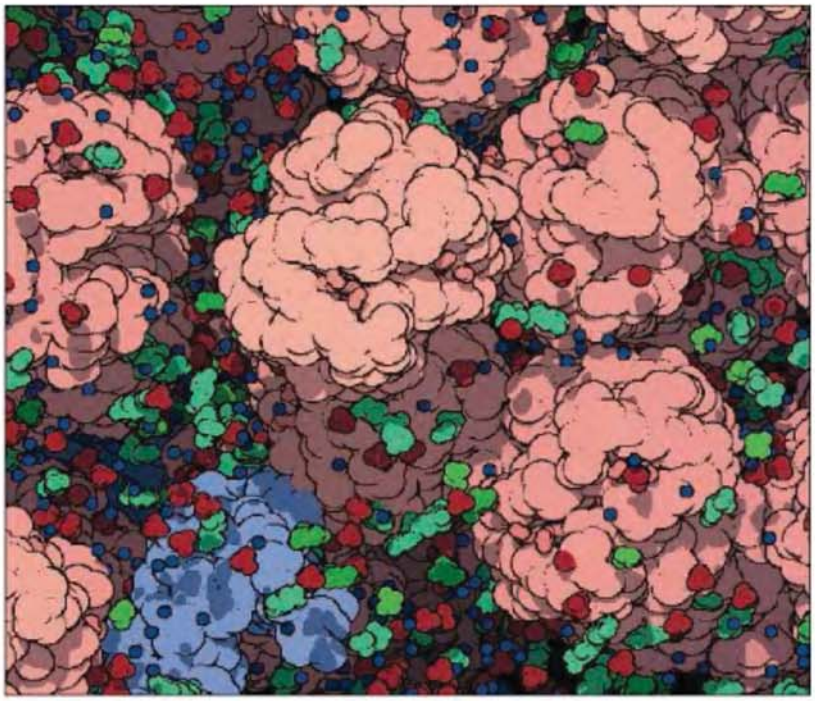
CHAPTER FIVE

Amino Acids, Peptides, and Proteins

OUTLINE



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 ions are red. The large blue molecule is an enzyme.

Proteins are essential constituents of all organisms. Most tasks performed by living cells require proteins. The variety of functions that they perform is astonishing. In animals, for example, proteins are the primary structural components of muscle, connective tissue, feathers, nails, and hair. In addition to serving as structural materials in all living organisms, 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.

As crucial as an uninterrupted flow of energy is to living systems, it is insufficient to maintain the organized complexity of life. Also required is a continuous flow of staggering amounts of timely, precise, and accurate information. Information is a measure of order and is often referred to as *negative entropy*. In general terms, information specifies the instructions required to create a specific organization. In living organisms, information is inherent in the three-dimensional atomic configuration of biomolecules. The information in genes is the instructions for making the proteins and ribonuclear proteins required to sustain life. The proteins and ribonuclear proteins constitute the machinery and structure of the cell. Proteins themselves are informational, each with a unique shape (Figure 5.1), allowing selective interactions with only one or a few other molecules. The enzyme glucokinase will only accept glucose as a substrate, whereas hexokinase will accept glucose, galactose, or mannose despite the fact that the two enzymes catalyze the same reaction in the same way. Glucokinase has high specificity; hexokinase has low specificity. In general, the larger the protein, the greater the potential for multifunctional capacities, so that an enzyme might bind modulating ligands in addition to the substrate.

