred to as supramolecular complexes. These are found in cellular skeletal structures, muscle and other types of cellular "engines," and virus coats.

General

Anfinsen, C.B. (1973) Principles that govern the folding of protein chains. *Science* **181,** 223–230.

The author reviews his classic work on ribonuclease.

Cantor, C.R. & Schimmel, P.R. (1980) *Biophysical Chemistry,* Part I: *The Conformation of Biological Macromolecules,* W.H. Freeman and Company, New York.

Evolution of Catalytic Function. (1987) *Cold Spring Harb. Symp. Quant. Biol.* **52.**

A source of excellent articles on many topics, including protein structure, folding, and function.

Creighton, T.E. (1984) *Proteins: Structures and Molecular Properties,* W.H. Freeman and Company, New York.

Oxender, D.L. (ed) (1987) *Protein Structure, Folding, and Design 2,* UCLA Symposia on Molecular and Cellular Biology, New Series, Vol. 69, Alan R. Liss, Inc., New York.

Summary papers from a major symposium on the title subject.

Structure and Function of Proteins. (1989) *Trends Biochem. Sci.* **14** (July).

A special issue devoted to reviews on protein chemistry and protein structure. Includes good summaries of protein folding, protein structure prediction, and many other topics.

Secondary, Tertiary, and Quaternary Structure

Dickerson, R.E. & Geis, I. (1982) *Hemoglobin: Structure, Function, Evolution, and Pathology,* The Benjamin/Cummings Publishing Company, Menlo Park, CA.

Ingram, V.M. (1957) Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature* ***180,*** *326–*328.

Discovery of the amino acid replacement in sickle-cell hemoglobin (hemoglobin S).

Kendrew, J.C. (1961) The three-dimensional structure of a protein molecule. *Sci. Am.* **205** (December), 96–111.

Describes how the structure of myoglobin was determined and what was learned from it.

Kim, P.S. & Baldwin, R.L. (1990) Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59,** 631–660.

Koshland, D.E., Jr. (1973) Protein shape and biological control. *Sci. Am.* **229** (October), 52–64.

A discussion of the importance of flexibility in protein structures.

McPherson, A. (1989) Macromolecular crystals, *Sci. Am.,* **260** (March), 62–69.

Describes how macromolecules such as proteins are crystallized.

Pace, C.N. (1990) Conformational stability of globular proteins. *Trends Biochem. Sci.* **15,** 14–17.

Perutz, M.F. (1978) Hemoglobin structure and respiratory transport. *Sci. Am.* **239** (December), 92–125.

Richards, F.M. (1991) The protein folding problem. *Sci. Am.* **264** (January), 54–63.

Richardson, J.S. (1981) The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* **34,** 167–339.

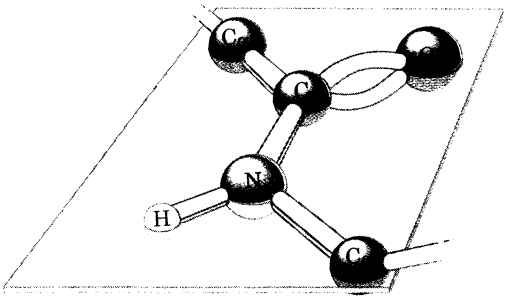
An outstanding summary of protein structural patterns and principles; the author originated the very useful "ribbon" representations of protein structure that are used in many places in this chapter.

Rothman, J.E. (1989) Polypeptide chain binding proteins: catalysts of protein folding and related processes in cells. *Cell* **59,** 591–601.

Shortle, D. (1989) Probing the determinants of protein folding and stability with amino acid substitutions. *J. Biol. Chem.* **264,** 5315–5318.

Problems

1. *Properties of the Peptide Bond* In x-ray studies of crystalline peptides Linus Pauling and Robert Corey found that the C—N bond in the peptide link is intermediate in length (0.132 nm) between a typical C—N single bond (0.149 nm) and a C=N double bond (0.127 nm). They also found that the peptide bond is planar (all four atoms attached to the C—N group are located in the same plane) and that the two α-carbon atoms attached to the C—N are always trans to each other (on opposite sides of the peptide bond):



(a) What does the length of the C—N bond in the peptide linkage indicate about its strength and its bond order, i.e., whether it is single, double, or triple?

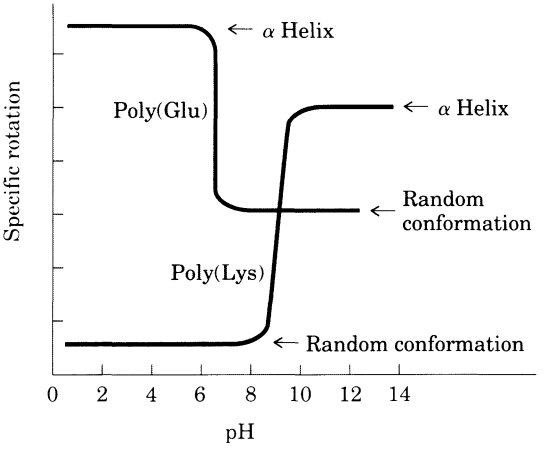
(b) In light of your answer to part (a), provide an explanation for the observation that such a C—N bond is intermediate in length between a double and single bond.

(c) What do the observations of Pauling and Corey tell us about the ease of rotation about the C—N peptide bond?

2. *Early Observations on the Structure of Wool* William Astbury discovered that the x-ray pattern of wool shows a repeating structural unit spaced about 0.54 nm along the direction of the wool fiber. When he steamed and stretched the wool, the x-ray pattern showed a new repeating structural unit at a spacing of 0.70 nm. Steaming and stretching the wool and then letting it shrink gave an x-ray pattern consistent with the original spacing of about 0.54 nm. Although these observations provided important clues to the molecular structure of wool, Astbury was unable to interpret them at the time. Given our current understanding of the structure of wool, interpret Astbury's observations.

3. *Rate of Synthesis of Hair α-Keratin* In human dimensions, the growth of hair is a relatively slow process, occurring at a rate of 15 to 20 cm/yr. All this growth is concentrated at the base of the hair fiber, where α-keratin filaments are synthesized inside living epidermal cells and assembled into ropelike structures (see Fig. 7–13). The fundamental structural element of α-keratin is the α helix, which has 3.6 amino acid residues per turn and a rise of 0.56 nm per turn (see Fig. 7–6). Assuming that the biosynthesis of α-helical keratin chains is the rate-limiting factor in the growth of hair, calculate the rate at which peptide bonds of α-keratin chains must be synthesized (peptide bonds per second) to account for the observed yearly growth of hair.

4. *The Effect of pH on the Conformations of Polyglutamate and Polylysine* The unfolding of the α helix of a polypeptide to a randomly coiled conformation is accompanied by a large decrease in a property called its specific rotation, a measure of a solution's capacity to rotate plane-polarized light. Polyglutamate, a polypeptide made up of only L-Glu residues, has the α-helical conformation at pH 3. However, when the pH is raised to 7, there is a large decrease in the specific rotation of the solution. Similarly, polylysine (L-Lys residues) is an α helix at pH 10, but when the pH is lowered to 7, the specific rotation also decreases, as shown by the following graph.



What is the explanation for the effect of the pH changes on the conformations of poly(Glu) and poly(Lys)? Why does the transition occur over such a narrow range of pH?

5. *The Disulfide-Bond Content Determines the Mechanical Properties of Many Proteins* A number of natural proteins are very rich in disulfide

bonds, and their mechanical properties (tensile strength, viscosity, hardness, etc.) are correlated with the degree of disulfide bonding. For example, glutenin, a wheat protein rich in disulfide bonds, is responsible for the cohesive and elastic character of dough made from wheat flour. Similarly, the hard, tough nature of tortoise shell is due to the extensive disulfide bonding in its α-keratin. What is the molecular basis for the correlation between disulfide-bond content and mechanical properties of the protein?

6. *Why Does Wool Shrink?* When wool sweaters or socks are washed in hot water and/or dried in an electric dryer, they shrink. From what you know of α-keratin structure, how can you account for this? Silk, on the other hand, does not shrink under the same conditions. Explain.

7. *Heat Stability of Proteins Containing Disulfide Bonds* Most globular proteins are denatured and lose their activity when briefly heated to 65 °C. Globular proteins that contain multiple disulfide bonds often must be heated longer at higher temperatures to denature them. One such protein is bovine pancreatic trypsin inhibitor (BPTI), which has 58 amino acid residues in a single chain and contains three disulfide bonds. On cooling a solution of denatured BPTI, the activity of the protein is restored. Can you suggest a molecular basis for this property?

8. *Bacteriorhodopsin in Purple Membrane Proteins* Under the proper environmental conditions, the salt-loving bacterium *Halobacterium halobium* synthesizes a membrane protein (M_r 26,000) known as bacteriorhodopsin, which is purple because it contains retinal. Molecules of this protein aggregate into "purple patches" in the cell membrane. Bacteriorhodopsin acts as a light-activated proton pump that provides energy for cell functions. X-ray analysis of this protein reveals that it consists of seven parallel α-helical segments, each of which traverses the bacterial cell membrane (thickness 4.5 nm). Calculate the minimum number of amino acids necessary for one segment of α helix to traverse the membrane completely. Estimate the fraction of the bacteriorhodopsin protein that occurs in α-helical form. Justify all your assumptions. (Use an average amino acid residue weight of 110.)

9. *Biosynthesis of Collagen* Collagen, the most abundant protein in mammals, has an unusual amino acid composition. Unlike most other proteins, collagen is very rich in proline and hydroxyproline *(see p. 172)*. *Hydroxyproline is not one of* the 20 standard amino acids, and its incorporation in collagen could occur by one of two routes: (1) proline is hydroxylated by enzymes *before* incorporation into collagen or (2) a Pro residue is hydroxylated *after* incorporation into collagen. To differentiate between these two possibilities, the following experiments were performed. When [^{14}C]proline was administered to a rat and the collagen from the tail isolated, the newly synthesized tail collagen was found to be radioactive. If, however, [^{14}C]hydroxyproline was administered to a rat, no radioactivity was observed in the newly synthesized collagen. How do these experiments differentiate between the two possible mechanisms for introducing hydroxyproline into collagen?

10. *Pathogenic Action of Bacteria That Cause Gas Gangrene* The highly pathogenic anaerobic bacterium *Clostridium perfringens* is responsible for gas gangrene, a condition in which animal tissue structure is destroyed. This bacterium secretes an enzyme that efficiently catalyzes the hydrolysis of the peptide bond indicated in red in the sequence:

$$\text{—X—Gly—Pro—Y—} \xrightarrow{H_2O} \text{—X—}COO^- + H_3\overset{+}{N}\text{—Gly—Pro—Y—}$$

where X and Y are any of the 20 standard amino acids. How does the secretion of this enzyme contribute to the invasiveness of this bacterium in human tissues? Why does this enzyme not affect the bacterium itself?

11. *Formation of Bends and Intrachain Cross-Linkages in Polypeptide Chains* In the following polypeptide, where might bends or turns occur? Where might intrachain disulfide cross-linkages be formed?

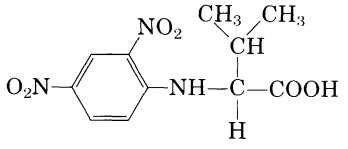
```
 1   2   3   4   5   6   7   8   9   10
Ile–Ala–His–Thr–Tyr–Gly–Pro–Phe–Glu–Ala–

 11  12  13  14  15  16  17  18  19  20
Ala–Met–Cys–Lys–Trp–Glu–Ala–Gln–Pro–Asp–

 21  22  23  24  25  26  27  28
Gly–Met–Glu–Cys–Ala–Phe–His–Arg–
```

12. *Location of Specific Amino Acids in Globular Proteins* X-ray analysis of the tertiary structure of myoglobin and other small, single-chain globular proteins has led to some generalizations about how the polypeptide chains of soluble proteins fold. With these generalizations in mind, indicate the probable location, whether in the interior or on the external surface, of the following amino acid residues in native globular proteins: Asp, Leu, Ser, Val, Gln, Lys. Explain your reasoning.

13. *The Number of Polypeptide Chains in an Oligomeric Protein* A sample (660 mg) of an oligomeric protein of M_r 132,000 was treated with an excess of 1-fluoro-2,4-dinitrobenzene under slightly alkaline conditions until the chemical reaction was complete. The peptide bonds of the protein were then completely hydrolyzed by heating it with concentrated HCl. The hydrolysate was found to contain 5.5 mg of the following compound:



However, 2,4-dinitrophenyl derivatives of the α-amino groups of other amino acids could not be found.

(a) Explain why this information can be used to determine the number of polypeptide chains in an oligomeric protein.

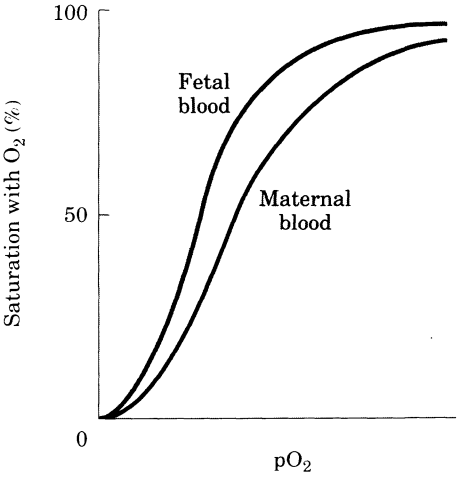
(b) Calculate the number of polypeptide chains in this protein.

14. *Molecular Weight of Hemoglobin* The first indication that proteins have molecular weights greatly exceeding those of the (then known) organic compounds was obtained over 100 years ago. For example, it was known at that time that hemoglobin contains 0.34% by weight of iron.

(a) From this information determine the minimum molecular weight of hemoglobin.

(b) Subsequent experiments indicated that the true molecular weight of hemoglobin is 64,500. What information did this provide about the number of iron atoms in hemoglobin?

15. *Comparison of Fetal and Maternal Hemoglobin* Studies of oxygen transport in pregnant mammals have shown that the O_2-saturation curves of fetal and maternal blood are markedly different when measured under the same conditions. Fetal erythrocytes contain a structural variant of hemoglobin, hemoglobin F, consisting of two α and two γ subunits ($\alpha_2\gamma_2$), whereas maternal erythrocytes contain the usual hemoglobin A ($\alpha_2\beta_2$).



(a) Which hemoglobin has a higher affinity for oxygen under physiological conditions, hemoglobin A or hemoglobin F? Explain.

(b) What is the physiological significance of the different oxygen affinities? Explain.

CHAPTER

Enzymes

We now come to the most remarkable and highly specialized proteins, the enzymes. Enzymes are the reaction catalysts of biological systems. They have extraordinary catalytic power, often far greater than that of synthetic catalysts. They have a high degree of specificity for their substrates, they accelerate specific chemical reactions, and they function in aqueous solutions under very mild conditions of temperature and pH. Few nonbiological catalysts show all these properties.

Enzymes are one of the keys to understanding how cells survive and proliferate. Acting in organized sequences, they catalyze the hundreds of stepwise reactions in metabolic pathways by which nutrient molecules are degraded, chemical energy is conserved and transformed, and biological macromolecules are made from simple precursors. Some of the many enzymes participating in metabolism are regulatory enzymes, which can respond to various metabolic signals by changing their catalytic activity accordingly. Through the action of regulatory enzymes, enzyme systems are highly coordinated to yield a harmonious interplay among the many different metabolic activities necessary to sustain life.

The study of enzymes also has immense practical importance. In some diseases, especially inheritable genetic disorders, there may be a deficiency or even a total absence of one or more enzymes in the tissues (see Table 6–6). Abnormal conditions can also be caused by the excessive activity of a specific enzyme. Measurements of the activity of certain enzymes in the blood plasma, erythrocytes, or tissue samples are important in diagnosing disease. Enzymes have become important practical tools, not only in medicine but also in the chemical industry, in food processing, and in agriculture. Enzymes play a part even in everyday activities in the home such as food preparation and cleaning.

The chapter begins with descriptions of the properties of enzymes and the principles underlying their catalytic power. Following is an introduction to enzyme kinetics, a discipline that provides much of the framework for any discussion of enzymes. Specific examples of enzyme mechanisms are then provided, illustrating principles introduced earlier in the chapter. We will end with a discussion of regulatory enzymes.

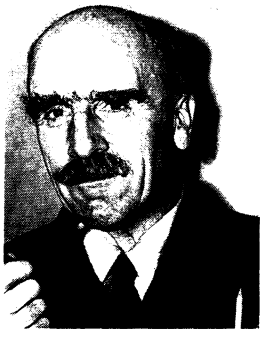
An Introduction to Enzymes

Much of the history of biochemistry is the history of enzyme research. Biological catalysis was first recognized and described in the early 1800s, in studies of the digestion of meat by secretions of the stomach

and the conversion of starch into sugar by saliva and various plant extracts. In the 1850s Louis Pasteur concluded that fermentation of sugar into alcohol by yeast is catalyzed by "ferments." He postulated that these ferments, later named **enzymes,** are inseparable from the structure of living yeast cells, a view that prevailed for many years. The discovery by Eduard Buchner in 1897 that yeast extracts can ferment sugar to alcohol proved that the enzymes involved in fermentation can function when removed from the structure of living cells. This encouraged biochemists to attempt the isolation of many different enzymes and to examine their catalytic properties.

James Sumner's isolation and crystallization of urease in 1926 provided a breakthrough in early studies of the properties of specific enzymes. Sumner found that the urease crystals consisted entirely of protein and postulated that all enzymes are proteins. Lacking other examples, this idea remained controversial for some time. Only later in the 1930s, after John Northrop and his colleagues crystallized pepsin and trypsin and found them also to be proteins, was Sumner's conclusion widely accepted. During this period, J.B.S. Haldane wrote a treatise entitled "Enzymes." Even though the molecular nature of enzymes was not yet fully appreciated, this book contained the remarkable suggestion that weak-bonding interactions between an enzyme and its substrate might be used to distort the substrate and catalyze the reaction. This insight lies at the heart of our current understanding of enzymatic catalysis. The latter part of the twentieth century has seen intensive research on the enzymes catalyzing the reactions of cell metabolism. This has led to the purification of thousands of enzymes (Fig. 8–1), elucidation of the structure and chemical mechanism of hundreds of these, and a general understanding of how enzymes work.

James Sumner
1887–1955



J.B.S. Haldane
1892–1964

Most Enzymes Are Proteins

With the exception of a small group of catalytic RNA molecules (Chapter 25), all enzymes are proteins. Their catalytic activity depends upon the integrity of their native protein conformation. If an enzyme is denatured or dissociated into subunits, catalytic activity is usually lost. If an enzyme is broken down into its component amino acids, its catalytic activity is always destroyed. Thus the primary, secondary, tertiary, and quaternary structures of protein enzymes are essential to their catalytic activity.

Enzymes, like other proteins, have molecular weights ranging from about 12,000 to over 1 million. Some enzymes require no chemical groups other than their amino acid residues for activity. Others require an additional chemical component called a **cofactor.** The cofactor may be either one or more inorganic ions, such as Fe^{2+}, Mg^{2+}, Mn^{2+}, or Zn^{2+} (Table 8–1), or a complex organic or metalloorganic molecule called a **coenzyme** (Table 8–2). Some enzymes require *both* a coenzyme and one or more metal ions for activity. A coenzyme or metal ion that is covalently bound to the enzyme protein is called a **prosthetic group.** A complete, catalytically active enzyme together with its coenzyme and/or metal ions is called a **holoenzyme.** The protein part of such an enzyme is called the **apoenzyme** or **apoprotein.** Coenzymes function as transient carriers of specific functional groups (Table 8–2). Many vitamins, organic nutrients required in small amounts in the diet, are precursors of coenzymes. Coenzymes will be considered in more detail as they are encountered in the discussion of metabolic pathways in Part III of this book.

Figure 8–1 Crystals of pyruvate kinase, an enzyme of the glycolytic pathway. The protein in a crystal is generally characterized by a high degree of purity and structural homogeneity.

Table 8–1 Some enzymes containing or requiring inorganic elements as cofactors

Fe^{2+} or Fe^{3+}	Cytochrome oxidase Catalase Peroxidase
Cu^{2+}	Cytochrome oxidase
Zn^{2+}	Carbonic anhydrase Alcohol dehydrogenase
Mg^{2+}	Hexokinase Glucose-6-phosphatase Pyruvate kinase
Mn^{2+}	Arginase Ribonucleotide reductase
K^+	Pyruvate kinase
Ni^{2+}	Urease
Mo	Dinitrogenase
Se	Glutathione peroxidase

Table 8–2 Some coenzymes serving as transient carriers of specific atoms or functional groups*

Coenzyme	Examples of some chemical groups transferred	Dietary precursor in mammals
Thiamine pyrophosphate	Aldehydes	Thiamin (vitamin B_1)
Flavin adenine dinucleotide	Electrons	Riboflavin (vitamin B_2)
Nicotinamide adenine dinucleotide	Hydride ion ($:H^-$)	Nicotinic acid (niacin)
Coenzyme A	Acyl groups	Pantothenic acid, plus other molecules
Pyridoxal phosphate	Amino groups	Pyridoxine (vitamin B_6)
5′-Deoxyadenosyl-cobalamin (coenzyme B_{12})	H atoms and alkyl groups	Vitamin B_{12}
Biocytin	CO_2	Biotin
Tetrahydrofolate	One-carbon groups	Folate
Lipoate	Electrons and acyl groups	Not required in diet

* The structure and mode of action of these coenzymes are described in Part III of this book.

Finally, some enzymes are modified by phosphorylation, glycosylation, and other processes. Many of these alterations are involved in the regulation of enzyme activity.

Enzymes Are Classified by the Reactions They Catalyze

Many enzymes have been named by adding the suffix "-ase" to the name of their substrate or to a word or phrase describing their activity. Thus urease catalyzes hydrolysis of urea, and DNA polymerase catalyzes the synthesis of DNA. Other enzymes, such as pepsin and trypsin, have names that do not denote their substrates. Sometimes the same enzyme has two or more names, or two different enzymes have the same name. Because of such ambiguities, and the ever-increasing number of newly discovered enzymes, a system for naming and classifying enzymes has been adopted by international agreement. This system places all enzymes in six major classes, each with subclasses, based on the type of reaction catalyzed (Table 8–3). Each enzyme is assigned a four-digit classification number and a systematic name, which identifies the reaction catalyzed. As an example, the formal systematic name of the enzyme catalyzing the reaction

$$\text{ATP} + \text{D-glucose} \longrightarrow \text{ADP} + \text{D-glucose-6-phosphate}$$

is ATP:glucose phosphotransferase, which indicates that it catalyzes the transfer of a phosphate group from ATP to glucose. Its enzyme classification number (E.C. number) is 2.7.1.1; the first digit (2) denotes the class name (transferase) (see Table 8–3); the second digit (7), the

Table 8–3 International classification of enzymes*

No.	Class	Type of reaction catalyzed
1	Oxidoreductases	Transfer of electrons (hydride ions or H atoms)
2	Transferases	Group-transfer reactions
3	Hydrolases	Hydrolysis reactions (transfer of functional groups to water)
4	Lyases	Addition of groups to double bonds, or formation of double bonds by removal of groups
5	Isomerases	Transfer of groups within molecules to yield isomeric forms
6	Ligases	Formation of C—C, C—S, C—O, and C—N bonds by condensation reactions coupled to ATP cleavage

* Most enzymes catalyze the transfer of electrons, atoms, or functional groups. They are therefore classified, given code numbers, and assigned names according to the type of transfer reaction, the group donor, and the group acceptor.

subclass (phosphotransferase); the third digit (1), phosphotransferases with a hydroxyl group as acceptor; and the fourth digit (1), D-glucose as the phosphate-group acceptor. When the systematic name of an enzyme is long or cumbersome, a trivial name may be used—in this case hexokinase.

