2.3 Protein Structure

Learning Objectives

- Draw the basic structure of an amino acid, specifying the amino group, carboxyl group, alpha carbon, and R group
- For each amino acid, tell whether it is polar, nonpolar, acidic, or basic
- Draw the reaction describing the formation of a peptide bond
- Describe what is meant by the primary structure of a protein
- Describe what is meant by the secondary, tertiary, and quaternary structure of a protein
- Describe the four kinds of noncovalent interactions that stabilize protein structure
- Define posttranslational modification
- List four major classes of posttranslational modification
- List the major kinds of chemical modification of proteins
- List the amino acids involved in glycosylation and its overall function in proteins
- Describe the consequence of gamma carboxylation of proteins and name the vitamin involved
- Describe two distinct ways of varying the activity of proteins in cells
- Describe three distinct ways membrane proteins can be anchored in the membrane by covalent modification

AMINO ACIDS MAKE UP PROTEINS

In Chapter 2.2, we described how proteins are made on ribosomes by linking amino acids together in long chains. Which amino acids make up a specific protein, and in which order, is determined by the sequence of triplet nucleotide codons residing on the mRNA, which in turn is produced on a DNA template in the nucleus. Because these constituent amino acids determine the detailed shape of the protein surface, we should find the key to protein activity in the three-dimensional arrangement of these amino acids.

The general chemical structure of the amino acids is shown in Figure 2.3.1. The amino acids are named for the two functional groups each of them possesses: an **amino group** (−NH) and a **carboxylic acid group** (−COOH). These two groups have widely different reactivities. The amino group is basic and will be positively charged at neutral pH, forming an ammonium ion ($-NH_3^+$). The carboxyl group is acidic and will form a negatively charged carboxyl ion ($-COO^-$) at neutral pH. Thus at neutral pH, many amino acids will have a positive charge on the amino end and a negative charge on the carboxyl end. Compounds possessing both positive and negative charges simultaneously are called **zwitterions**, as shown in Figure 2.3.2.

Although the amino and carboxyl groups are important for the structure of proteins, the tremendous variety of structure and function of proteins is produced by the R groups (R stands for "residue" and refers to the part of the amino acid other than the amino or carboxyl group covalently bonded to the α carbon). There are 20 different amino acids differing only in their R groups. The R groups confer properties on the amino acids that classify them as **nonpolar**, **polar**, **basic**, or **acidic**. Each of these groups is shown in Figures 2.3.3–2.3.5. Biochemists often use a three-letter abbreviation for the amino acids or a one-letter code for a further shorthand. These abbreviations are shown under each structural formula (see Figure 2.3.6).

As their name implies, the nonpolar amino acids contain side groups that are nonpolar and therefore not attracted to water. Because water is so attracted to itself, these groups naturally are repulsed by the water and are said to be hydrophobic, or water hating. For glycine and alanine, the nonpolar side groups are so small that the amino acid has little preference for either a watery (hydrophilic) or nonwatery (hydrophobic) environment. Phenylalanine and tryptophan are highly hydrophobic and seek environments away from water.

HYDROPHOBIC INTERACTIONS CAN BE ASSESSED FROM THE PARTITION COEFFICIENT

We can estimate the strength of hydrophobic interactions by measuring the partitioning of a material between the water phase and an immiscible solvent such as *N*-octanol or ethyl acetate, whose properties may be regarded as being similar to the interior of protein away from the water phase. The partition coefficient is

$$k_s = \frac{[x]_{\text{organic phase}}}{[x]_{\text{water phase}}} \quad [2.3.1]$$

DOI: http://dx.doi.org/10.1016/B978-0-12-800883-6.00012-4

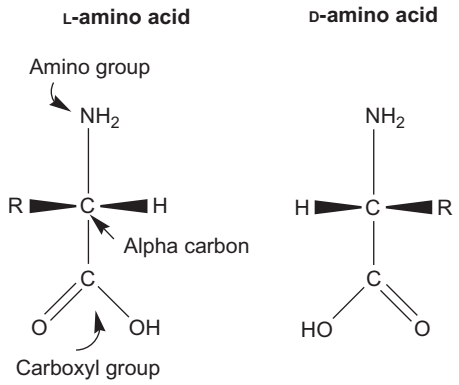

FIGURE 2.3.1 Chemical structure of the amino acids. Amino acids consist of a central carbon, the alpha carbon, to which are bonded a carboxyl group, an amino group, a hydrogen atom, and a variable group, called R. Because the alpha carbon has four different groups bonded to it, it is an asymmetric carbon that is capable of stereoisomerism. All of the amino acids can be made in L- and D-form (L and D originally described the ability of compounds to rotate the plane of polarized light; L for *levo*, meaning "left," and D for *dextro* meaning "right." The symbols L and D for the amino acids do not actually refer to the direction of rotation of the plane of polarized light). All naturally occurring amino acids in higher organisms are the L-form. The L and D forms are mirror images of each other and are not superimposable.

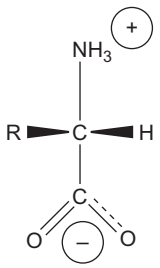

FIGURE 2.3.2 Zwitterion form of an amino acid. At neutral pH (pH = 7), the carboxyl group is dissociated to form the anionic $-COO^-$ group, and the amino group binds a H^+ ion to form $-NH_3^+$. Thus both ends of the amino acid are charged but with opposite polarity.

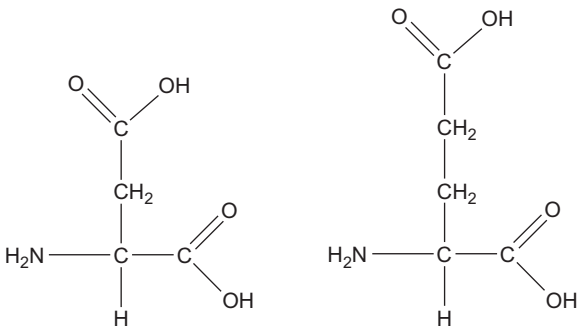

FIGURE 2.3.3 The acidic amino acids aspartic acid and glutamic acid. The R group in both cases contains another carboxylic acid group that dissociates to form an H^+ ion and a negatively charged carboxyl group ($-COO^-$). The presence of this group classifies these amino acids as acidic amino acids. The negatively charged carboxyl group imparts a negative charge to the region of any protein that contains these groups.

where the concentrations are the equilibrium concentrations, as shown in Figure 2.3.7. This equilibrium condition obeys Eqn [1.7.32]:

$$\Delta G^0 = -RT \ln K_{eq}$$

[2.3.2]

$$\Delta G_T^0 = -RT \ln \frac{[x]_{\text{organic phase}}}{[x]_{\text{water phase}}}$$

where ΔG_T^0 is the standard Gibbs free energy change for transfer of substance X from water to the organic phase. Although this can represent the strength of hydrophobic interaction of a small molecule, it cannot reliably predict the behavior of a polymer of the material or of the material when it participates in a heteropolymer such as a protein. This is true for amino acids in particular because some of their functional groups that determine the octanol/water partition are altered when the amino acids are linked up to make proteins. Nevertheless, parts of molecules can be described as being hydrophobic or hydrophilic, depending on whether or not they would partition themselves into the organic phase.