Proteins may be composed of as many as 20 different amino acids. In each protein the precise types and amounts of each amino acid are covalently linked in the

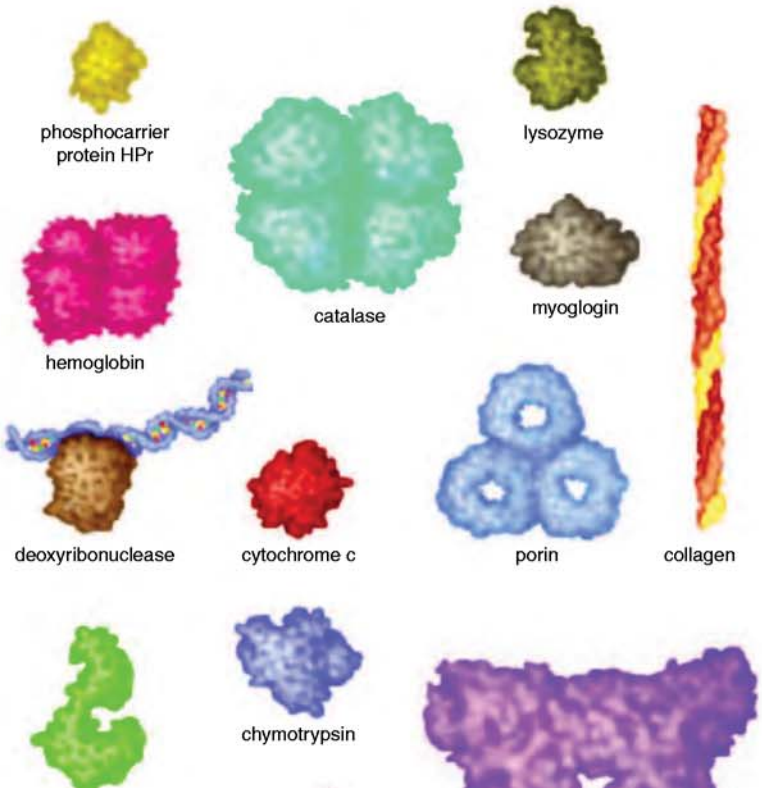


FIGURE 5.1

Protein Diversity.

Proteins occur in an enormous diversity of sizes and shapes.

linear sequence specified by the base sequence of the DNA-generated mRNA for that protein. The ability of each type of the tens of thousands of different proteins to perform its functions is specified by its unique amino acid sequence. During synthesis each polypeptide molecule bends in three-dimensional space as its amino acid components (called **amino acid residues**) interact with each other, largely through noncovalent interactions. The subsequent folding of a protein molecule into its own unique, complex, three-dimensional, and biologically active structure is a process dictated by information inherent in the structures of its amino acids.

Amino acid polymers are often 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 (D) 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. 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 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. Throughout this textbook the terms *peptide* and *protein* will be used as defined above. The term *polypeptide* will be used whenever the topic under discussion applies to both peptides and proteins.

Polypeptides may be broken into their constituent monomer molecules by hydrolysis. The amino acid products of the reaction constitute the *amino acid composition* of the polypeptide.

This chapter begins with a review of the structures and chemical properties of the amino acids. This is followed by descriptions of the structural features of peptides and proteins. The chapter ends with an examination of the structural and functional properties of several well-researched proteins. The emphasis throughout the chapter 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 and the folding process are covered in Chapter 19.

5.1 AMINO ACIDS

The structures of the 20 amino acids that are commonly found in proteins are shown in Figure 5.2. 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 they have been incorporated 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. Neutral molecules that bear an equal number of positive and negative charges simultaneously are called **zwitterions**. The R group, however, gives each amino acid its unique properties.

KEY CONCEPTS 5.1

Each protein is composed of building blocks called amino acids. Amino acids are amphoteric molecules; that is, they can act as either an acid or a base. In addition to their primary function as components of proteins, amino acids have several important biological roles.

