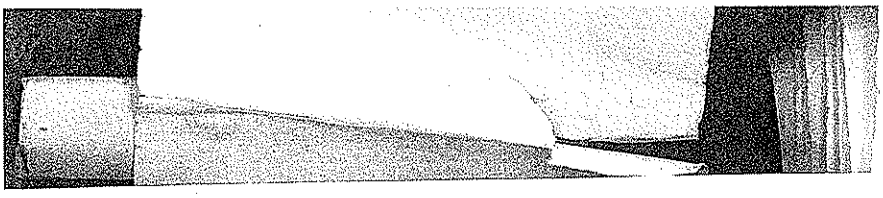
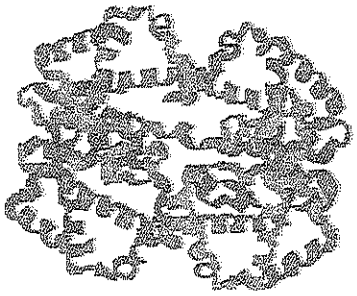




CHAPTER 16
Proteins

By far the most versatile biomolecules in living organisms are the amino acids which make up proteins. Proteins act as catalysts, structural support, protection, transport agents, chemical messengers, and cell recognition factors, to name only a few functions. As with other biopolymers, the organic structure and bonding in proteins give rise to their extraordinary features.

protein
polymer of amino acids

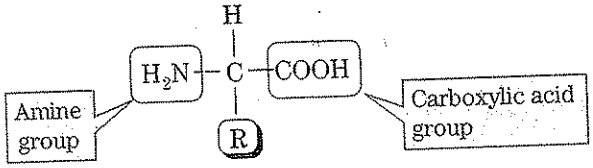
Proteins are polymers composed of monomer units known as *amino acids*. These amino acids are linked by amide bonds in macromolecules with molecular weights ranging from a few thousand to several million atomic mass units. The properties of proteins can be appreciated by considering the characteristics of their constituent amino acids.

16.1 Structure of Amino Acids

A. Fundamental Structure—An Amine and An Acid

amino acids
the monomer units of proteins

As the term *amino acid* suggests, every **amino acid** has an amine group and a carboxylic acid group. Both of these functional groups are attached to the same carbon atom, which usually also has a hydrogen atom and another variable group.



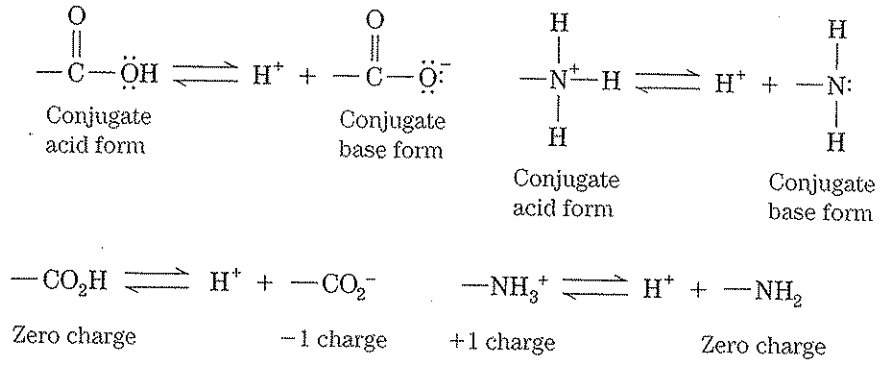
α- (alpha) amino acid
molecule with an amine group on the carbon adjacent to a carboxyl group

These monomers are sometimes referred to as **α- (alpha) amino acids** because the amine is on the carbon next to, or alpha to, the carboxylic acid group or vice versa.

B. Ionization of Amino Acids

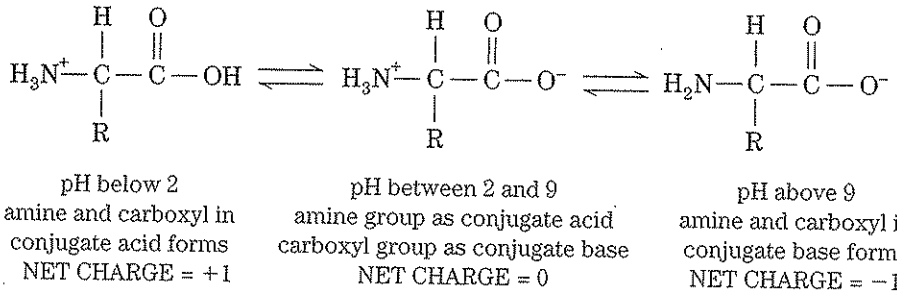
Recall that amine and carboxylic acid groups have conjugate acid–base forms in water that are dependent upon the pH of the solution in which they find themselves.

The ionization constants, K_as, for these groups are about 10^{-2} for the carboxyl and about 10^{-9} for the amine group. Therefore, the pK_as are 2 and 9, respectively. This means that at a pH of 2, 50% of the carboxyl groups are in the conjugate acid form and 50% are in the conjugate base form. When the pH is less than 2,



most of the carboxyls are in the uncharged conjugate acid form; above a pH of 2 most are in the −1 charged conjugate base form.

Overall then, an amino acid has several charged forms that are pH-dependent.



A titration curve for an amino acid shows at least two points of inflection accounting for the titration of the two ionizable groups in the molecule. This is seen in Figure 16.1.

Because amino acids are charged at certain pHs, they move if an electric field, with + and − electrical poles, is applied to the solution. The cationic (+1) form moves to the − pole or cathode, and the anionic (−) form migrates to the + pole or anode. The form with no net charge does not move at all.

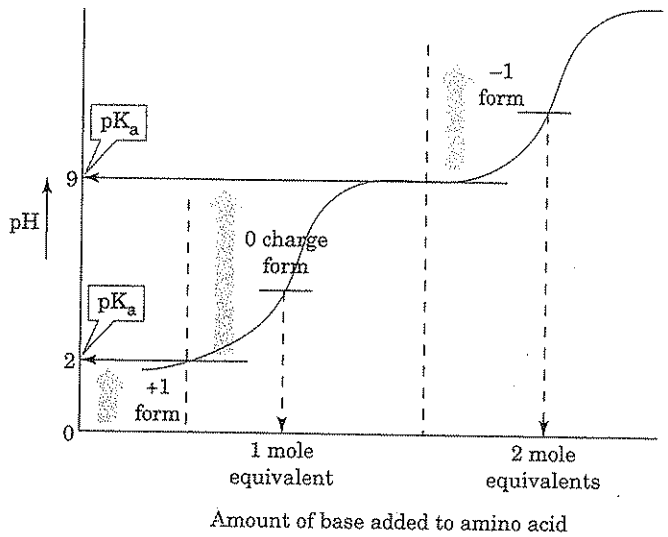
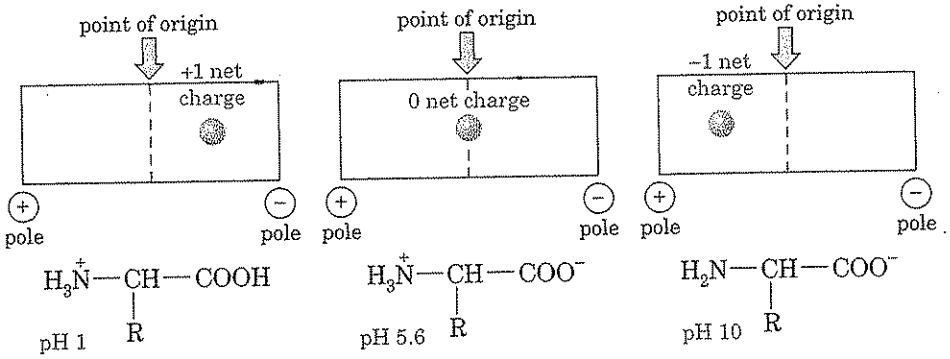


Figure 16.1 Titration curve for an amino acid.



electrophoresis
method of separating charged species in an electric field

zwitterion
the ionized form of an amino acid or peptide that has a net zero charge

isoelectric point (pI) or isoionic pH_I
The pH at which an amino acid or protein will not move in an electric field

The process of subjecting amino acids and proteins, or any charged species, to an electric field is known as **electrophoresis.**

The electrically neutral form is called the isoelectric form or **zwitterion** ("zwitter" is German for "both") and the pH at which the isoelectric form exists is called the **isoelectric point** or **isoionic pH_I**, the **pI.** A rough idea of the pI can be calculated by averaging the pK_a going from the +1 to the 0 form and the pK_a going from the 0 to the −1 form.

For our generic amino acid:

$$pI = \frac{pK_{a(+1 \rightarrow 0)} + pK_{a(0 \rightarrow -1)}}{2} \qquad pI = \frac{2 + 9}{2} = 5.5$$

This means that at pH 5.5 almost all of our generic amino acid molecules would be in the 0 net charge or zwitterion form, having an equal number of + and − charges.

If the R group contains a functional group that has conjugate acid–base properties, its ionization must be considered along with those of the amine and carboxyl groups. The pI is calculated in the same way as for the generic amino acid; the pK_a values used for the calculation must be those of the +1 ⟶ 0 transition and the 0 ⟶ −1 transition.

C. The Common Amino Acids

There are 20 amino acids that are commonly found in proteins. Their placement in the protein polymer chain is dependent upon the genetic code, that is, upon the DNA that is present in our genes. The DNA carries the code for the construction of proteins. Table 16.1 illustrates the structures of these 20 amino acids arranged according to the nature of the R group. They are shown in the forms present at very low pH. Table 16.2 lists the pK_a values for the amino, carboxyl, and R groups.

The amino acids designated with a superscript ([a]) are "essential" amino acids; that is, they cannot be made by the normal metabolic processes of the body and must therefore be provided in the diet. Not all food materials supply all of the essential amino acids. For example, corn and grains are deficient in lysine and tryptophan. A poor diet, low in protein and calories, can lead to severe nutritional disorders such as *kwashiorkor* and *marasmus*. These disorders frequently occur in developing or warring nations. Such a deficiency in developed countries can be evidence of anorexia nervosa.

The amino acids are most frequently represented by the three-letter abbreviations in Table 16.1 or by a one-letter format. This makes it easier to write long polymeric sequences.

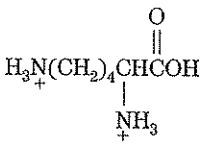
The two acidic amino acids, aspartic and glutamic, may also have amide forms

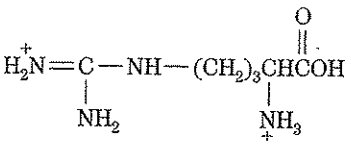
Table 16.1 The Common Amino Acids in Their Conjugate Acid Forms

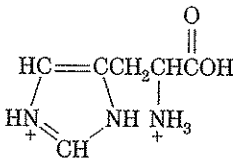
Basic Side Chains | Acidic Side Chains | Alkyl Side Chains | Aromatic Side Chains

Table 16.1 The Common Amino Acids in Their Conjugate Acid Forms

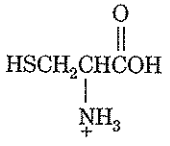
Basic Side Chains

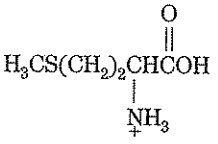

Lysine (Lys)


Arginine (Arg)[a]

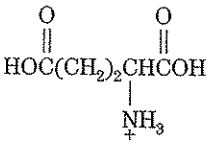

Histidine (His)[a]

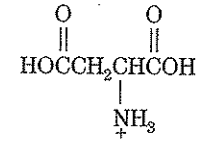
Sulfur-Containing Side Chains


Cysteine (CysH)[b]

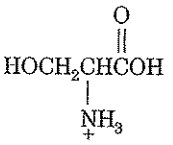

Methionine (Met)[a]

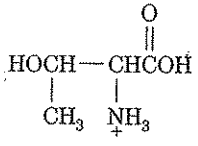
Acidic Side Chains

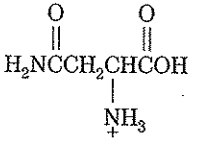

Glutamic acid (Glu)

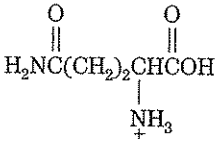

Aspartic acid (Asp)

Polar Side Chains


Serine (Ser)

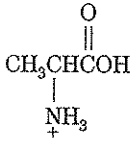

Threonine (Thr)[a]

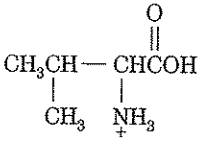

Asparagine (Asn)

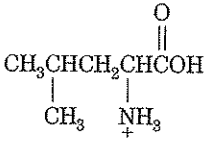

Glutamine (Gln)

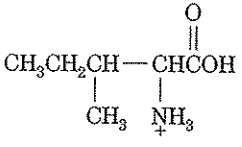
Alkyl Side Chains


Glycine (Gly)

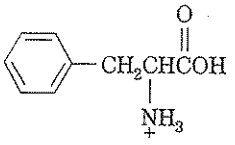

Alanine (Ala)

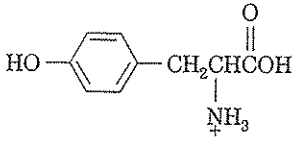

Valine (Val)[a]

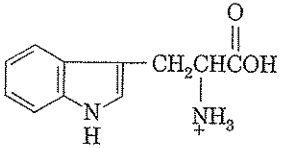

Leucine (Leu)[a]


Isoleucine (Ile)[a]

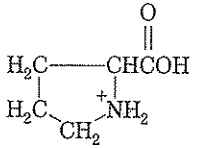
Aromatic Side Chains


Phenylalanine (Phe)[a]


Tyrosine (Tyr)


Tryptophan (Trp)[a]

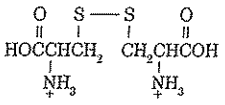
Cyclic Side Chains


Proline (Pro)

Note: Amino acid names are often represented by the three-letter abbreviations given in parentheses.

[a]An essential amino acid, which must be provided in the human diet.

[b]Often found as cystine, a dimer bonded through the sulfurs.



on the R side chain—$CONH_2$ rather than COOH. The amino acids are then called asparagine and glutamine, respectively. Amides do not accept or donate a proton under physiological conditions, and therefore the pK_a for the R group no longer exists. However, they are still polar and have the capacity to hydrogen-bond.

