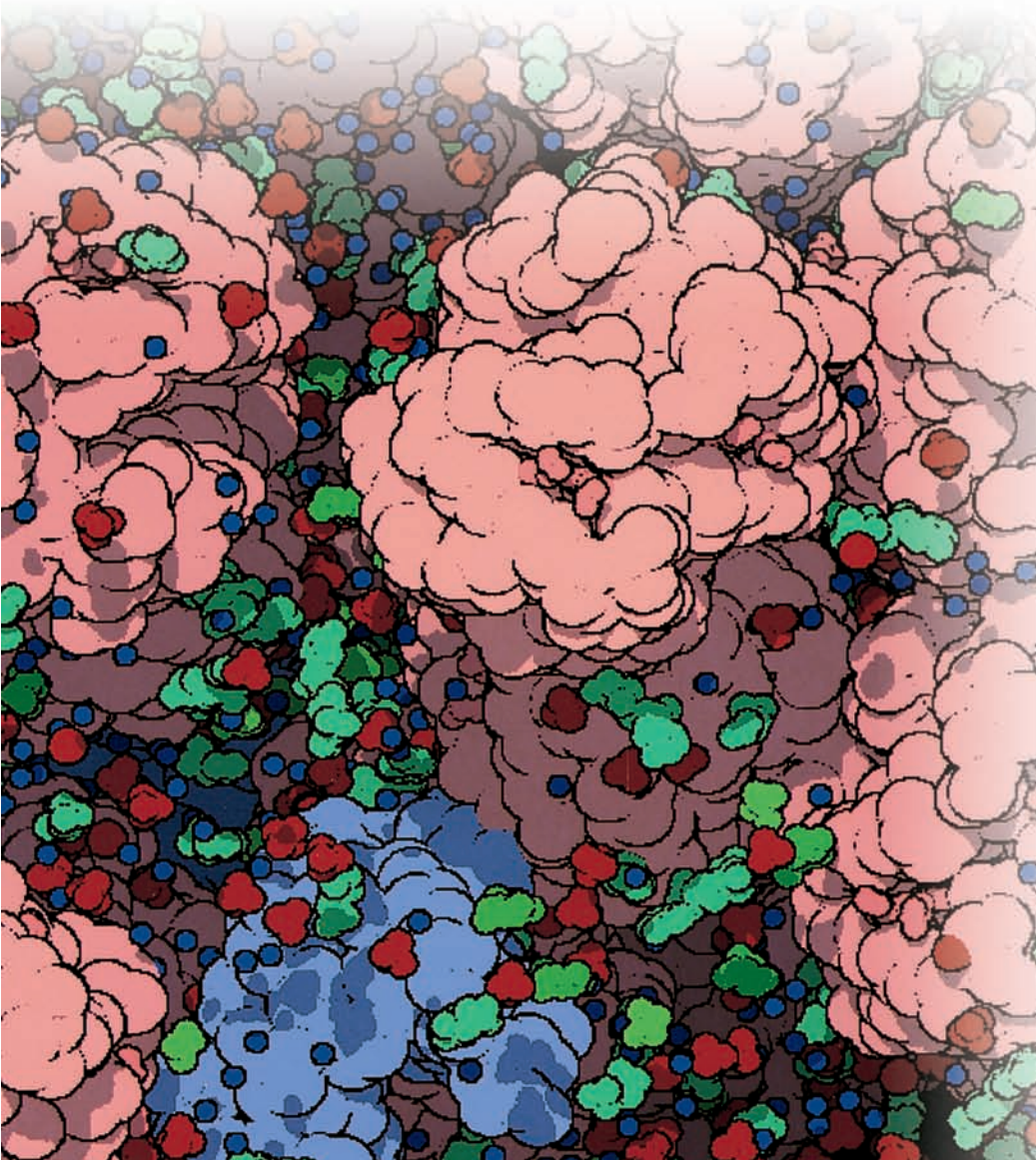
CHAPTER 5

Amino Acids, Peptides, and Proteins

Hemoglobin Within a Red Blood Cell
Human red blood cells are filled almost to bursting with the oxygen-carrying protein hemoglobin. The large pink structures are hemoglobin molecules. Sugar and amino acids are shown in green. Positive ions are blue. Negative are red. The large blue molecule is an enzyme.



OUTLINE

Overview

PROTEINS ARE ESSENTIAL CONSTITUENTS OF ALL ORGANISMS. MOST TASKS PERFORMED BY LIVING CELLS REQUIRE PROTEINS, WHICH PERFORM an astonishing variety of functions. In addition to serving as structural materials in all living organisms (e.g., structural components in the muscle and connective tissue of animals or cell wall components of prokaryotes), proteins are involved in such diverse functions as metabolic regulation, transport, defense, and catalysis. The functional diversity exhibited by this class of biomolecules is directly related to the combinatorial possibilities of the monomeric units, the 20 amino acids.

Proteins are molecular tools. They are a diverse and complex group of macromolecules that perform the thousands of tasks that sustain life. One measure of their importance is their abundance: at least 50% of the dry weight of cells is protein. Another is the vast numbers of unique protein molecules that living organisms produce. The genomes of most organisms code for thousands or tens of thousands of proteins. How can proteins be so diverse? The answer lies in their structural composition. Proteins are linear polymers composed of 20 different amino acids, linked by covalent bonds. Amino acids can theoretically be joined to form protein molecules in any imaginable size or sequence. Consider, for example, a protein composed of 100 amino acids. The total possible number of combinations for such a molecule is an astronomical 20^{100}. Not all protein sequences, however, code for useful proteins. Of the trillions of possible protein sequences, only a small fraction (recent estimates range from 650,000 to about 2 million) are actually produced by living organisms. An important reason for this remarkable discrepancy is demonstrated by the complex set of structural and functional properties of naturally occurring proteins that have evolved over billions of years in response to selection pressure. Among these are (1) structural features that make protein folding a relatively rapid and successful process, (2) the presence of binding sites that are specific for one or a small group of molecules, (3) an appropriate balance of structural flexibility and rigidity so that function is maintained, (4) surface structure that is appropriate for a protein's immediate environment (i.e., hydrophobic in membranes and hydrophilic in cytoplasm), and (5) vulnerability of proteins to degradation reactions when they become damaged or no longer useful.

Considering the vital importance of proteins in living organisms, the investigation of the structural and functional properties of proteins has always been a priority with biochemists. Proteins can be distinguished based on their number of amino acids (called **amino acid residues**), their overall amino acyl composition, and their amino acyl sequence. Selected examples of the diversity of proteins are illustrated in Figure 5.1.

Amino acid polymers are also differentiated according to their molecular weights or the number of amino acid residues they contain. Molecules with molecular weights ranging from several thousand to several million daltons are called **polypeptides**. Those with low molecular weights, typically consisting of fewer than 50 amino acids, are called **peptides**. The term **protein** describes molecules with more than 50 amino acids. Each protein consists of one or more polypeptide chains.

Throughout this textbook the terms *peptide* and *protein* will be used just as defined. In the literature, however, the distinction between proteins and peptides is often imprecise. For example, some biochemists define oligopeptides as polymers consisting of two to ten amino acids and polypeptides as having more than