Hydrophobicity has two components: the "squeezing out" of water-insoluble components due to the attraction of water for itself and the self-association of nonpolar materials due to dipole–dipole interactions, dipole–induced-dipole, and induced-dipole–induced-dipole interactions (London dispersion forces) discussed in Chapter 1.4. Water repels nonpolar materials because the surface of the nonpolar material cannot form hydrogen bonds with the water, and its shape therefore reduces the number of hydrogen bonds that the water can form at the surface—the water molecules at the surface are in a higher energetic state, having some of their hydrogen bonds broken. Thus it takes energy to insert a nonpolar material into the water phase.

The amino acids can be classified solely on the basis of their hydrophobicity or hydrophilicity. In this case, we have three categories: the hydrophilic amino acids include aspartic acid, glutamic acid, lysine, and arginine; the hydrophobic amino acids include valine, leucine, isoleucine, methionine, phenylalanine, and tryptophan; the "neutral" amino acids, those that are neither strongly hydrophilic nor strongly hydrophobic, include glycine, alanine, serine, histidine, proline, threonine, cysteine, glutamine, and asparagine.

PEPTIDE BONDS LINK AMINO ACIDS TOGETHER IN PROTEINS

Proteins are formed as a linear, unbranched chain of amino acids. The amino acids are covalently linked by a **peptide bond** formed between the amino group of one amino acid and the carboxyl group of the next. The formation of the peptide bond is a dehydration reaction, as shown in Figure 2.3.8.

Cells make proteins by the sequential addition of amino acids to the carboxyl terminus of the growing chain. This is accomplished on the ribosome when specific tRNAs (transfer RNAs) with their specific bound amino

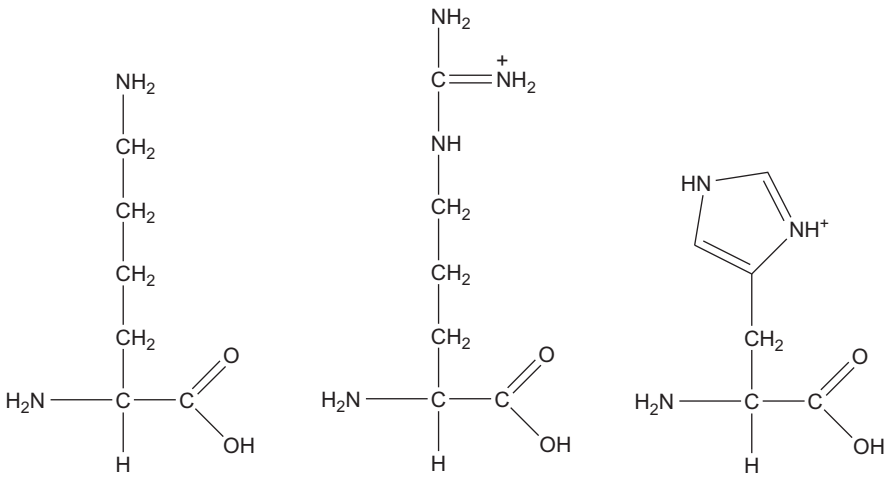

FIGURE 2.3.4 The chemical structures of the basic amino acids, lysine, arginine, and histidine. These contain ionizable NH groups so that at neutral pH, these residues would contribute positive charge to the protein.

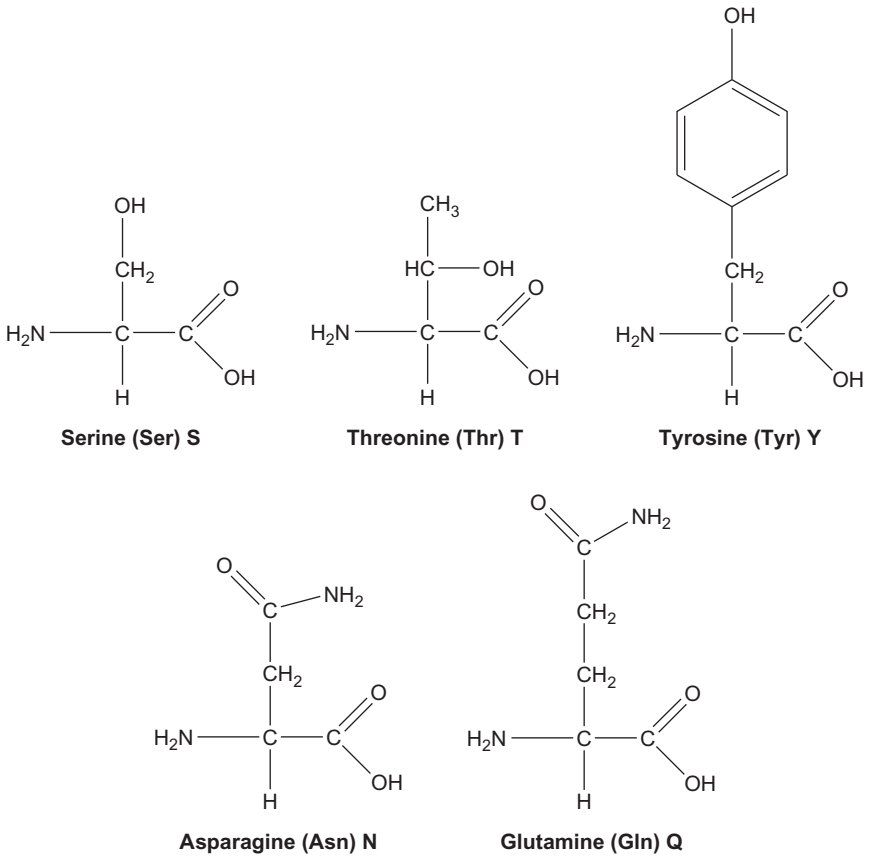

FIGURE 2.3.5 The chemical structure of the polar amino acids. The polar amino acids contain groups that can form hydrogen bonds with water. These groups are soluble in water. As a result, those portions of proteins which contain large numbers of polar residues will usually be exposed to water. They help solubilize proteins in solution, so we might expect soluble proteins to be coated with polar, acid, or basic R groups.