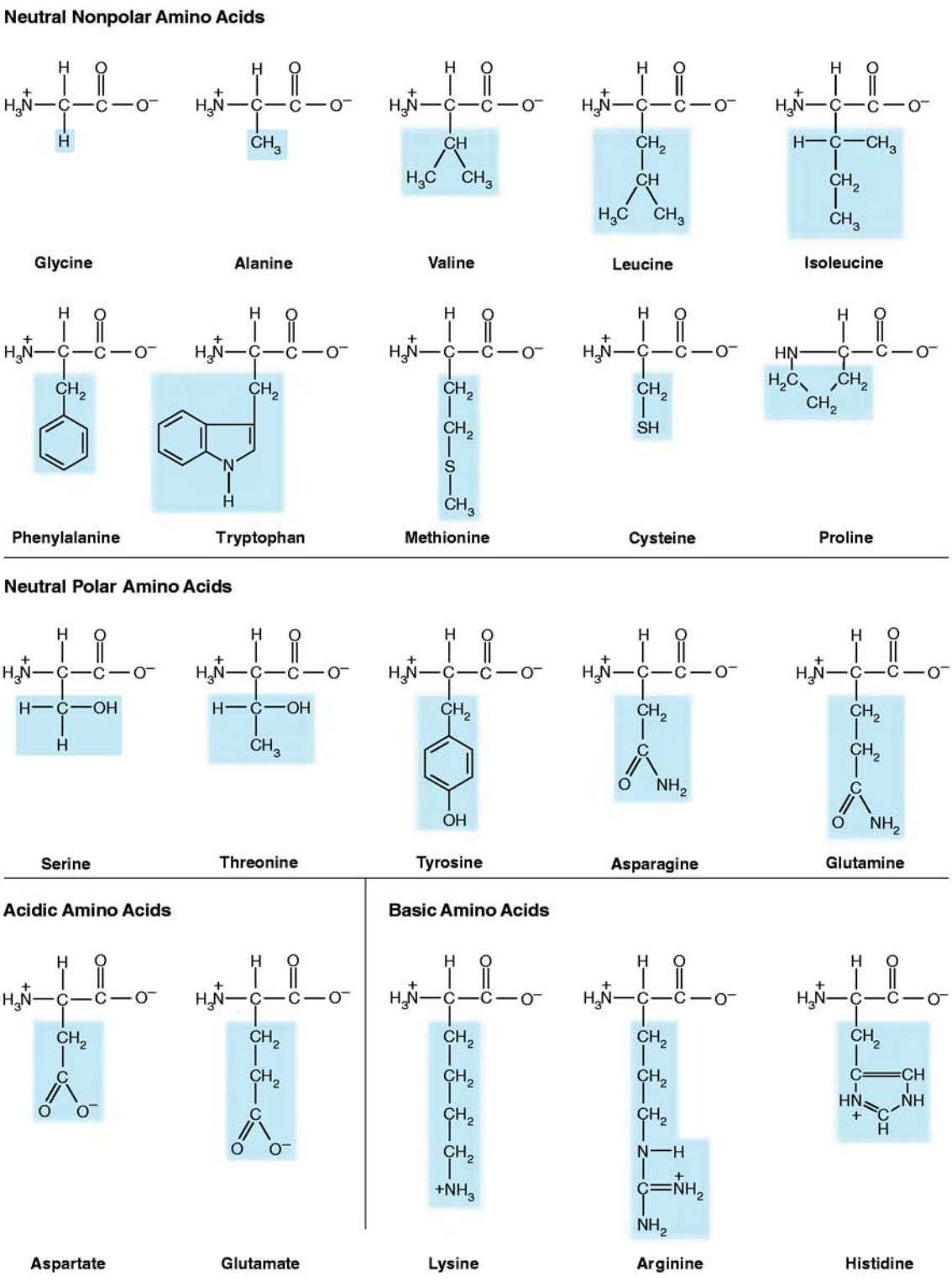


FIGURE 5.2

The Standard Amino Acids.

The side chain is indicated by the shaded box.

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
Gycine	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

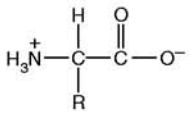


FIGURE 5.3
General Structure of the α-Amino Acids.

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 and neutral, (2) polar and neutral, (3) acidic, and (4) basic.