A complete list and description of the thousands of known enzymes would be well beyond the scope of this book. This chapter is instead devoted primarily to principles and properties common to all enzymes.

How Enzymes Work

The enzymatic catalysis of reactions is essential to living systems. Under biologically relevant conditions, uncatalyzed reactions tend to be slow. Most biological molecules are quite stable in the neutral-pH, mild-temperature, aqueous environment found inside cells. Many common reactions in biochemistry involve chemical events that are unfavorable or unlikely in the cellular environment, such as the transient formation of unstable charged intermediates or the collision of two or more molecules in the precise orientation required for reaction. Reactions required to digest food, send nerve signals, or contract muscle simply do not occur at a useful rate without catalysis.

An enzyme circumvents these problems by providing a specific environment within which a given reaction is energetically more favorable. The distinguishing feature of an enzyme-catalyzed reaction is that it occurs within the confines of a pocket on the enzyme called the **active site** (Fig. 8–2). The molecule that is bound by the active site and acted upon by the enzyme is called the **substrate.** The enzyme–substrate complex is central to the action of enzymes, and it is the starting point for mathematical treatments defining the kinetic behavior of enzyme-catalyzed reactions and for theoretical descriptions of enzyme mechanisms.

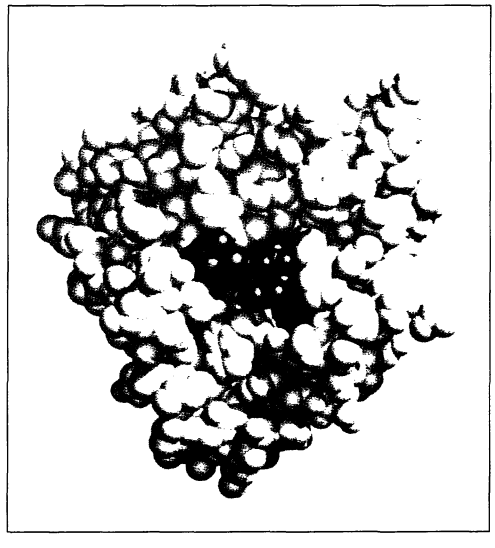


Figure 8–2 Binding of a substrate to an enzyme at the active site. The enzyme chymotrypsin is shown, bound to a substrate (in blue). Some key active-site amino acids are shown in red.

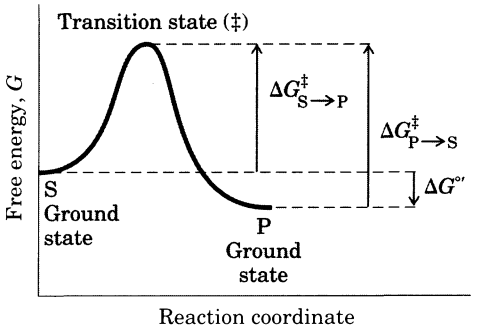


Figure 8–3 Reaction coordinate diagram for a chemical reaction. The free energy of the system is plotted against the progress of the reaction. A diagram of this kind is a description of the energetic course of the reaction, and the horizontal axis (reaction coordinate) reflects the progressive chemical changes (e.g., bond breakage or formation) as S is converted to P. The S and P symbols mark the free energies of the substrate and product ground states. The transition state is indicated by the symbol ‡. The activation energies, $\Delta G^{\ddagger}$, for the S → P and P → S reactions are indicated. $\Delta G^{\circ\prime}$ is the overall standard free-energy change in going from S to P.

Enzymes Affect Reaction Rates, Not Equilibria

A tour through an enzyme-catalyzed reaction serves to introduce some important concepts and definitions.

A simple enzymatic reaction might be written

$$E + S \rightleftharpoons ES \rightleftharpoons EP \rightleftharpoons E + P \qquad (8\text{–}1)$$

where E, S, and P represent the enzyme, substrate, and product, respectively. ES and EP are complexes of the enzyme with the substrate and with the product, respectively.

To understand catalysis, we must first appreciate the important distinction between reaction equilibria (discussed in Chapter 4) and reaction rates. The function of a catalyst is to increase the *rate* of a reaction. Catalysts do not affect reaction *equilibria*. Any reaction, such as S ⇌ P, can be described by a reaction coordinate diagram (Fig. 8–3). This is a picture of the energetic course of the reaction. As introduced in Chapters 1 and 3, energy in biological systems is described in terms of free energy, *G*. In the coordinate diagram, the free energy of the system is plotted against the progress of the reaction (reaction coordinate). In its normal stable form or **ground state,** any molecule (such as S or P) contains a characteristic amount of free energy. To describe the free-energy changes for reactions, chemists define a standard set of conditions (temperature 298 K; partial pressure of gases each 1 atm or 101.3 kPa; concentration of solutes each 1 M), and express the free-energy change for this reacting system as ΔG°, the **standard free-energy change.** Because biochemical systems commonly involve H^+ concentrations far from 1 M, biochemists define a constant $\Delta G^{\circ\prime}$, the standard free-energy change *at pH 7.0,* which we will employ throughout the book. A more complete definition of $\Delta G^{\circ\prime}$ is given in Chapter 13.

The equilibrium between S and P reflects the difference in the free energy of their ground states. In the example shown in Figure 8–3, the free energy of the ground state of P is lower than that of S, so $\Delta G^{\circ\prime}$ for the reaction is negative and the equilibrium favors P. This equilibrium is *not* affected by any catalyst.

A favorable equilibrium, however, does *not* mean that the S → P conversion is fast. The rate of a reaction is dependent on an entirely different parameter. There is an energetic barrier between S and P that represents the energy required for alignment of reacting groups, formation of transient unstable charges, bond rearrangements, and other transformations required for the reaction to occur in either direction. This is illustrated by the energetic "hill" in Figures 8–3 and 8–4. To undergo reaction, the molecules must overcome this barrier and therefore must be raised to a higher energy level. At the top of the energy hill is a point at which decay to the S or P state is equally probable (it is downhill either way). This is called the **transition state.** The transition state is not a chemical species with any significant stability and should not be confused with a reaction intermediate. It is simply a fleeting molecular moment in which events such as bond breakage, bond formation, and charge development have proceeded to the precise point at which a collapse to either substrate or product is equally likely. The difference between the energy levels of the ground state and the transition state is called the **activation energy** ($\Delta G^{\ddagger}$). The rate of a reaction reflects this activation energy; a higher activation energy corresponds to a slower reaction. Reaction rates can be increased by raising the temperature, thereby increasing the number of molecules with sufficient energy to overcome this energy barrier. Alternatively

the activation energy can be lowered by adding a catalyst (Fig. 8–4). *Catalysts enhance reaction rates by lowering activation energies.*

Enzymes are no exception to the rule that catalysts do not affect reaction equilibria. The bidirectional arrows in Equation 8–1 make this point: any enzyme that catalyzes the reaction S → P also catalyzes the reaction P → S. Its only role is to accelerate the interconversion of S and P. The enzyme is not used up in the process, and the equilibrium point is unaffected. However, the reaction reaches equilibrium much faster when the appropriate enzyme is present because the rate of the reaction is increased.

This general principle can be illustrated by considering the reaction of glucose and O_2 to form CO_2 and H_2O. This reaction has a very large and negative $\Delta G^{\circ\prime}$, and at equilibrium the amount of glucose present is negligible. Glucose, however, is a stable compound, and it can be combined in a container with O_2 almost indefinitely without reacting. Its stability reflects a high activation energy for reaction. In cells, glucose is broken down in the presence of O_2 to CO_2 and H_2O in a pathway of reactions catalyzed by enzymes. These enzymes not only accelerate the reactions, they organize and control them so that much of the energy released in this process is recovered in other forms and made available to the cell for other tasks. This is the primary energy-yielding pathway for cells (Chapters 14 and 18), and these enzymes allow it to occur on a time scale that is useful to the cells.

In practice, any reaction may have several steps involving the formation and decay of transient chemical species called **reaction intermediates.** When the S ⇌ P reaction is catalyzed by an enzyme, the ES and EP complexes are intermediates (Eqn 8–1); they occupy valleys in the reaction coordinate diagram (Fig. 8–4). When several steps occur in a reaction, the overall rate is determined by the step (or steps) with the highest activation energy; this is called the **rate-limiting step.** In a simple case the rate-limiting step is the highest-energy point in the diagram for interconversion of S and P (Fig. 8–4). In practice, the rate-limiting step can vary with reaction conditions, and for many enzymes several steps may have similar activation energies, which means they are all partially rate-limiting.

As described in Chapter 1, activation energies are energetic barriers to chemical reactions; these barriers are crucial to life itself. The stability of a molecule increases with the height of its activation barrier. Without such energetic barriers, complex macromolecules would revert spontaneously to much simpler molecular forms. The complex and highly ordered structures and metabolic processes in every cell could not exist. Enzymes have evolved to lower activation energies *selectively* for reactions that are needed for cell survival.

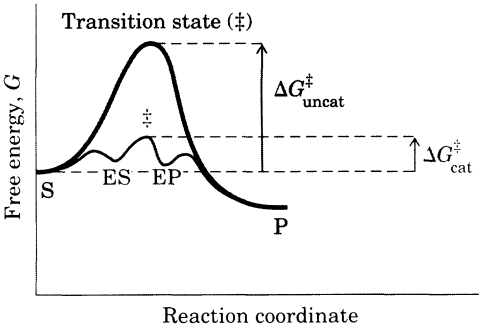


Figure 8–4 Reaction coordinate diagram comparing the enzyme-catalyzed and uncatalyzed reactions S → P. The ES and EP intermediates occupy minima in the energetic progress curve of the enzyme-catalyzed reaction. The terms $\Delta G^{\ddagger}_{uncat}$ and $\Delta G^{\ddagger}_{cat}$ correspond to the activation energies for the uncatalyzed and catalyzed reactions, respectively. The activation energy for the overall process is lower when the enzyme catalyzes the reaction.

Reaction Rates and Equilibria Have Precise Thermodynamic Definitions

Reaction *equilibria* are inextricably linked to $\Delta G^{\circ\prime}$ and reaction *rates* are linked to $\Delta G^{\ddagger}$. A basic introduction to these thermodynamic relationships is the next step in understanding how enzymes work.

As introduced in Chapter 4, an equilibrium such as S ⇌ P is described by an **equilibrium constant,** K_{eq}. Under the standard conditions used to compare biochemical processes, an equilibrium constant is denoted K_{eq}':

$$K_{eq}' = \frac{[P]}{[S]} \qquad (8\text{–}2)$$

From thermodynamics, the relationship between K_{eq}' and $\Delta G^{\circ\prime}$ can be described by the expression

$$\Delta G^{\circ\prime} = -RT \ln K_{eq}' \tag{8–3}$$

where R is the gas constant (8.315 J/mol • K) and T is the absolute temperature (298 K). This expression will be developed and discussed in more detail in Chapter 13. The important point here is that the equilibrium constant is a direct reflection of the overall standard free-energy change in the reaction (Table 8–4). A large negative value for $\Delta G^{\circ\prime}$ reflects a favorable reaction equilibrium, but as already noted this does not mean the reaction will proceed at a rapid rate.

Table 8–4 The relationship between K_{eq}' and $\Delta G^{\circ\prime}$ (see Eqn 8–3)

K_{eq}'	$\Delta G^{\circ\prime}$ (kJ/mol)
10^{-6}	34.2
10^{-5}	28.5
10^{-4}	22.8
10^{-3}	17.1
10^{-2}	11.4
10^{-1}	5.7
1	0.0
10^{1}	−5.7
10^{2}	−11.4
10^{3}	−17.1

The rate of any reaction is determined by the concentration of the reactant (or reactants) and by a **rate constant,** usually denoted by the symbol k. For the unimolecular reaction S → P, the rate or velocity of the reaction, V, representing the amount of S that has reacted per unit time, is expressed by a **rate law:**

$$V = k[S] \tag{8–4}$$

In this reaction, the rate depends only on the concentration of S. This is called a first-order reaction. The factor k is a proportionality constant that reflects the probability of reaction under a given set of conditions (pH, temperature, etc.). Here, k is a first-order rate constant and has units of reciprocal time (e.g., s^{-1}). If a first-order reaction has a rate constant k of 0.03 s^{-1}, this may be interpreted (qualitatively) to mean that 3% of the available S will be converted to P in 1 s. A reaction with a rate constant of 2,000 s^{-1} will be over in a small fraction of a second. If the reaction rate depends on the concentration of two different compounds, or if two molecules of the same compound react, the reaction is second order and k is a second-order rate constant (with the units $M^{-1}s^{-1}$). The rate law has the form

$$V = k[S_1][S_2] \tag{8–5}$$

From transition-state theory, an expression can be derived that relates the magnitude of a rate constant to the activation energy:

$$k = \frac{\mathbf{k}\,T}{h} e^{-\Delta G^{\ddagger}/RT} \tag{8–6}$$

where **k** is the Boltzmann constant and h is Planck's constant. The important point here is that the relationship between the rate constant, k, and the activation energy, $\Delta G^{\ddagger}$, is inverse and exponential. In simplified terms, this is the basis for the statement that a lower activation energy means a higher reaction rate, and vice versa.

Now we turn from *what* enzymes do to *how* they do it.

A Few Principles Explain the Catalytic Power and Specificity of Enzymes

Enzymes are extraordinary catalysts. The rate enhancements brought about by enzymes are often in the range of 7 to 14 orders of magnitude (Table 8–5). Enzymes are also very specific, readily discriminating between substrates with quite similar structures. How can these enormous and highly selective rate enhancements be explained? Where does the energy come from to provide a dramatic lowering of the activation energies for specific reactions?

Table 8–5 Some rate enhancements produced by enzymes

Carbonic anhydrase	10^{7}
Phosphoglucomutase	10^{12}
Succinyl-CoA transferase	10^{13}
Urease	10^{14}

Part of the explanation for enzyme action lies in well-studied chemical reactions that take place between a substrate and enzyme functional groups (specific amino acid side chains, metal ions, and coenzymes). Catalytic functional groups on enzymes can interact transiently with a substrate and activate it for reaction. In many cases, these groups lower the activation energy (and thereby accelerate the reaction) by providing a lower-energy reaction path. Common types of enzymatic catalysis are outlined later in this chapter.

Catalytic functional groups, however, are not the only contributor to enzymatic catalysis. The energy required to lower activation energies is generally derived from weak, noncovalent interactions between the substrate and the enzyme. The factor that really sets enzymes apart from most nonenzymatic catalysts is the formation of a specific ES complex. The interaction between substrate and enzyme in this complex is mediated by the same forces that stabilize protein structure, including hydrogen bonds and hydrophobic, ionic, and van der Waals interactions (Chapter 7). Formation of each weak interaction in the ES complex is accompanied by a small release of free energy that provides a degree of stability to the interaction. The energy derived from enzyme–substrate interaction is called **binding energy.** Its significance extends beyond a simple stabilization of the enzyme–substrate interaction. *Binding energy is the major source of free energy used by enzymes to lower the activation energies of reactions.*

Two fundamental and interrelated principles provide a general explanation for how enzymes work. First, the catalytic power of enzymes is ultimately derived from the free energy released in forming the multiple weak bonds and interactions that occur between an enzyme and its substrate. This binding energy provides specificity as well as catalysis. Second, weak interactions are optimized in the reaction transition state; enzyme active sites are complementary not to the substrates per se, but to the transition states of the reactions they catalyze. These themes are critical to an understanding of enzymes, and they now become the primary focus of the chapter.

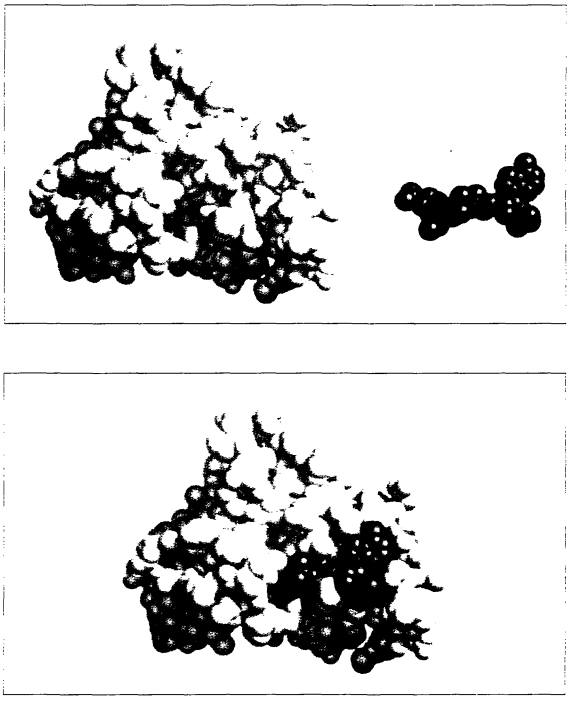


Figure 8–5 Complementary shapes of a substrate and its binding site on an enzyme. The enzyme dihydrofolate reductase is shown with its substrate, $NADP^+$ (red), unbound (top) and bound (bottom). Part of a tetrahydrofolate molecule (yellow), also bound to the enzyme, is visible. The $NADP^+$ binds to a pocket that is complementary to it in shape and ionic properties. Emil Fischer proposed that enzymes and their substrates have shapes that closely complement each other, like a lock and key. This idea can readily be extended to the interactions of other types of proteins with ligands or other proteins. In reality, the complementarity is rarely perfect, and the interaction of a protein with a ligand often involves changes in the conformation of one or both molecules. This *lack* of perfect complementarity between an enzyme and its substrate (not evident in this figure) is important to enzymatic catalysis.

Weak Interactions between Enzyme and Substrate Are Optimized in the Transition State

How does an enzyme use binding energy to lower the activation energy for reaction? Formation of the ES complex is not the explanation in itself, although some of the earliest considerations of enzyme mechanisms began with this idea. Studies on enzyme specificity carried out by Emil Fischer led him to propose, in 1894, that enzymes were structurally complementary to their substrates, so that they fit together like a "lock and key" (Fig. 8–5).

This elegant idea, that a specific (exclusive) interaction between two biological molecules is mediated by molecular surfaces with complementary shapes, has greatly influenced the development of biochemistry, and lies at the heart of many biochemical processes. However, the "lock and key" hypothesis can be misleading when applied to the question of enzymatic catalysis. An enzyme completely complementary to its substrate would be a very poor enzyme. Consider an imaginary reaction, the breaking of a metal stick. The uncatalyzed reaction is shown in Figure 8–6a. We will examine two imaginary enzymes to catalyze this reaction, both of which employ magnetic forces as a paradigm for the binding energy used by real enzymes. We first

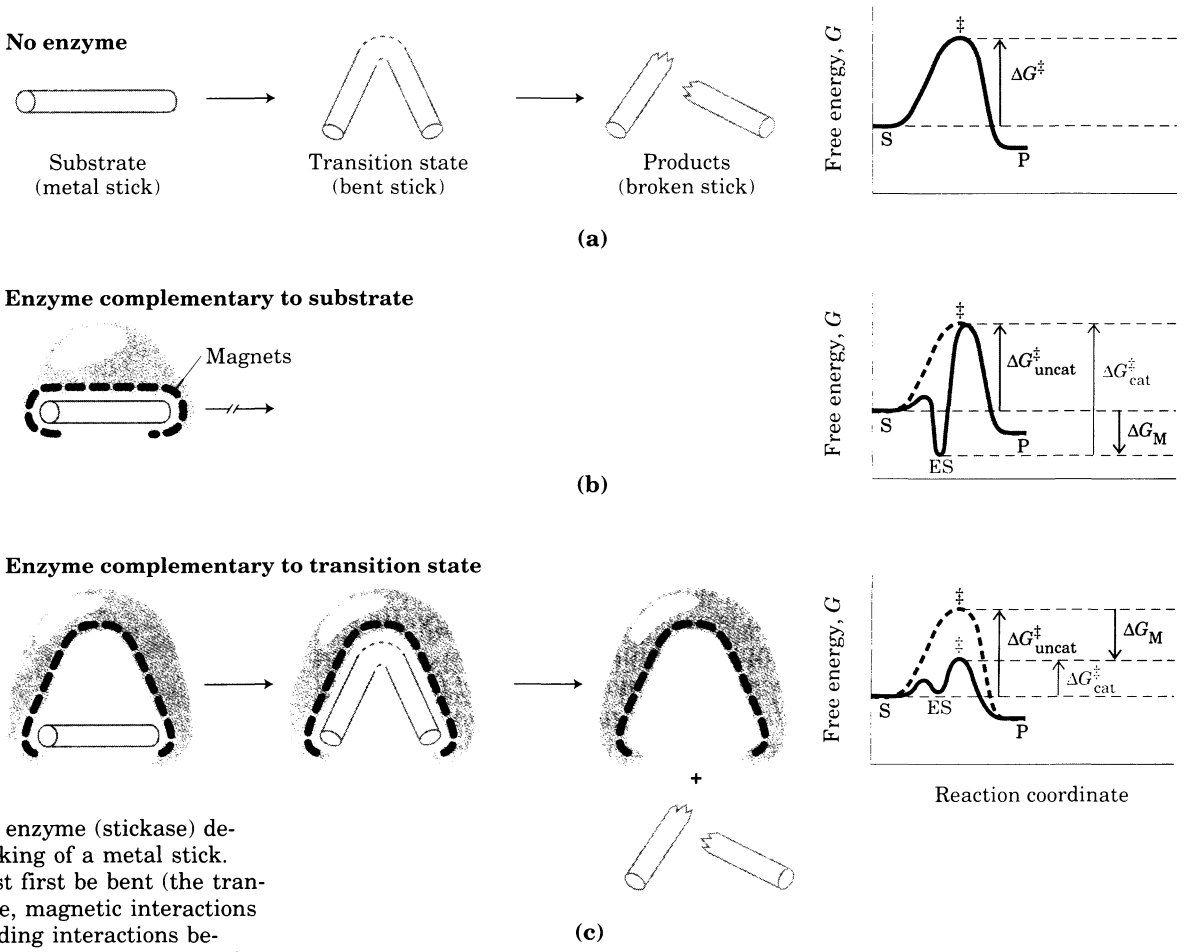


Figure 8–6 An imaginary enzyme (stickase) designed to catalyze the breaking of a metal stick. **(a)** To break, the stick must first be bent (the transition state). In the stickase, magnetic interactions take the place of weak-bonding interactions between enzyme and substrate. **(b)** An enzyme with a magnet-lined pocket complementary in structure to the stick (the substrate) will stabilize this substrate. Bending will be impeded by the magnetic attraction between stick and stickase. **(c)** An enzyme complementary to the reaction transition state will help to destabilize the stick, resulting in catalysis of the reaction. The magnetic interactions provide energy that compensates for the increase in free energy required to bend the stick. Reaction coordinate diagrams show the energetic consequences of complementarity to substrate versus complementarity to transition state. The term ΔG_M represents the energy contributed by the magnetic interactions between the stick and stickase. When the enzyme is complementary to the substrate, as in **(b)**, the ES complex is more stable and has less free energy in the ground state than substrate alone. The result is an *increase* in the activation energy. For simplicity, the EP complexes are not shown.

design an enzyme perfectly complementary to the substrate (Fig. 8–6b). The active site of this "stickase" enzyme is a pocket lined with magnets. To react (break), the stick must reach the transition state of the reaction. The stick fits so tightly in the active site that it cannot bend, because bending of the stick would eliminate some of the magnetic interactions between stick and enzyme. Such an enzyme *impedes* the reaction, stabilizing the substrate instead. In a reaction coordinate diagram (Fig. 8–6b), this kind of ES complex would correspond to an energy well from which it would be difficult for the substrate to escape. Such an enzyme would be useless.