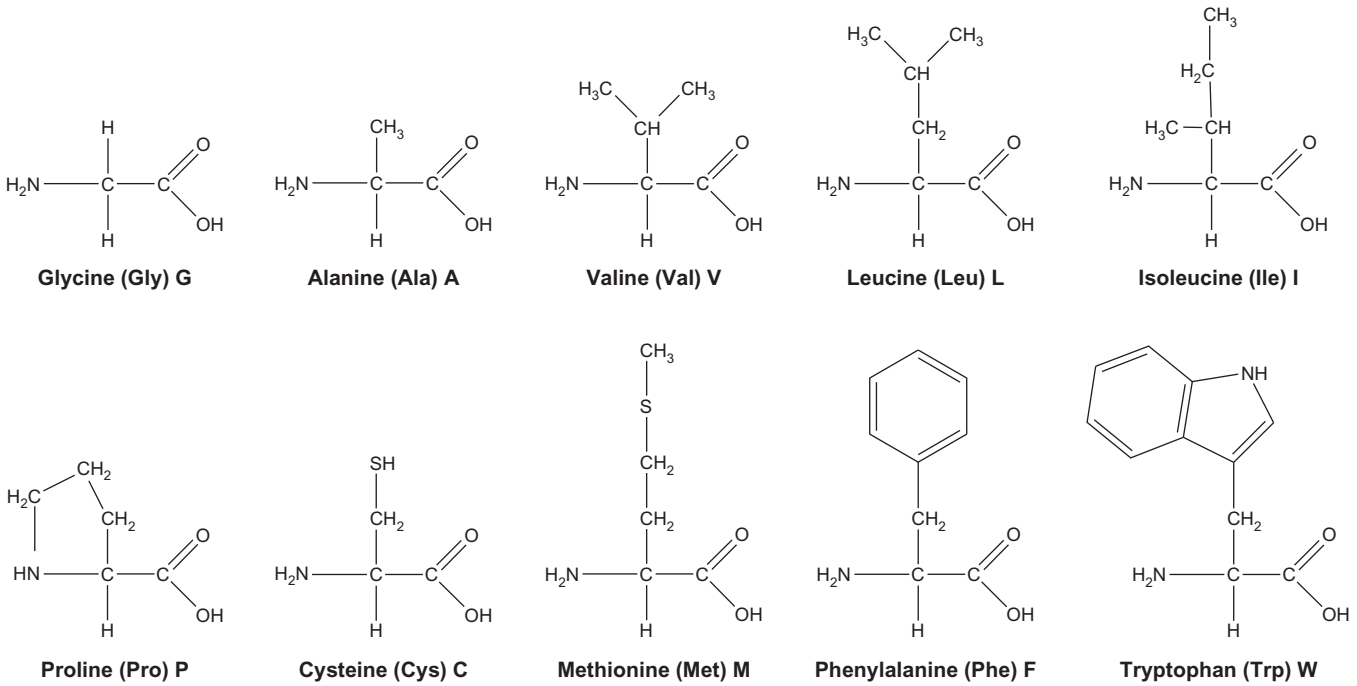

FIGURE 2.3.6 Chemical structures of the nonpolar amino acids.

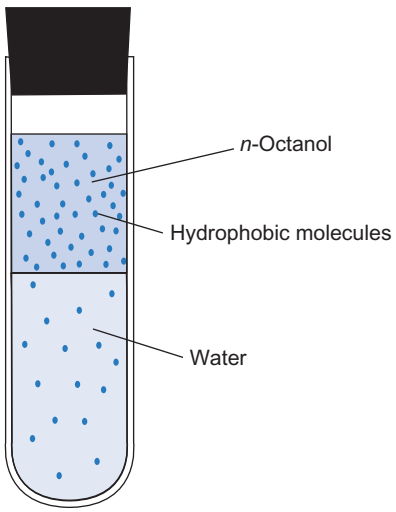

FIGURE 2.3.7 Assessment of hydrophobicity by the partition coefficient between octanol and water. A hydrophobic molecule is dissolved in *n*-octanol and then an aliquot of the solution is placed in contact with water and shaken. The material will distribute itself between the two phases. The ratio of the concentrations at equilibrium in the organic phase to the water phase defines the partition coefficient, which is an equilibrium constant that can be used to calculate the free energy of transfer from the organic phase to the water phase. Hydrophobic materials will have a higher concentration in the *n*-octanol phase; hydrophilic materials will have a higher concentration in the water phase.

acid bind to the appropriate triplet codon of the mRNA. The amino acid is attached to the tRNA via its carboxyl group. When the peptide bond is formed, the entire growing chain is transferred to the free amino terminal of the next amino acid. The ribosome then moves one frame (nucleotide triplet) and the process is repeated (see Chapter 2.2).

PROTEIN FUNCTION CENTERS ON THEIR ABILITY TO FORM REACTIVE SURFACES

Proteins perform a variety of functions in cells, and these can be broadly classified as **structural**, **catalytic**, **transport**, or **regulatory**. All of these functions require the surface of the protein to interact with other surfaces.

As discussed in Chapter 1.5, enzymes provide a surface on which biochemical reactions can occur. These reactions can occur in the water phase, but only slowly because of the large activation energy required. Binding of the substrates to the enzyme surface alters the path of the reaction, enabling it to proceed more quickly. This lower energy pathway is produced by the interaction of the protein surface with the substrates of the biochemical reaction. The details of the protein surface enable catalysis, the speeding up of a reaction without appearing in the overall stoichiometry of the reaction. The surface of the protein must closely match the surface of the substrates, which means that it cannot match other substrates well, because the protein surface cannot match two different surfaces. Thus the protein surface simultaneously enables catalysis while it confers **specificity**—the enzyme works only with specific substrates.

In the same way that protein enzymes interact with biochemical substrates on the surfaces of the proteins, structural proteins interact with other components by virtue of their surfaces. The proteins that make up connective tissue are sticky. They bind to themselves and to a variety of other proteins. Their surfaces make them sticky. Closely matching surfaces allow proteins to interact with other proteins, thereby allowing one protein to

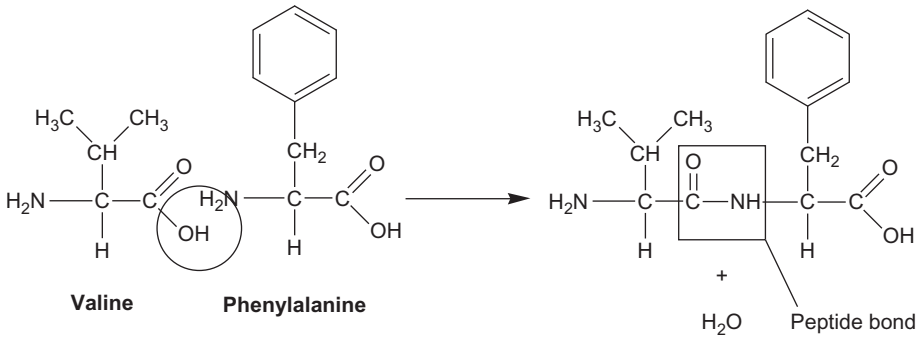

FIGURE 2.3.8 Formation of a peptide bond between two amino acids. The overall reaction is shown. The actual reaction involves many intermediary steps, catalyzed by enzymes.