NEUTRAL NONPOLAR AMINO ACIDS The neutral nonpolar amino acids contain mostly hydrocarbon R groups. The term *neutral* is used because these R groups do not bear positive or negative charges. Because they interact poorly with water, nonpolar (i.e., hydrophobic) amino acids play an important role maintaining the three-dimensional structure of proteins. Two types of hydrocarbon side chains are found in this group: aromatic and aliphatic. (Recall that **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. In methionine the nonbonding electrons of the sulfur atom can form bonds with electrophiles such as metal ions. Although the sulfhydryl (—SH) group of cysteine is nonpolar, it can form weak hydrogen bonds with oxygen and nitrogen. Sulfhydryl groups, which are highly reactive, are important components of many enzymes. Additionally, the sulfhydryl groups of two cysteine molecules may oxidize spontaneously to form a disulfide compound called cystine. (See p. 121 for a discussion of this reaction.)

FIGURE 5.4
Benzene.

NEUTRAL 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.

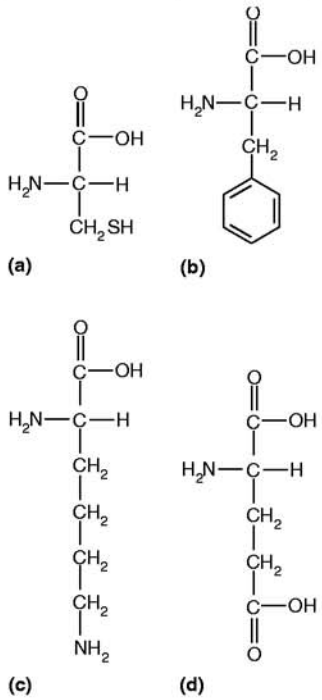
BASIC AMINO ACIDS Basic amino acids bear a positive charge at physiological pH. They can therefore form ionic bonds with acidic amino acids. Lysine, which has a side chain amino group, accepts a proton from water to form the conjugate acid ($-NH_3^+$). When lysine's side chain in proteins such as collagen is oxidized, strong intramolecular and intermolecular cross-linkages are formed. Because the guanidino group of arginine has a pK_a range of 11.5–12.5 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 CONCEPTS 5.2

Amino acids are classified according to their capacity to interact with water. By using this criterion, four classes may be distinguished: nonpolar, polar, acidic, and basic.

QUESTION 5.1

Shown are the structures of several standard amino acids. Classify them according to whether they are neutral nonpolar, neutral polar, acidic, or basic.

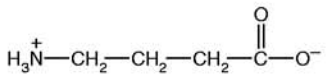


Biologically Active Amino Acids

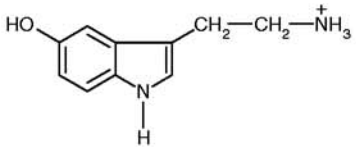
In addition to their primary function as components of protein, amino acids have several other biological roles.

1. Several α-amino acids or their derivatives act as chemical messengers (Figure 5.5). For example, glycine, 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 two examples of hormones. **Hormones** are chemical signal molecules produced in one cell that regulate the function of other cells.

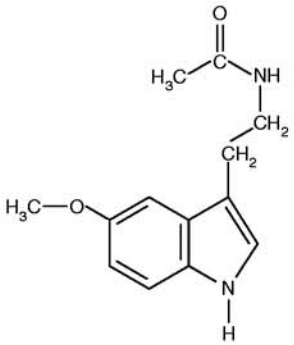
FIGURE 5.5
Some Derivatives of Amino Acids.



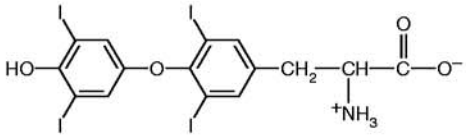
GABA



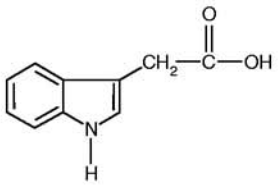
Serotonin



Melatonin



Thyroxine



Indole acetic acid

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, 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.