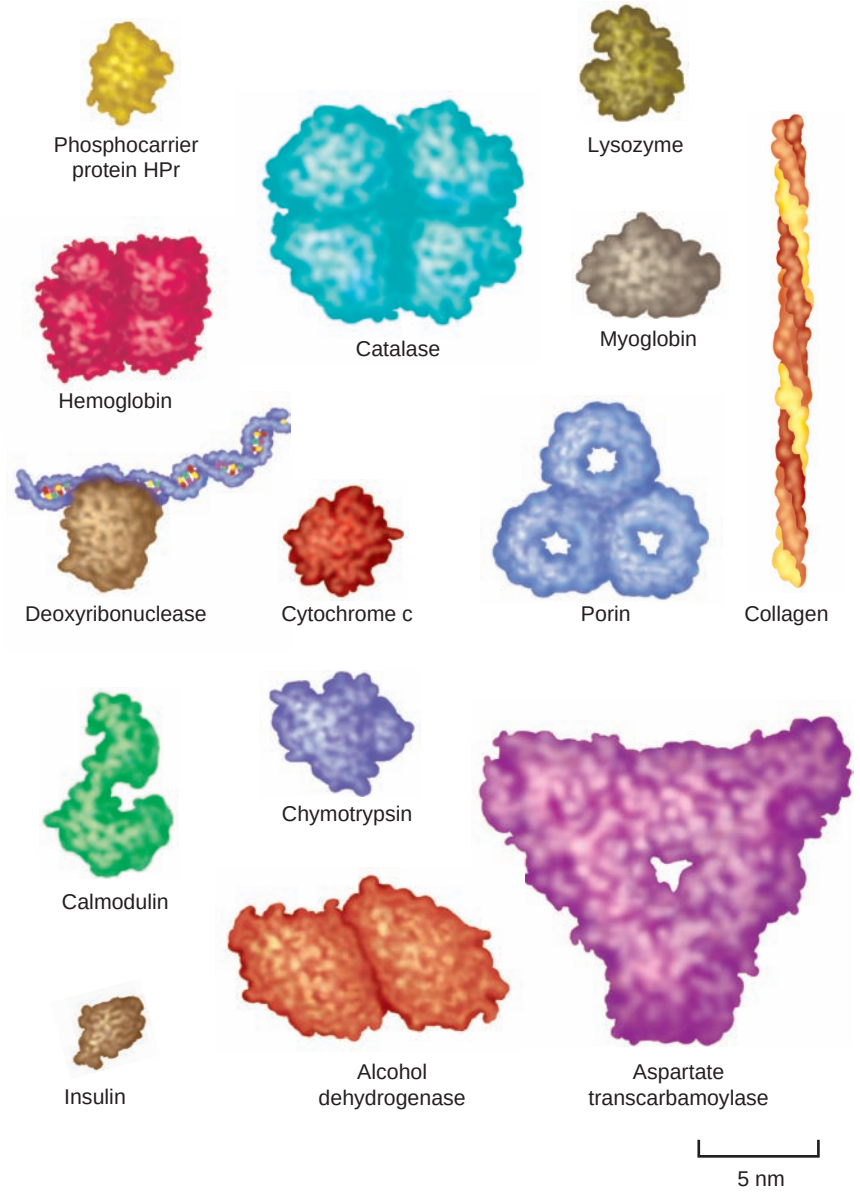


FIGURE 5.1

Protein Diversity

Proteins occur in an enormous diversity of sizes and shapes.

ten residues. Proteins, in this view, have molecular weights greater than 10,000 D. In addition, the terms *protein* and *polypeptide* are often used interchangeably. Here, the term *polypeptide* will be used whenever the topic under discussion applies to both peptides and proteins.

This chapter begins with a review of the structures and chemical properties of the amino acids. This is followed by descriptions of the structural and functional features of peptides and proteins and the protein folding process. The emphasis throughout is on the intimate relationship between the structure and function of polypeptides. In Chapter 6 the functioning of the enzymes, an especially important group of proteins, is discussed. Protein synthesis is covered in Chapter 19.

5.1 AMINO ACIDS

The hydrolysis of each polypeptide yields a set of amino acids, referred to as the molecule's *amino acid composition*. The structures of the 20 amino acids that are commonly found in naturally occurring polypeptides are shown in Figure 5.2.

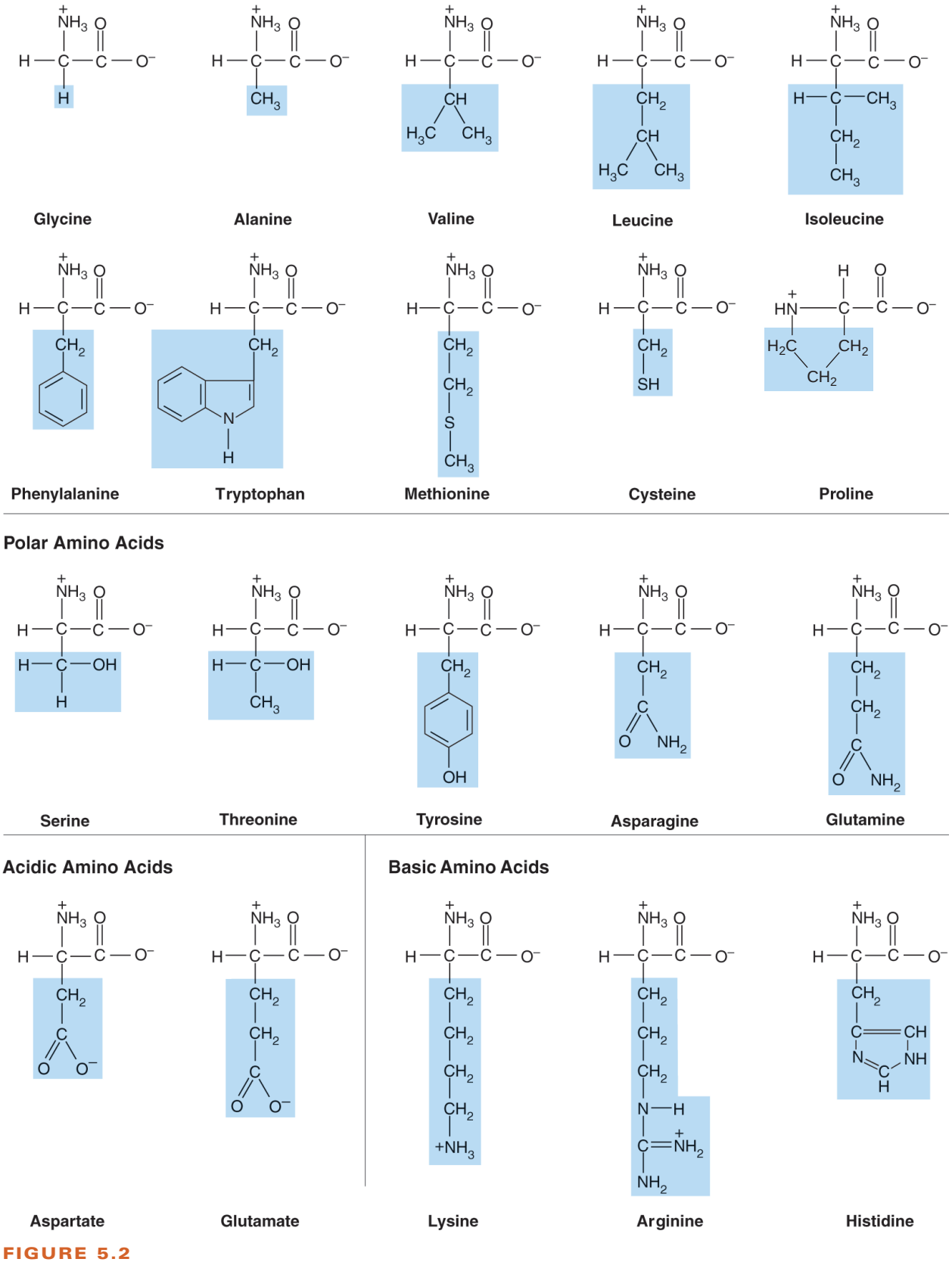


FIGURE 5.2

The Standard Amino Acids

The ionization state of the amino acid molecules in this illustration represents the dominant species that occur at a pH of 7. The side chains are indicated by shaded boxes.

TABLE 5.1 Names and Abbreviations of the Standard Amino Acids

Amino Acid	Three-Letter Abbreviation	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

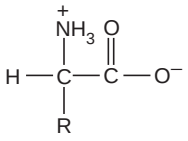


FIGURE 5.3
General Structure of the α-Amino Acids

These amino acids are referred to as *standard* amino acids. Common abbreviations for the standard amino acids are listed in Table 5.1. Note that 19 of the standard amino acids have the same general structure (Figure 5.3). These molecules contain a central carbon atom (the α-carbon) to which an amino group, a carboxylate group, a hydrogen atom, and an R (side chain) group are attached. The exception, proline, differs from the other standard amino acids in that its amino group is secondary, formed by ring closure between the R group and the amino nitrogen. Proline confers rigidity to the peptide chain because rotation about the α-carbon is not possible. This structural feature has significant implications in the structure and, therefore, the function of proteins with a high proline content.

Nonstandard amino acids consist of amino acid residues that have been chemically modified after incorporation into a polypeptide or amino acids that occur in living organisms but are not found in proteins.

At a pH of 7, the carboxyl group of an amino acid is in its conjugate base form (—COO^-), and the amino group is in its conjugate acid form (—NH_3^+). Thus each amino acid can behave as either an acid or a base. The term **amphoteric** is used to describe this property. Molecules that bear both positive and negative charges on different atoms are called **zwitterions**. The R group, however, gives each amino acid its unique properties.

Amino Acid Classes

Because the sequence of amino acids determines the final three-dimensional configuration of each protein, their structures are examined carefully in the next four subsections. Amino acids are classified according to their capacity to interact with water. By using this criterion, four classes may be distinguished: (1) nonpolar, (2) polar, (3) acidic, and (4) basic.