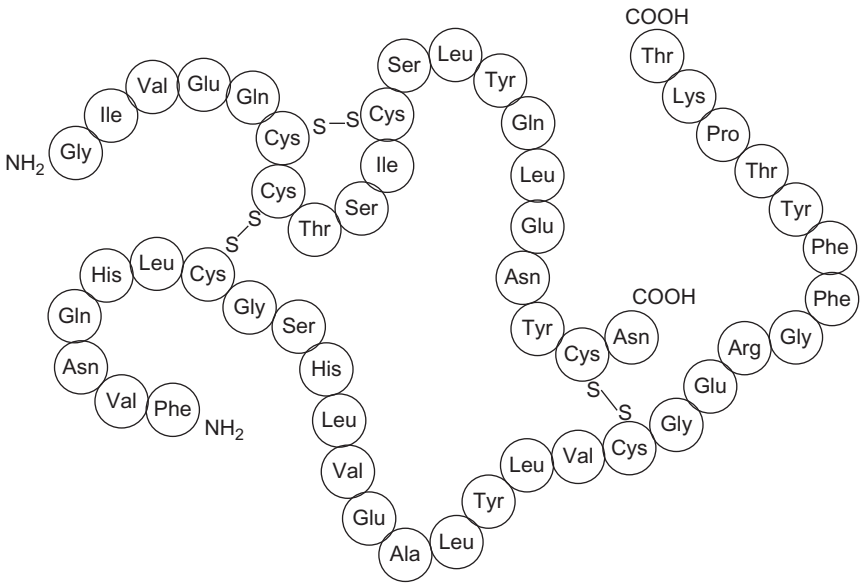

FIGURE 2.3.9 Primary structure of insulin. Insulin is a protein hormone secreted by the pancreas in response to high blood glucose. It causes peripheral tissues to take up the glucose, thereby lowering the plasma glucose concentration back toward normal. Insulin is synthesized as a larger, single polypeptide chain that is modified by excision of two peptides to form the A chain, with 21 amino acids, and the B chain with 30. This is one example of posttranslational modification that occurs with many proteins.

regulate another through the binding together of matching surfaces.

THERE ARE FOUR LEVELS OF DESCRIPTION FOR PROTEIN STRUCTURE

Proteins take on their remarkably diverse functions because they can fold to form specific shapes and because some of the amino acids that make up the proteins have inherent chemical reactivity. These shapes provide a surface for the binding of materials, and this binding originates all of the functions of proteins. The shape of proteins has four levels of description:

1. Primary
2. Secondary
3. Tertiary
4. Quaternary.

THE PRIMARY STRUCTURE OF A PROTEIN IS ITS AMINO ACID SEQUENCE

The primary structure of a protein refers to its amino acid sequence. It defines the chemical connectivity of the constituent amino acids. In 1953, Sanger determined the amino acid sequence of the A and B chain of insulin. This was the first protein whose entire amino acid sequence was determined and the work was a major milestone in biochemistry. The chemical structure of insulin is shown in Figure 2.3.9.

THE SECONDARY STRUCTURE OF PROTEIN REFERS TO THE FOLDING OF AMINO ACIDS IN ADJACENT SEQUENCES

Proteins have three regular secondary structures: the α-helix, the β-sheet, and the β-turn. These are complicated, three-dimensional structures. The α-helix was first postulated by Linus Pauling, Robert Corey, and Herman

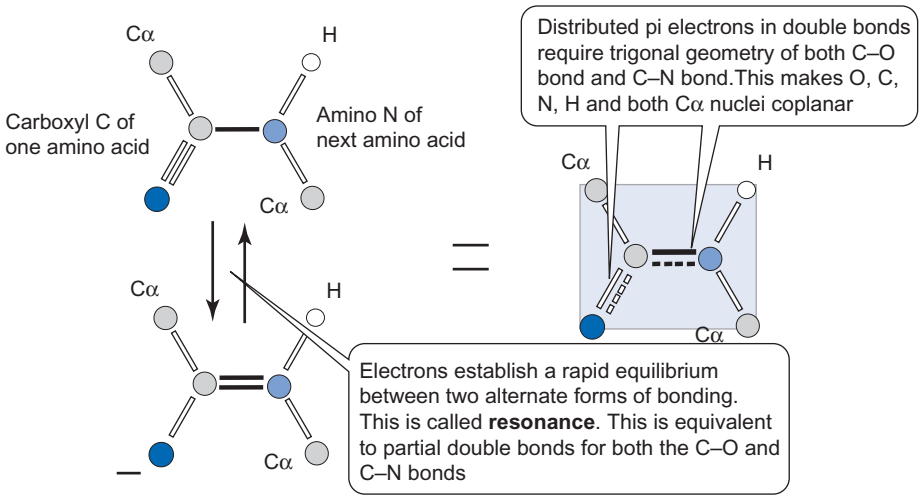

FIGURE 2.3.10 Planar peptide bond. The peptide bond involves a combination of an carboxyl group of one amino acid with the amino group of the next amino acid. The C–N bond and C = O bond **resonate**: the electrons alternate between a single C–N and double C = N bond, and simultaneously between a double C = O bond and a $C-O^-$ bond. This is equivalent to a distributed electron density over both bonds. This restricts free rotation about the C–N bond, locking all six nuclei into a single plane.

Branson (The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain, *Proc. Natl. Acad. Sci. USA* **37**:205–211, 1951) on the basis of a planar peptide bond and linear hydrogen bond length of 0.272 nm, and no requirement for an integral number of amino acids per turn of the helix. The planar nature of the peptide bond was key. Pauling realized that resonance between the carbonyl oxygen and the C–N bond would produce a partial double bond character to the C–N bond that would prohibit free rotation about its axis. The H bonding would be trigonal and therefore planar. The carboxyl C would be similarly planar, and because both the N and the carboxyl C were connected, all six nuclei connected to the peptide bond would be coplanar. This idea is shown in Figure 2.3.10.

Although the C–N bond cannot freely rotate, due to its partial double bond nature, the N–Cα bond is not so restricted. This Cα nucleus is connected to the next C that participates in a peptide bond, and this Cα–C bond can also rotate. The result is that the next peptide bond, which also forms a plane, is rotated relative to the first. Two dihedral angles are defined to describe this rotation: φ is the dihedral angle about the N–Cα bond and ψ is the dihedral angle about the Cα–C bond. These are shown diagrammatically in Figure 2.3.11.