GETTING INVOLVED

- ✓ What does the pK_a of an ionizable group tell us?
- ✓ Associate the structures of the ionized forms of a generic amino acid with the titration curve in Figure 16.1.
- ✓ If the carboxyl group was missing from an amino acid, what would be the appearance of the titration curve? Do the same for a structure missing the amino group but retaining the carboxyl group.
- ✓ Where will a zwitterion move in an electric field?
- ✓ Why is it important to know the isoionic pH? What is the other term for this factor?
- ✓ How do you calculate an approximate value for the pI of an amino acid?
- ✓ Where would you find the pI on the titration curve shown in Figure 16.1?
- ✓ Can amino acids have more than three ionized (conjugate acid-base) forms? Briefly explain your answer.

Example 16.1

Draw out the conjugate acid–base forms for aspartic acid. Find its pI and predict the movement of the ionized forms in an electric field at various pH values.

Solution

First draw aspartic acid with all three of its ionizable groups in their conjugate acid forms. Include the pK_a for each ionizable group as found in Table 16.2.

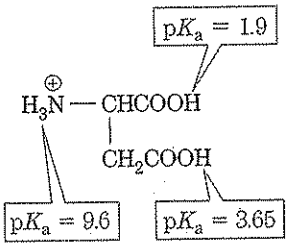
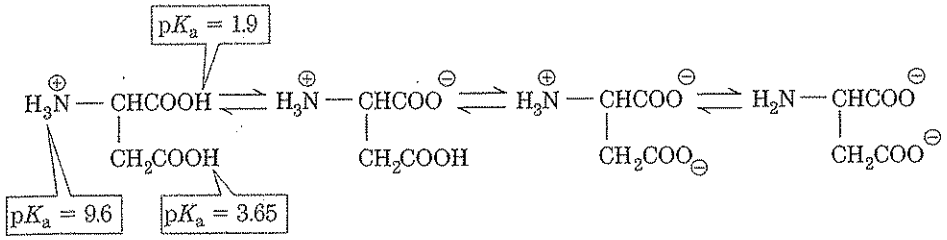


Table 16.2 The pK_a Values for the Common Amino Acids

Amino Acids	pK_as —COOH	—NH_3^+	—R	Amino Acids	pK_as —COOH	—NH_3^+	—R	Amino Acids	pK_as —COOH	—NH_3^+	—R
Ala	2.4	9.9		Gly	2.3	9.6		Pro	2.0	10.6	
Arg	2.2	9.1	11.8	His	1.8	9.0	6.0	Ser	2.2	9.2	
Asn	2.0	8.8		Ile	2.3	9.8		Thr	2.2	9.1	
Asp	1.9	9.6	3.65	Leu	2.4	9.6		Trp	2.4	9.4	
CySH	1.7	10.8	8.3	Lys	2.2	8.9	10.3	Tyr	2.2	9.1	10.1
Gln	2.2	9.1		Met	2.3	9.2		Val	2.3	9.7	
Glu	2.2	9.7	4.3	Phe	2.6	9.2					

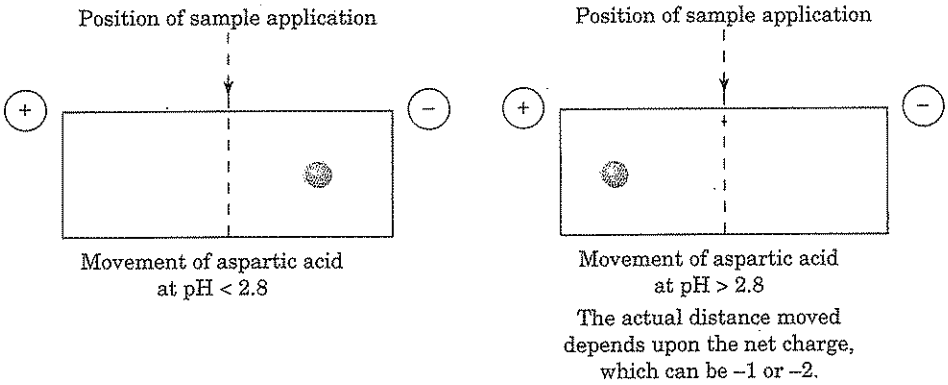
Find the net charge on this form and then remove protons in order of increasing pK_a. Calculate the net charge on each new form.



Find the zwitterion and use the pK_as on either side of it to calculate the pI.

$$\text{pI of Asp} = \frac{1.9 + 3.65}{2} = 2.8$$

The movement of aspartic acid in an electric field depends upon the pH. At a pH lower than 2.8 most of the molecules are in the cationic (+1) form and migrate to the − electrode. At pH values above 2.8 the aspartic acid molecules take on a negative charge, either −1 or −2, and migrate to the + pole.



Problem 16.1

Look at the R- groups in Table 16.2. Which could go from +1 → 0 and which could go from 0 → −1?

Problem 16.2

Draw out all of the possible ionized forms for the amino acids lysine, glutamic acid, alanine, and tyrosine. What is the net charge on each form?

Problem 16.3

Construct titration curves for aspartic acid, serine, and arginine. Indicate the pH range in which the various charged forms exist.

Problem 16.4

Toward which pole, + or −, would each of the following amino acids travel at pH 8.7 in an electric field: glutamic acid, arginine, threonine, tyrosine, and histidine?

Problem 16.5

What is the pI for histidine? isoleucine? cysteine?

Problem 16.6

Draw out the charged forms of glutamine and calculate its pI. How does the pI compare to that for glutamic acid?

Problem 16.7

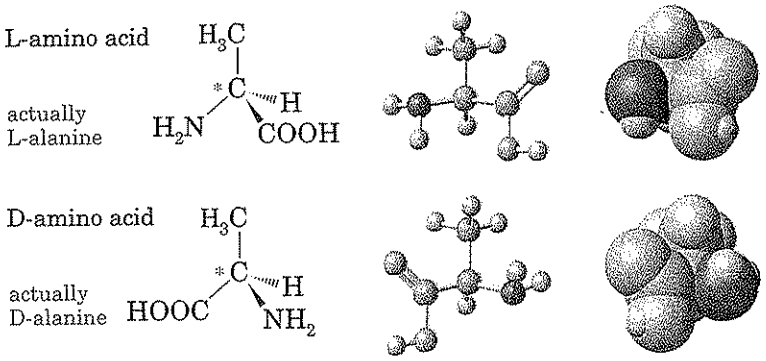
What is the most likely charged form that would exist for histidine at pH 6.8? for tyrosine at pH 13.4?

See related problems 16.25, 16.26, and 16.28.

D. Chirality of Amino Acids

If you inspect the structures of all of the amino acids except glycine, you can see that the attachment of a carboxyl, amine, R group, and hydrogen to a central carbon makes that carbon chiral, so the amino acid is optically active. With only one chiral carbon center there are 2^1 or 2 stereoisomers possible, related as nonsuperimposable mirror images or enantiomers. These are referred to as D- and L-amino acids. The genetic code uses only **L-amino acids** in constructing proteins, although D-amino acids may occur as modifications after the genetic code has been transcribed into protein, or they are formed by nongenetically directed processes. D-amino acids occur mainly in lower organisms such as bacteria.

L-amino acid
amino acid with the amine group on its primary chiral carbon oriented in the same way in three dimensions as the —OH in L-glyceraldehyde



GETTING INVOLVED

✓ There are two amino acids that have more than one chiral carbon. Identify them and draw out the optical isomers.

Problem 16.8

Use the structures of D- and L-alanine drawn above and determine which is R- and which is S-.

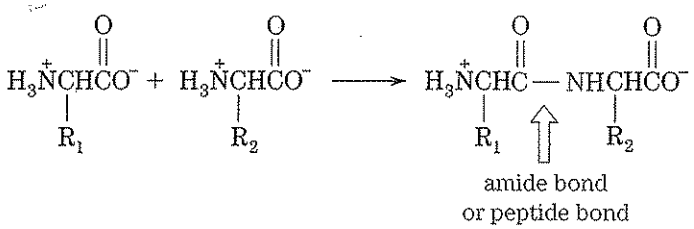
Problem 16.9
Why is glycine optically inactive?

See related problem 16.23.

16.2 The Peptide Bond: Formation of Polypeptides and Proteins

The protein polymer is made by linking together amino acids via an amide, or **peptide bond.** This occurs in a living organism through the transcription and translation of the genetic code. A summary of the reaction follows:

peptide bond
the amide bond formed between the carboxyl and amine groups of two amino acids



The formation of the peptide bond changes the ionization characteristics of the constituent amino acids. The carboxyl group of the first amino acid and the amine function of the second can no longer participate in conjugate acid–base behavior once they are joined by the peptide bond. That leaves the R side chains, as well as the terminal amino and carboxyl groups, as main sources of ionizable groups.

The amino acid chain, called a *polypeptide,* is usually drawn with the free amine group on the left and the free carboxyl group at the right. They are called the N- or amino-terminus and the C- or carboxy-terminus, respectively. As we start to add amino acids to the chain, the complete chemical structure becomes more cumbersome, and we resort to the abbreviations for the amino acids.

Polypeptides and proteins also have isoelectric points or pIs. As with an individual amino acid, if the pH is lower than the pI, the polypeptide has a net + charge; at a pH above the pI, the charge is −. If all of the molecules of a protein have the same net charge, they tend to repel each other. This keeps the protein dispersed in water. However, if the pH is adjusted to the pI, the net charge is zero and the protein molecules can aggregate, precipitating from solution. This is known as **isoelectric precipitation.**

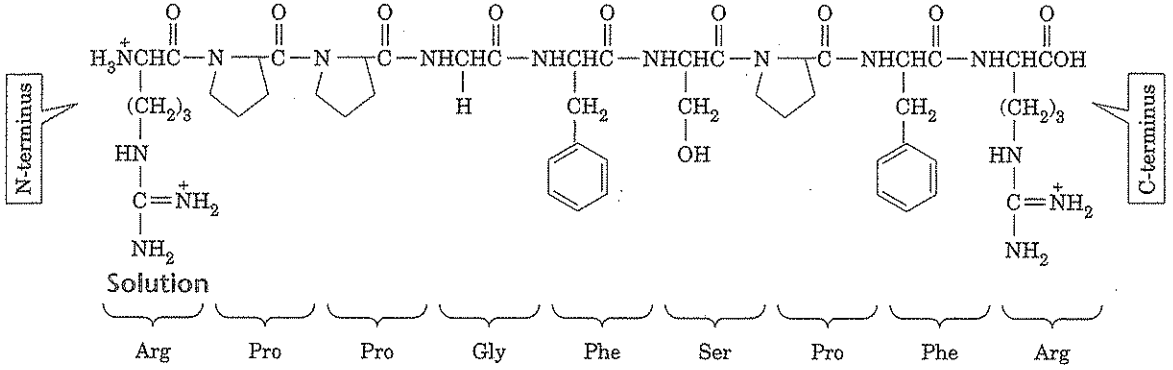
isoelectric precipitation
process of precipitating proteins at their isoelectric points, the pH of minimum solubility

GETTING INVOLVED

- ✓ Draw normal peptide bonds to link the amino acids serine, phenylalanine, and glutamic acid. There are several sequences of combination; draw only one sequence using the structures. Use the three-letter abbreviations to show the other combinations.
- ✓ Identify the N-terminus and C-terminus in each combination.
- ✓ What are the ionizable groups in your tripeptides?

Example 16.2

The structure of bradykinin appears below. This substance is a "tissue hormone" capable of dilating and increasing the permeability of blood vessels. It also causes intense pain. Identify the amino acids and determine the possible charged forms at low and high pH.



Solution

Arg ~ Pro ~ Pro ~ Gly ~ Phe ~ Ser ~ Pro ~ Phe ~ Arg

Arg~Pro~Pro~Gly~Phe~Ser~Pro~Phe~Arg

pK_a values	9.1	11.8						11.8	2.2
Charge at low pH	+1	+1						+1	0
Charge at pH 10	0	+1						+1	−1

Net charge +3

Net charge +1

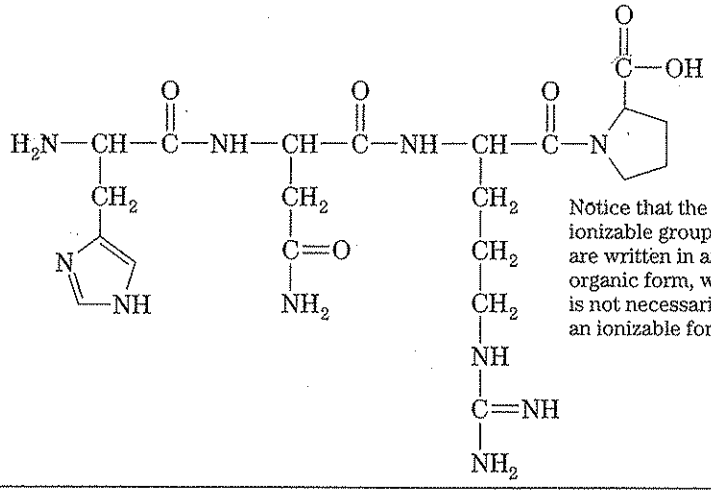
Problem 16.10

Find the net charge of the following polypeptide at pH 8.2 (approximate pH in the large intestine):

Ala ~ Lys ~ Asp ~ Tyr ~ Asp ~ His ~ CySH ~ Leu ~ Phe ~ Gln

Problem 16.11

For the polypeptide drawn below, identify the amino acids, Find the pI, and calculate the approximate net charge on the tetrapeptide at physiological pH, that is, pH 7.4.



Notice that the ionizable groups are written in an organic form, which is not necessarily an ionizable form.

16.3 The Hierarchy of Protein Structure

Because of their size and chemical nature, proteins exhibit three-dimensional structural organization. There are four formal levels of protein structure, each stabilized by specific intra-molecular interactions (primary, secondary, tertiary, and quaternary).