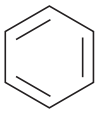


FIGURE 5.4
Benzene

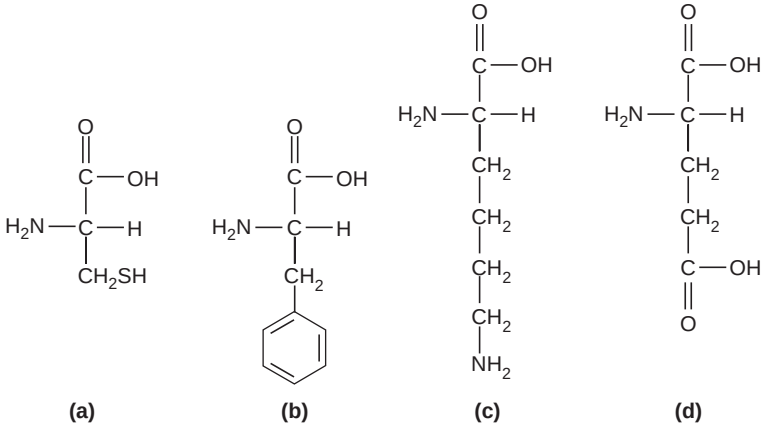
NONPOLAR AMINO ACIDS The nonpolar amino acids contain mostly hydrocarbon R groups that do not bear positive or negative charges. Nonpolar (i.e., hydrophobic) amino acids play an important role in maintaining the three-dimensional structures of proteins, because they interact poorly with water. Two types of hydrocarbon side chains are found in this group: aromatic and aliphatic. **Aromatic** hydrocarbons contain cyclic structures that constitute a class of unsaturated hydrocarbons with unique properties. Benzene is one of the simplest aromatic hydrocarbons (Figure 5.4). The term **aliphatic** refers to nonaromatic hydrocarbons such as methane and cyclohexane. Phenylalanine and tryptophan contain aromatic ring structures. Glycine, alanine, valine, leucine, isoleucine, and proline have aliphatic R groups. A sulfur atom appears in the aliphatic side chains of methionine and cysteine. Methionine contains a thioether group (—S—CH_3) in its side chain. Its derivative *S*-adenosyl methionine (SAM) is an important metabolite that serves as a methyl donor in numerous biochemical reactions. The sulfhydryl group (—SH) of cysteine is highly reactive and is an important component of many enzymes. It also binds metals (e.g., iron and copper ions) in proteins. Additionally, the sulfhydryl groups of two cysteine molecules oxidize easily in the extracellular compartment to form a disulfide compound called cystine. (See p. 136 for a discussion of this reaction.)

POLAR AMINO ACIDS Because polar amino acids have functional groups capable of hydrogen bonding, they easily interact with water. (Polar amino acids are described as hydrophilic, or "water-loving.") Serine, threonine, tyrosine, asparagine, and glutamine belong to this category. Serine, threonine, and tyrosine contain a polar hydroxyl group, which enables them to participate in hydrogen bonding, an important factor in protein structure. The hydroxyl groups serve other functions in proteins. For example, the formation of the phosphate ester of tyrosine is a common regulatory mechanism. Additionally, the —OH groups of serine and threonine are points for attaching carbohydrates. Asparagine and glutamine are amide derivatives of the acidic amino acids aspartic acid and glutamic acid, respectively. Because the amide functional group is highly polar, the hydrogen-bonding capability of asparagine and glutamine has a significant effect on protein stability.

ACIDIC AMINO ACIDS Two standard amino acids have side chains with carboxylate groups. Because the side chains of aspartic acid and glutamic acid are negatively charged at physiological pH, they are often referred to as aspartate and glutamate.

QUESTION 5.1

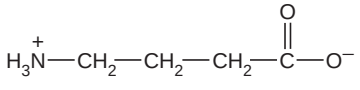
Classify these standard amino acids according to whether their structures are non-polar, polar, acidic, or basic.



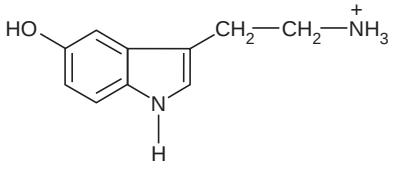
BASIC AMINO ACIDS Basic amino acids bear a positive charge at physiological pH. They can therefore form ionic bonds with acidic amino acids. Lysine, which has a side chain amino group, accepts a proton from water to form the conjugate acid ($—NH_3^+$). When lysine side chains in collagen fibrils, a vital structural component of ligaments and tendons, are oxidized and subsequently condensed, strong intramolecular and intermolecular cross-linkages are formed. Because the guanidino group of arginine has a pK_a range of 11.5 to 12.5 in proteins, it is permanently protonated at physiological pH and, therefore, does not function in acid-base reactions. Histidine, on the other hand, is a weak base because it is only partially ionized at pH 7. Consequently, histidine residues act as a buffer. They also play an important role in the catalytic activity of numerous enzymes.

KEY CONCEPT

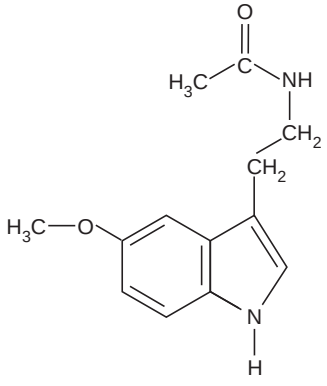
Amino acids are classified according to their capacity to interact with water. This criterion may be used to distinguish four classes: nonpolar, polar, acidic, and basic.



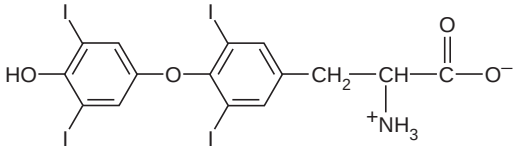
GABA



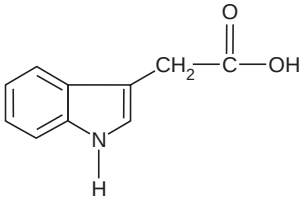
Serotonin



Melatonin



Thyroxine



Indole acetic acid

FIGURE 5.5
Some Derivatives of Amino Acids

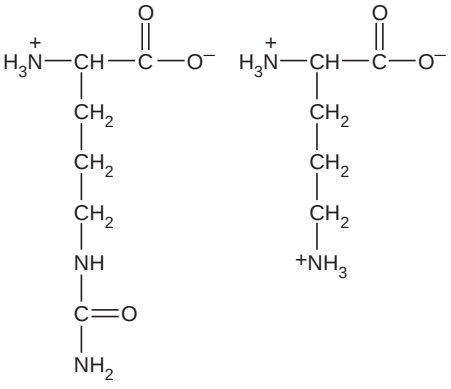


FIGURE 5.6
Citrulline and Ornithine

Biologically Active Amino Acids

In addition to their primary function as components of protein, amino acids have several other biological roles.

1. Several α-amino acids or their derivatives act as chemical messengers (Figure 5.5). For example, glycine, glutamate, γ-amino butyric acid (GABA, a derivative of glutamate), and serotonin and melatonin (derivatives of tryptophan) are **neurotransmitters**, substances released from one nerve cell that influence the function of a second nerve cell or a muscle cell. Thyroxine (a tyrosine derivative produced in the thyroid gland of animals) and indole acetic acid (a tryptophan derivative found in plants) are **hormones**—chemical signal molecules produced in one cell that regulate the function of other cells.
2. Amino acids are precursors of a variety of complex nitrogen-containing molecules. Examples include the nitrogenous base components of nucleotides and the nucleic acids, heme (the iron-containing organic group required for the biological activity of several important proteins), and chlorophyll (a pigment of critical importance in photosynthesis).
3. Several standard and nonstandard amino acids act as metabolic intermediates. For example, arginine (Figure 5.2), citrulline, and ornithine (Figure 5.6) are components of the urea cycle (Chapter 15). The synthesis of urea, a molecule formed in vertebrate livers, is the principal mechanism for the disposal of nitrogenous waste.

Modified Amino Acids in Proteins

Several proteins contain amino acid derivatives that are formed after a polypeptide chain has been synthesized. Among these modified amino acids is γ-carboxyglutamic acid (Figure 5.7), a calcium-binding amino acid residue found in the blood-clotting protein prothrombin. Both 4-hydroxyproline and 5-hydroxylysine are important structural components of collagen, the most abundant protein in connective tissue. Phosphorylation of the hydroxyl-containing amino acids serine, threonine, and tyrosine is often used to regulate the activity of proteins. For example, the synthesis of glycogen is significantly curtailed when the enzyme glycogen synthase is phosphorylated. Two other modified amino acids, selenocysteine and pyrolysine, are discussed in Chapter 19.

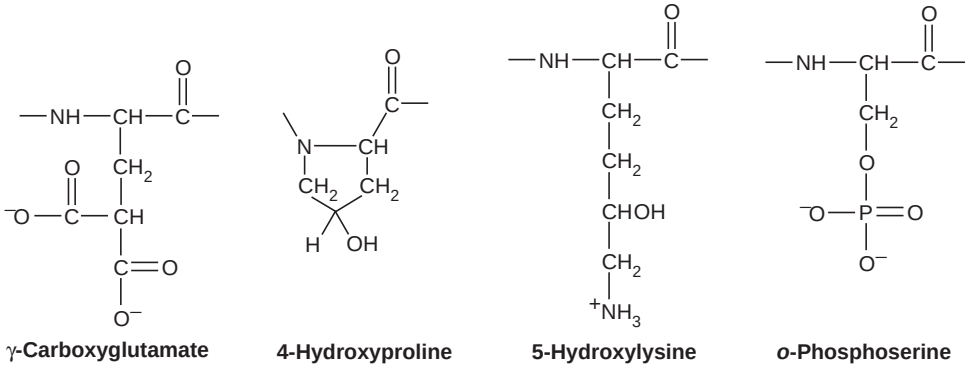


FIGURE 5.7
Some Modified Amino Acid Residues Found in Polypeptides

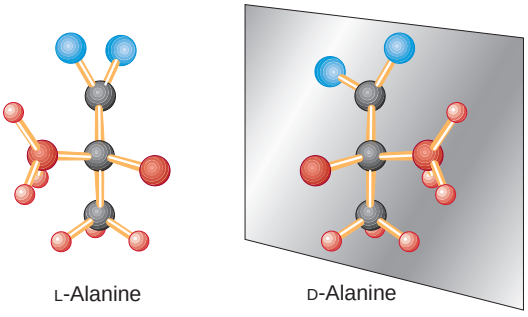


FIGURE 5.8

Two Enantiomers

L-Alanine and D-alanine are mirror images of each other. (Nitrogen = large red ball; Hydrogen = small red ball; Carbon = black ball; Oxygen = blue balls)