The modern notion of enzymatic catalysis was first proposed by Haldane in 1930, and elaborated by Linus Pauling in 1946. In order to catalyze reactions, an enzyme must be complementary to the *reaction transition state.* This means that the optimal interactions (through weak bonding) between substrate and enzyme can occur only in the transition state. Figure 8–6c demonstrates how such an enzyme *can* work. The metal stick binds, but only a few magnetic interactions are used in forming the ES complex. The bound substrate must still undergo the increase in free energy needed to reach the transition state. Now, however, the increase in free energy required to draw the stick into a bent and partially broken conformation is offset or "paid for" by the magnetic interactions that form between the enzyme and substrate in the transition state. Many of these interactions involve parts of the stick that are distant from the point of breakage; thus interactions

between the stickase and nonreacting parts of the stick provide some of the energy needed to catalyze stick breakage. This "energy payment" translates into a lower net activation energy and a faster reaction rate.

Real enzymes work on an analogous principle. Some weak interactions are formed in the ES complex, but the full complement of possible weak interactions between substrate and enzyme are formed only when the substrate reaches the transition state. The free energy (binding energy) released by the formation of these interactions partially offsets the energy required to get to the top of the energy hill. The summation of the unfavorable (positive) $\Delta G^{\ddagger}$ and the favorable (negative) binding energy (ΔG_B) results in a lower *net* activation energy (Fig. 8–7). Even on the enzyme, the transition state represents a brief point in time that the substrate spends atop an energy hill. The enzyme-catalyzed reaction is much faster than the uncatalyzed process, however, because the hill is much smaller. The important principle is that weak-bonding interactions between the enzyme and the substrate provide the major driving force for enzymatic catalysis. The groups on the substrate that are involved in these weak interactions can be at some distance from the bonds that are broken or changed. The weak interactions that are formed only in the transition state are those that make the primary contribution to catalysis.

The requirement for multiple weak interactions to drive catalysis is one reason why enzymes (and some coenzymes) are so large. The enzyme must provide functional groups for ionic interactions, hydrogen bonds, and other interactions, and also precisely position these groups so that binding energy is optimized in the transition state.

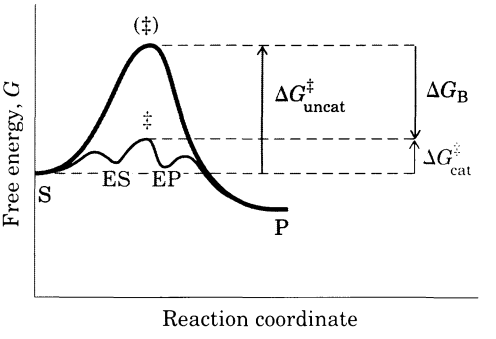


Figure 8–7 The role of binding energy in catalysis. To lower the activation energy for a reaction, the system must acquire an amount of energy equivalent to the amount by which $\Delta G^{\ddagger}$ is lowered. This energy comes largely from binding energy (ΔG_B) contributed by formation of weak noncovalent interactions between substrate and enzyme in the transition state. The role of ΔG_B is analogous to that of ΔG_M in Fig. 8–6.

Enzymes Use Binding Energy to Provide Reaction Specificity and Catalysis

Can binding energy account for the huge rate accelerations brought about by enzymes? Yes. As a point of reference, Equation 8–6 allows us to calculate that about 5.7 kJ/mol of free energy is required to accelerate a first-order reaction by a factor of ten under conditions commonly found in cells. The energy available from formation of a single weak interaction is generally estimated to be 4 to 30 kJ/mol. The overall energy available from formation of a number of such interactions can lower activation energies by the 60 to 80 kJ/mol required to explain the large rate enhancements observed for many enzymes.

The same binding energy that provides energy for catalysis also makes the enzyme specific. **Specificity** refers to the ability of an enzyme to discriminate between two competing substrates. Conceptually, this idea is easy to distinguish from the idea of catalysis. Catalysis and specificity are much more difficult to distinguish experimentally because they arise from the same phenomenon. If an enzyme active site has functional groups arranged optimally to form a variety of weak interactions with a given substrate in the transition state, the enzyme will not be able to interact as well with any other substrate. For example, if the normal substrate has a hydroxyl group that forms a specific hydrogen bond with a Glu residue on the enzyme, any molecule lacking that particular hydroxyl group will generally be a poorer substrate for the enzyme. In addition, any molecule with an extra functional group for which the enzyme has no pocket or binding site is likely to be excluded from the enzyme. In general, *specificity* is also derived from the formation of multiple weak interactions between the enzyme and many or all parts of its specific substrate molecule.

The general principles outlined above can be illustrated by a variety of recognized catalytic mechanisms. These mechanisms are not mutually exclusive, and a given enzyme will often incorporate several in its own complete mechanism of action. It is often difficult to quantify the contribution of any one catalytic mechanism to the rate and/or specificity of an enzyme-catalyzed reaction.

Binding energy is the dominant driving force in several mechanisms, and these can be the major, and sometimes the only, contribution to catalysis. This can be illustrated by considering what needs to occur for a reaction to take place. Prominent physical and thermodynamic barriers to reaction include (1) entropy, the relative motion of two molecules in solution; (2) the solvated shell of hydrogen-bonded water that surrounds and helps to stabilize most biomolecules in aqueous solution; (3) the electronic or structural distortion of substrates that must occur in many reactions; and (4) the need to achieve proper alignment of appropriate catalytic functional groups on the enzyme. Binding energy can be used to overcome all of these barriers.

A large reduction in the relative motions of two substrates that are to react, or **entropy reduction,** is one of the obvious benefits of binding them to an enzyme. Binding energy holds the substrates in the proper orientation to react—a major contribution to catalysis because productive collisions between molecules in solution can be exceedingly rare. Substrates can be precisely aligned on the enzyme. A multitude of weak interactions between each substrate and strategically located groups on the enzyme clamp the substrate molecules into the proper positions. Studies have shown that constraining the motion of two reactants can produce rate enhancements of as much as 10^8 M (a rate equivalent to that expected if the reactants were present at the impossibly high concentration of 100,000,000 M).

Formation of weak bonds between substrate and enzyme also results in **desolvation** of the substrate. Enzyme–substrate interactions replace most or all of the hydrogen bonds that may exist between the substrate and water in solution.

Binding energy involving weak interactions formed only in the reaction transition state helps to compensate thermodynamically for any **strain** or distortion that the substrate must undergo to react. Distortion of the substrate in the transition state may be electrostatic or structural.

The enzyme itself may undergo a change in conformation when the substrate binds, induced again by multiple weak interactions with the substrate. This is referred to as **induced fit,** a mechanism postulated by Daniel Koshland in 1958. Induced fit may serve to bring specific functional groups on the enzyme into the proper orientation to catalyze the reaction. The conformational change may also permit formation of additional weak-bonding interactions in the transition state. In either case the new conformation may have enhanced catalytic properties.

Specific Catalytic Groups Contribute to Catalysis

Once a substrate is bound, additional modes of catalysis can be employed by an enzyme to aid bond cleavage and formation, using properly positioned catalytic functional groups. Among the best characterized mechanisms are **general acid–base catalysis** and **covalent catalysis.** These are distinct from mechanisms based on binding energy because they generally involve *covalent* interaction with a substrate, or group transfer to or from a substrate.

General Acid–Base Catalysis Many biochemical reactions involve the formation of unstable charged intermediates that tend to break down rapidly to their constituent reactant species, thus failing to undergo reaction (Fig. 8–8). Charged intermediates can often be stabilized (and the reaction thereby catalyzed) by transferring protons to or from the substrate or intermediate to form a species that breaks down to products more readily than to reactants. The proton transfers can involve the constituents of water alone or may involve other weak proton donors or acceptors. Catalysis that simply involves the H^+ (H_3O^+) or OH^- ions present in water is referred to as **specific acid or base catalysis.** If protons are transferred between the intermediate and water faster than the intermediate breaks down to reactants, the intermediate will effectively be stabilized every time it forms. No additional catalysis mediated by other proton acceptors or donors will occur. In many cases, however, water is not enough. The term general acid–base catalysis refers to proton transfers mediated by other classes of molecules. It is observed in aqueous solutions only when the unstable reaction intermediate breaks down to reactants faster than

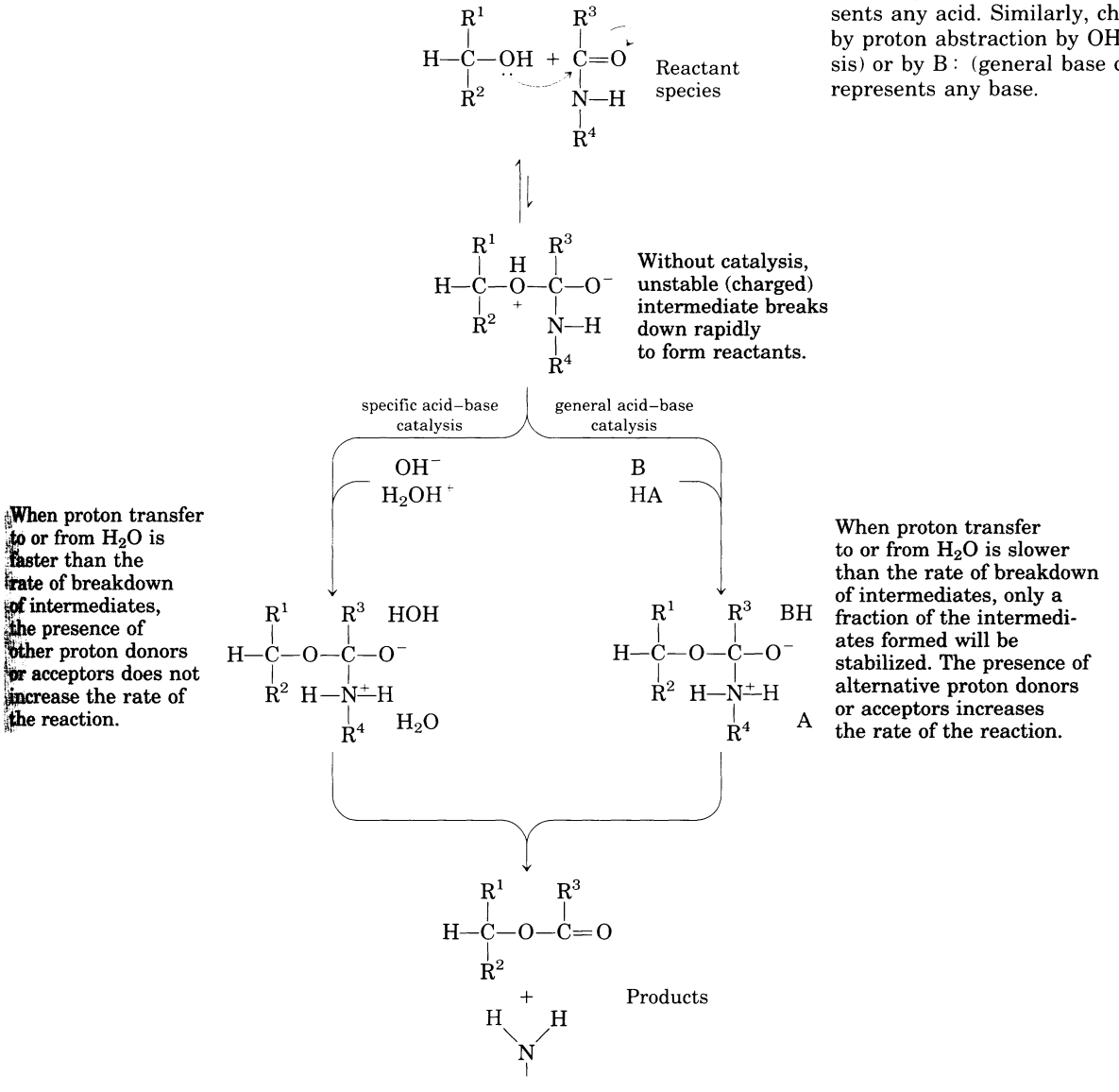


Figure 8–8 Unfavorable charge development during cleavage of an amide. This type of reaction is catalyzed by chymotrypsin and other proteases. Charge development can be circumvented by donation of a proton by H_3O^+ (specific acid catalysis) or by HA (general acid catalysis), where HA represents any acid. Similarly, charge can be neutralized by proton abstraction by OH^- (specific base catalysis) or by B: (general base catalysis), where B: represents any base.

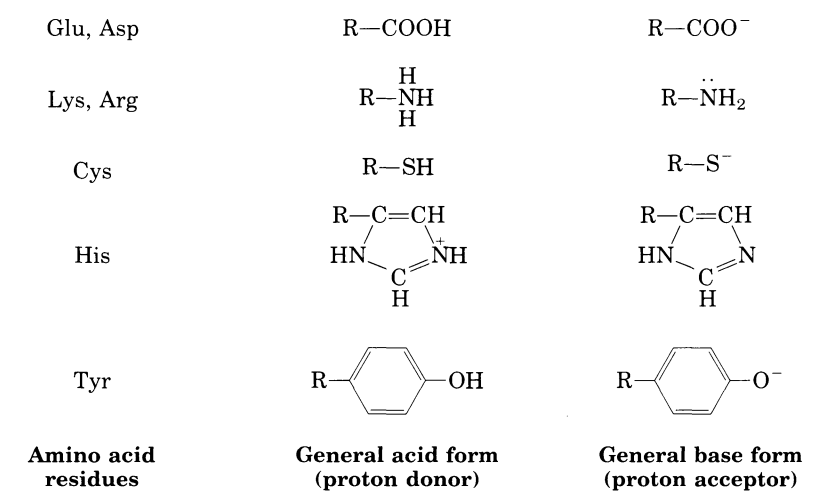


Figure 8–9 Many organic reactions are promoted by proton donors (general acids) or proton acceptors (general bases). The active sites of some enzymes contain amino acid functional groups, such as those shown here, that can participate in the catalytic process as proton donors or proton acceptors.

the rate of proton transfer to or from water. A variety of weak organic acids can supplement water as proton donors in this situation, or weak organic bases can serve as proton acceptors. A number of amino acid side chains can similarly act as proton donors and acceptors (Fig. 8–9). These groups can be precisely positioned in an enzyme active site to allow proton transfers, providing rate enhancements on the order of 10^2 to 10^5.

Covalent Catalysis This involves the formation of a transient covalent bond between the enzyme and substrate. Consider the hydrolysis of a bond between groups A and B:

$$\text{A—B} \xrightarrow{\text{H}_2\text{O}} \text{A} + \text{B}$$

In the presence of a covalent catalyst (an enzyme with a nucleophilic group X:) the reaction becomes

$$\text{A—B} + \text{X:} \longrightarrow \text{A—X} + \text{B} \xrightarrow{\text{H}_2\text{O}} \text{A} + \text{X:} + \text{B}$$

This alters the pathway of the reaction and results in catalysis only when the new pathway has a lower activation energy than the uncatalyzed pathway. Both of the new steps must be faster than the uncatalyzed reaction. A number of amino acid side chains (including all of those in Fig. 8–9), as well as the functional groups of some enzyme cofactors, serve as nucleophiles on some enzymes in the formation of covalent bonds with substrates. These covalent complexes always undergo further reaction to regenerate the free enzyme. The covalent bond formed between the enzyme and the substrate can activate a substrate for further reaction in a manner that is usually specific to the group or coenzyme involved. The chemical contribution to catalysis provided by individual coenzymes is described in detail as each coenzyme is encountered in Part III of this book.

Metal Ion Catalysis Metals, whether tightly bound to the enzyme or taken up from solution along with the substrate, can participate in catalysis in several ways. Ionic interactions between an enzyme-bound metal and the substrate can help orient a substrate for reaction or stabilize charged reaction transition states. This use of weak-bonding interactions between the metal and the substrate is similar to some of the uses of enzyme–substrate binding energy described earlier. Metals can also mediate oxidation–reduction reactions by reversible changes in the metal ion's oxidation state. Nearly a third of all known enzymes require one or more metal ions for catalytic activity.

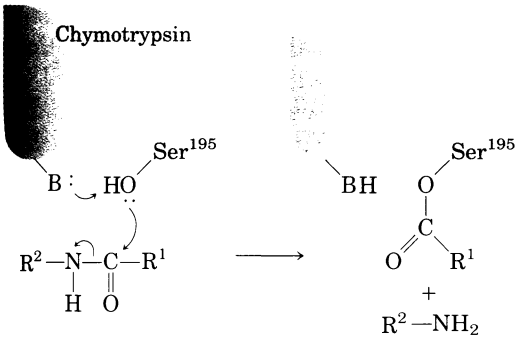


Figure 8–10 The first step in the reaction catalyzed by chymotrypsin, also called the acylation step. The hydroxyl group of Ser^{195} is the nucleophile in a reaction aided by general base catalysis (the base is the side chain of His^{57}). The chymotrypsin reaction is described in more detail in Fig. 8–19.

A combination of several catalytic strategies is usually employed by an enzyme to bring about a rate enhancement. A good example of the use of both covalent catalysis and general acid–base catalysis occurs in chymotrypsin. The first step in the reaction catalyzed by chymotrypsin is the cleavage of a peptide bond. This is accompanied by formation of a covalent linkage between a Ser residue on the enzyme and part of the substrate; this reaction is enhanced by general base catalysis by other groups on the enzyme (Fig. 8–10). The chymotrypsin reaction is described in more detail later in this chapter.

Enzyme Kinetics as an Approach to Understanding Mechanism

Multiple approaches are commonly used to study the mechanism of action of purified enzymes. A knowledge of the three-dimensional structure of a protein provides important information. The value of structural information is greatly enhanced by classical protein chemistry and modern methods of site-directed mutagenesis (changing the amino acid sequence of a protein in a defined way by genetic engineering; see Chapter 28) that permit enzymologists to examine the role of individual amino acids in structure and enzyme action. However, the *rate* of the catalyzed reaction can also reveal much about the enzyme. The study of reaction rates and how they change in response to changes in experimental parameters is known as **kinetics.** This is the oldest approach to understanding enzyme mechanism, and one that remains most important today. The following is a basic introduction to the kinetics of enzyme-catalyzed reactions. The more advanced student may wish to consult the texts and articles cited at the end of this chapter.

Substrate Concentration Affects the Rate of Enzyme-Catalyzed Reactions

A discussion of kinetics must begin with some fundamental concepts. One of the key factors affecting the rate of a reaction catalyzed by a purified enzyme in vitro is the amount of substrate present, [S]. But studying the effects of substrate concentration is complicated by the fact that [S] changes during the course of a reaction as substrate is converted to product. One simplifying approach in a kinetic experiment is to measure the initial rate (or initial velocity), designated V_0, when [S] is generally much greater than the concentration of enzyme. Then, if the time is sufficiently short following the start of a reaction, changes in [S] are negligible, and [S] can be regarded as a constant.

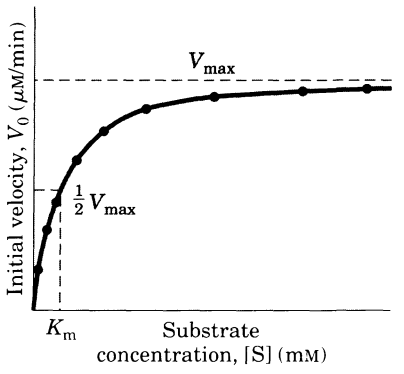


Figure 8–11 Effect of substrate concentration on the initial velocity of an enzyme-catalyzed reaction. V_{max} can only be approximated from such a plot, because V_0 will approach but never quite reach V_{max}. The substrate concentration at which V_0 is half maximal is K_m, the Michaelis–Menten constant. The concentration of enzyme E in an experiment such as this is generally so low that [S] $\gg$ [E] even when [S] is described as low or relatively low. The units given are typical for enzyme-catalyzed reactions and are presented only to help illustrate the meaning of V_0 and [S]. (Note that the curve describes *part* of a rectangular hyperbola, with one asymptote at V_{max}. If the curve were continued below [S] = 0, it would approach a vertical asymptote at [S] = $-K_m$.)

Leonor Michaelis
1875–1949

Maud Menten
1879–1960

The effect on V_0 of varying [S] when the enzyme concentration is held constant is shown in Figure 8–11. At relatively low concentrations of substrate, V_0 increases almost linearly with an increase in [S]. At higher substrate concentrations, V_0 increases by smaller and smaller amounts in response to increases in [S]. Finally, a point is reached beyond which there are only vanishingly small increases in V_0 with increasing [S] (Fig. 8–11). This plateau is called the maximum velocity, V_{max}.

The ES complex is the key to understanding this kinetic behavior, just as it represented a starting point for the discussion of catalysis. The kinetic pattern in Figure 8–11 led Victor Henri to propose in 1903 that an enzyme combines with its substrate molecule to form the ES complex as a necessary step in enzyme catalysis. This idea was expanded into a general theory of enzyme action, particularly by Leonor Michaelis and Maud Menten in 1913. They postulated that the enzyme first combines reversibly with its substrate to form an enzyme–substrate complex in a relatively fast reversible step:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \qquad (8\text{–}7)$$

The ES complex then breaks down in a slower second step to yield the free enzyme and the reaction product P:

$$ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P \qquad (8\text{–}8)$$

In this model the second reaction (Eqn 8–8) is slower and therefore limits the rate of the overall reaction. It follows that the overall rate of the enzyme-catalyzed reaction must be proportional to the concentration of the species that reacts in the second step, that is, ES.

At any given instant in an enzyme-catalyzed reaction, the enzyme exists in two forms, the free or uncombined form E and the combined form ES. At low [S], most of the enzyme will be in the uncombined form E. Here, the rate will be proportional to [S] because the equilibrium of Equation 8–7 will be pushed toward formation of more ES as [S] is increased. The maximum initial rate of the catalyzed reaction (V_{max}) is observed when virtually all of the enzyme is present as the ES complex and the concentration of E is vanishingly small. Under these conditions, the enzyme is "saturated" with its substrate, so that further increases in [S] have no effect on rate. This condition will exist when [S] is sufficiently high that essentially all the free enzyme will have been converted into the ES form. After the ES complex breaks down to yield the product P, the enzyme is free to catalyze another reaction. The saturation effect is a distinguishing characteristic of enzyme catalysts and is responsible for the plateau observed in Figure 8–11.

When the enzyme is first mixed with a large excess of substrate, there is an initial period called the **pre-steady state** during which the concentration of the ES complex builds up. The pre-steady state is usually too short to be easily observed. The reaction quickly achieves a **steady state** in which [ES] (and the concentration of any other intermediates) remains approximately constant over time. The measured V_0 generally reflects the steady state even though V_0 is limited to early times in the course of the reaction. Michaelis and Menten concerned themselves with the steady-state rate, and this type of analysis is referred to as **steady-state kinetics.**

The Relationship between Substrate Concentration and Enzymatic Reaction Rate Can Be Expressed Quantitatively

Figure 8–11 shows the relationship between [S] and V_0 for an enzymatic reaction. The curve expressing this relationship has the same general shape for most enzymes (it approaches a rectangular hyperbola). The hyperbolic shape of this curve can be expressed algebraically by the Michaelis–Menten equation, derived by these workers starting from their basic hypothesis that the rate-limiting step in enzymatic reactions is the breakdown of the ES complex to form the product and the free enzyme.