A. Primary Protein Structure—The Sequence of Amino Acids

The linear arrangement of amino acids in a protein from the free amino end to the carboxyl end is known as its **primary structure.** It is this sequence that is determined by the genetic code and that determines the overall shape and function of the macromolecule.

primary (1°) protein structure
the linear sequence of amino acids from N- to C-terminus

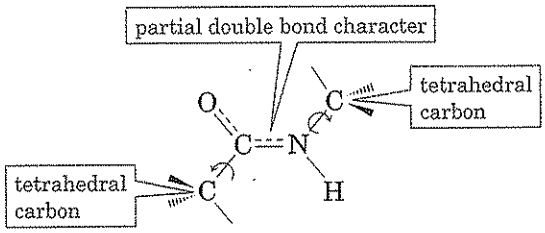
B. Secondary Protein Structure—Helices and Pleated Sheets

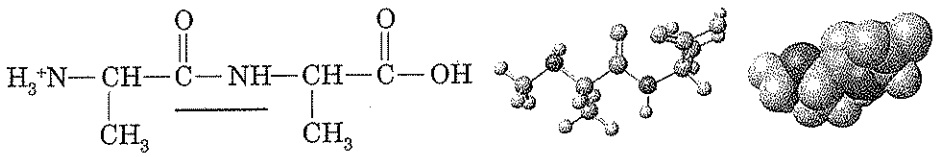
Secondary structure is the organization of regions or segments of the polypeptide chain that results from hydrogen-bonding between peptide bonds. The hydrogen-bonding produces structures that can be helical or sheetlike.

secondary (2°) protein structure
arrangement of a segment of a polypeptide chain into an organized structure, such as an α-helix or β-pleated sheet, stabilized by hydrogen-bonding between peptide bonds

1. ***Alpha and Beta Structures.*** Peptide bond geometry is trigonal planar due to a partial double bond formed by electron delocalization between the carbonyl carbon and the amide nitrogen. This means that there is restricted rotation about the amide bond and geometric isomers can exist. The predominant isomer is *trans;* that is, the oxygen of the carbonyl group and the hydrogen of the amide are across from each other. The α carbons of the attached amino acids have single bonds and are tetrahedral with free rotation about their bonds.

This gives rise to a polymer that looks like a series of flat plates attached by a swivel joint.





dialanine (Ala–Ala) models
peptide bond is underlined

The hydrogen attached to the amide nitrogen is electropositive (δ+), whereas the oxygen of the carbonyl group is electronegative (δ−). As a result, the amide hydrogen is said to be a hydrogen-bond donor and the carbonyl oxygen is a hydrogen-bond acceptor. The polypeptide chain rotates around the tetrahedral carbons in order to align amide hydrogens with carbonyl oxygens (hydrogen-bond donor–acceptor pairs).

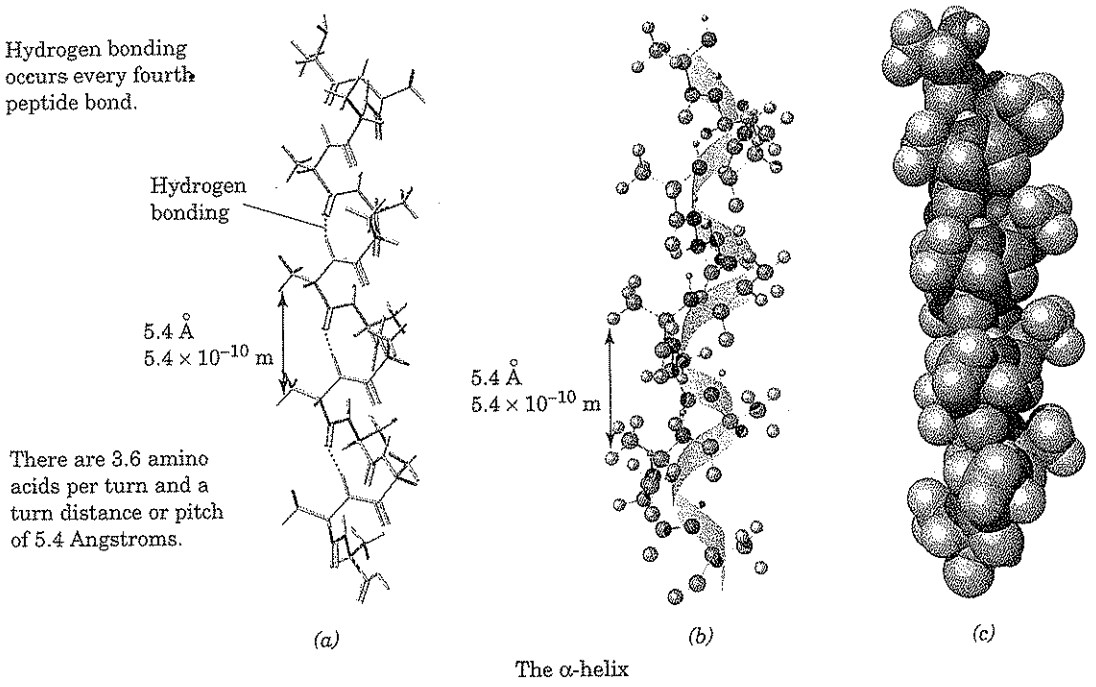
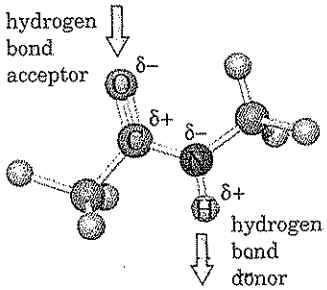


Figure 16.2 Various representations of the α-helix formed by $(Ala)_{12}$: (a) is a wireframe model with some of the dimensions and hydrogen bonding shown; (b) is a ball-and-stick model overlaid on a ribbon diagram used frequently in biochemistry to represent protein structure; (c) is the same polypeptide in a space-filling model.

α (alpha)-helix
spiral protein secondary structure stabilized by hydrogen-bonding between the peptide bonds of every four amino acids

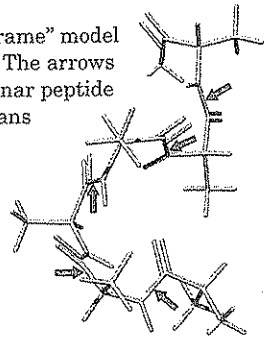
A partial rotation of about 45° allows the peptide bonds to arrange so that every fourth peptide bond occurs under another (see Figure 16.2). This sets up a spiral or helix, specifically a right-handed helix, known as the **α (alpha)-helix.** (Rotate your right hand in a clockwise direction.) Hydrogen-bonding can occur between the peptide bonds located above and below each other in a direction almost parallel to the long axis of the helix. The R groups protrude from the helix

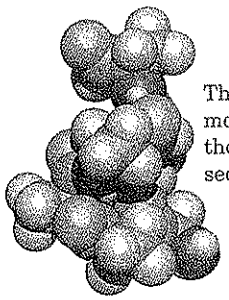


Notice that a hydrogen bond donor must be positioned above a hydrogen bond acceptor for this interaction to occur.

This means that the polypeptide chain has to somehow curve or bend back on itself to put these groups in proximity to each other.

This is a "wireframe" model of hexaalanine. The arrows indicate the planar peptide bonds with a trans configuration.





This space-filling model shows that the polypeptide chain seems to be twisting.

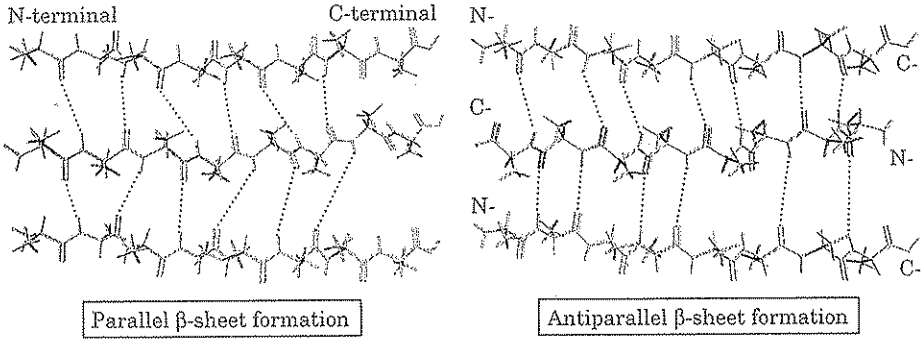


Figure 16.3 Parallel and antiparallel β-sheet formations. Notice the alignment of the N- and C- termini of the chain fragments shown as well as the orientation of the hydrogen bonding. Each chain fragment takes on a pleated ribbonlike appearance with one ribbon hydrogen bonding side-to-side with another.

in a manner analogous to the spokes protruding at almost right angles to the cylindrical hub of a bicycle wheel. (Hair is composed of the protein α-keratin, which is mainly α-helical in nature.)

Full rotation of the bonds to the α-carbons to 180° extends the chain and produces a pleated appearance with the hydrogen bond donors and acceptors located at the sides of the chain and the R groups directed up and down, perpendicular to the chain. If the polypeptide chain itself bends and comes back alongside itself, hydrogen-bonding can occur in a side-to-side arrangement. This is known as a **β (beta)-pleated sheet.**

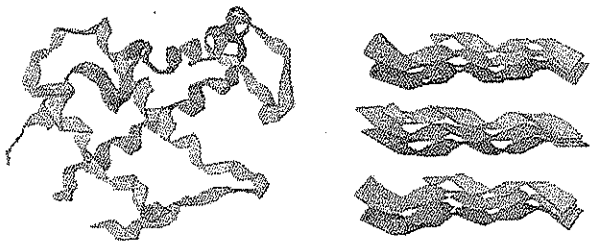
The polypeptide chain fragments may be oriented such that they are all progressing from N- to C-, called a **parallel sheet,** or they may alternate N- to C- aligned with C- to N-, an **antiparallel sheet.** Figure 16.3 illustrates the parallel and antiparallel β-pleated sheets.

Ribbon cartoons are frequently used to symbolize secondary protein structure. The α-helix is easily recognized as a spiral, whereas β-structure is shown with an arrowhead to indicate the N- to C- orientation of the chain. Figure 16.4 contains examples of such structures.

β (beta)-pleated sheet
layered protein secondary structure stabilized by side-to-side hydrogen-bonding between peptide bonds located in different chains or parts of a chain

parallel β-sheet
β-sheet with its polypeptide strands aligned N- to C-

antiparallel β-sheet
β-sheet with polypeptide strands aligned alternately N- to C- and C- to N-



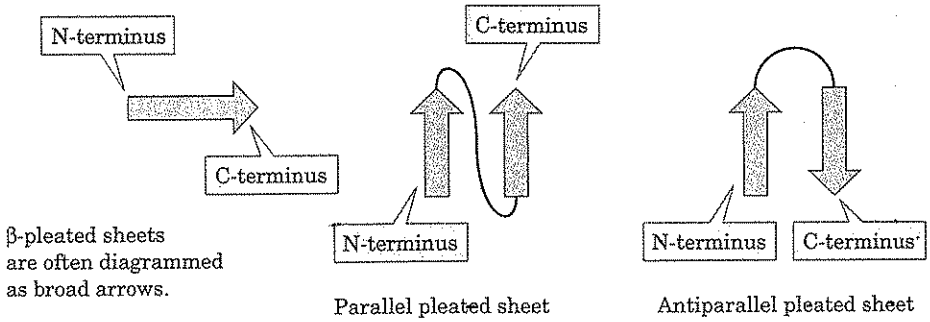


Figure 16.4
Ribbon diagrams of protein supersecondary structures or domains.

Spider webs and silk fibroin are formed by the protein β-keratin, which contains predominantly β structure. Other proteins have mixtures of α- and β-structures depending upon the nature of the amino acids present and the rotation about the α carbons. A limited number of rotational angles occur in proteins due to the presence of the R groups, which can interfere with the stability of a secondary structure.

As you look at the structures of the common amino acids, one stands out as being essentially different from the others in its backbone of amine, chiral carbon center, and carboxyl groups. It is proline, a cyclic amino acid. The constraint of its five-membered ring structure restricts the degree of rotation possible about the α carbon. It will not twist into an α-helix, nor will it extend to form a β-sheet. Rather it "kinks" or bends the polypeptide chain to disrupt potential α- and β-secondary structures.

The only nonoptically active amino acid, glycine, also interrupts α- and β-structures because it has no R group to form any bulk around the polypeptide chain. The R groups can actually help to stabilize or destabilize secondary structure. Glycine, therefore, frequently appears in positions of bends in the chain.

X-ray structural analysis of proteins has revealed that combinations of secondary structures occur in specific functional groupings known as **domains.** In fact, studies of evolution on a molecular level indicate that new proteins may have evolved by joining, deleting, or modifying the DNA sequences for domain supersecondary structures.

domain
combinations of secondary structures associated into functional units

2. ***The Collagen Triple-Helix.*** Collagen is the most abundant protein type in the human body. In its variations collagen contributes to the skin, bones, teeth, ligaments, cartilage, and tendons that cover, support, and hold us together. Collagen is a left-handed, triple, intertwined helix composed primarily of proline and glycine. In this case the structure depends upon the "kink" or bend that proline imposes on the chain. Located at approximately every third position in a chain 1000 amino acids long, one strand of the triple helix looks like an extended paper clip. The lack of a glycine side chain allows three of these strands to come into close proximity, forming a helix composed of three chains.

Bones are made of the protein collagen, which holds mineral deposits.

The helix is stabilized by hydrogen-bonding between the peptide bonds of glycines located on different, adjacent chains. The result is a left-handed triple helix (see Figure 16.5). The entire process of collagen assembly is complex and involves carbohydrate as well as protein.

It is important to emphasize that "protein secondary structure" refers to organized regions of protein chain of definite shape, stabilized by hydrogen-bonding between peptide bonds within (intra-) the polypeptide chain.

GETTING INVOLVED

- ✓ In your own words, define secondary structure in proteins.
- ✓ What molecular interaction is responsible for protein secondary structure?
- ✓ The α-helix, β-pleated sheet, and collagen triple helix are all examples of the principal types of protein secondary structure. How are these alike and how do they differ?
- ✓ Are secondary structure interactions limited to those within the same region of the chain of amino acids?

Problem 16.12

At physiological pH 7.4, polyaspartic acid and polylysine are known to destabilize an α-helix. Why does this occur?