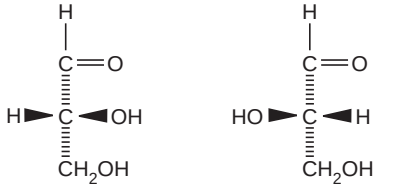
Amino Acid Stereoisomers

Because the α-carbons of 19 of the 20 standard amino acids are attached to four different groups (i.e., a hydrogen, a carboxyl group, an amino group, and an R group), they are referred to as **asymmetric**, or **chiral**, **carbons**. Glycine is a symmetrical molecule because its α-carbon is attached to two hydrogens. Molecules with chiral carbons can exist as **stereoisomers**, molecules that differ only in the spatial arrangement of their atoms. Three-dimensional representations of amino acid stereoisomers are illustrated in Figure 5.8. Notice in the figure that the atoms of the two isomers are bonded together in the same pattern except for the position of the ammonium group and the hydrogen atom. These two isomers are mirror images of each other. Such molecules, called **enantiomers**, cannot be superimposed on each other. Enantiomers have identical physical properties except that they rotate plane-polarized light in opposite directions. Plane-polarized light is produced by passing unpolarized light through a special filter; the light waves vibrate in only one plane. Molecules that possess this property are called **optical isomers**.

Glyceraldehyde is the reference compound for optical isomers (Figure 5.9). One glyceraldehyde isomer rotates the light beam in a clockwise direction and is said to be dextrorotatory (designated by +). The other glyceraldehyde isomer, referred to as levorotatory (designated by −), rotates the beam in the opposite direction to an equal degree. Optical isomers are often designated as D or L (e.g., D-glucose, L-alanine) to indicate the similarity of the arrangement of atoms around a molecule's asymmetric carbon to the asymmetric carbon in either of the glyceraldehyde isomers.

Most biomolecules have more than one chiral carbon. As a result, the letters D and L refer only to a molecule's structural relationship to either of the glyceraldehyde isomers, not to the direction in which it rotates plane-polarized light. Most asymmetric molecules found in living organisms occur in only one stereoisomeric form, either D or L. For example, with few exceptions, only L-amino acids are found in proteins.

Chirality has had a profound effect on the structural and functional properties of biomolecules. For example, the right-handed helices observed in proteins result from the exclusive presence of L-amino acids. Polypeptides synthesized in the laboratory from a mixture of both D- and L-amino acids do not form helices. In addition, because the enzymes are chiral molecules, most bind substrate (reactant) molecules in only one enantiomeric form. Proteases, enzymes that degrade proteins by hydrolyzing peptide bonds, cannot degrade artificial polypeptides composed of D-amino acids.



D-Glyceraldehyde **L-Glyceraldehyde**

FIGURE 5.9

D- and L-Glyceraldehyde

These molecules are mirror images of each other.

KEY CONCEPTS

- Molecules with an asymmetric or chiral carbon atom differ only in the spatial arrangement of the atoms attached to the carbon.
- The mirror-image forms of a molecule are called enantiomers.
- Most asymmetric molecules in living organisms occur in only one stereoisomeric form.

Certain bacterial species have outer layers composed of polymers made of D-amino acids. Immune system cells, whose task is to attack and destroy foreign cells, cannot destroy these bacteria. Suggest a reason for this phenomenon.

TABLE 5.2 pK_a Values for the Ionizing Groups of the Amino Acids

Amino Acid	pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R
Glycine	2.34	9.6	
Alanine	2.34	9.69	
Valine	2.32	9.62	
Leucine	2.36	9.6	
Isoleucine	2.36	9.6	
Serine	2.21	9.15	
Threonine	2.63	10.43	
Methionine	2.28	9.21	
Phenylalanine	1.83	9.13	
Tryptophan	2.83	9.39	
Asparagine	2.02	8.8	
Glutamine	2.17	9.13	
Proline	1.99	10.6	
Cysteine	1.71	10.78	8.33
Histidine	1.82	9.17	6.0
Aspartic acid	2.09	9.82	3.86
Glutamic acid	2.19	9.67	4.25
Tyrosine	2.2	9.11	10.07
Lysine	2.18	8.95	10.79
Arginine	2.17	9.04	12.48

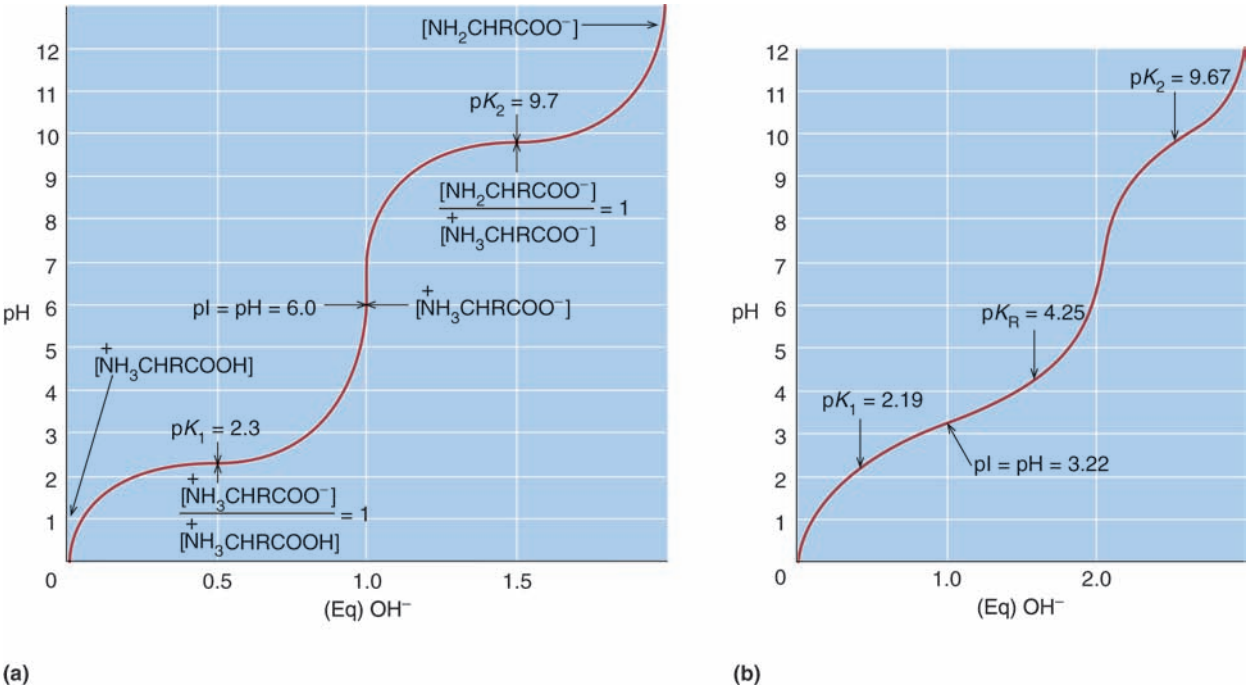


FIGURE 5.10

Titration of Two Amino Acids

(a) Alanine and (b) Glutamic Acids.

Titration of Amino Acids

Because amino acids contain ionizable groups (Table 5.2), the predominant ionic form of these molecules in solution depends on the pH. Titration of an amino acid illustrates the effect of pH on amino acid structure (Figure 5.10a). Titration is also a useful tool in determining the reactivity of amino acid side chains. Consider alanine, a simple amino acid, which has two titratable groups. During titration with a strong base such as NaOH, alanine loses two protons in stepwise fashion. In a strongly acidic solution (e.g., at pH 0), alanine is present mainly in the form in which the carboxyl group is uncharged. Under this circumstance the molecule's net charge is +1 because the ammonium group is protonated. If the H^+ concentration is lowered, the carboxyl group loses its proton to become a negatively charged carboxylate group. (In a polyprotic acid, the protons are first lost from the group with the lowest pK_a.) Once the carboxyl group has lost its proton, alanine has no net charge and is electrically neutral. The pH at which this occurs is called the **isoelectric point** (pI). The isoelectric point for alanine may be calculated as follows:

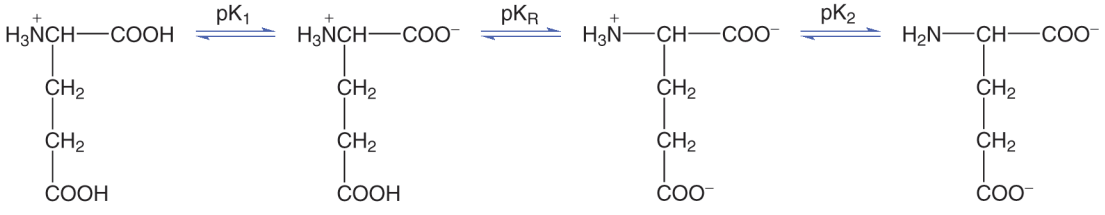
$$\text{pI} = \frac{\text{p}K_1 + \text{p}K_2}{2}$$

The pK_1 and pK_2 values for alanine are 2.34 and 9.7 respectively (see Table 5.2). The pI value for alanine is therefore

$$\text{pI} = \frac{2.34 + 9.67}{2} = 6.02$$

As the titration continues, the ammonium group loses its proton, leaving an uncharged amino group. The molecule then has a net negative charge because of the carboxylate group.