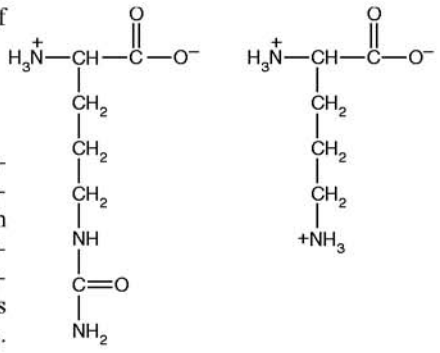


FIGURE 5.6
Citrulline and Ornithine.

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. The physical properties of enantiomers are identical except that they rotate plane-polarized light in opposite directions. (In plane-polarized light, 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**.

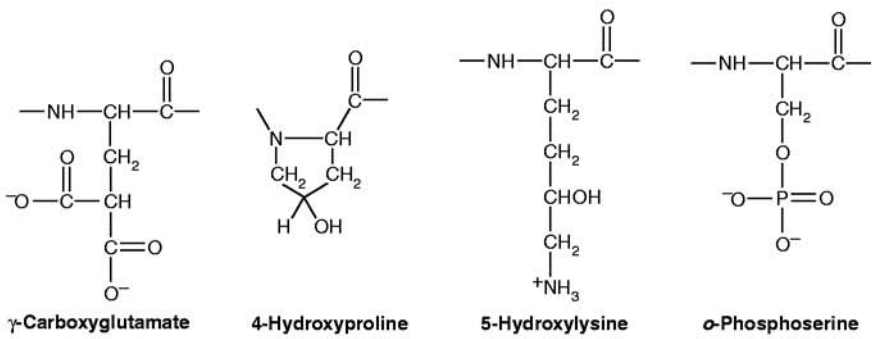
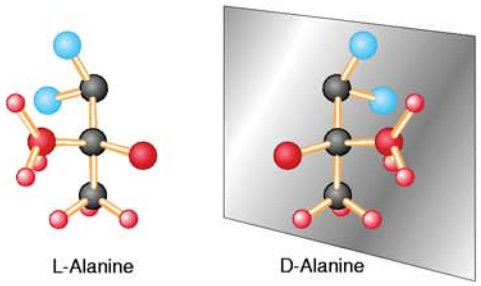


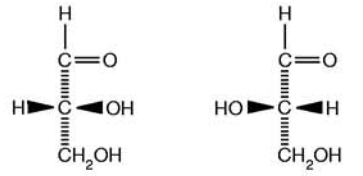
FIGURE 5.7
Some Modified Amino Acid Residues Found in Polypeptides.

FIGURE 5.8

Two Enantiomers.

L-Alanine and D-alanine are mirror images of each other.





D-Glyceraldehyde L-Glyceraldehyde

FIGURE 5.9

D- and L-Glyceraldehyde.

These molecules are mirror images of each other.

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 dextrorotary (designated by +). The other glyceraldehyde isomer, referred to as levorotary (designated by –), rotates the beam in the opposite direction to an equal degree. Optical isomers are often designated as D or L; for example, D-glucose and L-alanine. The D or L indicates the similarity of the arrangement of atoms around a molecule's asymmetric carbon to the asymmetric carbon in either of the glyceraldehyde isomers.

Because many biomolecules have more than one chiral carbon, 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 both D- and L- amino acids do not form helices. In addition, because the enzymes are chiral molecules, they only bind substrate molecules in one enantiomeric form. Proteases, enzymes that degrade proteins by hydrolyzing peptide bonds, cannot degrade artificial polypeptides composed of D-amino acids.

KEY CONCEPTS 5.3

Molecules that differ only in the spatial arrangement of some of their atoms are called stereoisomers. Stereoisomers with an asymmetric carbon atom have two nonsuperimposable mirror-image forms called enantiomers. Most asymmetric molecules in living organisms have only one stereoisomeric form.