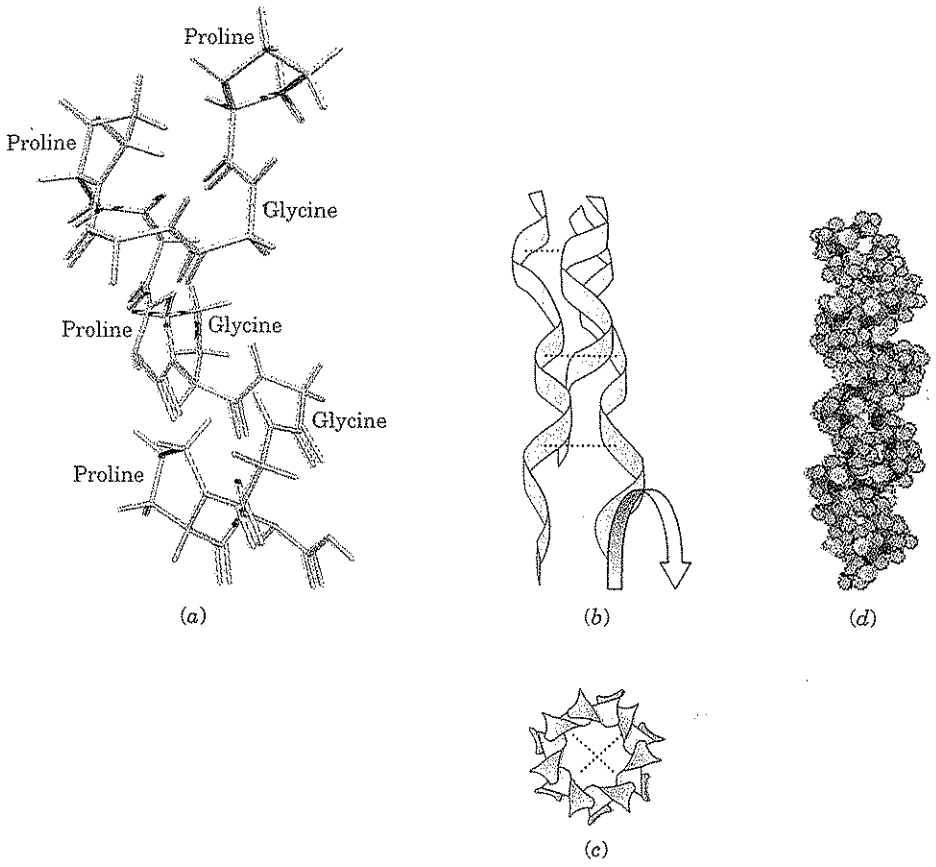


Figure 16.5
Models of collagen. (a) A single chain with the sequence $(Pro{\sim}Gly{\sim}X)_3$; the Pro and Gly are highlighted. (b) A ribbon diagram of the collagen triple helix. (c) An end-on view of the triple helix with the dashed lines indicating the directionality of the hydrogen bonding *between* strands. (d) The collagen triple helix in a space-filling model.

Problem 16.13

Are there any special amino acids that encourage or discourage the formation of specific kinds of secondary structure? Explain your answer.

Problem 16.14

Think about the structures of the following amino acids and determine the secondary structures in which they would be "comfortable": Leu, Ala, Ser, Pro, Gly, Tyr, Lys.

C. Protein Tertiary Structure

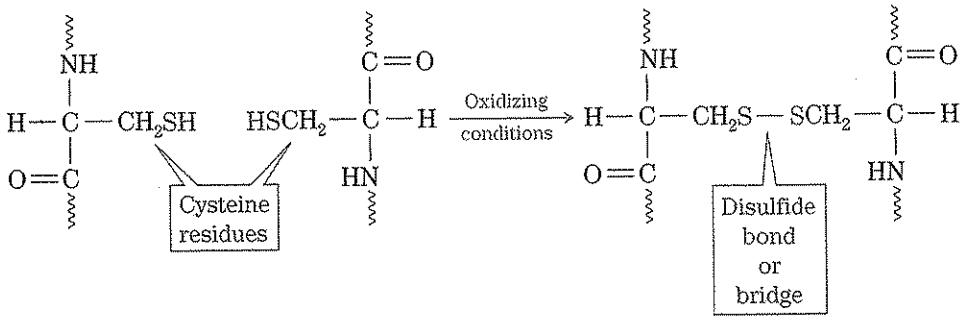
Proteins can bend and fold into overall structures that may be long and fibrous like hair and bone or more compact, that is, globular, like egg white (albumin). The R side chains participate in both covalent and noncovalent interactions in order to stabilize the protein in its final three-dimensional structure or **tertiary (3°) conformation.**

tertiary (3°) protein structure the folded, completely formed three-dimensional structure of a polypeptide chain that is stabilized by covalent and noncovalent forces

The common covalent side-chain bond that can hold together remote regions of the protein is the **disulfide bond** formed between two cysteine residues. The -SH groups on two cysteines are oxidized to form a covalent disulfide bond or bridge.

disulfide bridge covalent S — S bond formed between the side chains of cysteine residues that may be distant from each other in a polypeptide chain

There are three main noncovalent interactions: hydrogen-bonding, salt bridges (ionic interactions), and hydrophobic interactions. R groups that have a hydrogen atom bonded to an oxygen or nitrogen, such as histidine and serine, can hydrogen-bond with an electronegative group such as the oxygen of a carbonyl or the nitrogen of an amine. This is the same type of force we saw in



secondary structure, but now it is occurring between R side chains rather than between peptide bonds.

salt bridge
ionic interaction (+ to −) between the side chains of acidic and basic amino acids that stabilizes the tertiary and quaternary structures of proteins

Another noncovalent interaction is the formation of **salt bridges** between oppositely charged R groups.

Since most proteins are found in contact with the water that constitutes about 70% of our body weight, the surfaces of these macromolecules should exhibit amino acid side chains that form hydrogen bonds with water or associate via ion–water (ion-dipole) interactions.

hydrophobic interaction
weak attractive, nonpolar interactions between the hydrocarbon side chains of amino acids that stabilize tertiary and quaternary structures of proteins

The hydrocarbon side chains (valine, leucine, phenylalanine) do not interact with water or ions but rather aggregate in a **hydrophobic** environment, often forming a "waxy" core at the inside of a water-soluble protein. While there are very weak interactions between the atoms in these groups, the prevailing force is the avoidance of polarity.

Figure 16.6 illustrates examples of the major tertiary interactions in proteins.

GETTING INVOLVED

- ✓ What kinds of covalent and noncovalent interactions give rise to tertiary structure?
- ✓ Identify any amino acids that would probably not take part in tertiary interactions. Why is this the case?
- ✓ Why is this level of protein structure dependent upon the first two levels of structure?
- ✓ Look at all of the R groups of the common amino acids and identify the types of interactions that could occur. Be sure to consider all possibilities.
- ✓ Explain how repulsive forces could play a role in tertiary structure.

Example 16.3

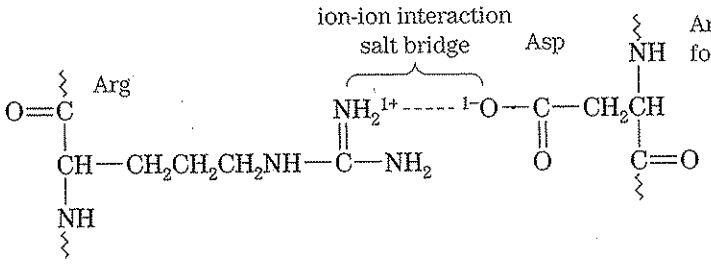
What type of tertiary interactions could occur between the side chains of the following pairs of amino acids under physiological conditions, that is pH 7.4?

(a) Arg and Asp **(b)** Phe and Leu **(c)** Ser and Gln

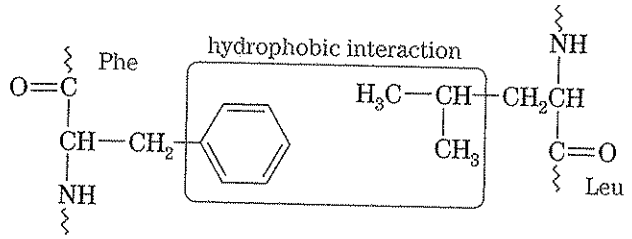
Solution

Draw out the side chains of each amino acid. Since we are discussing tertiary interactions, the α amino and carboxyl groups are involved in peptide bonds.

(a) Arg is arginine; its side chain should have a + charge at pH 7.4. Asp is aspartic acid; its side chain should have a − charge at pH 7.4.

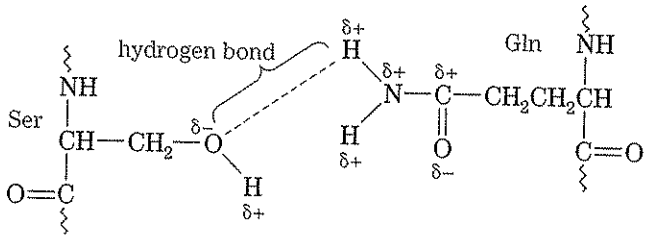


(b) Phe, phenylalanine, has an uncharged, hydrophobic side chain, as does Leu, leucine. Therefore a hydrophobic interaction takes place.



Phe and Leu would tend to exclude water and thereby interact hydrophobically

(c) Ser, serine, and Gln, glutamine, have uncharged, polar side chains. They can form a hydrogen bond.



The side chains of Ser and Gln present many possibilities for dipole-dipole interactions, specifically for hydrogen bonding.

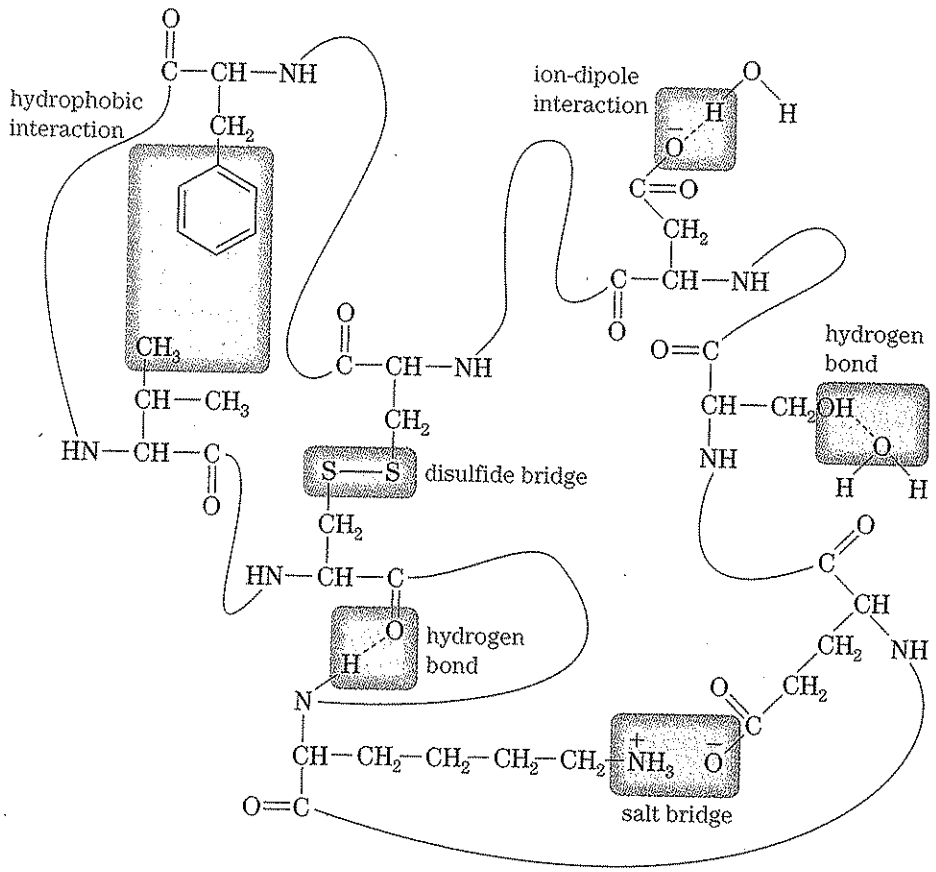


Figure 16.6 Examples of interactions in the tertiary structure of proteins.

Problem 16.15

The serine and glutamine shown in Example 16.3 have more than one possibility for the hydrogen-bonding between their side chains. What other possibilities exist?

Problem 16.16

What type of tertiary interactions could exist between the side chains of the following pairs of species?

(a) Thr and H_2O **(b)** Asn and Trp **(c)** Asp and Glu **(d)** His and Val

See related problems 16.28, 16.29, 16.30.

D. Quaternary Protein Structure—Association of Subunits

subunit of a protein
single polypeptide of a protein with tertiary structure that may or may not be functional

quaternary (4°) protein structure
noncovalent association of protein subunits to form a functional protein

A significant number of proteins contain more than one polypeptide chain, called a **subunit.** The subunits are held together by the same noncovalent forces of hydrogen-bonding, salt bridges, and hydrophobic interactions that give rise to tertiary conformation. This is called a protein's **quaternary (4°) structure.** The important fact to note is that most multisubunited proteins require all of their subunits in order to be fully functional.

E. Complex Proteins—Proteins Plus

complex protein
protein that requires one or more nonprotein portions, such as metal ions or organic groups, in order to function

simple protein
protein composed only of polymerized amino acids

Egg white or albumin is a relatively **simple protein** containing nothing but the polypeptide chain folded into its functional 2° and 3° forms. However, proteins sometimes require other types of molecules and ions in order to work. For example, the intestinal enzyme carboxypeptidase requires Zn^{2+} ion. Myoglobin is a muscle protein with one subunit that stores oxygen for use in times of oxygen starvation. It contains Fe^{2+} and a conjugated heterocyclic amine molecule called *heme* which actually binds the O_2 (see Figure 16.7).

Hemoglobin is related to myoglobin in that it, too, is an iron–heme protein, but it is composed of four subunits. The structure of hemoglobin is interesting by virtue of the cooperation that occurs between subunits in order to bind and release oxygen at the appropriate time and place in the body. The red blood cell, or erythrocyte, of a normal adult human contains a large concentration of hemoglobin. The tetramer is made up of two types of protein subunits called α and β; adult hemoglobin has an $\alpha_2\beta_2$ structure. Each subunit has a hole or crevasse in which can be found a heme group complexed with an Fe^{2+}. Molecular O_2 can complex with Fe^{2+} but not with Fe^{3+}. The protein crevasse provides a hydrophobic environment that excludes water. A water environment would facilitate the oxidation of Fe^{2+} to Fe^{3+}. As more O_2 is bound to the hemoglobin tetramer, it becomes easier to oxygenate (add O_2). This means that at the higher oxygen pressure of the lungs, hemoglobin is easily oxygenated, but at the low oxygen tension levels of the veins and capillaries, near respiring cells, it releases the O_2 readily. See Figure 16.9.

denaturation
process of disrupting the secondary, tertiary, and/or quaternary structures of a protein, usually resulting in irreversible loss of function

F. Denaturation

Formation of the complete and functional three-dimensional structure of a protein depends on optimal, physiological conditions. What happens when a protein is subjected to heat, extreme pH, organic solvents, or mechanical disturbance? As you might suspect, the forces holding the protein in its "native" conformation can be overcome. When this happens, the protein becomes denatured, a process that may be either reversible or irreversible.