Pauling and co-workers used the idea of the stiff plane of the peptide bond and the known length of the hydrogen bond to deduce the structure of the alpha helix, shown in Figure 2.3.12. In this structure, the polypeptide backbone traces a right-handed helix (going from the amino terminus to the carboxy terminus). Note that the C = O bonds point towards the carboxy terminus, where they hydrogen bond with the amino hydrogen. The protein forms a rod with the side groups of the amino acid sticking out more or less radially.

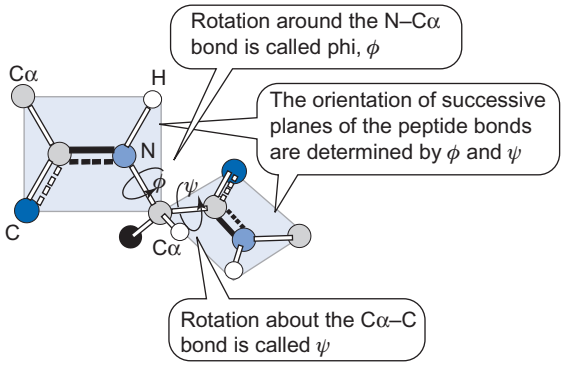

FIGURE 2.3.11 Rotation of successive peptide bonds. Each peptide bond defines a plane. Between each bond there is allowable rotation around the N–Cα bond (defined as φ) and rotation about the Cα–C bond (defined as ψ). The result is a change in direction of the polypeptide chain.

The second major secondary structure in proteins is the **beta sheet**, shown schematically in Figure 2.3.13. These result from sideways hydrogen bonding of a linear chain of amino acids whose peptide plane is bent at the Cα carbon. There are two ways for this to be accomplished: **parallel** beta sheets join segments of the polypeptide chain whose amino to carboxyl direction is going in the same direction. **Antiparallel** beta sheets join polypeptide chains of opposite N to C polarity.

Structures not fitting into these categories are often called "random coils," although they may not be random at all. There are four main principles involved in the formation of these secondary structures.

Hydrogen Bonding Stabilizes Structure

The polar centers of the carbonyl oxygen and the amide nitrogen are hydrogen bonded to other structures in the

protein, either another amide bond or a polar side group of an amino acid. These hydrogen bonds stabilize protein structure and are instrumental in forming the α-helix.

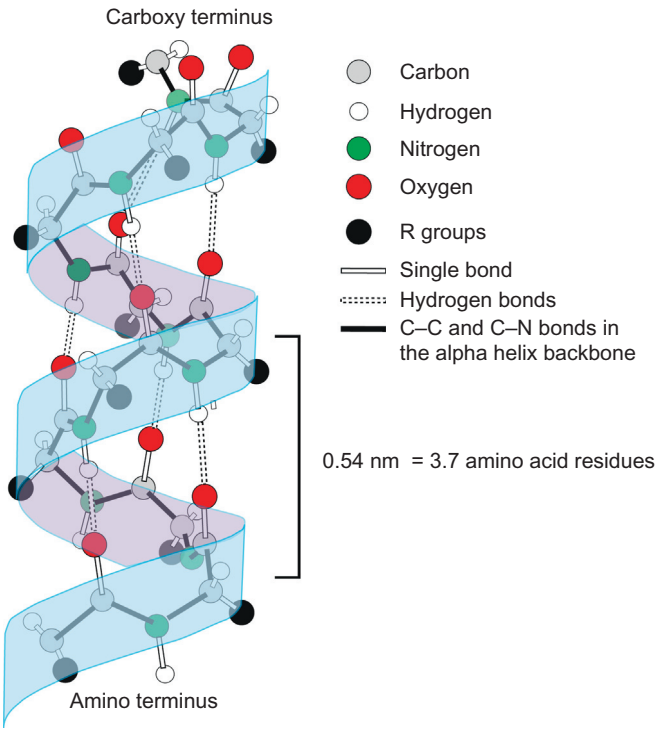

FIGURE 2.3.12 The alpha helix. The dark bonds indicate those in the polypeptide backbone that are directly involved in the peptide bonds. The dashed lines indicate hydrogen bonds that bond successive turns of the helix with each other. These hydrogen bonds connect the carbonyl of one amino acid with the amino hydrogen of the fourth amino acid down the chain.

Hydrophilic Groups Face Water; Hydrophobic Groups Face Away from Water

The protein will fold so that highly charged groups and polar groups are on the outside of the protein facing water. Here the interaction between these hydrophilic groups and the hydroxyl groups of water stabilizes the structure.

Side Groups Cannot Occupy the Same Space

This is a simple way of saying there is **steric hindrance** in the structure. Atoms in the amino acids exhibit repulsive forces when placed too close together. These forces ultimately arise from the interpenetration of atomic or molecular electron orbitals.

Electrostatic Interactions Stabilize Structures

Groups with opposite electric charge attract each other and this attractive force can stabilize the three-dimensional arrangement of amino acids. Similarly, groups with the same sign electric charge repel each other and this repulsive force can also stabilize the structure by preventing closer movement of the electrically charged groups.

TERTIARY STRUCTURE DEALS WITH THE THREE-DIMENSIONAL ARRANGEMENT OF ALL OF THE AMINO ACIDS

The tertiary structure of proteins deals with how the regional structures are put together in space. For example, the α-helices may be oriented parallel to each other or at right angles. So the tertiary structure refers to the folding of the different segments of helices, sheets, turns, and the remainder of the protein into its native three-dimensional structure. Commonly, membrane proteins are anchored in the membrane by hydrophobic