The important terms are [S], V_0, V_{max}, and a constant called the Michaelis–Menten constant or K_m. All of these terms are readily measured experimentally.

Here we shall develop the basic logic and the algebraic steps in a modern derivation of the Michaelis–Menten equation. The derivation starts with the two basic reactions involved in the formation and breakdown of ES (Eqns 8–7 and 8–8). At early times in the reaction, the concentration of the product [P] is negligible and the simplifying assumption is made that k_{-2} can be ignored. The overall reaction then reduces to

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \qquad (8\text{–}9)$$

V_0 is determined by the breakdown of ES to give product, which is determined by [ES]:

$$V_0 = k_2[ES] \qquad (8\text{–}10)$$

As [ES] in Equation 8–10 is not easily measured experimentally, we must begin by finding an alternative expression for [ES]. First, we will introduce the term $[E_t]$, representing the total enzyme concentration (the sum of the free and substrate-bound enzyme). Free or unbound enzyme can then be represented by $[E_t] - [ES]$. Also, because [S] is ordinarily far greater than $[E_t]$, the amount of substrate bound by the enzyme at any given time is negligible compared with the total [S]. With these in mind, the following steps will lead us to an expression for V_0 in terms of parameters that are easily measured.

Step 1. The rates of formation and breakdown of ES are determined by the steps governed by the rate constants k_1 (formation) and $k_{-1} + k_2$ (breakdown), according to the expressions

$$\text{Rate of ES formation} = k_1([E_t] - [ES])[S] \qquad (8\text{–}11)$$

$$\text{Rate of ES breakdown} = k_{-1}[ES] + k_2[ES] \qquad (8\text{–}12)$$

Step 2. An important assumption is now made that the initial rate of reaction reflects a steady state in which [ES] is constant, i.e., the rate of formation of ES is equal to its rate of breakdown. This is called the steady-state assumption. The expressions in Equations 8–11 and 8–12 can be equated at the steady state, giving

$$k_1([E_t] - [ES])[S] = k_{-1}[ES] + k_2[ES] \qquad (8\text{–}13)$$

Step 3. A series of algebraic steps is now taken to solve Equation 8–13 for [ES]. The left side is multiplied out and the right side is simplified to give

$$k_1[E_t][S] - k_1[ES][S] = (k_{-1} + k_2)[ES] \qquad (8\text{–}14)$$

Adding the term k_1[ES][S] to both sides of the equation and simplifying gives

$$k_1[E_t][S] = (k_1[S] + k_{-1} + k_2)[ES] \tag{8–15}$$

Solving this equation for [ES] gives

$$[ES] = \frac{k_1[E_t][S]}{k_1[S] + k_{-1} + k_2} \tag{8–16}$$

This can now be simplified further, in such a way as to combine the rate constants into one expression:

$$[ES] = \frac{[E_t][S]}{[S] + (k_2 + k_{-1})/k_1} \tag{8–17}$$

The term $(k_2 + k_{-1})/k_1$ is defined as the **Michaelis–Menten constant,** K_m. Substituting this into Equation 8–17 simplifies the expression to

$$[ES] = \frac{[E_t][S]}{K_m + [S]} \tag{8–18}$$

Step 4. V_0 can now be expressed in terms of [ES]. Equation 8–18 is used to substitute for [ES] in Equation 8–10, giving

$$V_0 = \frac{k_2[E_t][S]}{K_m + [S]} \tag{8–19}$$

This equation can be further simplified. Because the maximum velocity will occur when the enzyme is saturated and [ES] = [E_t], V_{max} can be defined as $k_2[E_t]$. Substituting this in Equation 8–19 gives

$$V_0 = \frac{V_{max}[S]}{K_m + [S]} \tag{8–20}$$

This is the **Michaelis–Menten equation,** the rate equation for a one-substrate, enzyme-catalyzed reaction. It is a statement of the quantitative relationship between the initial velocity V_0, the maximum initial velocity V_{max}, and the initial substrate concentration [S], all related through the Michaelis–Menten constant K_m. Does the equation fit the facts? Yes; we can confirm this by considering the limiting situations where [S] is very high or very low, as shown in Figure 8–12.

An important numerical relationship emerges from the Michaelis–Menten equation in the special case when V_0 is exactly one-half V_{max} (Fig. 8–12). Then

$$\frac{V_{max}}{2} = \frac{V_{max}[S]}{K_m + [S]} \tag{8–21}$$

On dividing by V_{max}, we obtain

$$\frac{1}{2} = \frac{[S]}{K_m + [S]} \tag{8–22}$$

Solving for K_m, we get K_m + [S] = 2[S], or

$$K_m = [S], \quad \text{when } V_0 = \tfrac{1}{2}V_{max} \tag{8–23}$$

This represents a very useful, practical definition of K_m: K_m is equivalent to that substrate concentration at which V_0 is one-half V_{max}. Note that K_m has units of molarity.

The Michaelis–Menten equation (8–20) can be algebraically transformed into forms that are useful in the practical determination of K_m and V_{max} (Box 8–1) and, as we will describe later, in the analysis of inhibitor action (see Box 8–2).

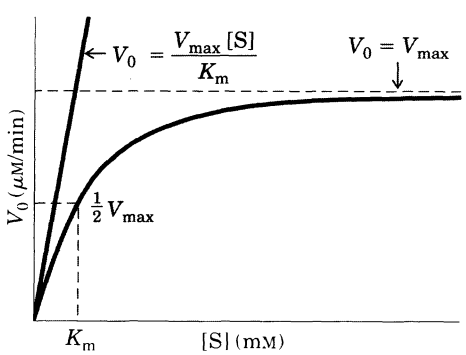


Figure 8–12 Dependence of initial velocity on substrate concentration, showing the kinetic parameters that define the limits of the curve at high and low [S]. At low [S], $K_m \gg$ [S], and the [S] term in the denominator of the Michaelis–Menten equation (Eqn 8–20) becomes insignificant; the equation simplifies to $V_0 = V_{max}[S]/K_m$, and V_0 exhibits a linear dependence on [S], as observed. At high [S], where [S] $\gg K_m$, the K_m term in the denominator of the Michaelis–Menten equation becomes insignificant, and the equation simplifies to $V_0 = V_{max}$; this is consistent with the plateau observed at high [S]. The Michaelis–Menten equation is therefore consistent with the observed dependence of V_0 on [S], with the shape of the curve defined by the terms V_{max}/K_m at low [S] and V_{max} at high [S].

BOX 8–1 Transformations of the Michaelis–Menten Equation: The Double-Reciprocal Plot

The Michaelis–Menten equation:

$$V_0 = \frac{V_{max}[S]}{K_m + [S]}$$

can be algebraically transformed into forms that are more useful in plotting experimental data. One common transformation is derived simply by taking the reciprocal of both sides of the Michaelis–Menten equation to give

$$\frac{1}{V_0} = \frac{K_m + [S]}{V_{max}[S]}$$

Separating the components of the numerator on the right side of the equation gives

$$\frac{1}{V_0} = \frac{K_m}{V_{max}[S]} + \frac{[S]}{V_{max}[S]}$$

which simplifies to

$$\frac{1}{V_0} = \frac{K_m}{V_{max}} \frac{1}{[S]} + \frac{1}{V_{max}}$$

This equation is a transform of the Michaelis–Menten equation called the **Lineweaver–Burk equation.** For enzymes obeying the Michaelis–Menten relationship, a plot of $1/V_0$ versus $1/[S]$ (the "double-reciprocal" of the V_0-versus-[S] plot we have been using to this point) yields a straight line (Fig. 1). This line will have a slope of K_m/V_{max}, an intercept of $1/V_{max}$ on the $1/V_0$ axis, and an intercept of $-1/K_m$ on the $1/[S]$ axis. The double-reciprocal presentation, also called a Lineweaver–Burk plot, has the great advantage of allowing a more accurate determination of V_{max}, which can only be *approximated* from a simple plot of V_0 versus [S] (see Fig. 8–12).

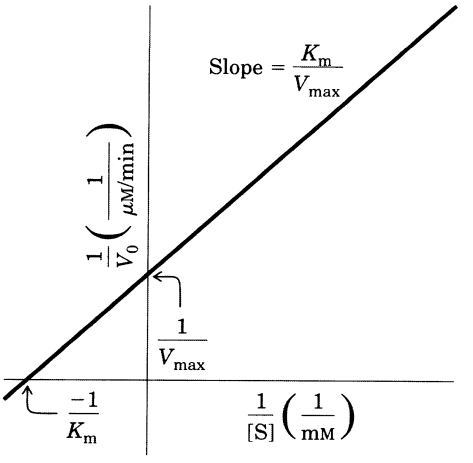


Figure 1
A double-reciprocal, or Lineweaver–Burk, plot.

Other transformations of the Michaelis–Menten equation have been derived and used. Each has some particular advantage in analyzing enzyme kinetic data.

The double-reciprocal plot of enzyme reaction rates is very useful in distinguishing between certain types of enzymatic reaction mechanisms (see Fig. 8–14) and in analyzing enzyme inhibition (see Box 8–2).

The Meaning of V_{max} and K_m Is Unique for Each Enzyme

It is important to distinguish between the Michaelis–Menten equation and the specific kinetic mechanism upon which it was originally based. The equation describes the kinetic behavior of a great many enzymes, and all enzymes that exhibit a hyperbolic dependence of V_0 on [S] are said to follow **Michaelis–Menten kinetics.** The practical rule that $K_m = [S]$ when $V_0 = \frac{1}{2}V_{max}$ (Eqn 8–23) holds for all enzymes that follow Michaelis–Menten kinetics (the major exceptions to Michaelis–Menten kinetics are the regulatory enzymes, discussed at the end of this chapter). However, this equation does not depend on the relatively simple two-step reaction mechanism proposed by Michaelis and Menten (Eqn 8–9). Many enzymes that follow Michaelis–Menten kinetics have quite different reaction mechanisms, and enzymes that catalyze

reactions with six or eight identifiable steps will often exhibit the same steady-state kinetic behavior. Even though Equation 8–23 holds true for many enzymes, both the magnitude and the real meaning of V_{max} and K_m can change from one enzyme to the next. This is an important limitation of the steady-state approach to enzyme kinetics. V_{max} and K_m are parameters that can be obtained experimentally for any given enzyme, but by themselves they provide little information about the number, rates, or chemical nature of discrete steps in the reaction. Steady-state kinetics nevertheless represents the standard language by which the catalytic efficiencies of enzymes are characterized and compared. We now turn to the application and interpretation of the terms V_{max} and K_m.

A simple graphical method for obtaining an approximate value for K_m is shown in Figure 8–12. A more convenient procedure, using a **double-reciprocal plot,** is presented in Box 8–1. The K_m can vary greatly from enzyme to enzyme, and even for different substrates of the same enzyme (Table 8–6). The term is sometimes used (inappropriately) as an indication of the affinity of an enzyme for its substrate.

Table 8–6 K_m for some enzymes

Enzyme	Substrate	K_m (mM)
Catalase	H_2O_2	25
Hexokinase (brain)	ATP	0.4
	D-Glucose	0.05
	D-Fructose	1.5
Carbonic anhydrase	HCO_3^-	9
Chymotrypsin	Glycyltyrosinylglycine	108
	N-Benzoyltyrosinamide	2.5
β-Galactosidase	D-Lactose	4.0
Threonine dehydratase	L-Threonine	5.0

The actual meaning of K_m depends on specific aspects of the reaction mechanism such as the number and relative rates of the individual steps of the reaction. Here we will consider reactions with two steps. On page 214 K_m is defined by the expression

$$K_m = \frac{k_2 + k_{-1}}{k_1} \qquad (8\text{–}24)$$

For the Michaelis–Menten reaction, k_2 is rate-limiting; thus $k_2 \ll k_{-1}$ and K_m reduces to k_{-1}/k_1, which is defined as the **dissociation constant,** K_S, for the ES complex. Where these conditions hold, K_m does represent a measure of the affinity of the enzyme for the substrate in the ES complex. However, this scenario does not apply to all enzymes. Sometimes $k_2 \gg k_{-1}$, and then $K_m = k_2/k_1$. In other cases, k_2 and k_{-1} are comparable, and K_m remains a more complex function of all three rate constants (Eqn 8–24). These situations were first analyzed by Haldane along with George E. Briggs in 1925. The Michaelis–Menten equation and the characteristic saturation behavior of the enzyme still apply, but K_m cannot be considered a simple measure of substrate affinity. Even more common are cases in which the reaction goes through multiple steps after formation of the ES complex; K_m can then become a very complex function of many rate constants.

V_{max} also varies greatly from one enzyme to the next. If an enzyme reacts by the two-step Michaelis–Menten mechanism, V_{max} is equivalent to $k_2[E_t]$, where k_2 is the rate-limiting step. However, the number of reaction steps and the identity of the rate-limiting step(s) can vary from enzyme to enzyme. For example, consider the quite common situation where product release, EP → E + P, is rate-limiting:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP \xrightarrow{k_3} E + P \qquad (8\text{–}25)$$

In this case, most of the enzyme is in the EP form at saturation, and $V_{max} = k_3[E_t]$. It is useful to define a more general rate constant, k_{cat}, to describe the limiting rate of any enzyme-catalyzed reaction at saturation. If there are several steps in the reaction, and one is clearly rate-limiting, k_{cat} is equivalent to the rate constant for that limiting step. For the Michaelis–Menten reaction, $k_{cat} = k_2$. For the reaction of Equation 8–25, $k_{cat} = k_3$. When several steps are partially rate-limiting, k_{cat} can become a complex function of several of the rate constants that define each individual reaction step. In the Michaelis–Menten equation, $k_{cat} = V_{max}/[E_t]$, and Equation 8–19 becomes

$$V_0 = \frac{k_{cat}[E_t][S]}{K_m + [S]} \qquad (8\text{–}26)$$

The constant k_{cat} is a first-order rate constant with units of reciprocal time, and is also called the **turnover number.** It is equivalent to the number of substrate molecules converted to product in a given unit of time on a single enzyme molecule when the enzyme is saturated with substrate. The turnover numbers of several enzymes are given in Table 8–7.

Table 8–7 Turnover numbers* (k_{cat}) of some enzymes

Enzyme	Substrate	k_{cat} (s^{-1})
Catalase	H_2O_2	40,000,000
Carbonic anhydrase	HCO_3^-	400,000
Acetylcholinesterase	Acetylcholine	14,000
β-Lactamase	Benzylpenicillin	2,000
Fumarase	Fumarate	800
RecA protein (ATPase)	ATP	0.4

* Number of substrate molecules transformed per second per molecule of enzyme.

The kinetic parameters k_{cat} and K_m are generally useful for the study and comparison of different enzymes, whether their reaction mechanisms are simple or complex. Each enzyme has optimum values of k_{cat} and K_m that reflect the cellular environment, the concentration of substrate normally encountered in vivo by the enzyme, and the chemistry of the reaction being catalyzed.

Comparison of the catalytic efficiency of different enzymes requires the selection of a suitable parameter. The constant k_{cat} is not entirely satisfactory. Two enzymes catalyzing different reactions may have the same k_{cat} (turnover number), yet the rates of the uncatalyzed reactions may be different and thus the rate enhancement brought about by the enzymes may differ greatly. Also, k_{cat} reflects the proper-

ties of an enzyme when it is saturated with substrate, and is less useful at low [S]. The constant K_m is also unsatisfactory by itself. As shown by Equation 8–23, K_m must have some relationship to the normal [S] found in the cell. An enzyme that acts on a substrate present at a very low concentration in the cell will tend to have a lower K_m than an enzyme that acts on a substrate that is normally abundant.

The most useful parameter for a discussion of catalytic efficiency is one that includes both k_{cat} and K_m. When $[S] \ll K_m$, Equation 8–26 reduces to the form

$$V_0 = \frac{k_{cat}}{K_m}[E_t][S] \qquad (8\text{–}27)$$

V_0 in this case depends on the concentration of two reactants, E_t and S; therefore this is a second-order rate law and the constant k_{cat}/K_m is a second-order rate constant. The factor k_{cat}/K_m is generally the best kinetic parameter to use in comparisons of catalytic efficiency. There is an upper limit to k_{cat}/K_m, imposed by the rate at which E and S can diffuse together in an aqueous solution. This diffusion-controlled limit is 10^8 to 10^9 $M^{-1}s^{-1}$, and many enzymes have a value of k_{cat}/K_m near this range (Table 8–8).

Table 8–8 Enzymes for which k_{cat}/K_m is close to the diffusion-controlled limit (10^8 to 10^9 $M^{-1}s^{-1}$)

Enzyme	Substrate	k_{cat} (s^{-1})	K_m (M)	k_{cat}/K_m ($M^{-1}s^{-1}$)
Acetylcholinesterase	Acetylcholine	1.4×10^4	9×10^{-5}	1.6×10^8
Carbonic anhydrase	CO_2	1×10^6	1.2×10^{-2}	8.3×10^7
	HCO_3^-	4×10^5	2.6×10^{-2}	1.5×10^7
Catalase	H_2O_2	4×10^7	1.1	4×10^7
Crotonase	Crotonyl-CoA	5.7×10^3	2×10^{-5}	2.8×10^8
Fumarase	Fumarate	8×10^2	5×10^{-6}	1.6×10^8
	Malate	9×10^2	2.5×10^{-5}	3.6×10^7
Triose phosphate isomerase	Glyceraldehyde-3-phosphate	4.3×10^3	4.7×10^{-4}	2.4×10^8
β-Lactamase	Benzylpenicillin	2.0×10^3	2×10^{-5}	1×10^8

Source: From Fersht, A. (1985) *Enzyme Structure and Mechanism*, p. 152, W.H. Freeman and Company, New York.

Many Enzymes Catalyze Reactions Involving Two or More Substrates

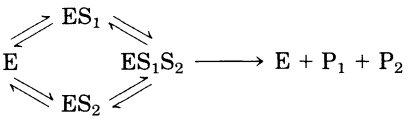
We have seen how [S] affects the rate of a simple enzyme reaction ($S \rightarrow P$) in which there is only one substrate molecule. In many enzymatic reactions, however, two (or even more) different substrate molecules bind to the enzyme and participate in the reaction. For example, in the reaction catalyzed by hexokinase, ATP and glucose are the substrate molecules, and ADP and glucose-6-phosphate the products:

$$\text{ATP} + \text{glucose} \longrightarrow \text{ADP} + \text{glucose-6-phosphate}$$

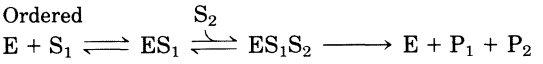
The rates of such bisubstrate reactions can also be analyzed by the Michaelis–Menten approach. Hexokinase has a characteristic K_m for each of its two substrates (Table 8–6).

Enzyme reaction involving a ternary complex

Random order

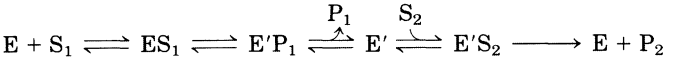


Ordered



(a)

Enzyme reaction in which no ternary complex is formed



(b)

Figure 8–13 Common mechanisms for enzyme-catalyzed bisubstrate reactions. In **(a)** the enzyme and both substrates come together to form a ternary complex. In ordered binding, substrate 1 must be bound before substrate 2 can bind productively. In **(b)** an enzyme–substrate complex forms, a product leaves the complex, the altered enzyme forms a second complex with another substrate molecule, and the second product leaves, regenerating the enzyme. Substrate 1 may transfer a functional group to the enzyme (forming E′), which is subsequently transferred to substrate 2. This is a ping-pong or double-displacement mechanism.

Enzymatic reactions in which there are two substrates (bisubstrate reactions) usually involve transfer of an atom or a functional group from one substrate to the other. Such reactions proceed by one of several different pathways. In some cases, both substrates are bound to the enzyme at the same time at some point in the course of the reaction, forming a ternary complex (Fig. 8–13a). Such a complex can be formed by substrates binding in a random sequence or in a specific order. No ternary complex is formed when the first substrate is converted to product and dissociates before the second substrate binds. An example of this is the ping-pong or double-displacement mechanism (Fig. 8–13b). Steady-state kinetics can often help distinguish among these possibilities (Fig. 8–14).

Pre-Steady State Kinetics Can Provide Evidence for Specific Reaction Steps

We have introduced kinetics as a set of methods used to study the steps in an enzymatic reaction, but have also outlined the limitations of the most common kinetic parameters in providing such information. The two most important experimental parameters provided by steady-state kinetics are k_{cat} and k_{cat}/K_m. Variation in these parameters with changes in pH or temperature can sometimes provide additional information about steps in a reaction pathway. In the case of bisubstrate reactions, steady-state kinetics can help determine whether a ternary complex is formed during the reaction (Fig. 8–14). A more complete picture generally requires more sophisticated kinetic methods that go beyond the scope of an introductory text. Here, we briefly introduce one of the most important kinetic approaches for studying reaction mechanisms, pre-steady state kinetics.

A complete description of an enzyme-catalyzed reaction requires direct measurement of the rates of individual reaction steps, for example the measurement of the association of enzyme and substrate to form the ES complex. It is during the pre-steady state that the rates of many reaction steps can be measured independently. Reaction conditions are adjusted to facilitate the measurement of events that occur during the reaction of a single substrate molecule. Because the pre-steady state phase of a reaction is generally very short, this often requires specialized techniques for very rapid mixing and sampling. One objective is to gain a complete and quantitative picture of the energetic course of a reaction. As we have already noted, reaction rates and equilibria are related to the changes in free energy that occur during the

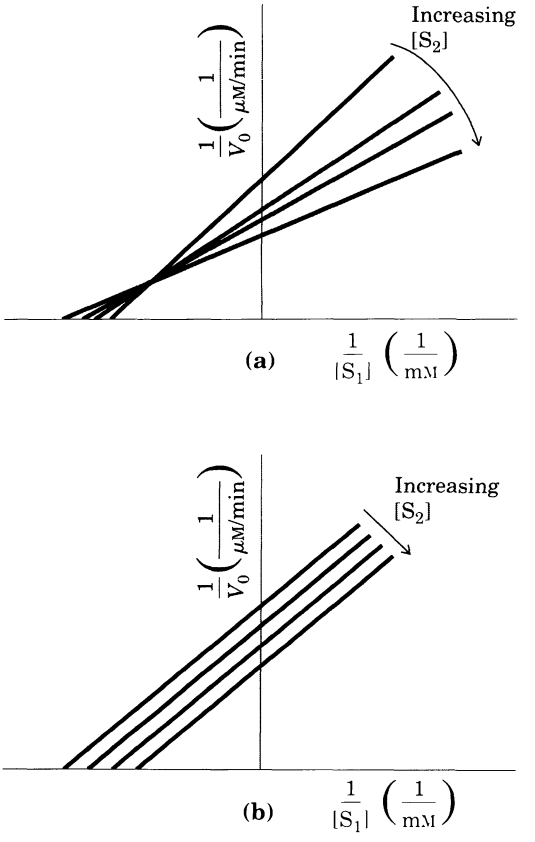
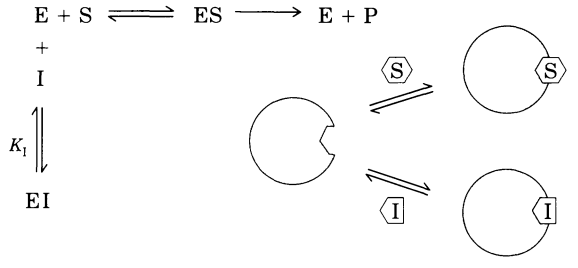
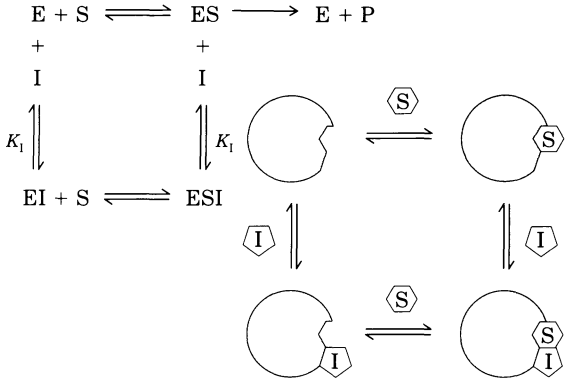


Figure 8–14 Steady-state kinetic analysis of bisubstrate reactions. In these double-reciprocal plots (see Box 8–1), the concentration of substrate 1 is varied while the concentration of substrate 2 is held constant. This is repeated for several values of $[S_2]$, generating several separate lines. The lines intersect if a ternary complex is formed in the reaction **(a)**, but are parallel if the reaction goes through a ping-pong or double-displacement pathway **(b)**.