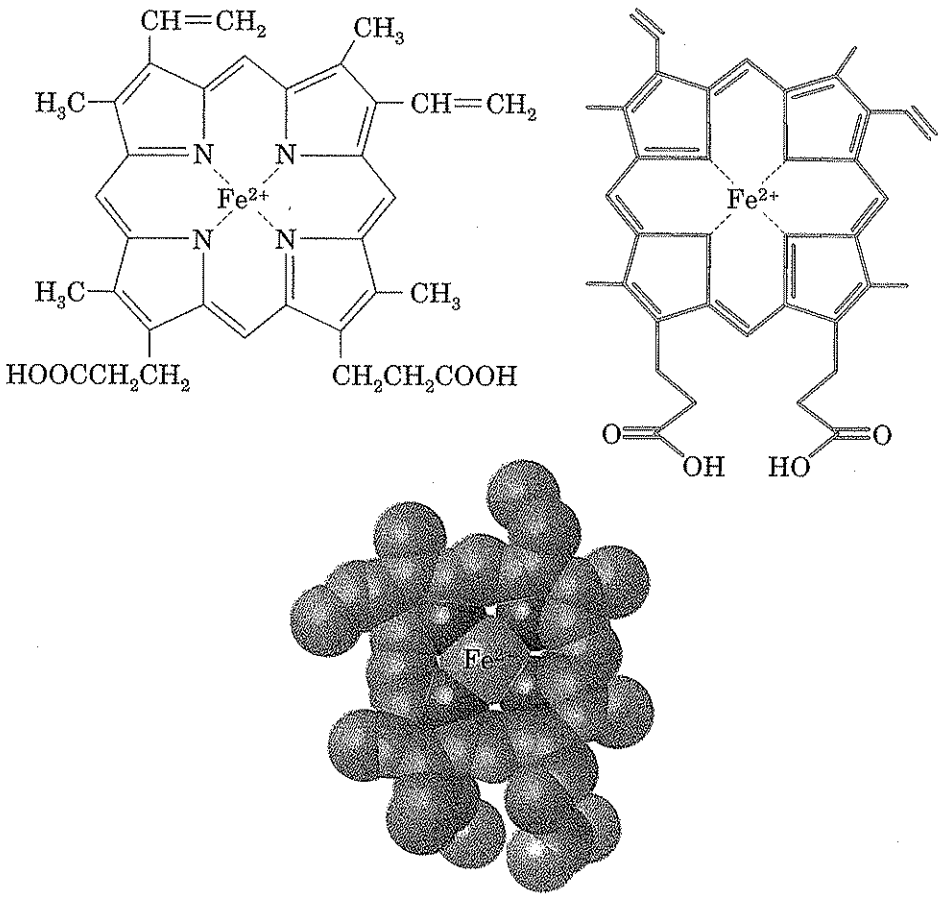


Figure 16.7 Three representations of the heme molecule. Notice that the iron is in the (II) oxidation state and that the heme is a planar structure.

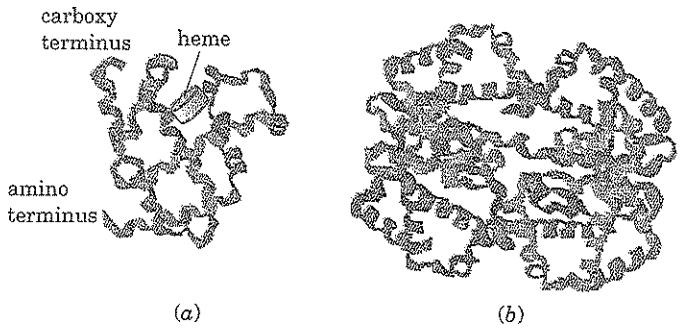


Figure 16.8 Ribbon diagrams of heme proteins. (a) Myoglobin structure illustrates the placement of the heme group in a slot or crevasse made by the secondary and tertiary protein structures. (b) Hemoglobin has four subunits of two different types, α and β, each of which contains a heme group. The subunits act in concert to capture and deliver molecular oxygen.

Consider boiling an egg, for example. As the temperature increases, the molecules of albumin (egg white) begin to vibrate more and more intensely until the tertiary forces as well as many of the secondary ones are negated by the vibrational energy of the unwinding molecule. Once the albumin is opened up, the hydrophobic amino acid core is exposed and aggregates with other exposed cores, forming a solid matrix of associated albumin molecules. We can see this in the conversion of the translucent, gelatinous raw egg white to the opaque hard-boiled egg white.

Connections 16.1 www.prenhall.com/bailey

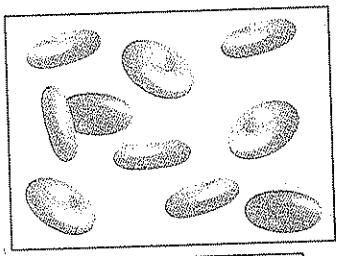
Sickle Cell Anemia—A Biochemical Disease

More than 400 natural variations in the primary amino acid sequence of hemoglobin are known. Most of these are inconsequential—the genetic code has substituted an amino acid quite similar in structure and properties to the one that should be present; for example, a leucine for an isoleucine.

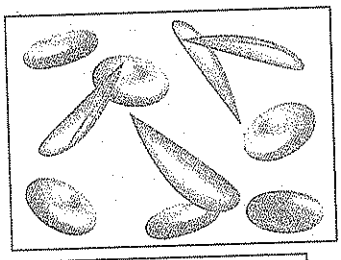
A devastating condition exists in which hemoglobin, after delivering its oxygen supply to tissues and starting its return trip to the lungs for reoxygenation, polymerizes into thick strands, and literally clogs up the smaller veins and capillaries. The red blood cells (erythrocytes) containing the hemoglobin change from their normal disclike shape to a collapsed, sickled shape. The plugging of blood vessels and the destruction of fragile blood cells lead to gangrene, heart disease, kidney disease, and brain damage. This condition is known as sickle cell anemia, and it occurs in about 0.3% of the African-American population.

The cause of this life-threatening condition turned out to be not as complicated as might have been anticipated. It was found by Linus Pauling that sickle cell hemoglobin, or HbS, had a different electrical charge (electrophoretic mobility) at physiological pH compared to normal hemoglobin (HbA). Then Vernon Ingram, using chemical reactivity and chromatography, discovered that there was but one change in the HbS molecule to distinguish it from normal adult hemoglobin, or HbA. The β-subunits of normal HbA have a glutamic acid at the sixth position from the amino end. However, HbS contains a valine at that position. Consider the side chains of glutamic acid and valine, shown in Table 16.1. The substitution of a valine (a hydrophobic amino acid) for a glutamic acid (a hydrophilic amino acid) is what is known as a *nonconservative change*. The HbA Glu is found on the outside of the β-subunit; and while Glu is "content" to be in a water environment, Val is not. Consequently, HbS molecules come together, or aggregate, because Val is attempting to find a compatible environment, away from water.

A person with sickle cell anemia must carry both genes (be homozygous) for HbS. About 10% of African Americans have only one gene for HbS—this is known as the *sickle cell trait*. The gene is also present in populations found in Africa, the Mediterranean, and Middle East countries. This makes the study and treatment of sickle cell disease an international health issue.



Normal red blood cells



Normal and sickled red blood cells

Persons with the trait show no overt symptoms of the anemia. The interesting fact about this genetic trait is that two parents, each possessing one gene for HbS, have a 2 in 4 chance of having children with the trait, a 1 in 4 chance of having children with anemia, and a 1 in 4 chance of having children with normal HbA. Is there any advantage in possessing the trait? Indeed, the parasite that causes a certain type of malaria cannot exist for long in the HbA/HbS blood of a trait carrier. The cultural heritage of those exhibiting the trait lies in the tropical, malaria-prone areas of Africa and Asia. It is a survival trait.

GETTING INVOLVED

- ✓ Which forces hold quaternary structures together?
- ✓ What is the difference between a simple and a complex protein?
- ✓ What are the functions of myoglobin and hemoglobin?
- ✓ What structural features do myoglobin and hemoglobin have in common? How do they differ?
- ✓ What is a "conservative" change in primary structure? Give examples of amino acids that could be involved in conservative changes.

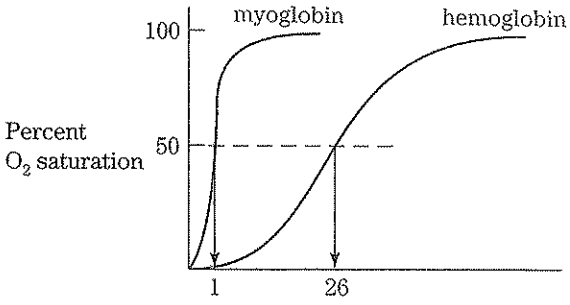


Figure 16.9 Comparative oxygen binding curves for myoglobin and hemoglobin.

Connections 16.2 www.prenhall.com/bailey

Mad Cow Disease

The title of the PBS special read like a science fiction movie ad: "The Brain Eaters." Even Oprah Winfrey got into the fray by stating she would not eat beef anymore for fear of getting something like "mad cow disease."

What was this scourge? Where did it start? How dangerous was it to humans? How could it be stopped?

Mad cow disease occurred in Great Britain after a change in laws permitting cattlefeed to contain meat and bonemeal of other animals such as sheep. The condition seen in the cattle is one of a group called the transmissible spongiform encephalopathies (TSE). Mad cow disease is a bovine spongiform encephalopathy (BSE). The symptoms in cattle are muscle and brain degeneration with premature death. Other types of animals such as minks and mule deer are known to exhibit similar degeneration. Brain material can literally become porous, and animals become incapable of even standing up. Obviously the practice of serving up sheep remains for consumption by cattle has ceased. Luckily, the United States as well as most of the world have not experienced mad cow disease.

BSE was linked to other human TSEs such as "kuru," a fatal dementia known to occur among aboriginal tribes practicing ritual cannibalism, which included eating the brain of a victim infected with the agent. The disappearance of cannibalism has coincided with the elimination of kuru. Crutzfeldt-Jakob disease (CJD), another TSE, was inadvertently spread to the uninfected by the practice of obtaining growth hormone from the pituitaries harvested from human cadavers. CJD and other human TSEs were found to have some hereditary properties.

The TSE in sheep and goats is called "scrapie" because the infected animals will rub themselves incessantly against objects, including barbed wire fences. It takes years or decades to develop symptoms of a TSE, depending upon the organism infected.

At first, the agent of transmission (vector) was thought to be a slow virus that could be passed on by the brain tissue of infected individuals. However, no genetic material could be found in the infectious particles. In the late 1970s and early 1980s Dr. Stanley Prusiner, then at the University of California at San Francisco, discovered that the infectious agent was protein in nature and coined the term *prion* for "proteinaceous infectious particles." The prion hypothesis of Prusiner states that brain tissue naturally produces PrP, a native *P*rion *P*rotein. However, subtle amino acid substitutions in a few PrP molecules eventually cause the alteration of the secondary structure of normal PrP molecules. For example, it was found that the replacement of a leucine an the α-helical native PrP with a proline reoriented the secondary structure to β-sheet. The aberrant PrP can bind to the native, natural prion and slowly cause the conversion of the native into an aberrant form. This could explain the genetic inheritance factors seen in some forms of TSE as well as the long incubation period for the symptoms. These results have been, and still are, under heavy dispute, even though Dr. Prusiner was awarded the Nobel Prize in Physiology/Medicine in 1997 for his discovery of a new kind of infectious agent.

- ✓ What is a "nonconservative" change in primary protein structure? Give examples of nonconservative changes.
- ✓ How does a nonconservative change in hemoglobin that results in hemoglobin S affect the function(s) of this protein?
- ✓ What is a denatured protein?
- ✓ List at least three factors that could lead to protein denaturation and be able to explain how these factors affect protein structure.
- ✓ Are all denaturation processes irreversible?

Problem 16.17

Considering that the molecules in air, O_2 and N_2, are nonpolar, how might you explain the formation of meringue by whipping egg whites?

Problem 16.18

What tertiary interactions in milk proteins would be upset by lowering the pH to about 3, as occurs during souring through the production of lactic acid by *lactobacilli*?

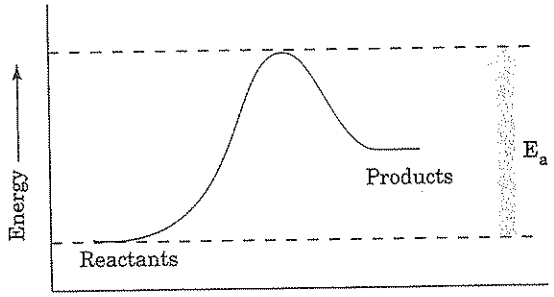
16.4 Functions of Proteins

Catalysis, protection, and regulation are a few of many protein functions.

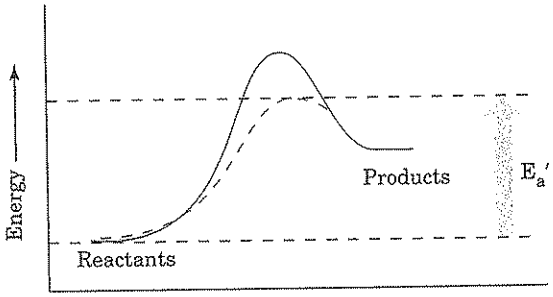
A. Enzymes—Biological Catalysts

catalysis
the process in which a chemical reaction rate is increased due to a lowering of the energy of activation

All chemical reactions must proceed through energy barriers, whether slight or huge, in order to form products from starting materials. This energy of activation, E_a, is due to many factors, including the need for the reactants to collide and orient themselves in space correctly and efficiently as well as follow the steps of the mechanism appropriate for the particular reaction. Anything that can enhance one or more of these factors will lower the energy of activation and make it easier for the reaction to occur. We refer to this as **catalysis.** Recall that the process of addition to alkenes, for example, may be acid-catalyzed or may require the presence of a metal such as nickel or platinum. In the case of acid catalysis, the H^+ ion actually participates in polarizing bonds and then is regenerated during the course of the reaction. For metal catalysis, the nickel or platinum provides a surface upon which the reactants may orient themselves to increase the probability of collision as well as to provide an atomic arrangement in space for efficient and productive contact.