Amino acids with ionizable side chains have more complex titration curves. Glutamic acid, for example, has a carboxyl side chain group (Figure 5.10b). At low pH, glutamic acid has net charge +1. As base is added, the α-carboxyl group loses a proton to become a carboxylate group. Glutamate now has no net charge.



Titration of Glutamic acid

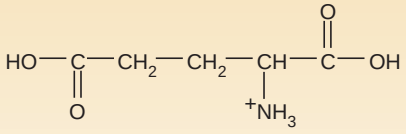
As more base is added, the second carboxyl group loses a proton, and the molecule has a –1 charge. Adding additional base results in the ammonium ion losing its proton. At this point, glutamate has a net charge of –2. The pI value for glutamate is the pH halfway between the pK_a values for the two carboxyl groups (i.e., the pK_a values that bracket the zwitterions):

$$\text{pI} = \frac{2.19 + 4.25}{2} = 3.22$$

The isoelectric point for histidine is the pH value halfway between the pK values for the two nitrogen-containing groups. Problem 5.1 is a sample titration problem.

WORKED PROBLEM 5.1

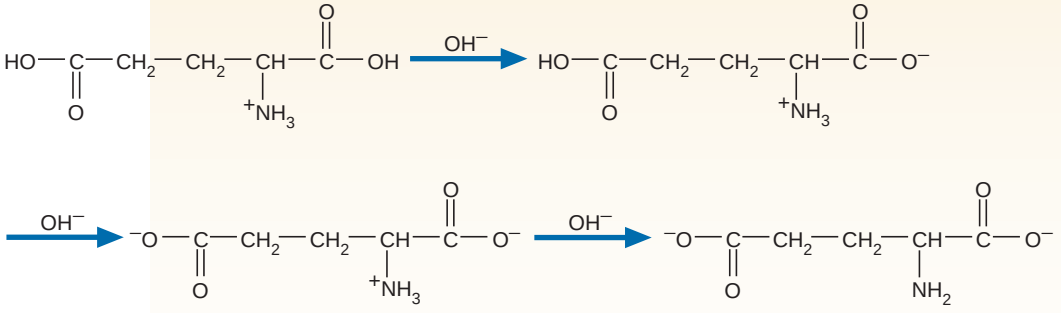
Consider the following amino acid and its pK_a values:



$pK_{a1} = 2.19$ $pK_{a2} = 9.67,$ $pK_{aR} = 4.25$

a. Draw the structure of the amino acid as the pH of the solution changes from highly acidic to strongly basic.

Solution (a)

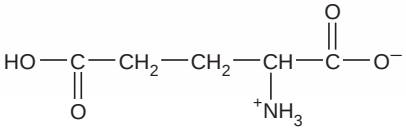


The ionizable hydrogens are lost in order of acidity, the most acidic ionizing first.

b. Which form of the amino acid is present at the isoelectric point?

Solution (b)

The form present at the isoelectric point is electrically neutral:



c. Calculate the isoelectric point.

Solution (c)

The isoelectric point is the average of the two pK_a's bracketing the isoelectric structure:

$$pI = \frac{pK_{a1} + pK_{aR}}{2} = \frac{2.19 + 4.25}{2} = 3.22$$

d. Sketch the titration curve for the amino acid.

Solution (d)

Plateaus appear at the pK_a and are centered about 0.5 equivalent (Eq), 1.5 Eq, and 2.5 Eq of base. There is a sharp rise at 1 Eq, 2 Eq, and 3 Eq. The isoelectric point is midway on the sharp rise between pK_{a1} and pK_{aR}.

e. In what direction does the amino acid move when placed in an electric field at the following pH values: 1, 3, 5, 7, 9, 12?

Solution (e)

At pH values below the pI, the amino acid is positively charged and moves to the cathode (negative electrode). At pH values above the pI, the amino acid is negatively charged and moves toward the anode (positive electrode). At the isoelectric point, the amino acid has no net charge and therefore does not move in the electric field. ■

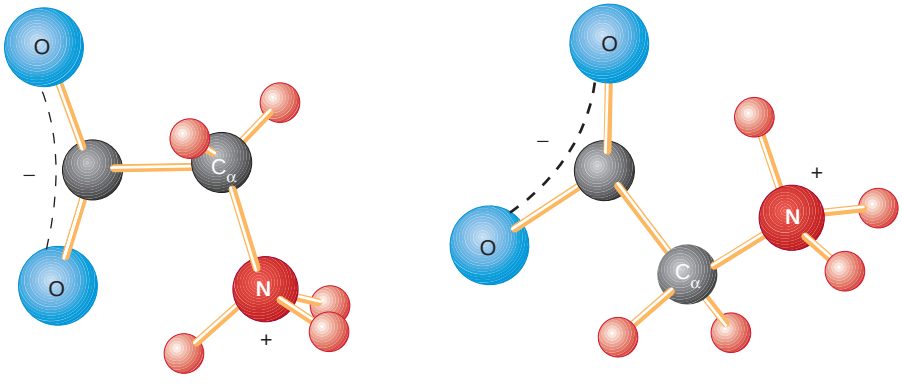
KEY CONCEPTS

- Titration is useful in determining the relative ionization potential of acidic and basic groups in an amino acid or peptide.
- The pH at which an amino acid has no net charge is called its isoelectric point.

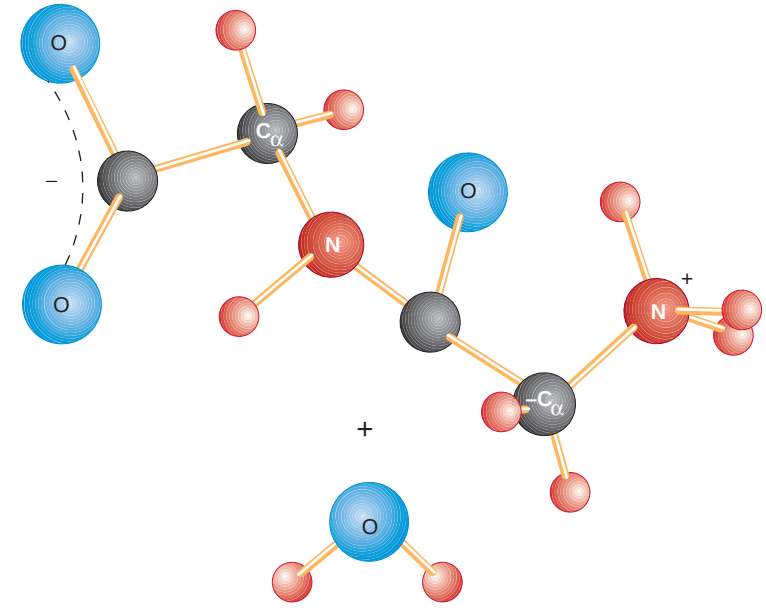
When amino acids are incorporated in polypeptides, the α-amino and α-carboxyl groups lose their charges. Consequently, except for the N- and C-terminal residues (amino acid residues at the beginning and end, respectively, of a polypeptide chain) all the ionizable groups of proteins are the side chain groups of seven amino acids: histidine, lysine, arginine, aspartate, glutamate, cysteine, and tyrosine. It should be noted that the pK_a values of these groups differ from those of free amino acids. The pK_a values of individual R groups are affected by their positions within protein microenvironments. For example, when the side chain groups of two aspartate residues are in close proximity, the pK_a of one of the carboxylate groups is raised. The significance of this phenomenon will become apparent in the discussion of enzyme catalytic mechanisms (Section 6.4).

Amino Acid Reactions

The functional groups of organic molecules determine which reactions they may undergo. Amino acids with their carboxyl groups, amino groups, and various R groups can undergo numerous chemical reactions. Peptide bond and disulfide bridge formation, however, are of special interest because of their effect on protein structure. Schiff base formation is another important reaction.



(a)



(b)

FIGURE 5.11
Formation of a Dipeptide

(a) A peptide bond forms when the α-carboxyl group of one amino acid reacts with the amino group of another. (b) A water molecule is formed in the reaction.