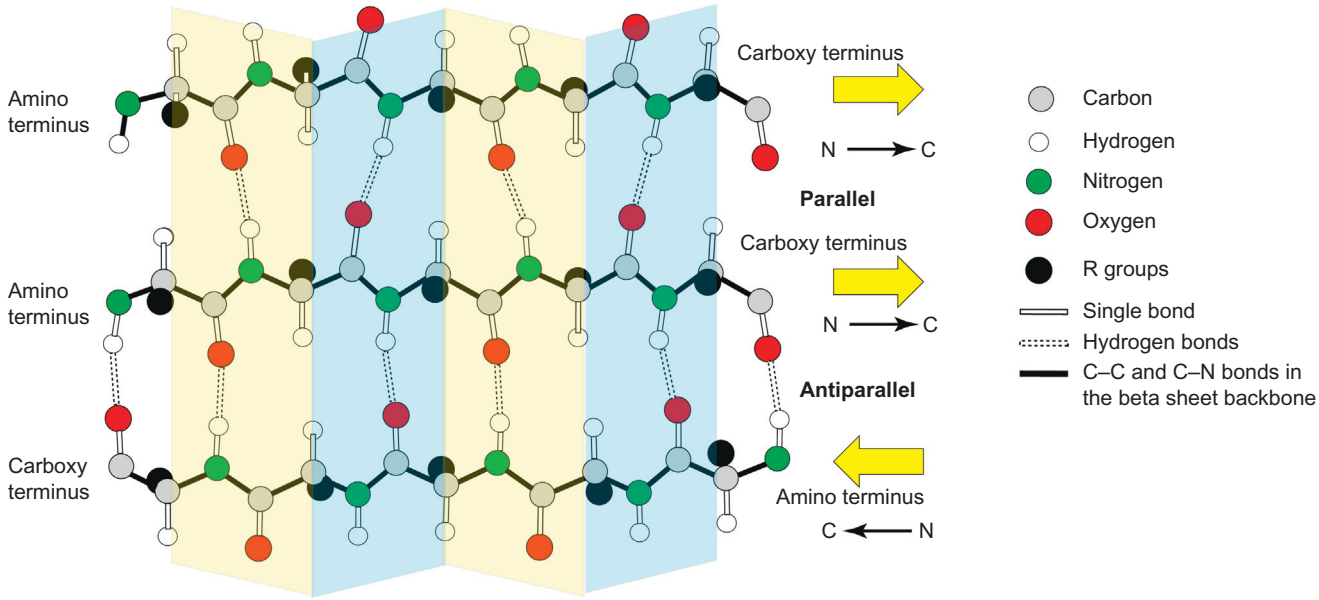

FIGURE 2.3.13 Structure of the beta sheet. Strings of amino acids are bonded laterally through hydrogen bonds. The arrangement can be between strings that have the same amino to carboxy orientation, called parallel, or between strings with opposite orientation, called antiparallel. The planar peptide bonds line up to form a kind of pleated sheet.

alpha helices that may be far removed in the primary sequence but closely apposed in the tertiary structure of the protein.

QUATERNARY STRUCTURE REFERS TO THE INTERACTION OF A PROTEIN WITH OTHER PROTEINS

The quaternary structure refers to the interaction of a protein with other proteins or other components of the cell. This association of the protein with other elements of the cell can alter its three-dimensional shape and its activities because the close proximity of another protein's surface can alter the shape of the protein. Many proteins exist in the cell in complex macromolecular assemblies in which quaternary structure enables or regulates the function of the component proteins.

POSTTRANSLATIONAL MODIFICATION REGULATES AND REFINES PROTEIN STRUCTURE AND FUNCTION

Proteins that come off the ribosome are not yet finished. The cell processes the newly made proteins in several ways, including:

- formation of disulfide bonds
- folding into the functional form
- cleaving the proteins at specific sites
- chemical modification.

PROTEIN DISULFIDE ISOMERASE CATALYZES DISULFIDE EXCHANGE

Proteins often are stabilized by disulfide bonds between cysteine residues in the protein. These cysteines are not necessarily close to each other on the primary sequence but must be close in the tertiary structure of the protein. The **protein disulfide isomerase** catalyzes the interconversion of disulfide bonds until the right ones are formed.

CHAPERONES AND CHAPERONINS HELP PROTEINS FOLD

As they are made on the ribosome, proteins begin to fold up. Sometimes the primary structure of the protein alone can determine the proper final shape, and **denaturing** the protein (adding materials or heat that disrupts its shape) is reversible. In other cases, proteins need help in determining their shape, and they are synthesized on a kind of scaffold that ensures that they fold properly. These scaffolds are generally other proteins called **chaperones**. There are two kinds of chaperones. **Molecular chaperones** bind to unfolded or partially unfolded proteins and stabilize their structure, preventing them from being degraded. **Chaperonins** directly facilitate the folding of proteins. Molecular chaperones are members of the **Hsp70** family of proteins; hsp means "heat shock protein" because these increase after heat stress to an animal that would denature proteins. Complexes of eight **Hsp60** molecules form a barrel-shaped chaperonin that aids protein folding.

PROTEOLYTIC CLEAVAGE

Cells make many proteins in precursor form with longer primary sequences than the finished product. Proteolytic cleavage forms the final protein by chopping off unwanted parts of the protein.

PROTEINS ARE CHEMICALLY MODIFIED IN A VARIETY OF WAYS

Cells chemically modify proteins after translation in a variety of ways, some of which are shown in Figure 2.3.14. These include:

- acetylation
- methylation

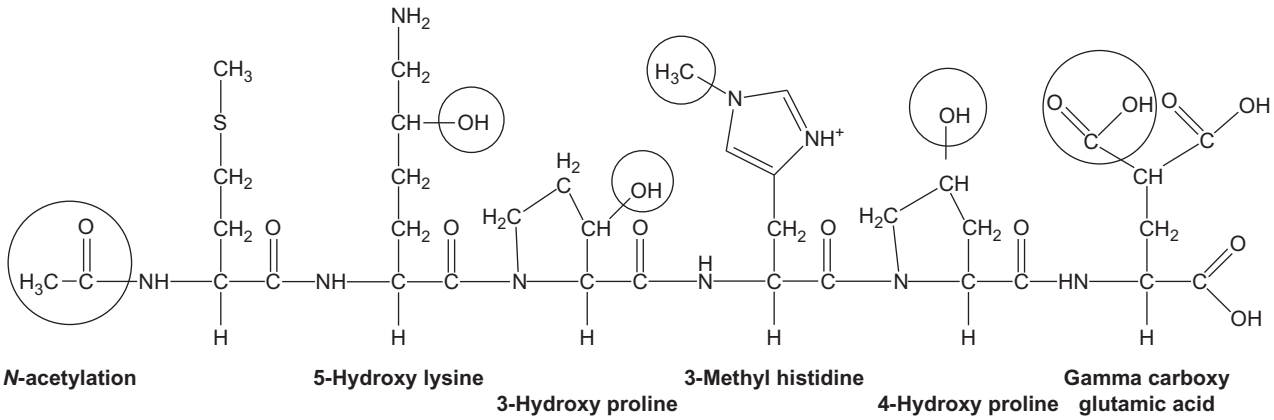

FIGURE 2.3.14 Examples of some chemical posttranslational modifications of proteins. The N-terminal amino acid of many proteins (about 80% of them) is acetylated. Lysine groups can also be acetylated (not shown). Acetylation may regulate the life span of proteins as nonacetylated proteins rapidly degrade. Hydroxylation of lysine in the 5 position occurs in collagen. In collagen, proline is converted to 3 hydroxyproline or 4 hydroxyproline. Histidine, particularly in actin, is methylated. Arginine can also be methylated (not shown) and this modification is part of the histone code (see Chapter 2.2). Glutamic acid residues in prothrombin (a protein involved in clotting of blood) and in bone proteins are carboxylated at the γ-side chain carbon to form gamma carboxy glutamic acid. The close proximity of the two carboxyl groups in gamma carboxy glutamic acid creates a binding site for calcium ions, and this confers Ca^{2+} sensitivity to the coagulation process. The gamma carboxylation reaction requires vitamin K, which is necessary for normal blood coagulation.