QUESTION 5.2

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.

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 a 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. Lowering of the H^+ concentration results in the carboxyl group losing 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.) At this point, alanine has no net charge and is electrically neutral. The pH at which this occurs is called the **isoelectric point**

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

Amino Acid	pK_1	pK_2	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

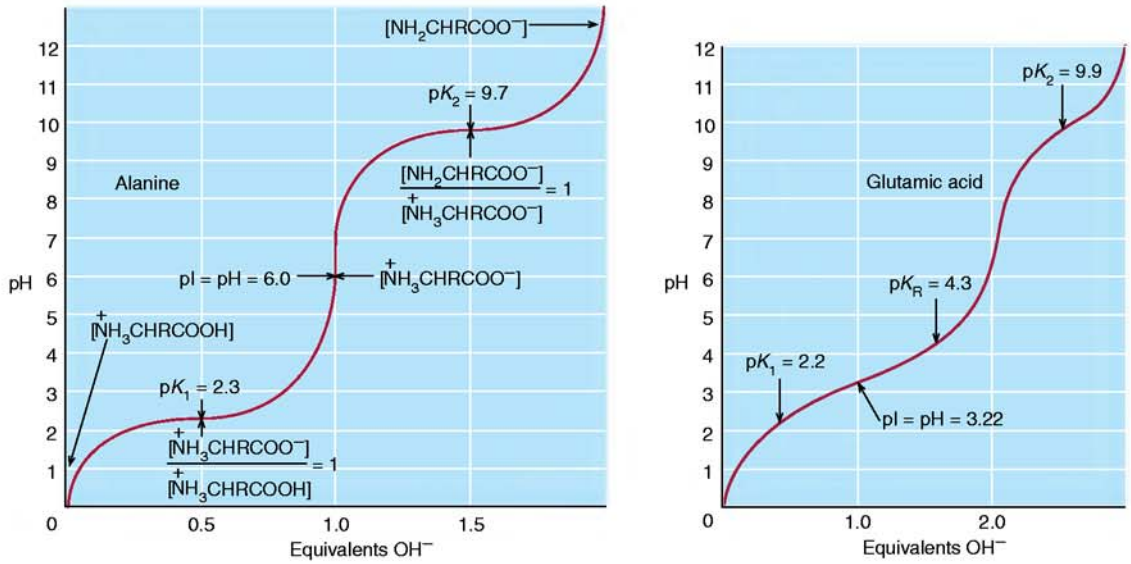


FIGURE 5.10
Titration of (a) Alanine and (b) Glutamic Acid.

(pI). Because there is no net charge at the isoelectric point, amino acids are least soluble at this pH. (Zwitterions crystallize relatively easily.) 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

$$pI = \frac{2.34 + 9.7}{2} = 6.0$$

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 (Figure 5.10b). Glutamic acid, for example, has a carboxyl side chain group. 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. 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:

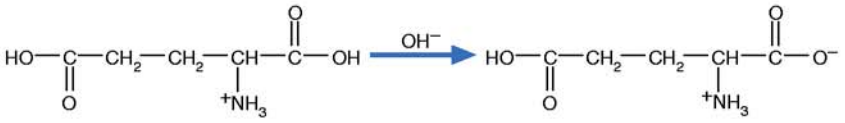
$$pI = \frac{2.19 + 4.25}{2} = 3.22$$

The isoelectric point for histidine is the pH value halfway between the pK values for the two nitrogen-containing groups. The pK_a and pI values of amino acids in peptides and proteins differ somewhat from those of free amino acids, principally because most of the α-amino and α-carboxyl groups are not ionized but are covalently joined in peptide bonds.

Problems 5.1 and 5.2 are sample titration problems, given with their solutions.