PEPTIDE BOND FORMATION Polypeptides are linear polymers composed of amino acids linked together by peptide bonds. **Peptide bonds** (Figure 5.11) are amide linkages formed when the unshared electron pair of the α-amino nitrogen atom of one amino acid attacks the α-carboxyl carbon of another in a nucleophilic acyl substitution reaction. A generalized acyl substitution reaction is shown:

Because peptide bond formation is a dehydration (i.e., a water molecule is removed), the linked amino acids are referred to as *amino acid residues*. When two amino acid molecules are linked, the product is called a dipeptide. For example, glycine and serine can form the dipeptides glycylserine or serylglycine. As amino acids are added and the chain lengthens, the prefix reflects the number of residues: a tripeptide contains three amino acid residues, a tetrapeptide four, and so on. By convention, the amino acid residue with the free amino group is called the *N-terminal* residue and is written to the left. The free carboxyl group on the *C-terminal* residue appears on the right. Peptides are named by using their amino acid sequences, beginning from their N-terminal residue. For example,

H_2N—Tyr—Ala—Cys—Gly—COOH

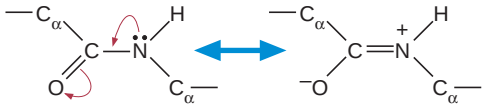
is a tetrapeptide named tyrosylalanylcysteinylglycine.

Large polypeptides have well-defined, three-dimensional structures. This structure, referred to as the molecule's native conformation, is a direct consequence of its *amino acid sequence* (the order in which the amino acids are linked together). Because all the linkages connecting the amino acid residues consist of single bonds, each polypeptide might be expected to undergo constant conformational changes caused by rotation around the single bonds. However, most polypeptides spontaneously fold into a single biologically active form. In the early 1950s, Linus Pauling (1901–1994, 1954 Nobel Prize in Chemistry) and his colleagues proposed an explanation. Using X-ray diffraction studies, they characterized the peptide bond as rigid and planar (flat) (Figure 5.12). Having discovered that the C—N bonds joining each two amino acids are shorter than other types of C—N bonds, Pauling deduced that peptide bonds have a partial double-bond character. (This indicates that peptide bonds are resonance hybrids.) The rigidity of the peptide bond has several consequences. Because fully one-third of the bonds in a polypeptide backbone chain cannot rotate freely, there are limits on the number of conformational possibilities.

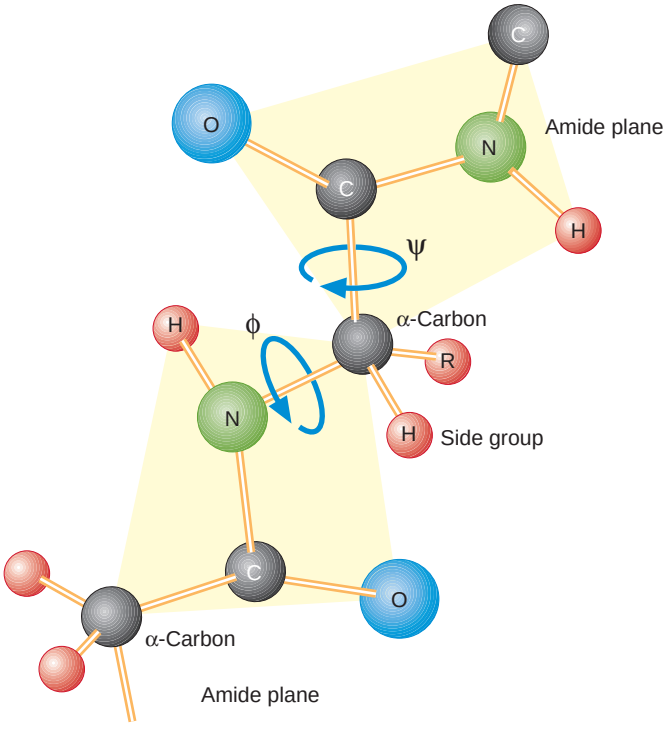
QUESTION 5.3

Considering only the 20 standard amino acids, calculate the total number of possible tetrapeptides.

CYSTEINE OXIDATION The sulfhydryl group of cysteine is highly reactive. The most common reaction of this group is a reversible oxidation that forms a disulfide. Oxidation of two molecules of cysteine forms cystine, a molecule that contains a disulfide bond (Figure 5.13). When two cysteine residues form such a bond, it is referred to as a **disulfide bridge**. This bond can occur in a single chain to form a ring or between two separate chains to form an intermolecular bridge. Disulfide bridges help stabilize many polypeptides and proteins.



(a)

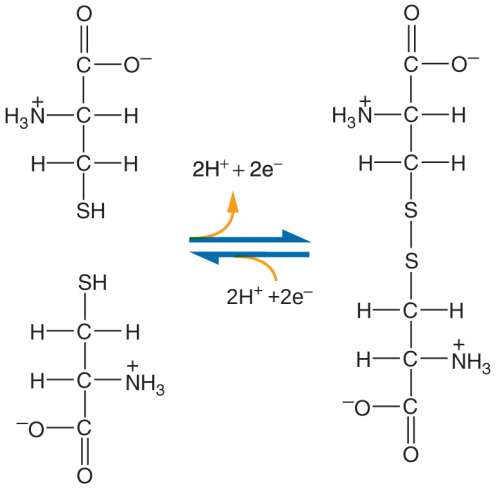


(b)

FIGURE 5.12

The Peptide Bond

(a) Resonance forms of the peptide bond. (b) Dimensions of a dipeptide. Because peptide bonds are rigid, the conformational degrees of freedom of a polypeptide chain are limited to rotations around the Cα—C and Cα—N bonds. The corresponding rotations are represented by ψ and ϕ, respectively.



Two cysteines Cystine

FIGURE 5.13

Oxidation of Two Cysteine Molecules to Form Cystine

The disulfide bond in a polypeptide is called a disulfide bridge.

QUESTION 5.4

In extracellular fluids such as blood (pH 7.2–7.4) and urine (pH 6.5), the sulfhydryl groups of cysteine (pK_a 8.1) are protonated and subject to oxidation to form cystine. In peptides and proteins the nucleophilic character of free protonated thiol groups are used to advantage in stabilizing protein structure and in thiol transfer reactions, but the free amino acid in tissue fluids can be problematic because of the low solubility of cystine. In a genetic disorder known as *cystinuria*, defective membrane transport of cystine results in excessive excretion of cystine into the urine. Crystallization of the amino acid results in formation of calculi (stones) in the kidney, ureter, or urinary bladder. The stones may cause pain, infection, and blood in the urine. Cystine concentration in the kidney is reduced by massively increasing fluid intake and administering D-penicillamine. It is believed that penicillamine (Figure 5.14) is effective because penicillamine–cysteine disulfide, which is substantially more soluble than cystine, is formed. What is the structure of the penicillamine–cysteine disulfide?

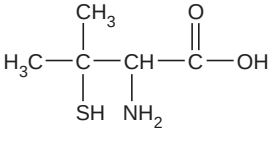


FIGURE 5.14

Structure of Penicillamine

KEY CONCEPTS

- Polypeptides are polymers composed of amino acids linked by peptide bonds. The order of the amino acids in a polypeptide is called the amino acid sequence.
- Disulfide bridges, formed by the oxidation of cysteine residues, are an important structural element in polypeptides and proteins.
- Schiff bases are imines that form when amine groups react reversibly with carbonyl groups.

SCHIFF BASE FORMATION Molecules such as amino acids that possess primary amine groups can reversibly react with carbonyl groups. The imine products of this reaction are often referred to as **Schiff bases**. In a nucleophilic addition reaction, an amine nitrogen attacks the electrophilic carbon of a carbonyl group to form an alkoxide product. The transfer of a proton from the amine group to the oxygen to form a carbinolamine, followed by the transfer of another proton from an acid catalyst, converts the oxygen into a good leaving group (OH_2^+). The subsequent elimination of a water molecule followed by loss of a proton from the nitrogen yields the imine product. The most important examples of Schiff base formation in biochemistry occur in amino acid metabolism. Schiff bases, referred to as **aldimines**, formed by the reversible reaction of an amino group with an aldehyde group, are *intermediates* (species formed during a reaction) in transamination reactions (pp. 508–510).

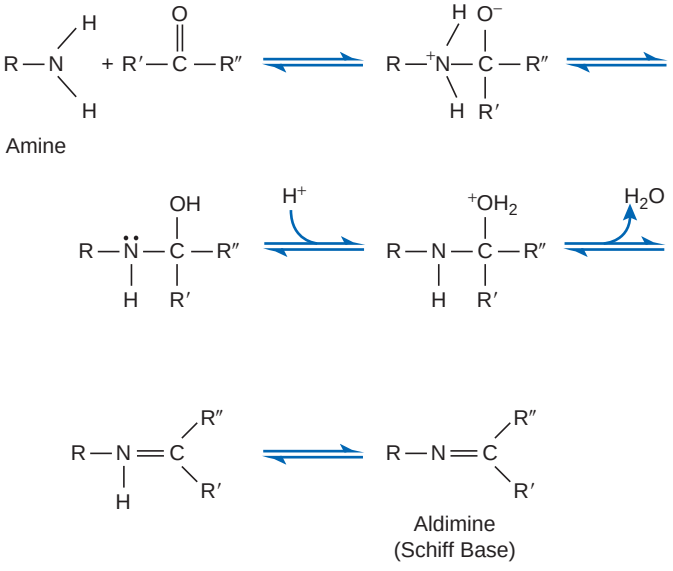
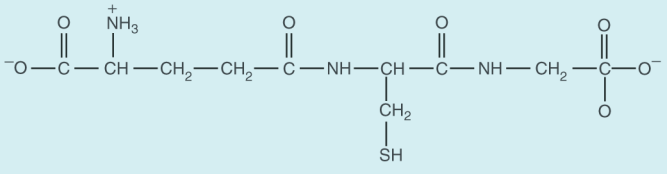


TABLE 5.3 Selected Biologically Important Peptides

Name	Amino Acid Sequence
Glutathione	
Oxytocin	Cys—Tyr—Ile—Gln—Asn—Cys—Pro—Leu—Gly—NH_2 (S—S)
Vasopressin	Cys—Tyr—Phe—Gln—Asn—Cys—Pro—Arg—Gly—NH_2 (S—S)
Atrial natriuretic factor	Ser^{1}—Leu—Arg—Arg—Ser—Ser—Cys—Phe—Gly—Gly^{10}—Arg—Met—Asp—Arg—Ile—Gly—Ala—Gln—Ser—Gly—Leu—Gly—Cys—Asn—Ser—Phe—Arg—Tyr^{28}

5.2 PEPTIDES

Although less structurally complex than the larger protein molecules, peptides have significant biological activities. The structure and function of several interesting examples, presented in Table 5.3, are now discussed.