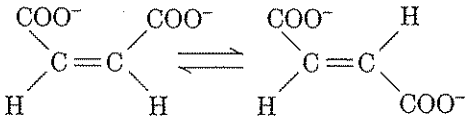
Course of a reaction *without* a catalyst



Course of a reaction *with* a catalyst

enzyme
biological catalyst, usually protein in nature

Enzymes are proteins that catalyze biological reactions. Enzymes are classified by the type of reaction that they catalyze: oxidation–reduction, hydrolysis, group transfer, bond breaking, isomerization, or bond making; and according to the reactants with which they interact (see Table 16.3). Technically, an enzyme's name should end in the suffix *-ase*. As an example, the enzyme that catalyzes the following reaction is called maleate *cis-trans*-isomerase.



Maleate Fumarate

However many enzymes were named before any convention directed such nomenclature, and they retain their common names, such as the stomach enzyme pepsin and the intestinal enzymes trypsin and chymotrypsin. Table 16.4 lists some common enzymes with typical uses.

An enzyme-catalyzed reaction has many advantages over an uncatalyzed reaction or one with a nonenzymatic catalyst. First, enzymes function at a rate thousands, if not millions, of times faster than uncatalyzed or normally catalyzed reactions. Second, enzymes can be very specific not only for the reactants, or substrates, in the reaction, but also for particular stereoisomers of those substrates. Third, enzymes function to produce specific products without the spurious by-products that can occur in organic reactions. These characteristics have led industry to the ever-increasing use of enzymes for the commercial production of natural and synthetic chemicals as well as for the preparation of foods and the cleanup of toxic waste.

Table 16.3 Enzyme Classification by International Enzyme Commission: A Summary

Class 1	*Oxido-reductases* carry out and influence oxidation-reduction reactions with alcohols, carbonyls, carbon-carbon double bonds, amines, etc.
Class 2	*Transferases* facilitate the transfer of certain functional groups, such as carbonyl, acyl, sugar, alkyl, and phosphate groups.
Class 3	*Hydrolases* catalyze the hydrolysis of esters, ethers, peptide bonds, glycosidic bonds, halides, acid anhydrides, and more.
Class 4	*Lyases* allow addition reactions with carbon-carbon double bonds, carbonyls, etc., or form such bonds themselves.
Class 5	*Isomerases* promote isomerization, optical and geometric, and also catalyze various intramolecular reactions, resulting in skeletal isomerization.
Class 6	*Ligases* (synthetases) aid in bond formation between carbon and sulfur, oxygen, nitrogen, or another carbon, and require ATP for energy.

Table 16.4 Some Common Enzymes

Enzyme	Typical Use	Enzyme	Typical Use
Rennin	milk coagulation for making cheese	Collagenase	removes tail from tadpoles when they become frogs
Bromelain	tenderizing meat; chill-proofing beer	Pepsin	begins protein digestion in stomach
Creatine kinase	provides metabolic energy in active muscle tissue	Streptokinase	dissolves blood clots
DNase	breaks up mucus in lungs of cystic fibrosis victims	Reverse transcriptase	responsible for the incorporation of viral genetic material into the host genome

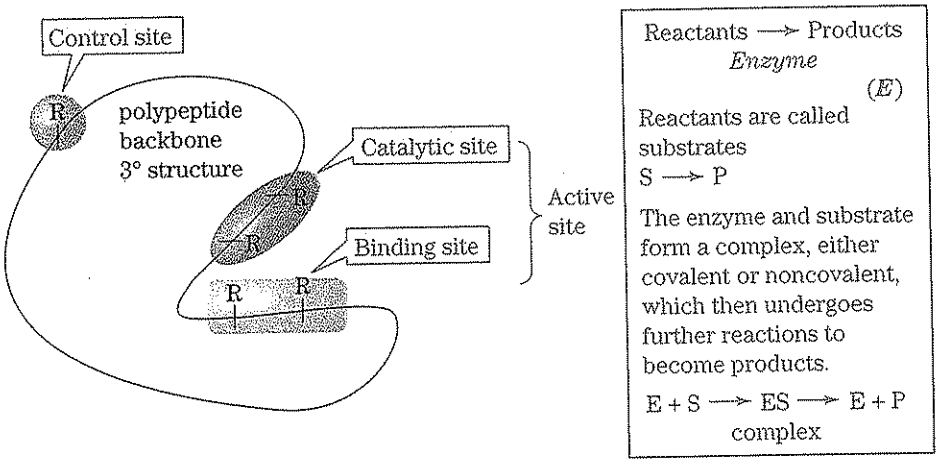


Figure 16.10
An enzyme and the parts contribute to the active and control sites.

Enzymes direct their remarkable feat of catalysis by presenting an interactive, three-dimensional environment to the reactants. Every enzyme molecule has an **active site,** where catalysis takes place. Within the active site is a **binding site,** which attracts and holds the **substrates,** and a **catalytic site,** which participates in the mechanism of the reaction (see Figure 16.10).

active site
functional portion of an enzyme

binding site
portion of an enzyme active site that attracts the substrate

substrates
molecules and/or ions on which an enzyme works

catalytic site
area within the active site of an enzyme that causes catalysis

B. Enzyme Control

The human body contains thousands of different enzymes working on different reactions with different substrates. How are all these reactions coordinated so that a single, coherent organism results? What keeps the body from digesting itself? The answers to these questions are of course very complicated, but we can discuss briefly how some enzymes can be turned on and off.

A common means by which enzymes are prevented from exerting their catalytic effects where they are not needed is their secretion in larger, inactive forms known as **zymogens.** An important example involves enzymes trypsin, chymotrypsin, and carboxypeptidase, which are responsible for protein digestion in the intestines. These proteins are produced in the pancreas as larger proteins; trypsinogen, chymotrypsinogen, and procarboxypeptidase. After biosynthesis, they are secreted through the bile duct into the small intestine, where trypsinogen is changed to trypsin by the action of another enzyme called enteropeptidase. The active trypsin can also convert trypsinogen to trypsin, and chymotrypsinogen and procarboxypeptidase to their active states. Should activation of the zymogens occur before they leave the pancreas, which can happen in certain disease states, then the pancreas will gradually be digested, a condition known as pancreatitis.

zymogen
inactive precursor of an enzyme

Other enzymes can exist in two forms, which differ only in the covalent modification of an amino acid in the protein. For example, the enzyme glycogen phosphorylase is responsible for the first step in the conversion of the storage carbohydrate glycogen to glucose. Glycogen phosphorylase itself needs to be phosphorylated, that is, to be derivatized with two phosphate groups, in order to be enzymatically active. Can you guess what catalyzes the phosphorylation of phosphorylase? That's right—another enzyme, phosphorylase kinase. The active form of phosphorylase can be inactivated (dephosphorylated) by a third enzyme, a phosphatase.

Other materials, natural and synthetic, can slow down or completely stop the action of enzymes. These species are called **inhibitors.** The pancreas, in its role of zymogen secretion, also produces another protein, pancreatic trypsin inhibitor,

enzyme inhibitor
molecule or ion that reversibly or irreversibly slows down or stops the activity of an enzyme

which helps to keep trypsin in check. Heavy metals such as mercury, lead, and arsenic will inhibit enzymes to such an extent that the organism can die. This is the fundamental premise behind the development of many pesticides and poisons.

Because enzymes are organic molecules, they can be manipulated for commercial use. For example, they may be compounded with detergents in order to remove grease or blood stains from clothing or they may be attached to a solid support to convert glucose to fructose in the production of high-fructose corn syrup. The applications of enzyme chemistry are virtually limitless.

GETTING INVOLVED

✓ What are some of the problems that arise from attempting to use enzymes on a large scale for industrial processes, considering that enzymes are proteins?

C. Antibodies—Immune System Protection

The immune system is a complex network of cells, proteins, and chemicals that act in concert to thwart the invasion of anything that is not part of the organism, sometimes referred to as "nonself," or in immunological terms as the antigen. **Antibodies** are part of this protective arsenal.

antibody
glycoprotein produced by the B-cells of the immune system as protection

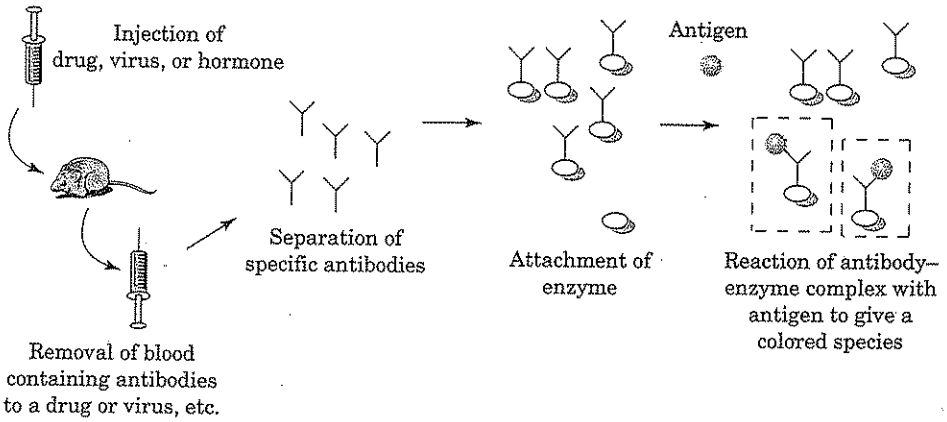
Connections 16.3 www.prenhall.com/bailey

Testing for Drugs, Pregnancy, and AIDS

Antibodies have proved invaluable in the clinical determination of disease, drug intoxication, and pregnancy. Because they are so specific, antibodies can be generated in an animal to specific substances such as a virus, hormone, or drug. As proteins, the antibodies can be linked chemically to enzymes. The enzymes, in turn, may catalyze a reaction involving a color change, which can be detected by using a single or multiple wavelength spectrophotometer.

This type of assay is called an enzyme-linked immunosorbent assay, or ELISA. The actual process is a little more complex, but the diagram below contains the fundamental concept. This technique can be modified by using radioactive antibody or antigen complexes, which may increase the sensitivity and thereby enhance the limits of detection.

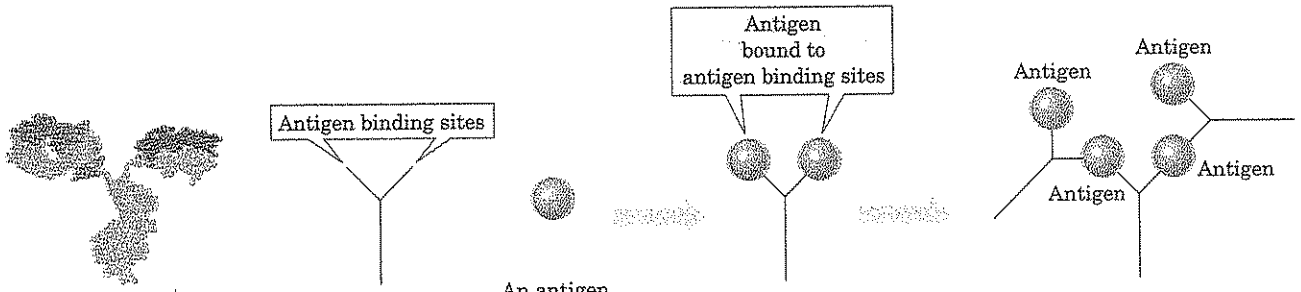
A variety of drugs, from morphine to amphetamine, can be assayed quantitatively in this manner. Pregnancy is determined in over-the-counter kits that detect the presence of the hormone chorionic gonadotropin, which is excreted by a woman during the first few weeks after conception. The AIDS virus has a protein capsule or coat that can be detected via an ELISA assay.



glycoprotein
protein with carbohydrate attached

antigen
the material to which the immune system responds

Antibodies are **glycoproteins,** that is, proteins to which carbohydrates are covalently attached. An antibody is produced by the B-cells of the immune system in response to a foreign substance or **antigen.** Formation of an antibody–antigen complex can result in precipitation or in identification to the other immune system components that can help to destroy the invader.



The process of immunization against toxins and disease is based upon the fact that repeated challenges by the same type of antigen result in increasingly intense antibody responses by the cells of the immune system.

Poliomyelitis, a paralytic viral disease, has been almost eradicated from the United States by the immunization of babies with small amounts of the virus, that have been treated to be less dangerous (attenuated). Booster immunizations keep the amount of defensive antibodies high and ready to respond. We can also be immunized against the toxins produced by various bacteria. Antibodies to the venoms of some poisonous animals and insects can be directly administered following the bite or sting. Table 16.5 lists some of the common immunizations available in the United States.

GETTING INVOLVED

- ✓ What is the function of an enzyme? An antibody?
- ✓ Write a generic chemical equation for the action of an enzyme on a chemical reaction.
- ✓ Are enzymes and antibodies simple or complex proteins?
- ✓ How do antibodies interact with antigens?
- ✓ Describe the biochemical basis for immunization.
- ✓ Outline the general steps in an ELISA assay.

Table 16.5 Immunizations

Standard Immunizations		Other Available Immunizations
DPT	Rubella (German measles)	Typhoid
diphtheria	Mumps	Current type of influenza
pertussis (whooping cough)	Hepatitis B	Cholera
tetanus (lockjaw)	Chicken pox	Rabies
Polio (Sabin vaccine)	Influenza	Smallpox
		Tuberculosis
		Hepatitis A
		Pneumonia

D. Polypeptide and Protein Hormones—Metabolic Regulation

The ability of living things to grow, reproduce, and respond to stress is regulated by secretions of biochemicals known as **hormones.** The structures of hormones may be simple, such as those for epinephrine (adrenalin) and cortisone (a steroid), or they may be quite large and complex, such as growth hormone (see Connections 16.4). Most known hormones are steroids (lipids), amino-acidlike molecules, polypeptides, and proteins. Table 16.6 lists some key hormones, their biochemical classes, and one or more primary actions.

hormone
compound secreted by an organ or gland that controls metabolism

16.5 Determination of Protein Structure

The primary structure of a protein can be determined chemically and through molecular biology. The secondary, tertiary, and quaternary structures depend upon instrumental techniques such as X-ray crystallography, nuclear magnetic resonance, and computer modeling for high resolution. The upper levels of protein structure can also be probed chemically, and this information can be used in conjunction with the methods mentioned to get a complete picture of the molecule. For now we shall concentrate on the chemical methods of analysis.

A. Amino Acid Composition

The peptide bond, although stable under physiological conditions, can be broken through the process of acid or base hydrolysis. Subjected to boiling in 6M hydrochloric acid for 18–24 hours, most proteins will break down into their constituent amino acids. The sample of amino acids can then be separated through liquid chromatography. As the separated amino acids leave, or elute from, the column, they can be mixed with a color reagent and assayed with the use of a spectrophotomer (Figure 16.12).