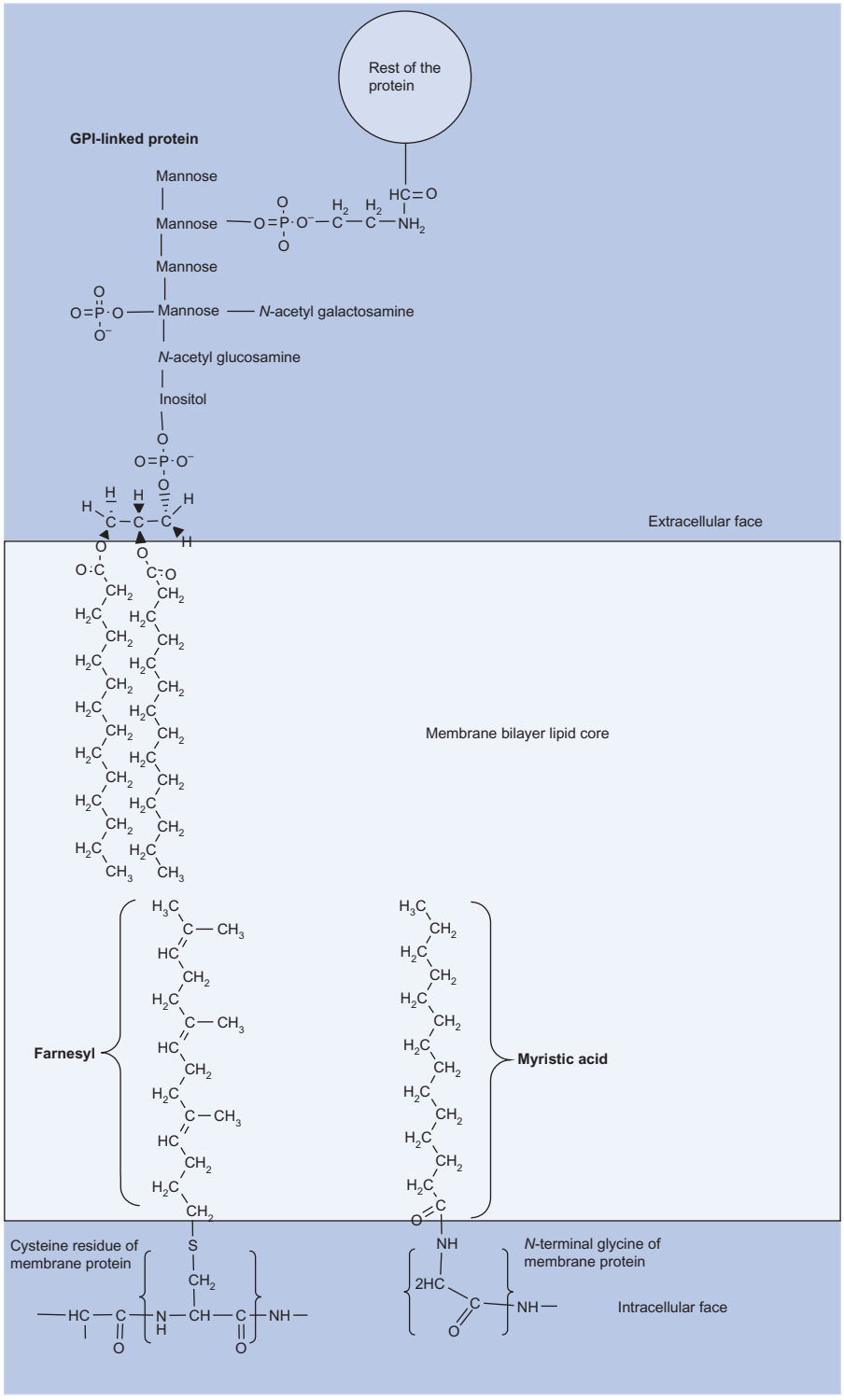

FIGURE 2.3.15 Anchors of membrane proteins. Some membrane proteins are anchored in the membrane by attachment of hydrophobic parts such as myristic acid or palmitic acid. Myristic acid is a 14-carbon hydrocarbon chain with a carboxyl group on one end that can covalently attach to the N-terminal glycine of membrane proteins. Palmitic acid is a 16-carbon saturated monocarboxylic fatty acid and attaches to proteins similar to myristic acid. The process of attaching myristic acid or palmitic acid is called myristoylation or palmitoylation. Farnesyl is a polymer of three 5-carbon units called isoprene. Farnesyl attaches covalently via a thioester bond to cysteine residues somewhere in the middle of the protein and helps anchor some proteins in the membranes. Other proteins link to **GPI** or glycosylphosphatidylinositol, a glycosylated membrane lipid that helps keep proteins in the membrane.

- hydroxylation
- gamma carboxylation
- glycosylation
- myristoylation or palmitoylation
- phosphorylation.

Proteins in the endoplasmic reticulum are often **glycosylated**, meaning that sugars or sugar derivatives are covalently attached to the proteins. *N*-linked glycosylation occurs when carbohydrate branches are added to the side chain NH_2 of asparagine through *N*-acetylglucosamine. *O*-glycosylation occurs on the side chain OH of serine, threonine, or hydroxylysine and the connecting carbohydrate is *N*-acetylgalactosamine.

Proteins stick in membranes because they are anchored there by a variety of posttranslational modifications, as shown in Figure 2.3.15.

PROTEIN ACTIVITY IS REGULATED BY THE NUMBER OF MOLECULES OR BY REVERSIBLE ACTIVATION/ INACTIVATION

The cell can alter the activity of its component proteins by altering the number of copies of the protein, or by activating or inactivating the proteins that are already present.

Regulating the number of protein molecules requires synthesis of new protein molecules or degradation of existing ones. This takes time and degrading existing proteins wastes energy. Reversible activation or inactivation of proteins can achieve rapid regulation that also conserves energy. Cells use phosphorylation/dephosphorylation of proteins to regulate their activities. Attachment of a phosphate group changes the charge on a local region of the protein, which alters its three-dimensional shape and changes its activity. Serine, threonine, and tyrosine residues are the targets for these phosphorylation reactions (see Figure 2.3.16). An enzyme that phosphorylates proteins is called a **protein kinase**; one that dephosphorylates proteins is called a **protein phosphatase**. Cells contain a variety of protein kinases and protein phosphatases (see Figure 2.3.17).