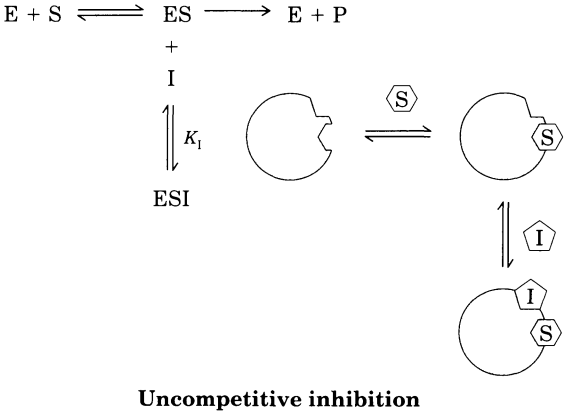
reaction. Measuring the rate of individual reaction steps defines how energy is used by a specific enzyme, which represents an important component of the overall reaction mechanism. In a number of cases it has proven possible to measure the rates of every individual step in a multistep enzymatic reaction. Some examples of the application of pre-steady state kinetics are included in the descriptions of specific enzymes later in this chapter.



Competitive inhibition



Noncompetitive inhibition



Uncompetitive inhibition

Figure 8–15 Three types of reversible inhibition. Competitive inhibitors bind to the enzyme's active site. Noncompetitive inhibitors generally bind at a separate site. Uncompetitive inhibitors also bind at a separate site, but they bind only to the ES complex. K_I is the equilibrium constant for inhibitor binding.

Enzymes Are Subject to Reversible and Irreversible Inhibition

Enzymes catalyze virtually every process in the cell, and it should not be surprising that enzyme inhibitors are among the most important pharmaceutical agents known. For example, aspirin (acetylsalicylate) inhibits the enzyme that catalyzes the first step in the synthesis of prostaglandins, compounds involved in many processes including some that produce pain. The study of enzyme inhibitors also has provided valuable information about enzyme mechanisms and has helped define some metabolic pathways. There are two broad classes of enzyme inhibitors: reversible and irreversible.

One common type of reversible inhibition is called competitive (Fig. 8–15). A **competitive inhibitor** competes with the substrate for the active site of an enzyme, but a reaction usually does not occur once the inhibitor (I) is bound. While the inhibitor occupies the active site it prevents binding by the substrate. Competitive inhibitors are often compounds that resemble the substrate and combine with the enzyme to form an EI complex (Fig. 8–15). This type of inhibition can be analyzed quantitatively by steady-state kinetics (Box 8–2). Because the inhibitor binds reversibly to the enzyme, the competition can be biased to favor the substrate simply by adding more substrate. When enough substrate is present the probability that an inhibitor molecule will bind is minimized, and the reaction exhibits a normal V_{max}. However, the [S] at which $V_0 = \frac{1}{2}V_{max}$, the K_m, will increase in the presence of inhibitor. This effect on the apparent K_m and the absence of an effect on V_{max} is diagnostic of competitive inhibition, and is readily revealed in a double-reciprocal plot (Box 8–2). The equilibrium constant for inhibitor binding, K_I, can be obtained from the same plot.

Competitive inhibition is used therapeutically to treat patients who have ingested methanol, a solvent found in gas-line antifreeze. Methanol is converted to formaldehyde by the action of the enzyme alcohol dehydrogenase. Formaldehyde damages many tissues, and blindness is a common result because the eyes are particularly sensitive. Ethanol competes effectively with methanol as a substrate for alcohol dehydrogenase. The therapy for methanol poisoning is intravenous infusion of ethanol, which slows the formation of formaldehyde sufficiently so that most of the methanol can be excreted harmlessly in the urine.

Two other types of reversible inhibition, noncompetitive and uncompetitive, are often defined in terms of one-substrate enzymes but in practice are only observed with enzymes having two or more substrates. A **noncompetitive inhibitor** is one that binds to a site distinct from that which binds the substrate (Fig. 8–15); inhibitor binding does not block substrate binding (or vice versa). The enzyme is inactivated when inhibitor is bound, whether or not substrate is also present. The inhibitor effectively lowers the concentration of active enzyme and hence lowers the apparent V_{max} ($V_{max} = k_{cat}[E_t]$). There is often

BOX 8–2 Kinetic Tests for Determining Inhibition Mechanisms

The double-reciprocal plot (see Box 8–1) offers an easy way of determining whether an enzyme inhibitor is competitive or noncompetitive. Two sets of rate experiments are carried out, in both of which the enzyme concentration is held constant. In the first set, [S] is also held constant, permitting measurement of the effect of increasing inhibitor concentration [I] on the initial rate V_0 (not shown). In the second set, [I] is held constant but [S] is varied. In the double-reciprocal plot $1/V_0$ is plotted versus 1/[S].

Figure 1 shows a set of double-reciprocal plots obtained in the absence of the inhibitor and with two different concentrations of a competitive inhibitor. Increasing [I] results in the production of a family of lines with a common intercept on the $1/V_0$ axis but with different slopes. Because the intercept on the $1/V_0$ axis is equal to $1/V_{max}$, we can see that V_{max} is unchanged by the presence of a competitive inhibitor. That is, regardless of the concentration of a competitive inhibitor, there is always some high substrate concentration that will displace the inhibitor from the enzyme's active site.

In noncompetitive inhibition, similar plots of the rate data give the family of lines shown in Figure 2, having a common intercept on the 1/[S] axis. This indicates that K_m for the substrate is not altered by a noncompetitive inhibitor, but V_{max} decreases.

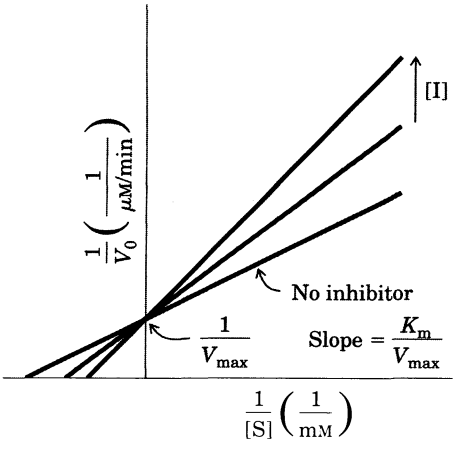


Figure 1 Competitive inhibition.

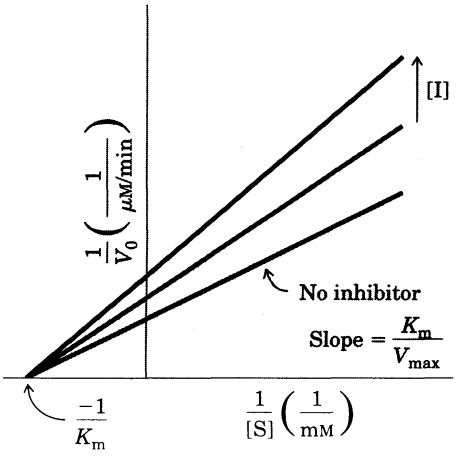


Figure 2 Noncompetitive inhibition.

little or no effect on K_m. These characteristic effects of a noncompetitive inhibitor are further analyzed in Box 8–2. An **uncompetitive inhibitor** (Fig. 8–15) also binds at a site distinct from the substrate. However, an uncompetitive inhibitor will bind only to the ES complex. (The noncompetitive inhibitor binds to *either* free enzyme or the ES complex.)

With these definitions in mind, consider a bisubstrate enzyme with separate binding sites within the active site for two substrates, S_1 and S_2, and suppose an inhibitor (I) binds to the site for S_2. If S_1 and S_2 normally bind to the enzyme independently (in random order), I may act as a competitive inhibitor of S_2. However, since I binds at a site distinct from the site for S_1, but will exclude S_2 and thereby block the reaction of S_1, I may act as a noncompetitive inhibitor of S_1. Alternatively, if S_1 normally binds to the enzyme before S_2 (ordered binding), then I may bind only to the ES_1 complex and act as an uncompetitive inhibitor of S_1. These are only a few of the scenarios that can be encountered with reversible inhibition of bisubstrate enzymes, and the effects of these inhibitors can provide much information about reaction mechanisms.

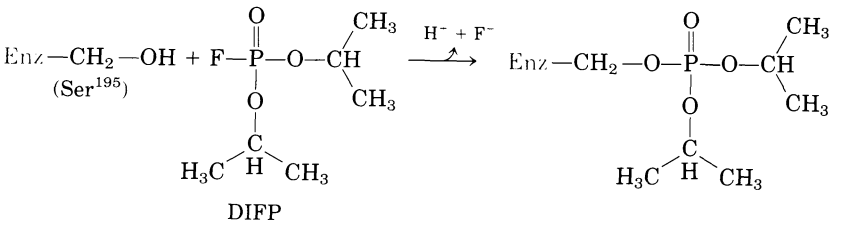


Figure 8–16 Reaction of chymotrypsin with diisopropylfluorophosphate (DIFP). This reaction led to the discovery that Ser^{195} is the key active-site serine. DIFP also acts as a poison nerve gas because it irreversibly inactivates the enzyme acetylcholinesterase by a mechanism similar to that shown here. Acetylcholinesterase cleaves the neurotransmitter acetylcholine, an essential step in normal functioning of the nervous system.

Irreversible inhibitors are those that combine with or destroy a functional group on the enzyme that is essential for its activity. Formation of a covalent link between an irreversible inhibitor and an enzyme is common. Irreversible inhibitors are very useful in studying reaction mechanisms. Amino acids with key catalytic functions in the active site can sometimes be identified by determining which amino acid is covalently linked to an inhibitor after the enzyme is inactivated. An example is shown in Figure 8–16.

A very special class of irreversible inhibitors are the **suicide inhibitors.** These compounds are relatively unreactive until they bind to the active site of a specific enzyme. A suicide inhibitor is designed to carry out the first few chemical steps of the normal enzyme reaction. Instead of being transformed into the normal product, however, the inhibitor is converted to a very reactive compound that combines irreversibly with the enzyme. These are also called **mechanism-based inactivators,** because they utilize the normal enzyme reaction mechanism to inactivate the enzyme. These inhibitors play a central role in the modern approach to obtaining new pharmaceutical agents, a process called rational drug design. Because the inhibitor is designed to be specific for a single enzyme and is unreactive until within that enzyme's active site, drugs based on this approach are often very effective and have few side effects (see Box 21–1).

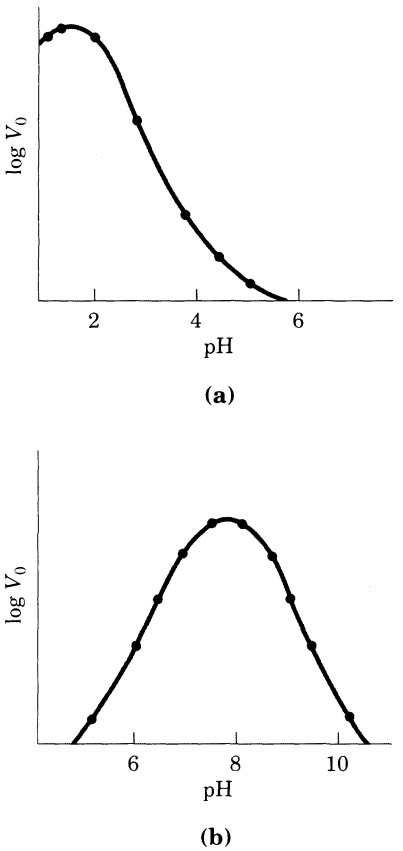


Figure 8–17 pH-activity profiles of two enzymes. Such curves are constructed from measurements of initial velocities when the reaction is carried out in buffers of different pH. The pH optimum for the activity of an enzyme generally reflects the cellular environment in which it is normally found. **(a)** Pepsin, which hydrolyzes certain peptide bonds of proteins during digestion in the stomach, has a pH optimum of about 1.6. The pH of gastric juice is between 1 and 2. **(b)** Glucose-6-phosphatase of hepatocytes, with a pH optimum of about 7.8, is responsible for releasing glucose into the blood. The normal pH of the cytosol of hepatocytes is about 7.2.

Enzyme Activity Is Affected by pH

Enzymes have an optimum pH or pH range in which their activity is maximal (Fig. 8–17); at higher or lower pH their activity decreases. This is not surprising because some amino acid side chains act as weak acids and bases that perform critical functions in the enzyme active site. The change in ionization state (titration) of groups in the active site is a common reason for the activity change, but it is not the only one. The group being titrated might instead affect some critical aspect of the protein structure. Removing a proton from a His residue outside the active site might, for example, eliminate an ionic interaction essential for stabilization of the active conformation of the enzyme. Less common are cases in which the group being titrated is on the substrate.

The pH range over which activity changes can provide a clue as to what amino acid is involved (see Table 5–1). A change in enzyme activity near pH 7.0, for example, often reflects titration of a His residue. The effects of pH must be interpreted with some caution, however. In the closely packed environment of a protein, the pK of amino acid side chains can change significantly. For example, a nearby positive charge can lower the pK of a Lys residue, and a nearby negative charge can increase its pK. Such effects sometimes result in a pK that is perturbed by 2 or more pH units from its normal value. One Lys residue in the enzyme acetoacetate decarboxylase has a pK of 6.6 (10.5 is normal) due to electrostatic effects of nearby positive charges.

Examples of Enzymatic Reactions

This chapter has focused on the general principles of catalysis and an introduction to some of the kinetic parameters used to describe enzyme action. Principles and kinetics are combined in Box 8–3, which describes some of the evidence that reinforces the notion that binding energy and transition-state complementarity are central to enzymatic catalysis. We now turn to several examples of specific enzyme reaction mechanisms.

An understanding of the complete mechanism of action of a purified enzyme requires a knowledge of (1) the temporal sequence in which enzyme-bound reaction intermediates occur, (2) the structure of each intermediate and transition state, (3) the rates of interconversion between intermediates, (4) the structural relationship of the enzyme with each intermediate, and (5) the energetic contributions of all reacting and interacting groups with respect to intermediate complexes and transition states. There is probably no enzyme for which current understanding meets this standard exactly. Many decades of research, however, have produced mechanistic information about hundreds of enzymes, and in some cases this information is highly detailed.

Reaction Mechanisms Illustrate Principles

Mechanisms are presented for three enzymes: chymotrypsin, hexokinase, and tyrosyl-tRNA synthetase. These are chosen not necessarily because they are the best-understood enzymes or cover all possible classes of enzyme chemistry, but because they help to illustrate some general principles outlined in this chapter. The discussion concentrates on selected principles, along with some key experiments that have helped to bring them into focus. Much mechanistic detail and experimental evidence is omitted, and in no instance do the mechanisms described below provide a complete explanation for the catalytic rate enhancements brought about by these enzymes.

Chymotrypsin This enzyme is a protease (M_r 25,000) specific for peptide bonds adjacent to aromatic amino acid residues (see Table 6–7). The three-dimensional structure of chymotrypsin is shown in Figure 8–18, with functional groups in the active site emphasized. This enzyme reaction illustrates the principle of transition-state stabilization by an enzyme, and also provides a classic example of the use of general acid–base catalysis and covalent catalysis (Fig. 8–19, p. 226).

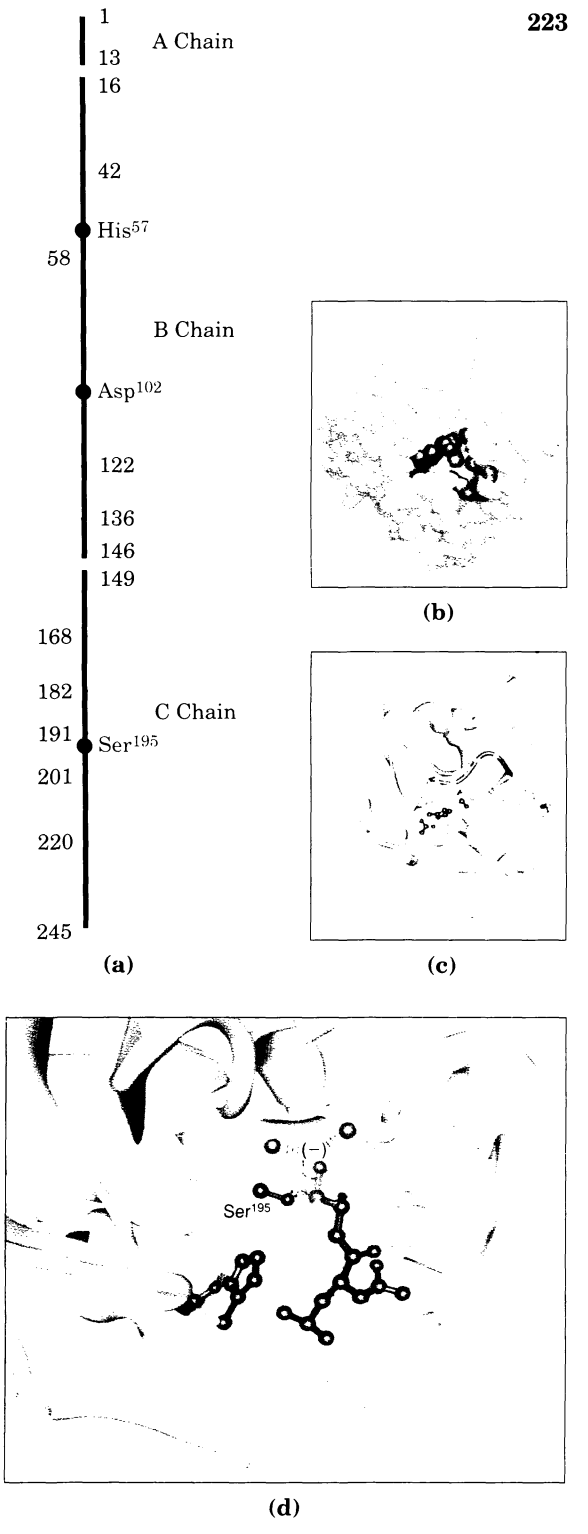


Figure 8–18 The structure of chymotrypsin. **(a)** A representation of primary structure, showing disulfide bonds and the location of key amino acids. Note that the protein consists of three polypeptide chains. The active-site amino acids are found grouped together in the three-dimensional structure. **(b)** A space-filling model of chymotrypsin. The pocket in which the aromatic amino acid side chain is bound is shown in green. The amide nitrogens of Gly^{193} and Ser^{195} in the polypeptide backbone make up the oxyanion hole (see Fig. 8–19), and are shown in orange. The side chains of other key active site residues, including Ser^{195}, His^{57}, and Asp^{102} are shown in red, and are explained in Fig. 8–19. **(c)** The polypeptide backbone of chymotrypsin shown as a ribbon structure. Disulfide bonds are shown in yellow; the A, B, and C chains are shown in dark blue, light blue, and white, respectively. **(d)** A close-up of the chymotrypsin active site with a substrate bound. Ser^{195} attacks the carbonyl group of the substrate (shown in purple); the developing negative charge on the oxygen is stabilized by the oxyanion hole (amide nitrogens shown in orange), as explained in Fig. 8–19. In the substrate, the aromatic amino acid side chain and the amide nitrogen of the peptide bond to be cleaved are shown in light blue.

BOX 8–3 Evidence for Enzyme–Transition State Complementarity

The transition state of a reaction is difficult to study, because by definition it has no finite lifetime. To understand enzymatic catalysis, however, we must dissect the interaction between the enzyme and this ephemeral moment in the course of a reaction. The idea that an enzyme is complementary to the transition state is virtually a requirement for catalysis, because the energy hill upon which the transition state sits is what the enzyme must lower if catalysis is to occur. How can we obtain evidence that the idea of enzyme–transition state complementarity is really correct? Fortunately, there are a variety of approaches, old and new, to this problem. Each has provided compelling evidence in support of this general principle of enzyme action.

Structure–Activity Correlations

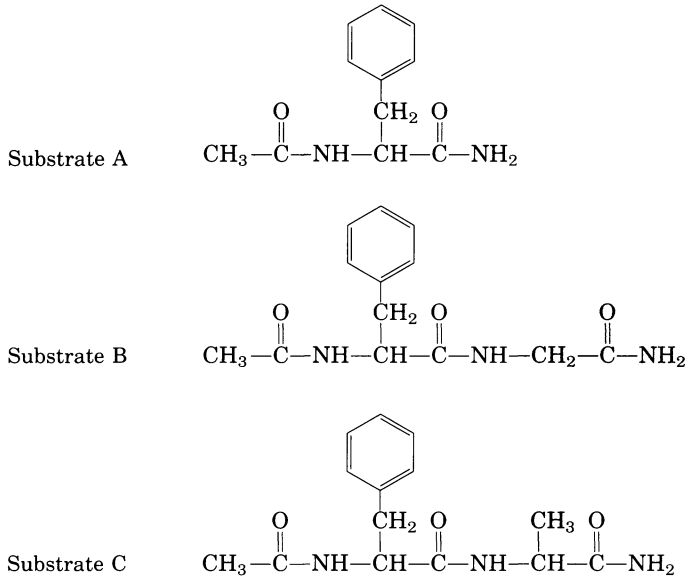
If enzymes are complementary to reaction transition states, then some functional groups in the substrate and in the enzyme must interact preferentially with the transition state rather than the ES complex. Altering these groups should have little effect on formation of the ES complex, and hence should not affect kinetic parameters (K_S, or sometimes K_m if $K_S = K_m$) that reflect the $E + S \rightleftharpoons ES$ equilibrium. Changing the same groups, however, should have a large effect on the overall rate (k_{cat} or k_{cat}/K_m) of the reaction, because the bound substrate lacks potential binding interactions needed to lower the activation energy.

An excellent example is seen in a series of substrates for the enzyme chymotrypsin (Fig. 1). Chymotrypsin normally catalyzes the hydrolysis of peptide bonds next to aromatic amino acids, and the substrates shown in Figure 1 are convenient smaller models for the natural substrates (long polypeptides and proteins; see Chapter 6). The additional chemical groups added in going from A to B to C are shaded in red. Note that the interaction between the enzyme and these added functional groups has a minimal effect on K_m (which is taken here as a reflection of K_S), but a large, positive effect on k_{cat} and k_{cat}/K_m. This is what we would expect if the interaction occurred only in the transition state. Chymotrypsin is described in more detail beginning on page 223.

A complementary experimental approach to this problem is to modify the enzyme, eliminating certain enzyme–substrate interactions, by replacing specific amino acids through site-directed mutagenesis (Chapters 7 and 28). A good example is found in tyrosyl-tRNA synthetase (p. 227).