Connections 16.4 www.prenhall.com/bailey

Growth Hormone

Somatotropin or growth hormone is a protein of molecular weight 22,000, having 191 amino acids. It is secreted by the pituitary gland and has its effects on most organs and tissues in the body, notably the muscles and bones. A deficiency in growth hormone results in short stature or the extreme of dwarfism. Hypersecretion will cause elongation of the bones and a coarsening of the skin and facial features, a condition known as gigantism or acromegaly.

Until the early 1980s the only source of functional growth hormone for humans was humans—human cadavers. It took approximately 40 human pituitaries to supply the hormone needed for one child for one year. In contrast to hormones like insulin, which can be harvested from pigs and sheep for human use, growth hormone is very species-specific. In addition there was a risk of developing a fatal brain disorder called Creutzfeldt-Jakob disease from contamination of brain tissue. The advent of recombinant DNA biotechnology opened the door for expressing human proteins in bacterial hosts. Human somatotropin has been available since 1985. It is injected several times per week, causing a child to catch up rapidly and maintain normal growth throughout adolescence.

Three areas of social concern have arisen over the availability of growth hormone. One is its illegal and undetectable use by athletes and a second is its administration to normal children in order to increase stature or athletic potential. The third issue concerns the use of bovine (beef) growth hormone (bGH) to increase milk production in cows. Consumer-group objections have discouraged the introduction of recombinant bGH into agriculture. There is no doubt that advances in the understanding and production of proteins present a challenge not only to science but also to the fabric of society.

Table 16.6 Hormones

Hormone	Source	Type of Biochemical Action
Polypeptide 1–50 amino acids		
• epinephrine and norepinephrine (modified tyrosine)	adrenal medulla	regulate stimulation of heart function, contraction of blood vessels and smooth muscle, control of metabolism
• thyroxine (an iodinated tyrosine dimer)	thyroid	stimulates general cell growth
• releasing and inhibiting factors	hypothalamus	affect secretions of the pituitary
• oxytocin	pituitary	stimulates mammary gland and uterine muscle
• vasopressin	pituitary	regulates blood pressure and water retention
• melanocyte-stimulating hormones	pituitary	control pigmentation
• corticotropin	pituitary	stimulates adrenal steroid synthesis
• calcitonin	thyroid	calcium and phosphorus metabolism
• glucagon	pancreas	increases blood glucose levels
• gastrin	GI tract	stimulates production of acid in stomach and pancreas
• vasoactive intestinal peptide	GI tract	inhibits acid and pepsin secretion
• motilin	GI tract	controls GI muscle
• somatostatin	GI tract	inhibits gastrin and glucagon secretion
• angiotensin	liver	regulates water retention and excretion
Proteins > 50 amino acids		
• insulin	pancreas	lowers blood glucose levels
• growth hormone	pituitary	stimulates general growth and metabolism
• prolactin	pituitary	controls milk secretion
• lutenizing and follicle-stimulating hormones	pituitary	stimulate male and female hormone and cell development
Steroids		
• testosterone	testes and adrenals	regulates male secondary sex characteristics and metabolism
• estradiol	ovaries	regulates female secondary sex characteristics and metabolism
• progesterone	ovaries and placenta	affects egg implantation and pregnancy
• glucocorticoids	adrenal cortex	control protein and carbohydrate metabolism, inflammation
• mineralocorticoids	adrenal cortex	regulate water and salt balance

Overall, the only data obtainable by these means are the types and amounts of the individual amino acids that make up the protein in question.

B. Sequence of Amino Acids—Determination of Primary Structure

There are several organic reagents, such as dansyl chloride, that can react with intact proteins to derivatize the N-terminal amino acid (see Figure 16.11). The "tagged" amino acid can then be separated and identified. However, the procedure destroys the rest of the polypeptide chain and only the N-terminus has been determined. It would be advantageous to have a method in which the rest of the chain remains intact during the course of the experimental procedure.

Edman degradation
nondestructive, sequential method of determining polypeptide primary structure

1. *Edman Degradation.* Pehr Edman was responsible for developing the sequential method that bears his name. The reagent is phenylisothiocyanate, or

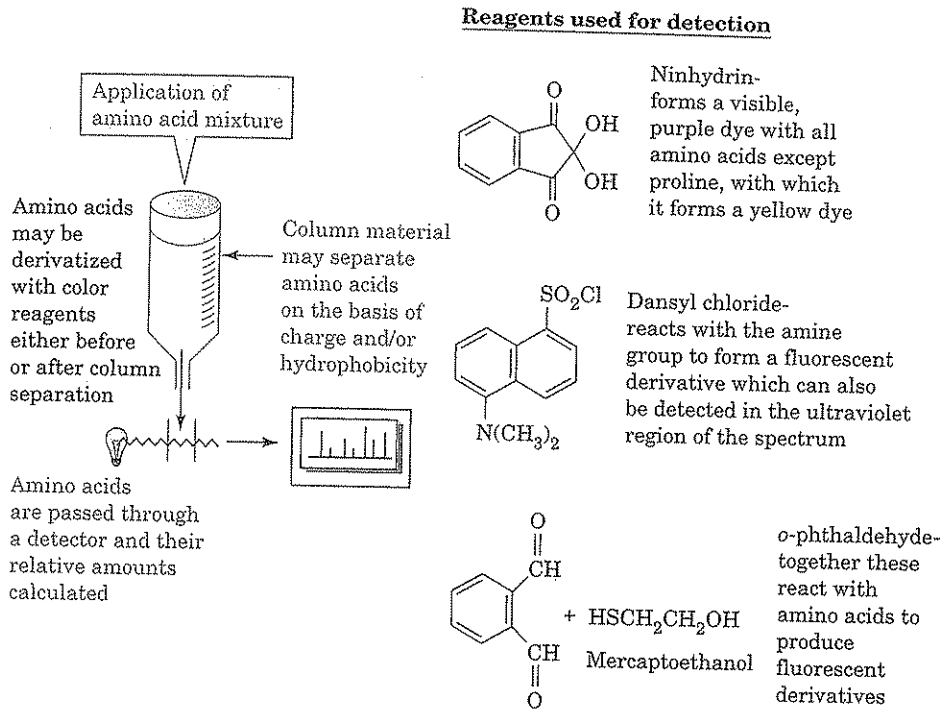
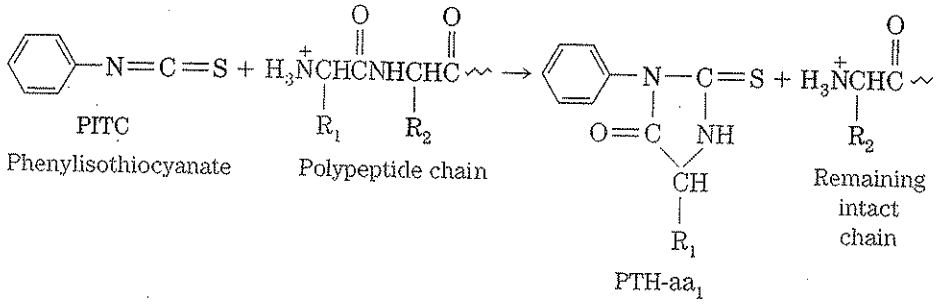


Figure 16.11 Schematic of amino acid analysis.

PITC. PITC derivatizes the N-terminal amino acid and leaves the rest of the chain sequence intact. After separating the PTH (phenylthiohydantoin) amino acid, the remaining chain can once again be treated by the Edman reagent.



The PTH amino acids can be separated on a chromatographic column and identified by their ultraviolet absorption spectra.

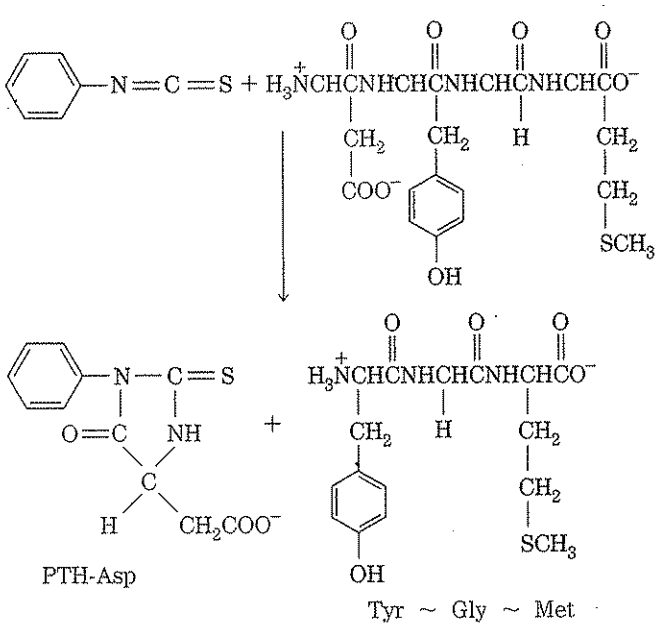
GETTING INVOLVED

Example 16.4

Draw the products of one cycle of the Edman degradation with the polypeptide, Asp ~ Tyr ~ Gly ~ Met.

Solution

The Edman reagent, phenylisothiocyanate, reacts with the N-terminal amino group of Asp and leaves the Tyr ~ Gly ~ Met tripeptide intact.



Problem 16.19

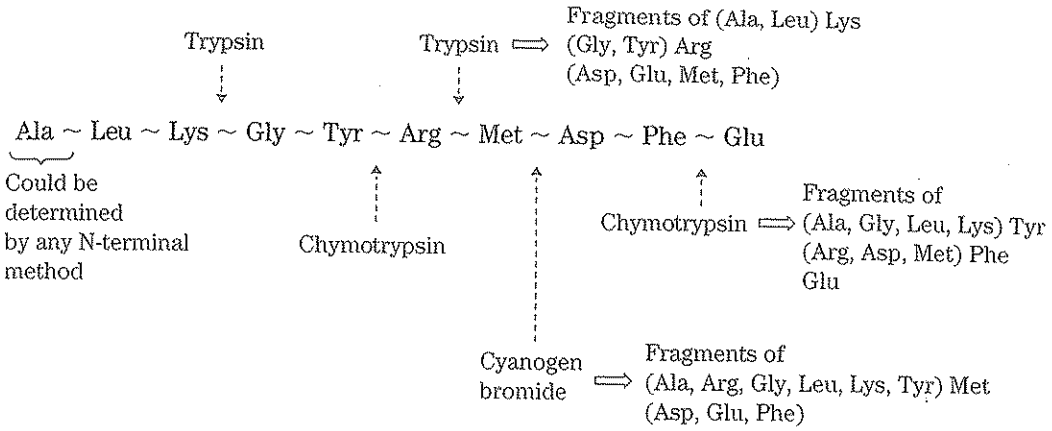
Draw the products of two more cycles of the Edman degradation on the tripeptide remaining in Example 16.4.

Problem 16.20

Assume that an N-terminal sequential method has an 85% yield at each of five steps in degrading a polypeptide chain. What is the theoretical yield of the desired amino acid six positions from the N-terminus?

2. ***Fragmenting the Chain.*** The Edman degradation is limited in terms of the length of chain it can successfully sequence, as well as the types of amino acids that can be readily derivatized. Therefore, it is necessary to fragment a long protein chain into pieces manageable for the sequencing routine. There are chemical reagents to do this, such as cyanogen bromide, which breaks the chain at methionine. The easiest and most specific cleavages can be effected by enzymes.

Trypsin, an intestinal protease (peptide bond hydrolase), has a specificity for breaking peptide bonds in a chain at the carboxy end of basic amino acids, that is, lysines and arginines. Chymotrypsin, also found in the small intestine, will hydrolyze peptide bonds contributed by the aromatic, hydrophobic amino acids—phenylalanine, tyrosine, and tryptophan. By performing digestions of the protein to be analyzed with each of these enzymes, perhaps doing an additional chemical cleavage, then separating and finding the amino acid content of the resulting peptides, an overlapping picture of the primary structure can be ascertained.



Carboxypeptidase and aminopeptidase are enzymes that will cleave amino acids sequentially from the C- and N-termini, respectively. Timed assays of the released products are taken. This may seem a convenience at first, but incomplete hydrolysis at one or more steps in the breakdown will contaminate subsequent releases of amino acids farther on in the chain.

GETTING INVOLVED

Example 16.5

A hexapeptide was analyzed by using the organic and enzymatic methods described. The results are shown below. Find the primary sequence of the hexapeptide.

(a) Acid hydrolysis and amino acid analysis gave the following content (subscripts refer to the relative quantities of the amino acids; the listing is alphabetical):

$$Gly_2, His_1, Ile_1, Phe_1$$

(b) One cycle of the Edman degradation produced PTH-Gly.

(c) Digestion of the hexapeptide with trypsin produced two fragments with the following amino acid compositions:

Fragment 1: Gly_1, Ile_1, Phe_1 *Fragment 2:* Gly_1, His_1, Lys_1

(d) Timed carboxypeptidase digestion of the hexapeptide gave a release of

$$Ile > Phe > Gly$$

Solution

- Part (a) tells us the amino acid content only.
- In (b) we find out the identity of ht N-terminal amino acid—Gly.
- For part (c) remember that trypsin cleaves the chain at the carboxyl end of Lys and Arg. The fact that two fragments were found indicates that the Lys is somewhere inside of the peptide and not at the C-terminal. Notice that Lys cannot be the N-terminal because we have already identified the N-terminal as Gly *and* we did not find free Lys.
- Therefore, we have the sequence of the first three amino acids—Gly ~ His ~ Lys.
- The carboxypeptidase digestion releases amino acids in sequence from the C-terminal. Therefore, Ile must be the C-terminus preceded by Phe; Phe is proceeded by Gly. The complete sequence is Gly ~ His ~ Lys ~ Gly ~ Phe ~ Ile.

Problem 16.21

What are the number of fragments and their amino acid composition if chymotrypsin digestion of the hexapeptide in Example 16.5 is used? What are the number and composition of fragments if trypsin digestion is followed by chymotrypsin?

See related problems 16.31, 16.32, 16.33.