The tripeptide glutathione (γ-glutamyl-L-cysteinylglycine) contains an unusual γ-amide bond. (Note that the γ-carboxyl group of the glutamic acid residue, not the α-carboxyl group, contributes to the peptide bond.) Found in almost all organisms, glutathione is involved in protein and DNA synthesis, drug and environmental toxin metabolism, amino acid transport, and other important biological processes. One group of glutathione's functions exploits its effectiveness as a reducing agent. Glutathione protects cells from the destructive effects of oxidation by reacting with substances such as peroxides (R–O–O–R), by-products of O_2 metabolism. For example, in red blood cells, hydrogen peroxide (H_2O_2) oxidizes the iron of hemoglobin to its ferric form (Fe^{3+}). Methemoglobin, the product of this reaction, is incapable of binding O_2. Glutathione protects against the formation of methemoglobin by reducing H_2O_2 in a reaction catalyzed by the enzyme glutathione peroxidase. In the oxidized product GSSG, two tripeptides are linked by a disulfide bond:

$$2\ GSH + H_2O_2 \rightarrow GSSG + 2H_2O$$

Because of the high GSH:GSSG ratio normally present in cells, glutathione is an important intracellular antioxidant. The abbreviation GSH is used because the reducing component of the molecule is the —SH group of the cysteine residue.

Peptides are one class of signal molecules that multicellular organisms use to regulate their complex activities. The dynamic interplay between opposing processes maintains a stable internal environment, a condition called *homeostasis*. Peptide molecules with opposing functions are now known to affect numerous processes (e.g., blood pressure regulation). The roles of selected peptides in each of these processes are briefly described.

Blood pressure, the force exerted by blood against the walls of blood vessels, is influenced by several factors such as blood volume and viscosity. Two peptides known to affect blood volume are vasopressin and atrial natriuretic factor. Vasopressin, also called antidiuretic hormone, contains nine amino acid residues. It is synthesized in the hypothalamus, a small structure in the brain that regulates a wide variety of functions including water balance, appetite, body temperature, and sleep. In response to low blood pressure or a high blood Na^+ concentration, osmoreceptors in the hypothalamus trigger vasopressin secretion. Vasopressin stimulates water reabsorption in the kidneys by initiating a signal transduction mechanism that inserts aquaporins (water channels) into kidney tubule membrane. Water then flows down its concentration gradient through the tubule cells and back into the blood. The structure of vasopressin is remarkably similar to that of another peptide produced in the hypothalamus called oxytocin, the signal molecule that stimulates the ejection of milk by mammary glands during lactation. Oxytocin produced in the uterus stimulates the contraction of uterine muscle during childbirth. Because vasopressin and oxytocin have similar structures, it is not surprising that the functions of the two molecules overlap. Oxytocin has mild antidiuretic activity and vasopressin has some oxytocin-like activity. Atrial natriuretic factor (ANF), a peptide produced by specialized cells in the heart in response to stretching and in the nervous system, stimulates the production of a dilute urine, an effect opposite to that of vasopressin. ANF exerts its effect, in part, by increasing the excretion of Na^+, a process that causes increased excretion of water, and by inhibiting the secretion of renin by the kidney. (Renin is an enzyme that catalyzes the formation of angiotensin, a hormone that constricts blood vessels.)

QUESTION 5.5

Write out the complete structure of oxytocin. What would be the net charge on this molecule at the average physiological pH of 7.3? At pH 4? At pH 9? Indicate which atoms in oxytocin can potentially form hydrogen bonds with water molecules.

QUESTION 5.6

The structural features of vasopressin that allow binding to vasopression receptors are the rigid hexapeptide ring and the amino acid residues at positions 3 (Phe) and 8 (Arg). The aromatic phenylalanine side chain, which fits into a hydrophobic pocket in the receptor, and the large positively charged arginine side chain are especially important structural features. Compare the structures of vasopressin and oxytocin and explain why their functions overlap. Can you suggest what will happen to the binding properties of vasopressin if the arginine at position 8 is replaced by lysine?

KEY CONCEPT

Although small in comparison to larger protein molecules, peptides have significant biological activity. They are involved in a variety of signal transduction processes.

5.3 PROTEINS

Of all the molecules encountered in living organisms, proteins have the most diverse functions, as the following list suggests.

1. **Catalysis**. *Enzymes* are proteins that direct and accelerate thousands of biochemical reactions in such processes as digestion, energy capture, and biosynthesis. These molecules have remarkable properties. For example, enzymes can increase reaction rates by factors of between 10^6 and 10^{12}. They can perform this feat under mild conditions of pH and temperature because they can induce or stabilize strained reaction intermediates. For example, ribulose bisphosphate carboxylase is an important enzyme in photosynthesis, and the protein complex nitrogenase is responsible for nitrogen fixation.
2. **Structure**. Some proteins provide structural support. Structural proteins often have very specialized properties. For example, collagen (the major components of connective tissues) and fibroin (silk protein) have significant mechanical strength. Elastin, the rubberlike protein found in elastic fibers, is found in several tissues in the body (e.g., blood vessels and skin) that must be elastic to function properly.
3. **Movement**. Proteins are involved in all cell movements. For example, actin, tubulin, and other proteins comprise the cytoskeleton. Cytoskeletal proteins are active in cell division, endocytosis, exocytosis, and the ameboid movement of white blood cells.
4. **Defense**. A wide variety of proteins are protective. In vertebrates, for example, keratin, a protein found in skin cells, aids in protecting the organism against mechanical and chemical injury. The blood-clotting proteins fibrinogen and thrombin prevent blood loss when blood vessels are damaged. The immunoglobulins (or antibodies) are produced by lymphocytes when foreign organisms such as bacteria invade an organism. Binding antibodies to an invading organism is the first step in its destruction. Many organisms protect themselves by producing toxic proteins that either kill or deter predators or competitors. Examples include the neurotoxin α-bungarotoxin produced by certain venomous snakes and ricin, a protein synthesis inhibitor found in the seeds of the castor bean plant.
5. **Regulation**. Binding a hormone molecule or a growth factor to cognate receptors on its target cell changes cellular function. For example, insulin and glucagon are peptide hormones that regulate blood glucose levels. Growth hormone stimulates cell growth and division. Growth factors are

COMPANION WEBSITE **Visit the companion website at www.oup.com/us/mckee to read the Biochemistry in Perspective box on protein poisons.**

polypeptides that control animal cell division and differentiation. Examples include platelet-derived growth factor (PDGF) and epidermal growth factor (EGF).

6. **Transport**. Many proteins function as carriers of molecules or ions across membranes or between cells. Examples of membrane proteins include the enzyme Na^+-K^+ ATPase and the glucose transporter. Other transport proteins include hemoglobin, which carries O_2 to the tissues from the lungs, and the lipoproteins LDL and HDL, which transport lipids from the liver and intestines to other organs. Transferrin and ceruloplasmin are serum proteins that transport iron and copper, respectively.
7. **Storage**. Certain proteins serve as a reservoir of essential nutrients. For example, ovalbumin in bird eggs and casein in mammalian milk are rich sources of organic nitrogen during development. Plant proteins such as zein perform a similar role in germinating seeds.
8. **Stress response**. The capacity of living organisms to survive a variety of abiotic stresses is mediated by certain proteins. Examples include cytochrome P_{450}, a diverse group of enzymes found in animals and plants that usually convert a variety of toxic organic contaminants into less toxic derivatives, and metallothionein, a cysteine-rich intracellular protein found in virtually all mammalian cells that binds to and sequesters toxic metals such as cadmium, mercury, and silver. Excessively high temperatures and other stresses result in the synthesis of a class of proteins called the **heat shock proteins** (hsps) that promote the correct refolding of damaged proteins. If such proteins are severely damaged, hsps promote their degradation. (Certain hsps function in the normal process of protein folding.) Cells are protected from radiation by DNA repair enzymes.