PROBLEM 5.1

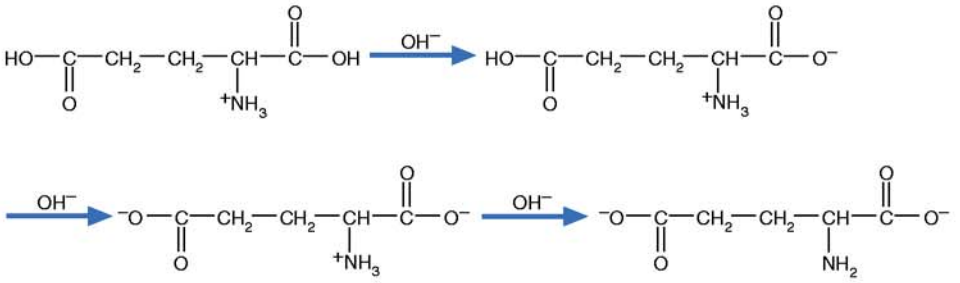
Consider the following amino acid and its pK_a values:



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

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

Solution

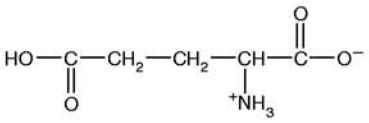


The ionizable hydrogens are lost in order of acidity, the most acidic ionizing first.

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

Solution

The form present at the isoelectric point is electrically neutral:



c. Calculate the isoelectric point.

Solution

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

Plateaus appear at the pK_a's and are centered about 0.5 equivalents (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

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.

PROBLEM 5.2

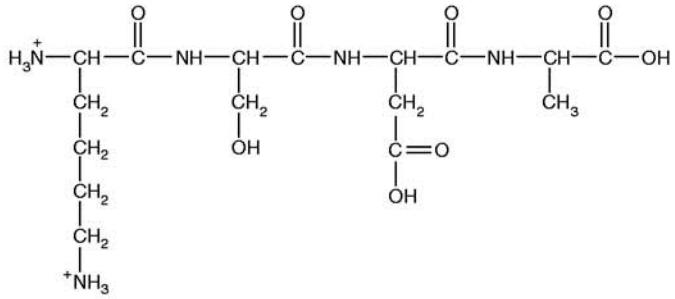
Consider the following tetrapeptide:

Lys—Ser—Asp—Ala

a. Determine the pI for the peptide.

Solution

The structure of the tetrapeptide in its most acidic form is shown below.



Refer to Table 5.2 for the pK_a values for lysine and aspartic acid, both of which have ionizable side chains. Lysine also contains a terminal α-amino and alanine a terminal α-carboxyl group. These values are as follows:

Lysine: α-amino = 8.95, amino side chain = 10.79
Aspartic acid: carboxyl side chain = 3.86
Alanine: α-carboxyl = 2.34

(These values are approximations, because the behavior of amino acids is affected by the presence of other groups.) The electrically neutral peptide is formed after

KEY CONCEPTS 5.4

Titration is useful to determine the reactivity of amino acid side chains. The pH at which an amino acid has no net charge is called its isoelectric point.

both carboxyl groups have lost their protons but before either ammonium group has lost any protons. The isoelectric point is calculated as follows:

$$\text{pI} = \frac{3.86 + 8.95}{2} = \frac{12.81}{2} = 6.4$$

b. In what direction does the peptide move when placed in an electric field at the following pHs: 4 and 9?

Solution

At pH = 4 the peptide is positively charged and moves toward the negative electrode (cathode). At pH = 9 the peptide is negatively charged and will move toward the positive electrode (anode).

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. However, two reactions (i.e, peptide bond and disulfide bridge formation) are of special interest because of their effect on protein structure.

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 this reaction 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. For example, 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,

$$H_2N\text{—Tyr—Ala—Cys—Gly—COOH}$$

is a tetrapeptide named tyrosylalanylcysteinylglycine.

QUESTION 5.3

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

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, it might be expected that each polypeptide undergoes constant conformational changes caused by rotation around the single bonds. However, most polypeptides spontaneously fold into a single biologically active form. In the early