Transition-State Analogs

Even though transition states cannot be observed directly, chemists can often predict the approximate structure of a transition state based on accumulated knowledge about reaction mechanisms. The transition state by definition is transient and so unstable that direct measurement of the binding interaction between this species and the enzyme is impossible. In some cases, however, stable molecules can be designed that resemble transition states. These are called **transition-state analogs.** In principle, they should bind to an enzyme



	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($M^{-1}\ s^{-1}$)
Substrate A	0.06	31	2
Substrate B	0.14	15	10
Substrate C	2.8	25	114

Figure 1 Effects of small structural changes in the substrate on kinetic parameters for chymotrypsin-catalyzed amide hydrolysis.

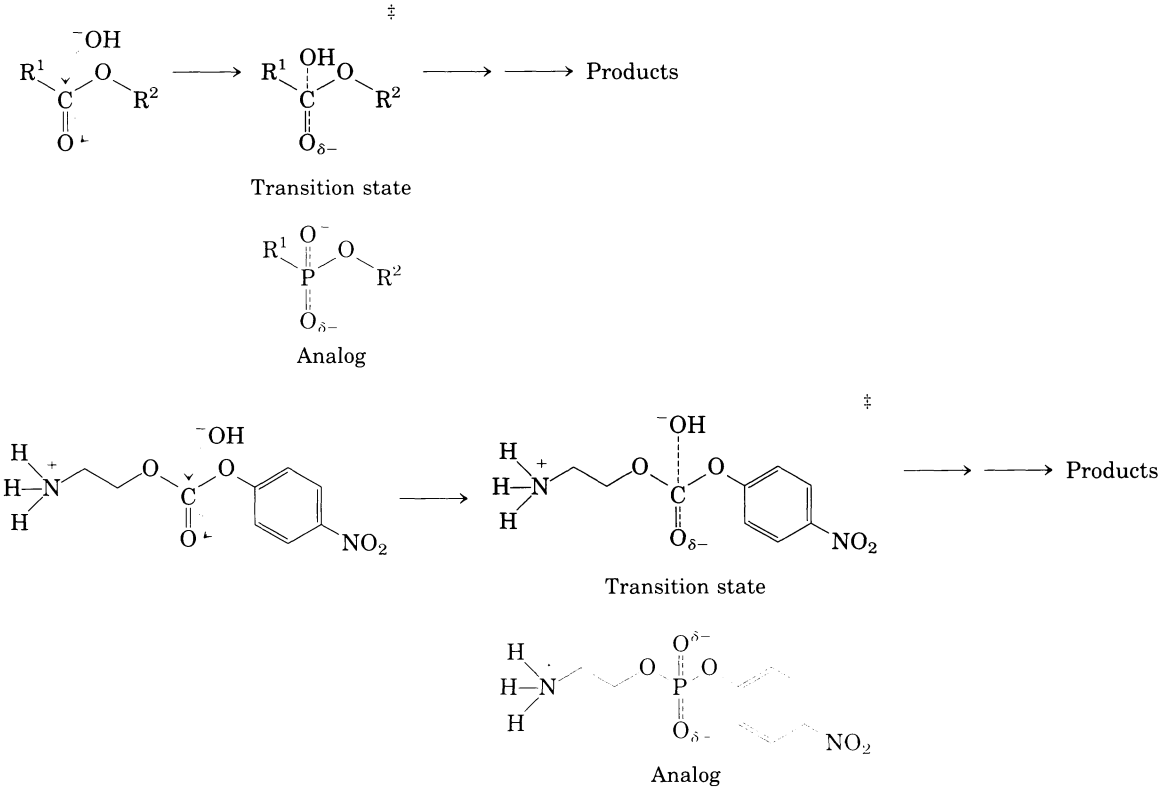


Figure 2 The expected transition states for ester or carbonate hydrolysis reactions. Phosphonate and phosphate compounds, respectively, make good transition-state analogs for these reactions.

more tightly than the substrate binds in the ES complex, because they should fit in the active site better (i.e., form more weak interactions) than the substrate itself. The idea of transition-state analogs was suggested by Pauling in the 1940s, and it has been used for a number of enzymes. These experiments have the limitation that a transition-state analog can never mimic a transition state perfectly. Analogs have been found, however, that bind an enzyme 10^2 to 10^6 times more tightly than the normal substrate, providing good evidence that enzyme active sites are indeed complementary to transition states.

Catalytic Antibodies

If a transition-state analog can be designed for the reaction S → P, then an antibody that binds tightly to the transition-state analog might catalyze S → P. Antibodies (immunoglobulins; see Fig. 6–8) are key components of the immune response. A molecule or chemical group that is bound tightly and specifically by a given antibody is referred to as an **antigen.** When a transition-state analog is used as an antigen to stimulate the production of antibodies, the antibodies that bind it are potential catalysts of the corresponding reaction. This approach, first suggested by William P. Jencks in 1969, has become practical with the development of laboratory techniques to produce antibodies that are all identical and bind one specific antigen (these are known as monoclonal antibodies; see Chapter 6).

Pioneering work in the laboratories of Richard Lerner and Peter Schultz has resulted in the isolation of a number of monoclonal antibodies that catalyze the hydrolysis of esters or carbonates (Fig. 2). In these reactions, the attack by water (OH^-) on the carbonyl carbon produces a tetrahedral transition state in which a partial negative charge has developed on the carbonyl oxygen. Phosphonate compounds mimic the structure and charge distribution of this transition state in ester hydrolysis, making them good transition-state analogs; phosphate compounds are used for carbonate reactions. Antibodies that bind the phosphonate or phosphate tightly have been found to catalyze the corresponding ester or carbonate hydrolysis reaction by factors of 10^3 to 10^4. Structural analyses of a few of these catalytic antibodies have shown that the catalytic amino acid side chains are arranged where they could interact with the substrate only in the transition state. These studies provide additional evidence for enzyme–transition state complementarity and suggest that new classes of antibody catalysts might be developed for research and industry.

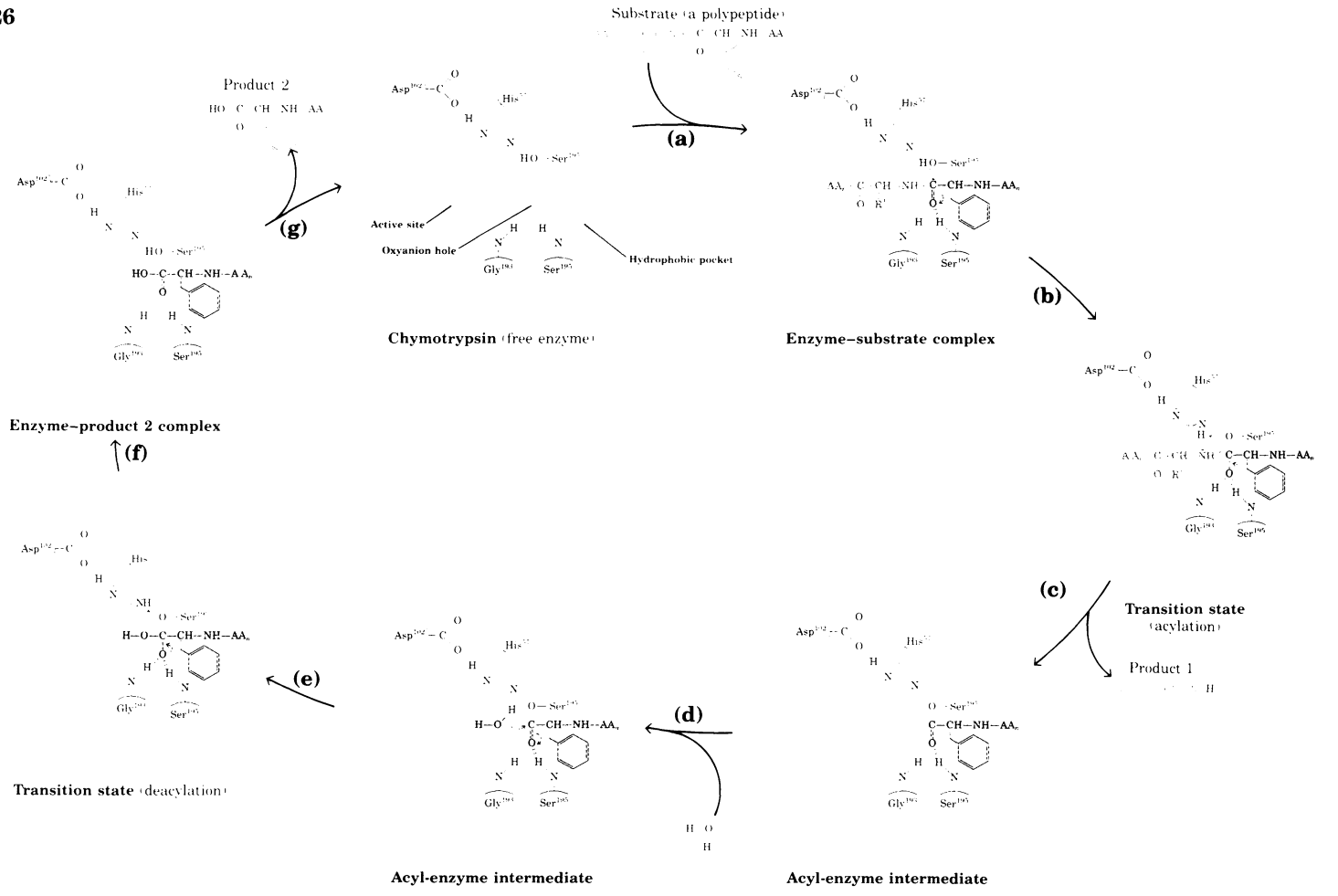
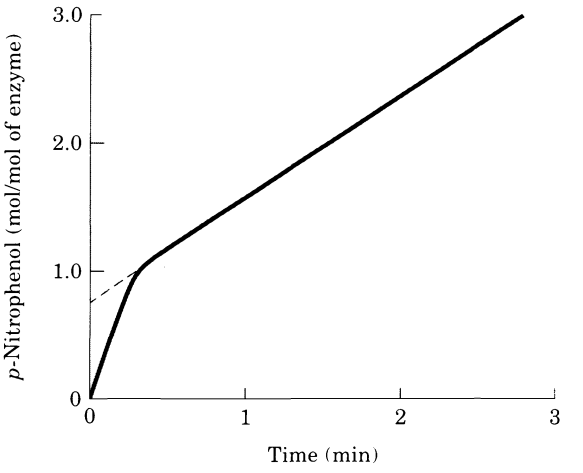


Figure 8–19 Steps in the cleavage of a peptide bond by chymotrypsin. The substrate (a polypeptide or protein) is bound at the active site. The peptide bond to be cleaved is positioned by the binding of the adjacent hydrophobic amino acid side chain (a Phe residue in this example) in a special hydrophobic pocket on the enzyme, as shown. The reaction consists of two phases: **(a)** to **(c)** formation of a covalent acyl-enzyme intermediate coupled to cleavage of the peptide bond (the acylation phase) and **(d)** to **(g),** deacylation to regenerate the free enzyme (the deacylation phase). In both phases, the carbonyl oxygen of the substrate acquires a negative charge in the transition state. The charge is stabilized by a hydrogen bond to the amide nitrogens of Gly^{193} and Ser^{195}; the hydrogen bond to Gly^{193} forms only in the transition state. Deacylation is essentially the reverse of acylation, with water serving in place of the amine component of the substrate. The His and Asp residues cooperate in a catalytic triad, providing general base catalysis of steps **(b)** and **(e)** and general acid catalysis of steps **(c)** and **(f).**

The enzyme reaction of chymotrypsin has two major phases: acylation, in which the peptide bond is cleaved and an ester linkage is formed between the peptide carbonyl carbon and the enzyme; and deacylation, in which the ester linkage is hydrolyzed and the enzyme regenerated. The nucleophile in the acylation phase is the oxygen of Ser^{195}. A serine hydroxyl is normally protonated at neutral pH, but in the enzyme Ser^{195} is hydrogen-bonded to His^{57}, which is further hydrogen-bonded to Asp^{102}. These three amino acids are often referred to as a catalytic triad. As the serine oxygen attacks the carbonyl carbon of a peptide bond, the hydrogen-bonded His^{57} functions as a general base to abstract the serine proton, and the negatively charged Asp^{102} stabilizes the positive charge that forms on the His residue. This prevents the development of a very unstable positive charge on the serine hydroxyl and increases its nucleophilicity. His^{57} can also act as a proton donor to protonate the amino group in the displaced portion of the substrate (the leaving group). A similar set of proton transfers occurs in the deacylation step (Fig. 8–19).

As the serine oxygen attacks the carbonyl group in the substrate, a transition state is reached in which the carbonyl oxygen acquires a negative charge. This charge is formed within a pocket on the enzyme called the oxyanion hole, and it is stabilized by hydrogen bonds contributed by the amide nitrogens of two peptide bonds in the protein backbone. One of these hydrogen bonds occurs only in the transition state and thereby reduces the energy required to reach the transition state. This represents an example of the use of binding energy in catalysis. The importance of binding energy in catalysis by chymotrypsin is discussed further in Box 8–3.

The first evidence for a covalent acyl-enzyme intermediate came from a classic application of pre-steady state kinetics. In addition to its action on polypeptides, chymotrypsin will catalyze the hydrolysis of small ester and amide compounds. These reactions are much slower because less binding energy is available with these substrates, but they are easier to study. Studies by B.S. Hartley and B.A. Kilby found that the hydrolysis of *p*-nitrophenylacetate by chymotrypsin, as measured by release of *p*-nitrophenol, proceeded with a rapid burst before leveling off to a slower rate (Fig. 8–20). By extrapolating back to zero time, they concluded that the burst phase corresponded to just under one molecule of *p*-nitrophenol released for every enzyme molecule present. They suggested that this reflected a rapid acylation of all the enzyme molecules (with release of *p*-nitrophenol) but that subsequent turnover of the enzyme was limited in rate by a slow deacylation step. Similar results have been obtained with many enzymes.



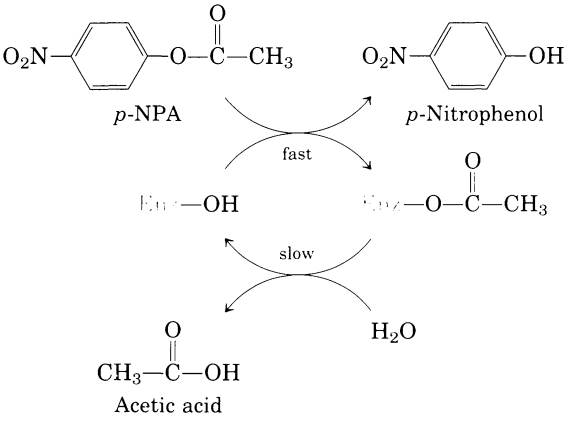
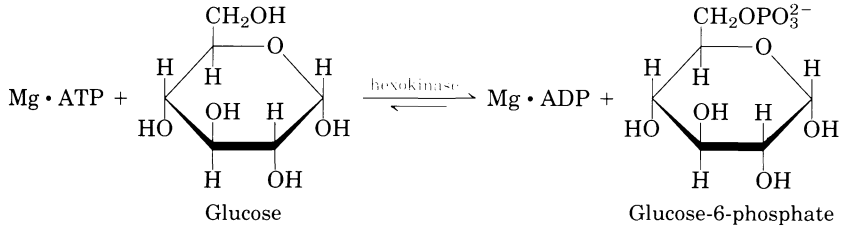


Figure 8–20 Observed kinetics of the hydrolysis of *p*-nitrophenylacetate (*p*-NPA) by chymotrypsin as measured by release of *p*-nitrophenol (a colored product). A rapid release (burst) of an amount of *p*-nitrophenol nearly stoichiometric with the amount of enzyme present is observed. This reflects the fast acylation phase of the reaction. The subsequent rate is slower because enzyme turnover is limited by the rate of the slower deacylation phase.

Hexokinase This is a bisubstrate enzyme (M_r 100,000), catalyzing the interconversion of glucose and ATP with glucose-6-phosphate and ADP. The hydroxyl at position 6 of the glucose molecule (to which the γ-phosphate of ATP is transferred) is similar in chemical reactivity to water, and water freely enters the enzyme active site. Yet hexokinase discriminates between glucose and water, with glucose favored by a factor of 10^6.



Hexokinase can discriminate between glucose and water because of a conformational change in the enzyme that occurs when the correct substrates are bound (Fig. 8–21). The enzyme thus provides a good example of induced fit. When glucose is not present, the enzyme is in an inactive conformation with the active-site amino acid side chains out of position for reaction. When glucose (but not water) and ATP bind, the binding energy derived from this interaction induces a change to the catalytically active enzyme conformation.

Tyrosyl-tRNA Synthetase This enzyme (M_r 95,000) catalyzes the attachment of tyrosine to an RNA molecule called a transfer RNA, activating the amino acid to form a precursor for protein synthesis (described in Chapter 26). The reaction proceeds in two phases:

$$\text{Enz} + \text{Tyr} + \text{ATP} \rightleftharpoons \text{Enz} \cdot \text{Tyr-AMP} + \text{PP}_i$$

$$\text{Enz} \cdot \text{Tyr-AMP} + \text{tRNA} \rightleftharpoons \text{Tyr-tRNA} + \text{Enz} + \text{AMP}$$

(PP_i is the abbreviation for inorganic pyrophosphate. P_i, used later in this chapter, is the abbreviation for inorganic phosphate.) Kinetic studies have shown that ATP and tyrosine bind to the enzyme in random order. The tyrosyl-AMP intermediate is not released by the enzyme, and is sufficiently stable to allow study of the first reaction phase in isolation. The following discussion focuses on this phase.

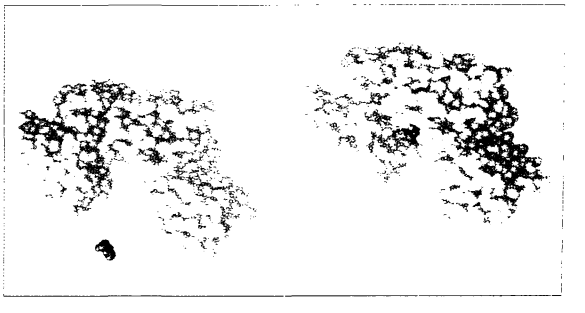


Figure 8–21 The conformational change induced in hexokinase by the binding of a substrate (D-glucose, shown in blue).

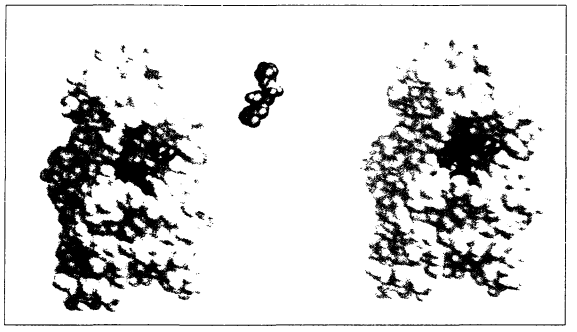


Figure 8–22 The structure of tyrosyl-tRNA synthetase. At left, the enzyme is shown without substrate bound. Active site residues that hydrogen-bond to the tyrosyl–AMP intermediate (Fig. 8–23a) are shown in red, and two residues (Thr^{40} and His^{45}) that contribute hydrogen bonds in the reaction transition state (Fig. 8–23b) are shown in orange. The amino acid side chains of Lys^{82} and Arg^{86}, which cover part of the active site, are not shown in order to expose more of the active site residues. At right, the same view is shown with the tyrosyl–AMP intermediate (shown in blue, with a phosphorus atom in yellow) bound.

The structure of tyrosyl-tRNA synthetase in its complex with tyrosyl-AMP is shown in Figure 8–22. This structure indicates a number of potential hydrogen bonds between enzyme and substrate (Fig. 8–23). Alan Fersht and colleagues have used this information to create, by site-directed mutagenesis (Chapter 28), a series of mutant enzymes lacking one or another of the amino acid side chains contributing to these hydrogen bonds.

Hydrogen bonds formed in the ES complex affect K_S, which can be directly measured for this enzyme. Hydrogen bonds formed only in the transition state affect k_{cat}. In several cases, substitution of a non-hydrogen-bonding amino acid at key positions in the active site affects

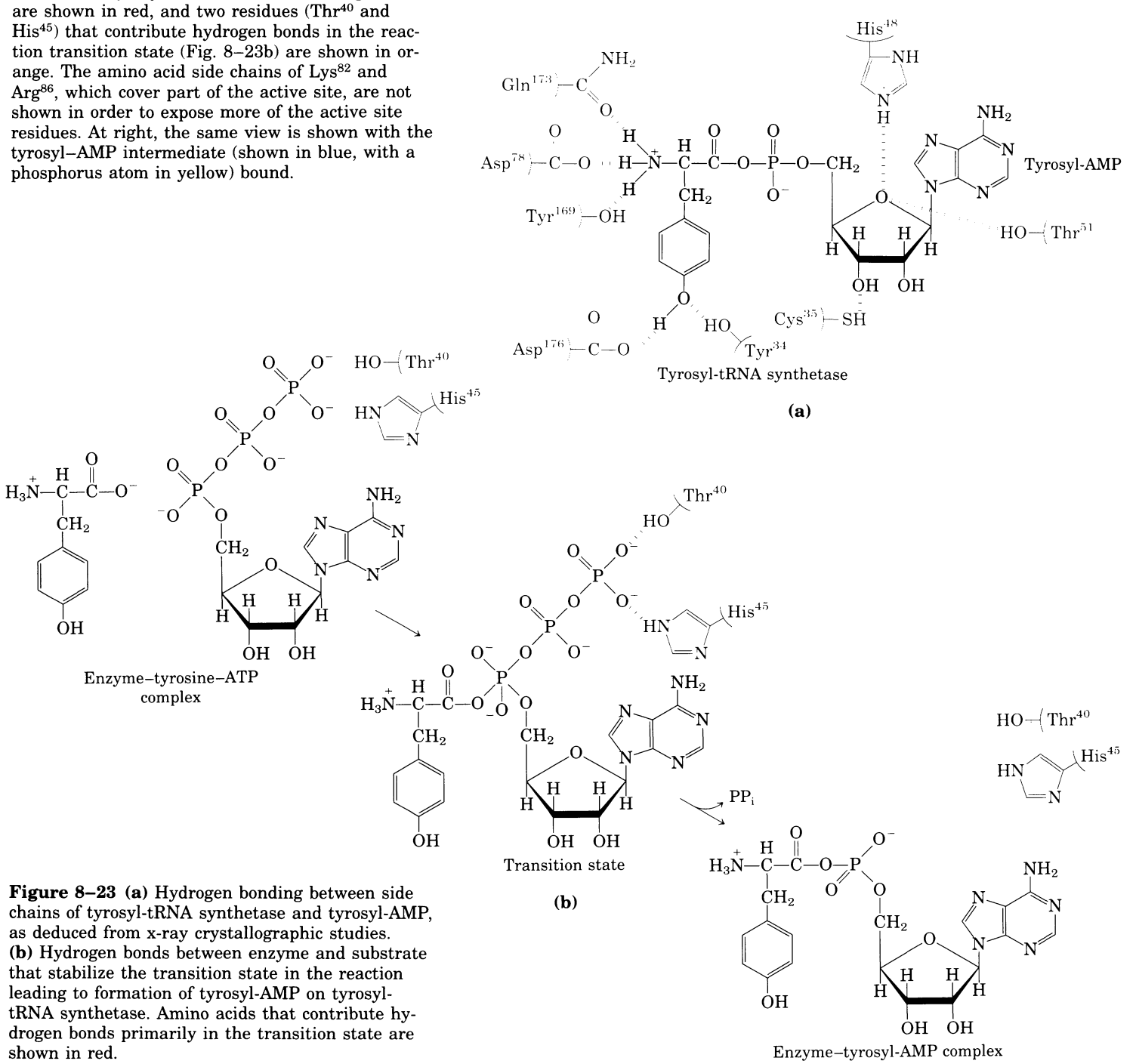


Figure 8–23 **(a)** Hydrogen bonding between side chains of tyrosyl-tRNA synthetase and tyrosyl-AMP, as deduced from x-ray crystallographic studies. **(b)** Hydrogen bonds between enzyme and substrate that stabilize the transition state in the reaction leading to formation of tyrosyl-AMP on tyrosyl-tRNA synthetase. Amino acids that contribute hydrogen bonds primarily in the transition state are shown in red.

k_{cat} but does not affect K_S. For example, substitution of an alanine for Thr^{40} and a glycine for His^{45} has little effect on the observed K_S for ATP. However, these substitutions lower the k_{cat} for the reaction by a factor of 300,000. In other words, this altered enzyme can still bind its substrates to form the ES complex, but once bound, the substrates do not react as rapidly because the enzyme can no longer form two essential hydrogen bonds that normally help to lower the activation energy required to reach the transition state (Fig. 8–23).