Protein research efforts in recent years have revealed that numerous proteins have multiple and often unrelated functions. Once thought to be a rare phenomenon, **multifunction proteins** (sometimes referred to as *moonlighting proteins*) are a diverse class of molecules. Prominent examples include glyceraldehyde-3-phosphate dehydrogenase (GAPD) and the crystallins. As the name suggests, GAPD (p. 273) is an enzyme that catalyzes the oxidation of glyceraldehyde-3-phosphate, an intermediate in glucose catabolism. The GAPD protein is now known to have roles in such diverse processes as DNA replication and repair, endocytosis, and membrane fusion events. The crystallins are the major water-soluble protein components of the transparent cells in the lens of the vertebrate eye. To perform their structural function, crystallin proteins must be soluble in the liquid-crystal cytoplasm and exceptionally stable. Most importantly, they must prevent the scattering of visible light, often by binding to small numbers of *chromophores*, substances that absorb light energy. Crystallins appear to have been genetically "recruited" from a variety of metabolic enzymes and molecular chaperones. They have retained some of their original functions, although often with some loss of activity and/or specificity. For example, the α-crystallins are heat shock proteins that protect cells from physiological stress. Another crystallin (ε-crystallin), found in the lens of ducks, is a lactate dehydrogenase (p. 277). η-Crystallin, found in elephant shrews, is also a retinal dehydrogenase, an essential enzyme in the metabolism of retinol (vitamin A).

In addition to their functional classifications, proteins are categorized on the basis of amino acid sequence similarities and overall three-dimensional shape. **Protein families** are composed of protein molecules that are related by amino acid sequence similarity. Such proteins share an obvious common ancestry. Classic protein families include the hemoglobins (blood oxygen transport proteins, pp. 168–171) and the immunoglobulins, the antibody proteins produced by the immune system in response to antigens (foreign substances). Proteins more distantly related are often classified into **superfamilies**. For example, the globin superfamily includes a variety of heme-containing proteins that serve in the

binding and/or transport of oxygen. In addition to the hemoglobins and myoglobins (oxygen-binding proteins in muscle cells), the globin superfamily includes neuroglobin and cytoglobin (oxygen-binding proteins in brain and other tissues, respectively) and the leghemoglobins (oxygen-sequestering proteins in the root nodules of leguminous plants).

Because of their diversity, proteins are often classified in two additional ways: shape and composition. Proteins are classified into two major groups based on their shape. As the name suggests, **fibrous proteins** are long, rod-shaped molecules that are insoluble in water and physically tough. Fibrous proteins, such as the keratins found in skin, hair, and nails, have structural and protective functions. **Globular proteins** are compact spherical molecules that are usually water-soluble. Typically, globular proteins have dynamic functions. For example, nearly all enzymes have globular structures. Other examples include the immunoglobulins and the transport proteins hemoglobin and albumin (a carrier of fatty acids in blood).

On the basis of composition, proteins are classified as simple or conjugated. Simple proteins, such as serum albumin and keratin, contain only amino acids. In contrast, each **conjugated protein** consists of a simple protein combined with a nonprotein component. The nonprotein component is called a **prosthetic group**. (A protein without its prosthetic group is called an **apoprotein**. A protein molecule combined with its prosthetic group is referred to as a **holoprotein**.) Prosthetic groups typically play an important, even crucial, role in the function of proteins. Conjugated proteins are classified according to the nature of their prosthetic groups. For example, **glycoproteins** contain a carbohydrate component, **lipoproteins** contain lipid molecules, and **metalloproteins** contain metal ions. Similarly, **phosphoproteins** contain phosphate groups, and **hemoproteins** possess heme groups (p. 168).

Protein Structure

Proteins are extraordinarily complex molecules. Complete models depicting even the smallest of the polypeptide chains are almost impossible to comprehend. Simpler images that highlight specific features of a molecule are useful. Two methods of conveying structural information about proteins are presented in Figure 5.15. Another structural representation, referred to as a ball-and-stick model, is presented later (Figures 5.37 and 5.39).

Biochemists have distinguished several levels of the structural organization of proteins. **Primary structure**, the amino acid sequence, is specified by genetic information. As the polypeptide chain folds, it forms certain localized arrangements of adjacent (but not necessarily contiguous) amino acids that constitute **secondary structure**. The overall three-dimensional shape that a polypeptide assumes is called the **tertiary structure**. Proteins that consist of two or more polypeptide chains (or subunits) are said to have a **quaternary structure**.

PRIMARY STRUCTURE Every polypeptide has a specific amino acid sequence. The interactions between amino acid residues determine the protein's three-dimensional structure and its functional role and relationship to other proteins. Polypeptides that have similar amino acid sequences and have arisen from the same ancestral gene are said to be **homologous**. Sequence comparisons among homologous polypeptides have been used to trace the genetic relationships of different species. For example, the sequence homologies of the mitochondrial redox protein cytochrome c have been used extensively in the study of evolution of species. Sequence comparisons of cytochrome c, an essential molecule in energy production, among numerous species reveal a significant amount of sequence conservation. The amino acid residues that are identical in all homologues of a protein, referred to as *invariant*, are presumed to be essential for the protein's function. (In cytochrome c the invariant residues interact with heme, a prosthetic group, or certain other proteins involved in energy generation.)

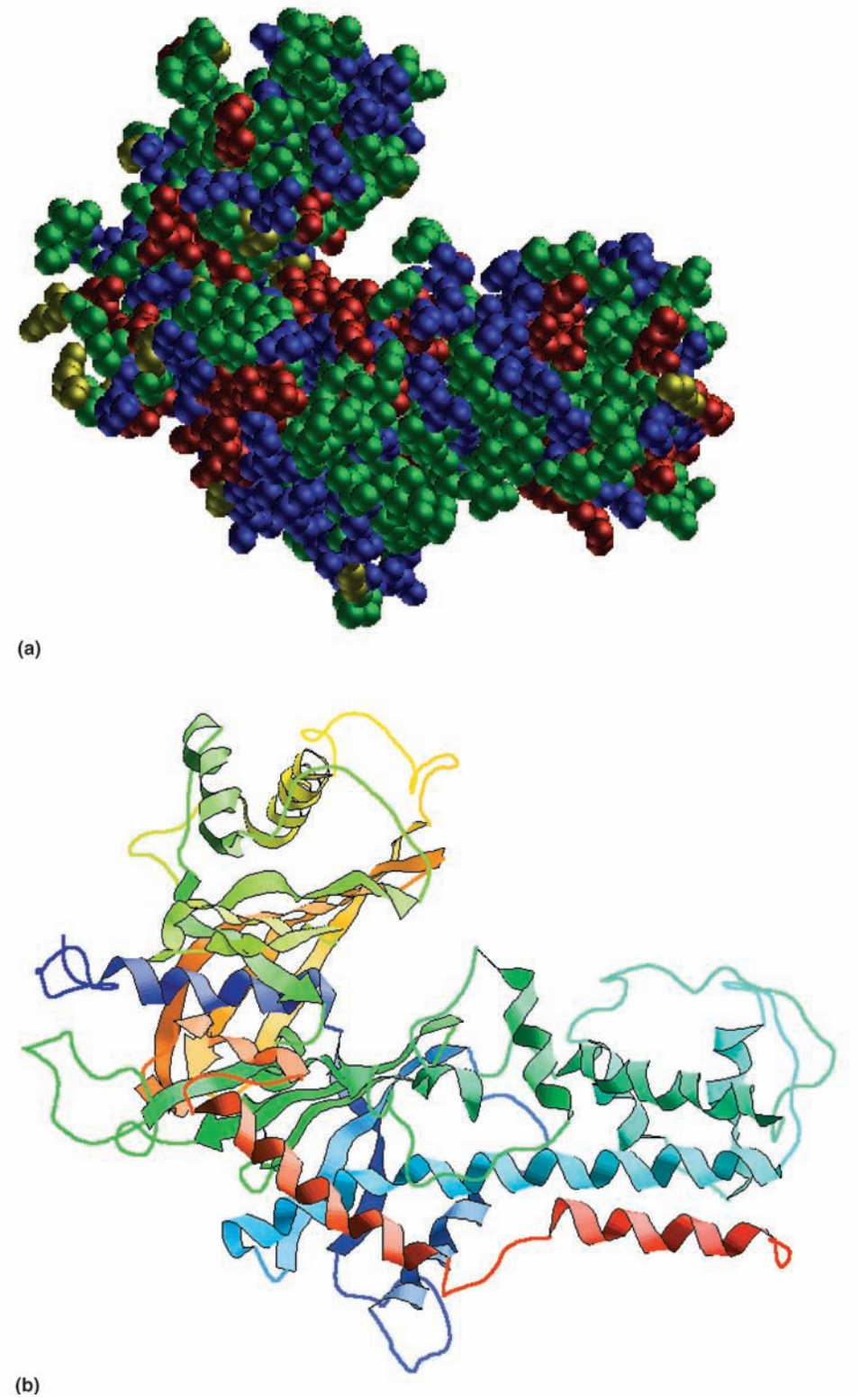


FIGURE 5.15

The Enzyme Adenylate Kinase

(a) This space-filling model illustrates the volume occupied by molecular components and overall shape. (b) In a ribbon model β-pleated segments are represented by flat arrows. The α-helices appear as spiral ribbons.