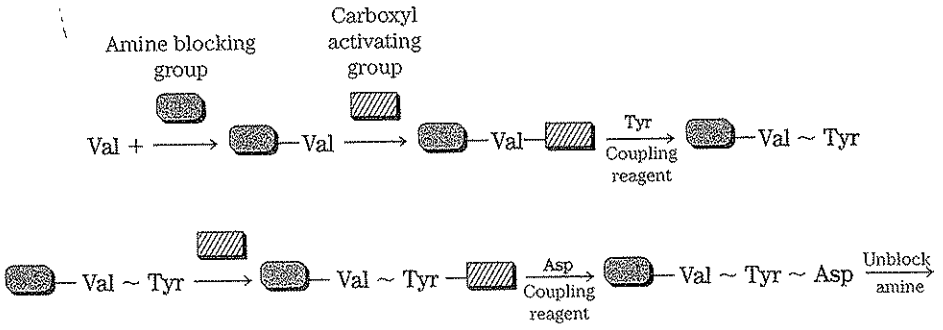
16.6 Organic Synthesis of Polypeptides

The importance of polypeptides and proteins has been an impetus to attempt to synthesize them for practical pharmacological purposes and for study.

A. General Considerations

Making something as simple as the tripeptide Val ~ Tyr ~ Asp does not involve just mixing the three amino acids together. Even if they could react, this would give a mixture of polypeptides: the Val ~ Tyr ~ Asp desired as well as Tyr ~ Asp ~ Val, Asp ~ Val ~ Tyr, Asp ~ Tyr ~ Val, Asp ~Asp ~ Tyr, $(Val)_3$, etc.

Then we must consider that the carboxyl groups are not reactive enough to form peptide bonds readily. Using an activating group such as an acid chloride greatly enhances the carbonyl reactivity. In addition, the amine group of the amino acids that you do not wish to react must be derivatized reversibly. Finally you must add the amino acids sequentially, isolating the first dipeptide product before putting in the third amino acid. A generic scheme for synthesizing the tripeptide might be as follows:



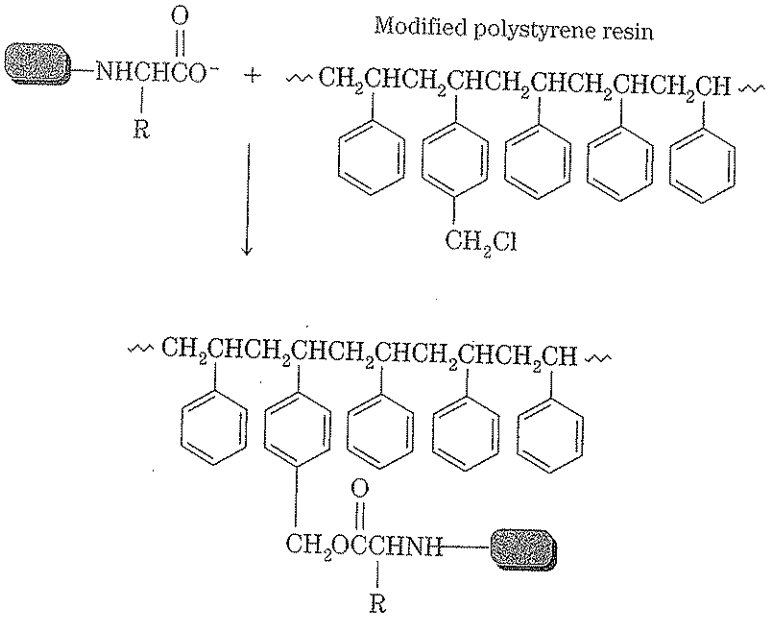
This procedure involves an extensive set of blocking, activating, coupling, and deblocking steps with purification of the desired intermediates along the route. Reactive R groups must be protected and then deblocked in a similar manner. The more amino acids in the polypeptide, the more steps, the lower the yield.

B. Solid-State Synthesis

Merrifield synthesis
method of synthesizing polypeptides using a solid phase

Dr. R. Bruce **Merrifield** proposed a novel method of synthesis in 1965. For this discovery, he was awarded the Nobel Prize in 1984. The procedure uses

a solid polystyrene resin in which about 10% of the aromatic rings have been derivatized with chloromethyl groups. The carboxyl group of an amino acid reacts with this group via an S_{N2} mechanism to become covalently attached to the resin.



The other desired amino acids are coupled to it; all reagents and solvents can be washed over the growing polypeptide chain as it is held on a solid support. This method works so efficiently that it has been automated. Two common reagents are the *t*-butyloxycarbonyl amino protecting group (Boc) and the coupling agent dicyclohexylcarbodiimide (DCC). The entire process is illustrated in Figure 16.12.

There is a limit to the number and types of amino acids that can be put together in this way. For long chains, that is, for proteins, a biosynthetic process using DNA as a guide is more specific, accurate, and efficient.

GETTING INVOLVED

Problem 16.22

Write out the possible combinations of any four amino acids, assuming that each occurs only once in a tetrapeptide.

Protein chemistry is a complex and fascinating field related to both biology and organic chemistry. It requires research endeavors that involve the talents and coordination of almost every aspect of the physical and medical sciences. We hope the background given here will encourage those interested to study further in the area of biochemistry.

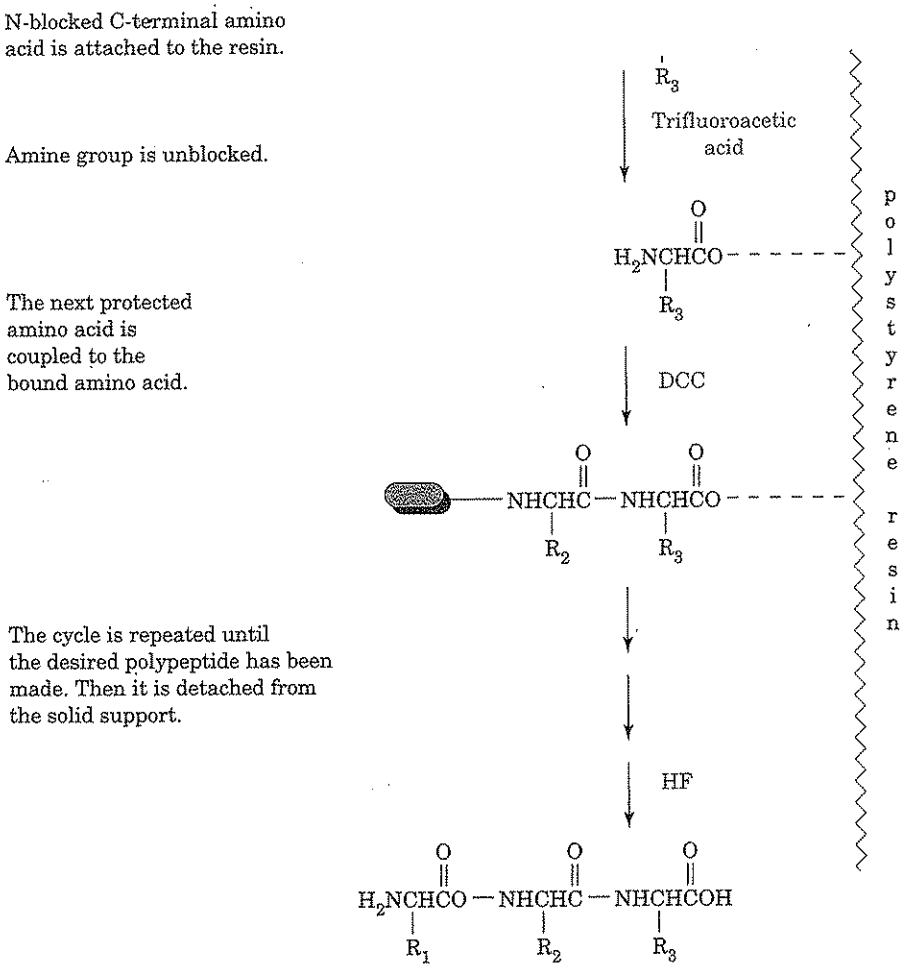


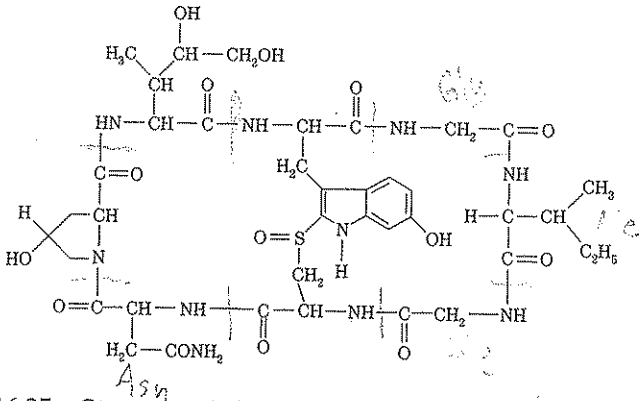
Figure 16.12 The Merrifield solid-phase synthesis of polypeptides.

Problems

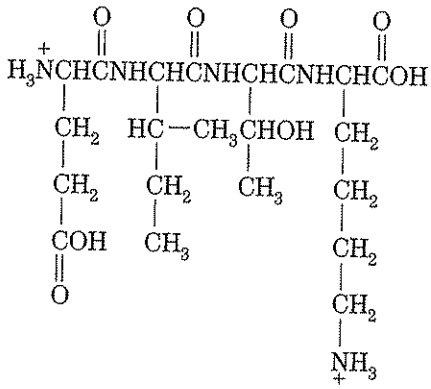
16.23 **Structure:** Identify the amino acids having the following characteristics:

(a) optical inactivity
(b) a phenolic group
(c) involvement in covalent bridging
(d) two optical isomers
(e) responsibility for bending a peptide chain and "breaking" a helical structure
(f) hydrogen-bonding through an R side-chain group
(g) more than two possible optical isomers

16.24 **Structure:** α-Amanitine is a polypeptide analogue that is the deadly component of a type of poisonous mushroom, *Amanita phalloides*. From its structure, try to identify the component of amino acids and any novel linkage (besides the α-aminocarboxyl peptide bond).



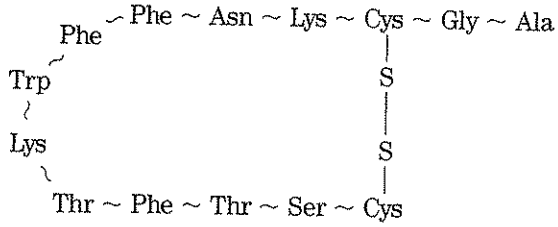
16.25 **Structure:** What are the products of the acid-catalyzed hydrolysis of the following peptide? Name the resulting amino acids, using the three-letter abbreviations.



16.26 **Structure:** Find the pI of the following polypeptides:

(a) met-enkephalin—an opiate neurotransmitter:
Tyr ~ Gly ~ Gly ~ Phe ~ Met

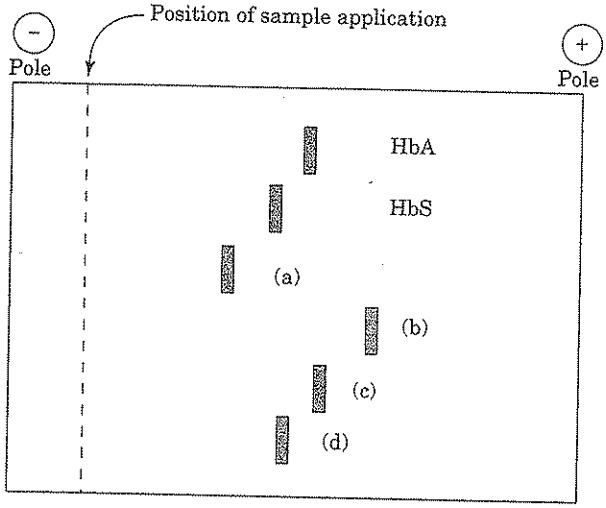
(b) somatostatin (growth hormone inhibiting factor)



16.27 **Structure:** Histones are proteins associated with nucleic acids. As phosphoric acid derivatives, nucleic acids have a negative charge under physiological pH conditions. What should be the net charge on the histones? Which amino acids should be found to a large extent in the primary structures of histones?

16.28 **Structure:** About 400 variations have been identified in the primary structure of hemoglobin. Some of these variations are conservative, that is, they will not make a difference to the physical properties or functions of the molecule. Others are nonconservative and can be fatal. Shown below is a diagram of a pH 8.0 electrophoresis of normal hemoglobin (HbA) and five variants, including sickle cell hemoglobin. Using the changes listed in the following table and the relative migrations of the variants at pH 8.0, match the variant to its position of the electrophoretogram. HbS is given as an example (remember that there are two α and β chains).

Hb variant	Changes in Primary Sequence				
	Chain	Position from N-Terminus	Amino Acid in HBA	Amino Acid in Variant	
S	β	6	Glu	Val	change of +2
C	β	6	Glu	Lys	
Chesapeake	α	92	Arg	Leu	
Hasharon	α	47	Asp	His	
Koln	β	98	Val	Met	



16.29 **Protein Structure:** Where do the following terms fit in a protein's hierarchy of structure (as 1°, 2°, 3°, or 4°)?

(a) the α and β subunits of hemoglobin
(b) Phe-Val side-chain interactions
(c) intrachain hydrogen-bonding
(d) linear sequence of amino acids
(e) salt bridges
(f) disulfide bridges

16.30 **Protein Structure:** What type of tertiary interaction might the side chains of each of the following pairs of amino acids be capable of?

(a) Ser and His **(b)** Phe and Leu
(c) Arg and Glu **(d)** Thr and Val

16.31 **Peptide Sequence:** Draw out the structures for end products of three cycles of Edman degradation for the polypeptide Leu ~ Met ~ His ~ Ser.

16.32 **Peptide Sequence:** A polypeptide, on acid hydrolysis, contained the amino acids Arg (1), Ala (1), Ile (1), Leu (2), Lys (1), Phe (2), Tyr (1). Treating the intact peptide with dansyl chloride and subsequent hydrolysis gave dansyl-Leu. Reaction with carboxypeptidase gave varying amounts of free amino acids, Phe > Leu > Ala. Digestion of the intact polypeptide with trypsin gave the following fragments: Tyr ~ Ile ~ Phe ~ Lys, Leu ~ Arg, and Ala ~ Leu ~ Phe. Chymotryptic treatment of the intact polypeptide produced Ile ~ Phe, Lys ~ Ala ~ Leu ~ Phe, and Leu ~ Arg ~ Tyr. What is the sequence of the nonapeptide? (*Note:* The lines between amino acids represent a peptide bond.)

16.33 **Peptide Sequence:** A polypeptide with the indicated amino acid composition (listed alphabetically) was analyzed as shown below. What is the primary sequence of the peptide?