Regulatory Enzymes

We now turn to a special class of enzymes that represent exceptions to some of the rules outlined so far in this chapter. In cell metabolism, groups of enzymes work together in sequential pathways to carry out a given metabolic process, such as the multireaction conversion of glucose into lactate in skeletal muscle or the multireaction synthesis of an amino acid from simpler precursors in a bacterial cell. In such enzyme systems, the reaction product of the first enzyme becomes the substrate of the next, and so on (Figure 8–24).

Most of the enzymes in each system follow kinetic patterns already described. In each enzyme system, however, there is at least one enzyme that sets the rate of the overall sequence because it catalyzes the slowest or rate-limiting reaction. These **regulatory enzymes** exhibit increased or decreased catalytic activity in response to certain signals. By the action of such regulatory enzymes, the rate of each metabolic sequence is constantly adjusted to meet changes in the cell's demands for energy and for biomolecules required in cell growth and repair. In most multienzyme systems the first enzyme of the sequence is a regulatory enzyme. Catalyzing even the first few reactions of a pathway that leads to an unneeded product diverts energy and metabolites from more important processes. An excellent place to regulate a metabolic pathway, therefore, is at the point of commitment to the pathway. The other enzymes in the sequence are usually present in amounts providing a large excess of catalytic activity; they can promote their reactions only as fast as their substrates are made available from preceding reactions.

The activity of regulatory enzymes is modulated through various types of signal molecules, which are generally small metabolites or cofactors. There are two major classes of regulatory enzymes in metabolic pathways. **Allosteric enzymes** function through reversible, noncovalent binding of a regulatory metabolite called a modulator. The term allosteric derives from Greek *allos,* "other," and *stereos,* "solid" or "shape." Allosteric enzymes are those having "other shapes" or conformations induced by the binding of modulators. The second class includes enzymes regulated by reversible covalent modification. Both classes of regulatory enzymes tend to have multiple subunits, and in some cases the regulatory site(s) and the active site are on separate subunits.

There are at least two other mechanisms by which enzyme activity is regulated. Some enzymes are stimulated or inhibited by separate control proteins that bind to them and affect their activity. Others are activated by proteolytic cleavage, which unlike the other mechanisms is irreversible. Important examples of both these mechanisms are found in physiological processes such as digestion, blood clotting, hormone action, and vision.

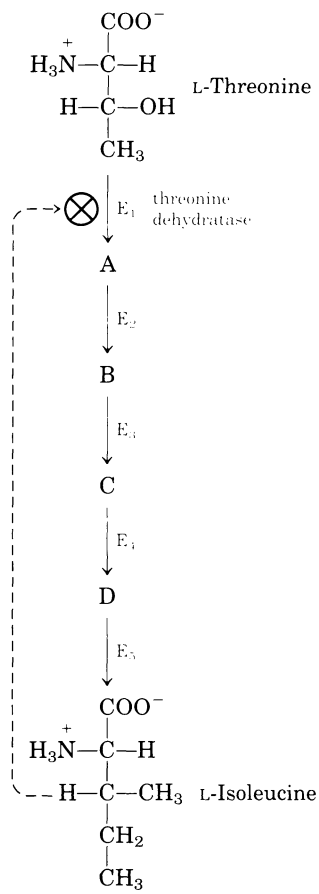


Figure 8–24 Feedback inhibition of the conversion of L-threonine into L-isoleucine, catalyzed by a sequence of five enzymes (E_1 to E_5). Threonine dehydratase (E_1) is specifically inhibited allosterically by L-isoleucine, the end product of the sequence, but not by any of the four intermediates (A to D). Such inhibition is indicated by the dashed feedback line and the ⊗ symbol at the threonine dehydratase reaction arrow.

No single rule governs the occurrence of different types of regulation in different systems. To a degree, allosteric (noncovalent) regulation may permit fine-tuning of metabolic pathways that are required continuously but at different levels of activity as cellular conditions change. Regulation by covalent modification tends to be all-or-none. However, both types of regulation are observed in a number of regulatory enzymes.

Allosteric Enzymes Are Regulated by Noncovalent Binding of Modulators

In some multienzyme systems the regulatory enzyme is specifically inhibited by the end product of the pathway, whenever the end product increases in excess of the cell's needs. When the regulatory enzyme reaction is slowed, all subsequent enzymes operate at reduced rates because their substrates are depleted by mass action. The rate of production of the pathway's end product is thereby brought into balance with the cell's needs. This type of regulation is called **feedback inhibition.** Buildup of the pathway's end product ultimately slows the entire pathway.

One of the first discovered examples of such allosteric feedback inhibition was the bacterial enzyme system that catalyzes the conversion of L-threonine into L-isoleucine (Fig. 8–24). In this system, the first enzyme, threonine dehydratase, is inhibited by isoleucine, the product of the last reaction of the series. Isoleucine is quite specific as an inhibitor. No other intermediate in this sequence of reactions inhibits threonine dehydratase, nor is any other enzyme in the sequence inhibited by isoleucine. Isoleucine binds not to the active site, but to another specific site on the enzyme molecule, the regulatory site. This binding is noncovalent and thus readily reversible; if the isoleucine concentration decreases, the rate of threonine dehydratase activity increases. Thus threonine dehydratase activity responds rapidly and reversibly to fluctuations in the concentration of isoleucine in the cell.

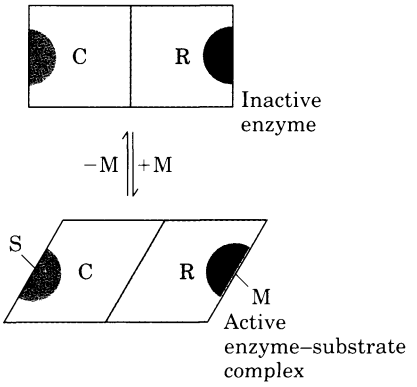


Figure 8–25 Schematic model of the subunit interactions in an allosteric enzyme, and interactions with inhibitors and activators. In many allosteric enzymes the substrate binding site and the modulator binding site(s) are on different subunits, the catalytic (C) and regulatory (R) subunits, respectively. Binding of the positive modulator (M) to its specific site on the regulatory subunit is communicated to the catalytic subunit through a conformational change. This change renders the catalytic subunit active and capable of binding the substrate (S) with higher affinity. On dislocation of the modulator from the regulatory subunit, the enzyme reverts to its inactive or less active form.

Allosteric Enzymes Are Exceptions to Many General Rules

The modulators for allosteric enzymes may be either inhibitory or stimulatory. An activator is often the substrate itself, and regulatory enzymes for which substrate and modulator are identical are called **homotropic.** When the modulator is a molecule other than the substrate the enzyme is **heterotropic.** Some enzymes have two or more modulators.

As already noted, the properties of allosteric enzymes are significantly different from those of simple nonregulatory enzymes discussed earlier in this chapter. Some of the differences are structural. In addition to active or catalytic sites, allosteric enzymes generally have one or more regulatory or allosteric sites for binding the modulator (Fig. 8–25). Just as an enzyme's active site is specific for its substrate, the allosteric site is specific for its modulator. Enzymes with several modulators generally have different specific binding sites for each. In homotropic enzymes the active site and regulatory site are the same.

Allosteric enzymes are also generally larger and more complex than simple enzymes. Most of them have two or more polypeptide chains or subunits. Aspartate transcarbamoylase, which catalyzes the first reaction in the biosynthesis of pyrimidine nucleotides (Chapter 21), has 12 polypeptide chains organized into catalytic and regulatory

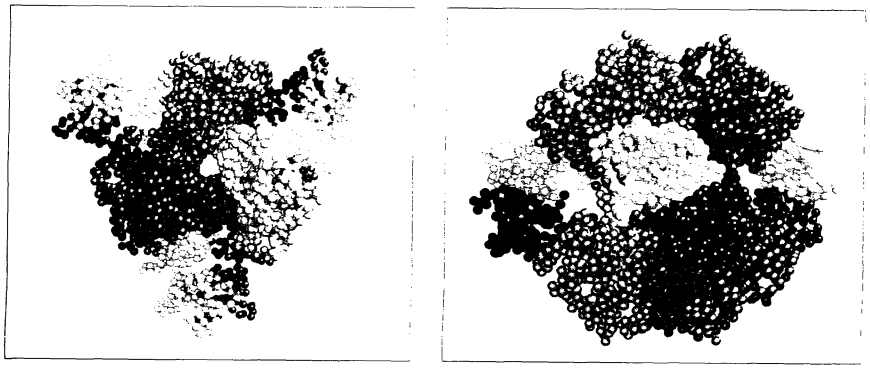


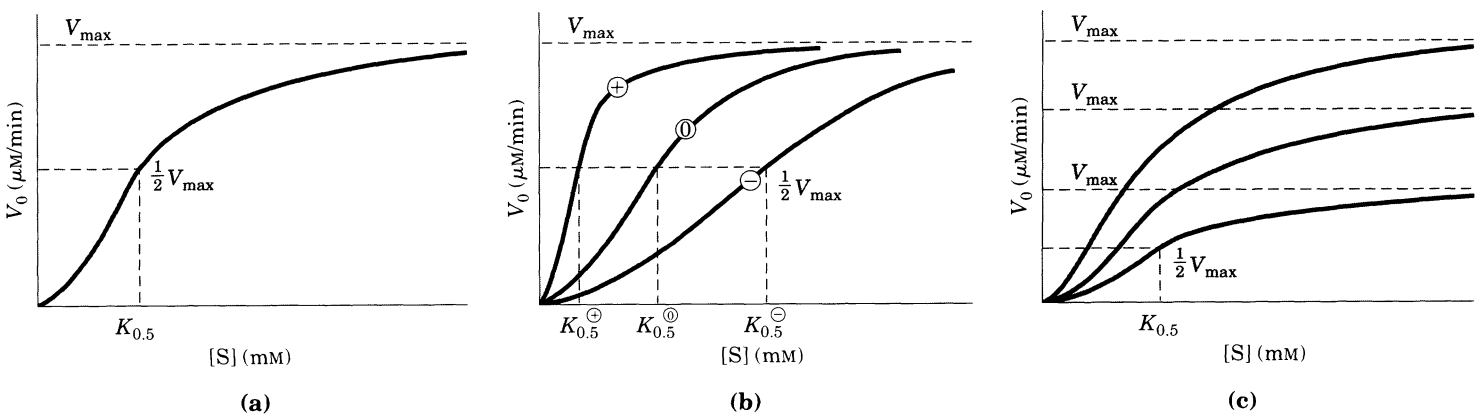
Figure 8–26 The three-dimensional subunit architecture of the regulatory enzyme aspartate transcarbamoylase; two different views. This allosteric regulatory enzyme has two catalytic clusters, each with three catalytic polypeptide chains, and three regulatory clusters, each with two regulatory polypeptide chains. The catalytic polypeptides in each cluster are shown in shades of blue and purple. Binding sites for allosteric modulators are found on the regulatory subunits (shown in white and red). Modulator binding produces large changes in enzyme conformation and activity. The role of this enzyme in nucleotide synthesis, and details of its regulation, will be discussed in Chapter 21.

subunits. Figure 8–26 shows the quaternary structure of this enzyme, deduced from x-ray analysis.

Other differences between nonregulated enzymes and allosteric enzymes involve kinetic properties. Allosteric enzymes show relationships between V_0 and [S] that differ from normal Michaelis–Menten behavior. They do exhibit saturation with the substrate when [S] is sufficiently high, but for some allosteric enzymes, when V_0 is plotted against [S] (Fig. 8–27) a sigmoid saturation curve results, rather than the hyperbolic curve shown by nonregulatory enzymes. Although we can find a value of [S] on the sigmoid saturation curve at which V_0 is half-maximal, we cannot refer to it with the designation K_m because the enzyme does not follow the hyperbolic Michaelis–Menten relationship. Instead the symbol $[S]_{0.5}$ or $K_{0.5}$ is often used to represent the substrate concentration giving half-maximal velocity of the reaction catalyzed by an allosteric enzyme (Fig. 8–27).

Sigmoid kinetic behavior generally reflects cooperative interactions between multiple protein subunits. In other words, changes in the structure of one subunit are translated into structural changes in adjacent subunits, an effect that is mediated by noncovalent interactions at the subunit–subunit interface. The principles are similar to those discussed for cooperativity in oxygen binding to the nonenzyme protein hemoglobin (p. 188). Homotropic allosteric enzymes generally have multiple subunits. In many cases the same binding site on each subunit functions as both the active site and the regulatory site. The substrate can function as a positive modulator (an activator) because the subunits act cooperatively. The binding of one molecule of the substrate to one binding site alters the enzyme's conformation and greatly enhances the binding of subsequent substrate molecules. This accounts for the sigmoid rather than hyperbolic increase in V_0 with increasing [S].

Figure 8–27 Substrate-activity curves for representative allosteric enzymes. Three examples of complex responses given by allosteric enzymes to their modulators. **(a)** The sigmoid curve given by a homotropic enzyme, in which the substrate also serves as a positive (stimulatory) modulator. Note that a relatively small increase in [S] in the steep part of the curve can cause a very large increase in V_0. Note also the resemblance to the oxygen-saturation curve of hemoglobin (see Fig. 7–28). **(b)** The effects of a positive modulator ⊕, a negative modulator ⊖, and no modulator ⓪ on an allosteric enzyme in which $K_{0.5}$ is modulated without a change in V_{max}. **(c)** A less common type of modulation, in which V_{max} is modulated with $K_{0.5}$ nearly constant.



With heterotropic enzymes, in which the modulator is a metabolite other than the substrate itself, it is difficult to generalize about the shape of the substrate-saturation curve. An activator may cause the substrate-saturation curve to become more nearly hyperbolic, with a decrease in $K_{0.5}$ but no change in V_{max}, thus resulting in an increased reaction velocity at a fixed substrate concentration (V_0 is higher for any value of [S]) (Fig. 8–27b). Other allosteric enzymes respond to an activator by an increase in V_{max}, with little change in $K_{0.5}$ (Fig. 8–27c). A negative modulator (an inhibitor) may produce a *more* sigmoid substrate-saturation curve, with an increase in $K_{0.5}$ (Fig. 8–27b). Allosteric enzymes therefore show different kinds of responses in their substrate-activity curves because some have inhibitory modulators, some have activating modulators, and some have both.

Two Models Explain the Kinetic Behavior of Allosteric Enzymes

The sigmoidal dependence of V_0 on [S] reflects subunit cooperativity, and has inspired two models to explain these cooperative interactions.

In the first model (the symmetry model), proposed by Jacques Monod and colleagues in 1965, an allosteric enzyme can exist in only two conformations, active and inactive (Fig. 8–28a). All subunits are in the active form or all are inactive. Every substrate molecule that binds increases the probability of a transition from the inactive to the active state.

In the second model (the sequential model) (Fig. 8–28b), proposed by Koshland in 1966, there are still two conformations, but subunits can undergo the conformational change individually. Binding of substrate increases the probability of the conformational change. A conformational change in one subunit makes a similar change in an adjacent subunit, as well as the binding of a second substrate molecule, more likely. There are more potential intermediate states in this model than in the symmetry model. The two models are not mutually exclusive; the symmetry model may be viewed as the "all-or-none" limiting case of the sequential model. The precise mechanism of allosteric interaction has not been established. Different allosteric enzymes may have different mechanisms for cooperative interactions.

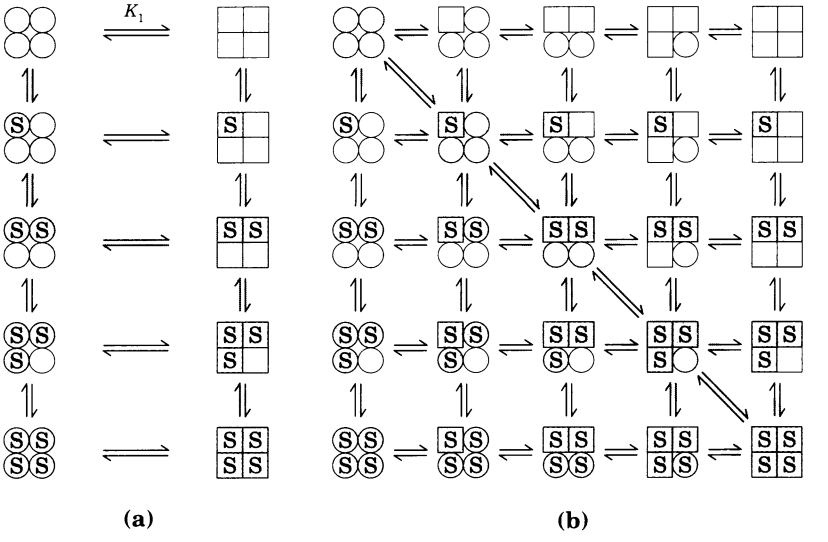


Figure 8–28 Two general models for the interconversion of inactive and active forms of allosteric enzymes. Four subunits are shown because the model was originally proposed for the oxygen-carrying protein hemoglobin. In the symmetry, or all-or-none, model **(a)** all the subunits are postulated to be in the same conformation, either all ◯ (low affinity or inactive) or all □ (high affinity or active). Depending on the equilibrium, K_1, between ◯ and □ forms, the binding of one or more substrate (S) molecules will pull the equilibrium toward the □ form. Subunits with bound S are shaded. A possible pathway is given by the gray shading. In the sequential model **(b)** each individual subunit can be in either the ◯ or □ form. A very large number of conformations is thus possible, but the shaded pathway (diagonal arrows) is the most probable route.

Other Mechanisms of Enzyme Regulation

In another important class of regulatory enzymes activity is modulated by covalent modification of the enzyme molecule. Modifying groups include phosphate, adenosine monophosphate, uridine monophosphate, adenosine diphosphate ribose, and methyl groups. These are generally covalently linked to and removed from the regulatory enzyme by separate enzymes (some examples are given in Box 8–4).

An important example of regulation by covalent modification is glycogen phosphorylase (M_r 94,500) of muscle and liver (Chapter 14), which catalyzes the reaction

$$\underset{\text{Glycogen}}{(\text{Glucose})_n} + P_i \longrightarrow \underset{\substack{\text{Shortened}\\ \text{glycogen}\\ \text{chain}}}{(\text{glucose})_{n-1}} + \text{glucose-1-phosphate}$$

The glucose-1-phosphate so formed can then be broken down into lactate in muscle or converted to free glucose in the liver. Glycogen phosphorylase occurs in two forms: the active form phosphorylase *a* and the relatively inactive form phosphorylase *b* (Fig. 8–29). Phosphorylase *a* has two subunits, each with a specific Ser residue that is phosphorylated at its hydroxyl group. These serine phosphate residues are required for maximal activity of the enzyme. The phosphate groups can be hydrolytically removed from phosphorylase *a* by a separate enzyme called phosphorylase phosphatase:

$$\underset{\text{(More active)}}{\text{Phosphorylase } a} + 2H_2O \longrightarrow \underset{\text{(Less active)}}{\text{phosphorylase } b} + 2P_i$$

In this reaction phosphorylase *a* is converted into phosphorylase *b* by the cleavage of two serine–phosphate covalent bonds.

Phosphorylase *b* can in turn be reactivated—covalently transformed back into active phosphorylase *a*—by another enzyme, phosphorylase kinase, which catalyzes the transfer of phosphate groups from ATP to the hydroxyl groups of the specific Ser residues in phosphorylase *b*:

$$2\text{ATP} + \underset{\text{(Less active)}}{\text{phosphorylase } b} \longrightarrow 2\text{ADP} + \underset{\text{(More active)}}{\text{phosphorylase } a}$$

The breakdown of glycogen in skeletal muscles and the liver is regulated by variations in the ratio of the two forms of the enzyme. The *a* and *b* forms of phosphorylase differ in their quaternary structure; the active site undergoes changes in structure and, consequently, changes in catalytic activity as the two forms are interconverted.

Some of the more complex regulatory enzymes are located at particularly crucial points in metabolism, so that they respond to multiple regulatory metabolites through both allosteric and covalent modification. Glycogen phosphorylase is an example. Although its primary regulation is through covalent modification, it is also modulated in a noncovalent, allosteric manner by AMP, which is an activator of phosphorylase *b*, and several other molecules that are inhibitors.

Glutamine synthetase of *E. coli,* one of the most complex regulatory enzymes known, provides examples of regulation by allostery, reversible covalent modification, and regulating proteins. It has at least eight allosteric modulators. The glutamine synthetase system is described in more detail in Chapter 21.

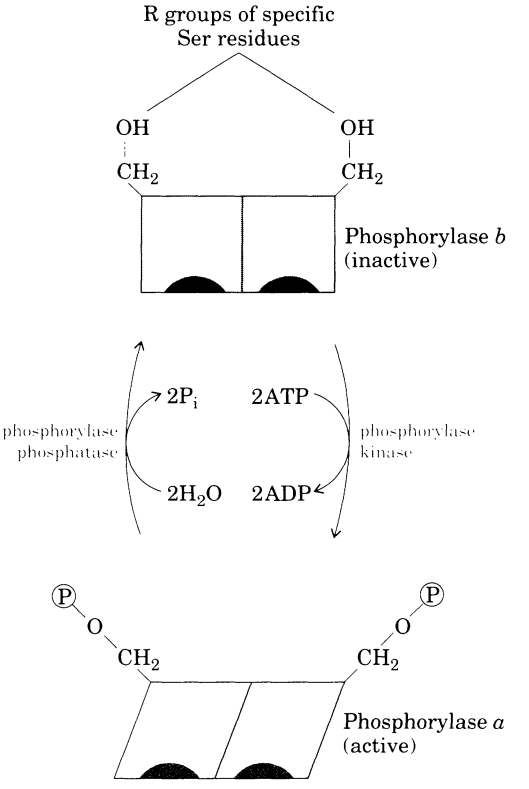
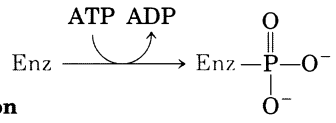
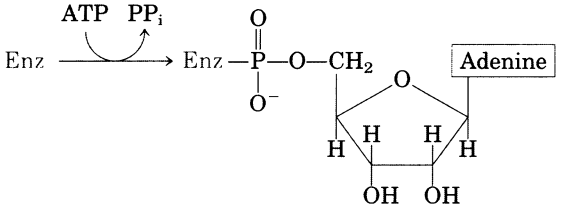
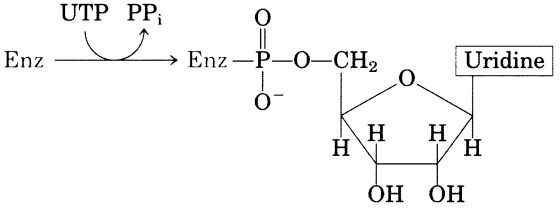
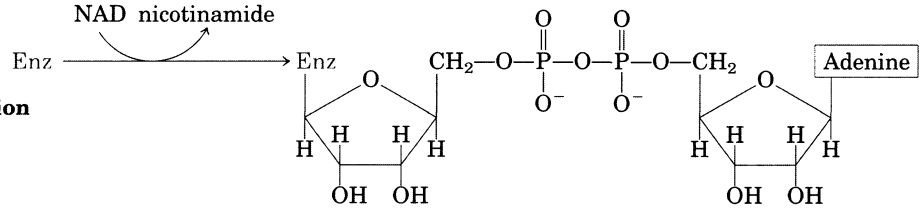
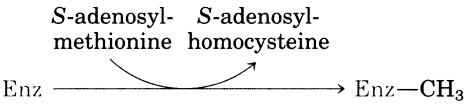


Figure 8–29 Regulation of glycogen phosphorylase activity by covalent modification. In the active form of the enzyme, phosphorylase *a*, specific Ser residues, one on each subunit, are in the phosphorylated state. Phosphorylase *a* is converted into phosphorylase *b*, which is relatively inactive, by enzymatic loss of these phosphate groups, promoted by phosphorylase phosphatase. Phosphorylase *b* can be reactivated to form phosphorylase *a* by the action of phosphorylase kinase.

BOX 8–4 Regulation of Protein Activity by Reversible Covalent Modification

A variety of chemical groups are used in reversible covalent modification of regulatory proteins to produce activity changes (Fig. 1). An example of phosphorylation (glycogen phosphorylase) is given in the text. An excellent example of methylation involves the methyl-accepting chemotaxis protein of bacteria. This protein is part of a system that permits a bacterium to swim toward an attractant in solution (such as a sugar) and away from repellent chemicals. The methylating agent is *S*-adenosylmethionine (adoMet), described in Chapter 17. ADP-ribosylation is an especially interesting reaction observed in only a few proteins. ADP-ribose is derived from nicotinamide adenine dinucleotide (see Fig. 12–41). This type of modification occurs for dinitrogenase reductase, resulting in the regulation of the important process of biological nitrogen fixation. In addition, both diphtheria toxin and cholera toxin are enzymes that catalyze the ADP-ribosylation (and inactivation) of key cellular enzymes or proteins. Diphtheria toxin acts on and inhibits elongation factor 2, a protein involved in protein biosynthesis. Cholera toxin acts on a specific G protein (Chapter 22) leading ultimately to several physiological responses including a massive loss of body fluids and sometimes death.

Figure 1 Some well-studied examples of enzyme modification reactions.

Covalent modification	Amino acid residues known to accept covalent modification
Phosphorylation 	Tyr, Ser, Thr, His
Adenylylation 	Tyr
Uridylylation 	Tyr
ADP-Ribosylation 	Arg, Gln, Cys, diphthamide (a modified His)
Methylation 	Glu

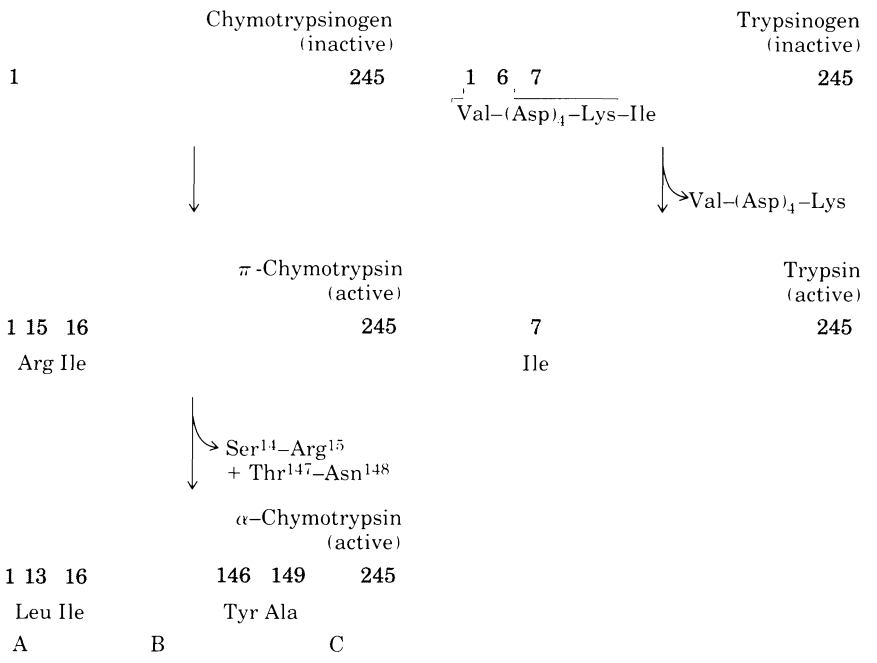


Figure 8–30 Activation of the zymogens of chymotrypsin and trypsin by proteolytic cleavage. The bars represent the primary sequence of the polypeptide chains. Amino acids at the termini of the polypeptide fragments generated by cleavage are indicated below the bars. The numbers represent the positions of the amino acids in the primary sequence of chymotrypsinogen or trypsinogen. (The amino-terminal amino acid is number 1.)

Activation of an enzyme by proteolytic cleavage is a somewhat different type of regulatory mechanism. An inactive precursor of the enzyme, called a **zymogen,** is cleaved to form the active enzyme. Many proteolytic enzymes (proteases) of the stomach and pancreas are regulated this way. Chymotrypsin and trypsin are initially synthesized as chymotrypsinogen and trypsinogen, respectively (Fig. 8–30). Specific cleavage causes conformational changes that expose the enzyme active site. Because this type of activation is irreversible, other mechanisms are needed to inactivate these enzymes. Proteolytic enzymes are inactivated by inhibitor proteins that bind very tightly to the enzyme active site. Pancreatic trypsin inhibitor (M_r 6,000) binds to and inhibits trypsin; α_1-antiproteinase (M_r 53,000) primarily inhibits elastase. An insufficiency of α_1-antiproteinase, believed to be caused by exposure to cigarette smoke, leads to lung damage and the condition known as emphysema.

Other examples of zymogen activation occur in hormones, connective tissue, and the blood-clotting system. The hormone insulin is produced by cleavage of proinsulin, collagen is initially synthesized as a soluble precursor called procollagen, and blood clotting is mediated by a complicated cascade of zymogen activations.

Summary

Virtually every biochemical reaction is catalyzed by enzymes. With the exception of a few catalytic RNAs, all known enzymes are proteins. Enzymes are extraordinarily effective catalysts, commonly producing reaction rate enhancements of 10^7 to 10^{14}. To be active, some enzymes require a chemical cofactor, which can be loosely or tightly bound. Each enzyme is classified according to the specific reaction it catalyzes.

Enzyme-catalyzed reactions are characterized by the formation of a complex between substrate and enzyme (an ES complex). The binding occurs in a pocket on the enzyme called the active site. The function of enzymes and other catalysts is to lower the activation energy for the reaction and thereby enhance the reaction rate. The equilibrium of a reaction is unaffected by the enzyme.

The energy used for enzymatic rate enhancements is derived from weak interactions (hydrogen bonds and van der Waals, hydrophobic, and ionic interactions) between the substrate and enzyme. The enzyme active site is structured so that many of these weak interactions occur only in the reaction transition state, thus stabilizing the transition state. The energy available from the numerous weak interactions between enzyme and substrate (the binding energy) is substantial and can generally account for observed rate enhancements. The need for multiple interactions is one reason for the large size of enzymes. Binding energy can be used to lower substrate entropy, to strain the substrate, or to cause a conformational change in the enzyme (induced fit). This same binding energy accounts for the exquisite specificity exhibited by enzymes for their substrates. Other catalytic mechanisms include general acid–base catalysis and covalent catalysis. Details of the reaction mechanisms have been worked out for many enzymes.

Kinetics is an important method for the study of enzyme mechanisms. Most enzymes have some common kinetic properties. As the concentration of the substrate is increased, the catalytic activity of a fixed concentration of an enzyme will increase in a hyperbolic fashion to approach a characteristic maximum rate V_{max}, at which essentially all the enzyme is in the form of the ES complex. The substrate concentration giving one-half V_{max} is the Michaelis–Menten constant K_m, which is characteristic for each enzyme acting on a given substrate. The Michaelis–Menten equation

$$V_0 = \frac{V_{max}[S]}{K_m + [S]}$$

relates the initial velocity of an enzymatic reaction to the substrate concentration and V_{max} through the constant K_m. Both K_m and V_{max} can be measured; they have different meanings for different enzymes. The limiting rate of an enzyme-catalyzed reaction at saturation is described by the constant k_{cat}, also called the turnover number. The ratio k_{cat}/K_m provides a good measure of catalytic efficiency. The Michaelis–Menten equation is also applicable to bisubstrate reactions, which occur by either ternary complex or double-displacement (ping-pong) pathways. Each enzyme has an optimum pH, as well as a characteristic specificity for the substrates on which it acts.

Enzymes can be inactivated by irreversible modification of a functional group essential for catalytic activity. They can also be reversibly inhibited, competitively or noncompetitively. Competitive inhibitors compete reversibly with the substrate for binding to the active site, but they are not transformed by the enzyme. Noncompetitive inhibitors bind to some other site on the free enzyme or to the ES complex.

Some enzymes regulate the rate of metabolic pathways in cells. In feedback inhibition, the end product of a pathway inhibits the first enzyme of that pathway. The activity of some regulatory enzymes, called allosteric enzymes, is adjusted by reversible, noncovalent binding of a specific modulator to a regulatory or allosteric site. Such modulators may be inhibitory or stimulatory and may be either the substrate itself or some other metabolite. The kinetic behavior of allosteric enzymes reflects cooperative interactions among the enzyme subunits. Other regulatory enzymes are modulated by covalent modification of a specific functional group necessary for activity, or by proteolytic cleavage of a zymogen.

General

Evolution of Catalytic Function. (1987) *Cold Spring Harb. Symp. Quant. Biol.* **52.**

A collection of excellent papers on many topics discussed in this chapter.

Fersht, A. (1985) *Enzyme Structure and Mechanism,* W.H. Freeman and Company, New York.

A clearly written, concise introduction. More advanced.

Friedmann, H. (ed) (1981) *Benchmark Papers in Biochemistry,* Vol. 1: *Enzymes,* Hutchinson Ross Publishing Company, Stroudsburg, PA.

A collection of classic papers in enzyme chemistry, with historical commentaries by the editor. Extremely interesting.

Jencks, W.P. (1987) *Catalysis in Chemistry and Enzymology,* Dover Publications, Inc., New York.

A new printing of an outstanding book on the subject. More advanced.

Principles of Catalysis

Hansen, D.E. and Raines, R.T. (1990) Binding energy and enzymatic catalysis. *J. Chem. Educ.* **67,** 483–489.

A good place for the beginning student to acquire a better understanding of principles.

Jencks, W.P. (1975) Binding energy, specificity, and enzymic catalysis: the Circe effect. *Adv. Enzymol.* **43,** 219–410.

Kraut, J. (1988) How do enzymes work? *Science* **242,** 533–540.

Lerner, R.A., Benkovic, S.J., & Schulz, P.G. (1991) At the crossroads of chemistry and immunology: catalytic antibodies. *Science* **252,** 659–667.

Schultz, P.G. (1988) The interplay between chemistry and biology in the design of enzymatic catalysts. *Science* **240,** 426–433.

Kinetics

Cleland, W.W. (1977) Determining the chemical mechanisms of enzyme-catalyzed reactions by kinetic studies. *Adv. Enzymol.* **45,** 273–387.

Raines, R. T. & Hansen, D.E. (1988) An intuitive approach to steady-state kinetics. *J. Chem. Educ.* **65,** 757–759.

Segel, I.H. (1975) *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady State Enzyme Systems,* John Wiley & Sons, Inc., New York.

A more advanced treatment.

Enzyme Examples

Anderson, C.M., Zucker, F.H., & Steitz, T.A. (1979) Space-filling models of kinase clefts and conformation changes. *Science* **204,** 375–380.

Structure of hexokinase and other enzymes utilizing ATP.

Fersht, A.R. (1987) Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. *Biochemistry* **26,** 8031–8037.

Warshel, A., Naray-Szabo, G., Sussman, F., & Hwang, J.-K. (1989) How do serine proteases really work? *Biochemistry* **28,** 3629–3637.

Regulatory Enzymes

Dische, Z. (1976) The discovery of feedback inhibition. *Trends Biochem. Sci.* **1,** 269–270.

Koshland, D.E., Jr. & Neet, K.E. (1968) The catalytic and regulatory properties of enzymes. *Annu. Rev. Biochem.* **37,** 359–410.

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

A classic paper introducing the concept of allosteric regulation.

Problems

1. *Keeping the Sweet Taste of Corn* The sweet taste of freshly picked corn is due to the high level of sugar in the kernels. Store-bought corn (several days after picking) is not as sweet, because about 50% of the free sugar of corn is converted into starch within one day of picking. To preserve the sweetness of fresh corn, the husked ears are immersed in boiling water for a few minutes ("blanched") and then cooled in cold water. Corn processed in this way and stored in a freezer maintains its sweetness. What is the biochemical basis for this procedure?

2. *Intracellular Concentration of Enzymes* To approximate the actual concentration of enzymes in a bacterial cell, assume that the cell contains 1,000 different enzymes in solution in the cytosol, that each protein has a molecular weight of 100,000, and that all 1,000 enzymes are present in equal concentrations. Assume that the bacterial cell is a cylinder (diameter 1 μm, height 2.0 μm). If the cytosol (specific gravity 1.20) is 20% soluble protein by weight, and if the soluble protein consists entirely of different enzymes, calculate the *average* molar concentration of each enzyme in this hypothetical cell.

3. *Rate Enhancement by Urease* The enzyme urease enhances the rate of urea hydrolysis at pH 8.0 and 20 °C by a factor of 10^{14}. If a given quantity of urease can completely hydrolyze a given quantity of urea in 5 min at 20 °C and pH 8.0, how long will it take for this amount of urea to be hydrolyzed under the same conditions in the absence of urease? Assume that both reactions take place in sterile systems so that bacteria cannot attack the urea.

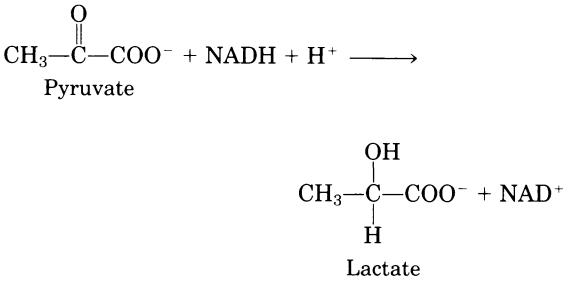
4. *Requirements of Active Sites in Enzymes* The active site of an enzyme usually consists of a pocket on the enzyme surface lined with the amino acid side chains necessary to bind the substrate and catalyze its chemical transformation. Carboxypeptidase, which sequentially removes the carboxyl-terminal amino acid residues from its peptide substrates, consists of a single chain of 307 amino acids. The two essential catalytic groups in the active site are furnished by Arg^{145} and Glu^{270}.

(a) If the carboxypeptidase chain were a perfect α helix, how far apart (in nanometers) would Arg^{145} and Glu^{270} be? (*Hint:* See Fig. 7–6.)

(b) Explain how it is that these two amino acids, so distantly separated in the sequence, can catalyze a reaction occurring in the space of a few tenths of a nanometer.

(c) If only these two catalytic groups are involved in the mechanism of hydrolysis, why is it necessary for the enzyme to contain such a large number of amino acid residues?

5. *Quantitative Assay for Lactate Dehydrogenase* The muscle enzyme lactate dehydrogenase catalyzes the reaction



NADH and NAD^+ are the reduced and oxidized forms, respectively, of the coenzyme NAD. Solutions of NADH, but *not* NAD^+, absorb light at 340 nm. This property is used to determine the concentration of NADH in solution by measuring spectrophotometrically the amount of light absorbed at 340 nm by the solution. Explain how these properties of NADH can be used to design a quantitative assay for lactate dehydrogenase.

6. *Estimation of* V_{max} *and* K_m *by Inspection* Although graphical methods are available for accurate determination of the values of V_{max} and K_m of an enzyme-catalyzed reaction (see Box 8–1), these values can be quickly estimated by inspecting values of V_0 at increasing [S]. Estimate the approximate value of V_{max} and K_m for the enzyme-catalyzed reaction for which the following data were obtained:

[S] (M)	V_0 (μM/min)
2.5×10^{-6}	28
4.0×10^{-6}	40
1×10^{-5}	70
2×10^{-5}	95
4×10^{-5}	112
1×10^{-4}	128
2×10^{-3}	139
1×10^{-2}	140

7. *Relation between Reaction Velocity and Substrate Concentration: Michaelis–Menten Equation*

(a) At what substrate concentration will an enzyme having a k_{cat} of 30 s^{-1} and a K_m of 0.005 M show one-quarter of its maximum rate?

(b) Determine the fraction of V_{max} that would be found in each case when [S] = $\frac{1}{2}K_m$, $2K_m$, and $10K_m$.

8. *Graphical Analysis of* V_{max} *and* K_m *Values* The following experimental data were collected during a study of the catalytic activity of an intestinal peptidase capable of hydrolyzing the dipeptide glycylglycine:

$$\text{Glycylglycine} + H_2O \longrightarrow 2 \text{ glycine}$$

[S] (mM)	Product formed (μmol/min)
1.5	0.21
2.0	0.24
3.0	0.28
4.0	0.33
8.0	0.40
16.0	0.45

From these data determine by graphical analysis (see Box 8–1) the values of K_m and V_{max} for this enzyme preparation and substrate.

9. *The Turnover Number of Carbonic Anhydrase* Carbonic anhydrase of erythrocytes (M_r 30,000) is among the most active of known enzymes. It catalyzes the reversible hydration of CO_2:

$$H_2O + CO_2 \rightleftharpoons H_2CO_3$$

which is important in the transport of CO_2 from the tissues to the lungs.

(a) If 10 μg of pure carbonic anhydrase catalyzes the hydration of 0.30 g of CO_2 in 1 min at 37 °C under optimal conditions, what is the turnover number (k_{cat}) of carbonic anhydrase (in units of min^{-1})?

(b) From the answer in (a), calculate the activation energy of the enzyme-catalyzed reaction (in kJ/mol).

(c) If carbonic anhydrase provides a rate enhancement of 10^7, what is the activation energy for the uncatalyzed reaction?

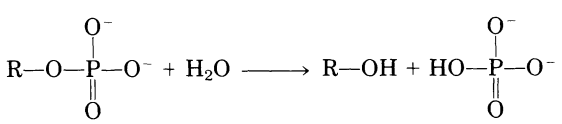
10. *Irreversible Inhibition of an Enzyme* Many enzymes are inhibited irreversibly by heavy-metal ions such as Hg^{2+}, Cu^{2+}, or Ag^+, which can react with essential sulfhydryl groups to form mercaptides:

$$\text{Enz—SH} + Ag^+ \longrightarrow \text{Enz—S—Ag} + H^+$$

The affinity of Ag^+ for sulfhydryl groups is so great that Ag^+ can be used to titrate —SH groups quantitatively. To 10 mL of a solution containing 1.0 mg/mL of a pure enzyme was added just enough $AgNO_3$ to completely inactivate the enzyme. A total of 0.342 μmol of $AgNO_3$ was required. Calculate the *minimum* molecular weight of the enzyme. Why does the value obtained in this way give only the minimum molecular weight?

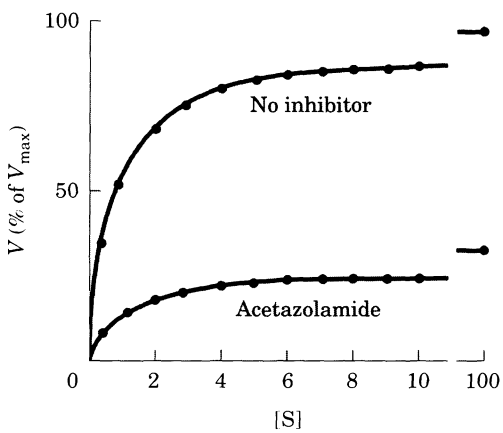
11. *Protection of an Enzyme against Denaturation by Heat* When enzyme solutions are heated, there is a progressive loss of catalytic activity with time. This loss is the result of the unfolding of the native enzyme molecule to a randomly coiled conformation, because of its increased thermal energy. A solution of the enzyme hexokinase incubated at 45 °C lost 50% of its activity in 12 min, but when hexokinase was incubated at 45 °C in the presence of a very large concentration of one of its substrates, it lost only 3% of its activity. Explain why thermal denaturation of hexokinase was retarded in the presence of one of its substrates.

12. *Clinical Application of Differential Enzyme Inhibition* Human blood serum contains a class of enzymes known as acid phosphatases, which hydrolyze biological phosphate esters under slightly acidic conditions (pH 5.0):

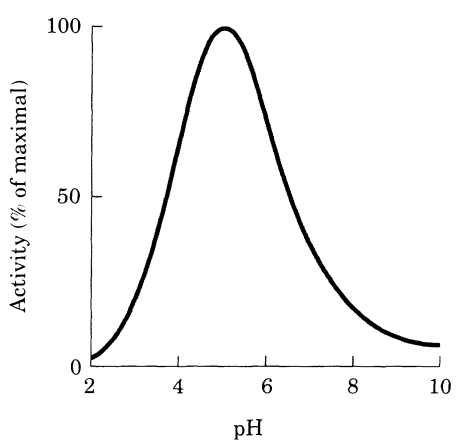


Acid phosphatases are produced by erythrocytes, the liver, kidney, spleen, and prostate gland. The enzyme from the prostate gland is clinically important because an increased activity in the blood is frequently an indication of cancer of the prostate gland. The phosphatase from the prostate gland is strongly inhibited by the tartrate ion, but acid phosphatases from other tissues are not. How can this information be used to develop a specific procedure for measuring the activity of the acid phosphatase of the prostate gland in human blood serum?

13. *Inhibition of Carbonic Anhydrase by Acetazolamide* Carbonic anhydrase is strongly inhibited by the drug acetazolamide, which is used as a diuretic (increases the production of urine) and to treat glaucoma (reduces excessively high pressure within the eyeball). Carbonic anhydrase plays an important role in these and other secretory processes, because it participates in regulating the pH and bicarbonate content of a number of body fluids. The experimental curve of reaction velocity (given here as percentage of V_{max}) versus [S] for the carbonic anhydrase reaction is illustrated below (upper curve). When the experiment is repeated in the presence of acetazolamide, the lower curve is obtained. From an inspection of the curves and your knowledge of the kinetic properties of competitive and noncompetitive enzyme inhibitors, determine the nature of the inhibition by acetazolamide. Explain.



14. *pH Optimum of Lysozyme* The enzymatic activity of lysozyme is optimal at pH 5.2.



The active site of lysozyme contains two amino acid residues essential for catalysis: Glu^{35} and Asp^{52}. The pK_a values of the carboxyl side chains of these two residues are 5.9 and 4.5, respectively. What is the ionization state (protonated or deprotonated) of each residue at the pH optimum of lysozyme? How can the ionization states of these two amino acid residues explain the pH-activity profile of lysozyme shown above?