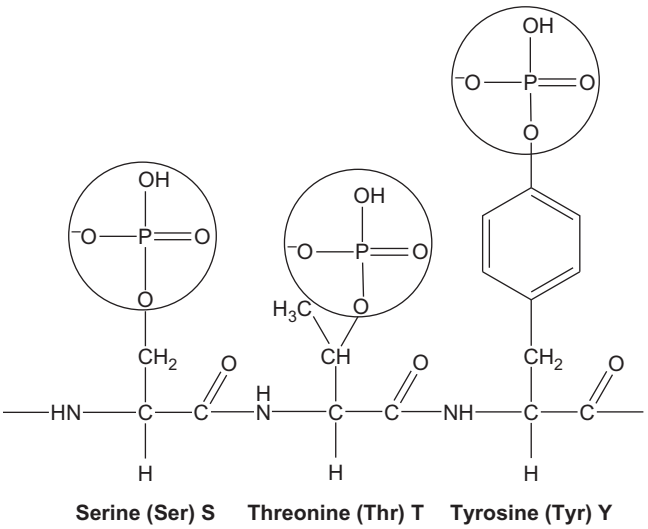

FIGURE 2.3.16 Phosphorylation sites of proteins. Serine, threonine, and tyrosine all have hydroxyl groups that can be esterified with phosphate. This alters the shape and charge of the protein surface in that area, leading to changes in protein activity.

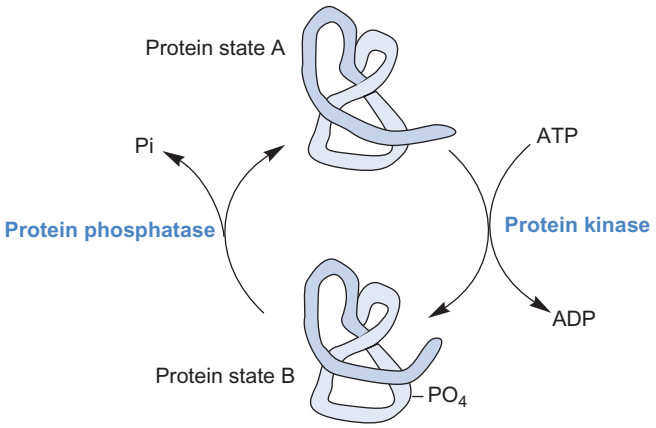

FIGURE 2.3.17 Phosphorylation cycle for protein regulation. Some proteins have phosphorylation sites that can be covalently linked to phosphate from ATP through the action of a variety of protein kinases. This alters the local charge of the protein, which in turn changes its shape and its activity. The protein returns to its dephosphorylated state by the action of protein phosphatases.

Clinical Applications: Protein Folding Diseases

With the advent of the microscope in the mid-1800s, Pasteur, Koch, and many others formulated the germ theory of infectious diseases. We now know that microscopic viral, bacterial, protozoan, or parasitic agents cause a long list of diseases: smallpox, polio, rabies, HIV, yellow fever, anthrax, bubonic plague, syphilis, tuberculosis, cholera, gonorrhea, malaria, sleeping sickness, schistosomiasis, to name but a few. Each of these infectious agents contains DNA or RNA that codes for the organism's own proteins and enables replication of its nucleic acid. Viruses do this by using the host's machinery. Bacteria, protozoans, and parasites are self-contained organisms that use the host's environment for their own reproduction. Investigation of a group of diseases called **transmissible spongiform encephalopathies (TSE)** required revolutionarily new thinking about infection. These diseases include **Creutzfeldt–Jakob disease**, **kuru**, Gerstmann–Straussler syndrome (GSS), and fatal familial insomnia (FFI) in humans, and **scrapies** and **bovine spongiform encephalopathy (BSE)** in animals.

Kuru is a neurological disease among the Fore, a linguistic group of people in Papua New Guinea. In ritual cannabalism, these people ate the bodies of dead relatives. After long incubations, they developed partial paralysis and loss of motor control and eventually died. Because the disease seemed to run in families, a genetic cause was first proposed, but later rejected. Carleton Gajdusek, a pediatrician and virologist, was unable to transmit the disease to any animal, including primates. Igor Klatzo, a neuropathologist, examined tissues sent by Gajdusek and found that kuru was a unique disease without precedent, the closest condition being Creutzfeldt–Jakob disease. In a bizarre twist, William Hadlow, a veterinarian neuropathologist, saw some of Klatzo's photomicrographs in an exhibit at the Wellcome Medical Museum in London, and noted a startling resemblance between neurohistological changes in kuru and those in scrapie, a neurological disease in sheep first described in 1732. It was known to be infectious, but the infectious agent was not yet identified. Hadlow wrote to Gajdusek and published a letter in *The Lancet*

(*Continued*)

Clinical Applications: Protein Folding Diseases (Continued)

in 1959. The similarity of kuru to infectious scrapie prompted Gajdusek to further attempt inoculating animals with kuru. In 1966, he and his co-workers transmitted kuru to chimpanzees and then Creutzfeldt–Jakob disease to chimpanzees in 1968. Gajdusek earned the 1976 Nobel Prize in Medicine.

Although it was initially assumed that the infectious agent must be some virus, investigators failed to identify any virus or any immunological response to one. The agent was not destroyed by UV irradiation or nucleases that typically inactivate nucleic acids. Tikvah Alper and her co-workers in 1966 found that the infectious agent of scrapie was too small to provide a nucleic acid code. In 1967, J.S. Griffith proposed that a protein alone could be the infectious agent of TSE if it was a pathogenic form of a host protein that could convert normal host protein to the pathogenic form. Finally, in 1982, Stanley Prusiner identified the infectious agent as a protein devoid of nucleic acids. He coined the term "**prion**" which stands for proteinaceous infectious particle. Prusiner earned the 1997 Nobel Prize in Medicine.

Prions are infectious proteins. They "reproduce" by causing normal cellular prion protein (PrP^C) to fold up differently, converting it into the pathogenic, or scrapie, isoform (PrP^{Sc}). The native PrP^C appears to have three α-helices and two short β-strands; PrP^{Sc} has two α-helices and much more β-sheet. This transition from α-helix to β-sheet is the fundamental event underlying prion diseases. Proteolysis of PrP^{Sc} produces a smaller, protease-resistant molecule of about 142 amino acids (Prp 27–30) which polymerizes into **amyloid** that presents itself in the disease state.

Ritualistic cannabalism transmitted kuru among the Fore people of New Guinea; industrial cannabalism spread BSE ("mad cow disease") in Europe. There is more than one bad way to fold a protein. Increasing numbers of patients have contracted a new variant of Creutzfeldt–Jakob disease (vCJD) from prion-tainted beef. Because of its long incubation time, we do not yet know the price of mad cow disease. Current thinking that the disease can be eradicated completely by control of infection is wrong. The disease can spontaneously appear without infection.

SUMMARY

The core of amino acids is their asymmetric alpha carbon. Attached to one side is an amino group. On the other side is a carboxyl group. The third group is hydrogen and the fourth group is a variable group, usually referred to as a side chain or R group. These variable groups define the class of the amino acids. Glutamic acid and aspartic acid have a carboxyl group on the side chain. At neutral pH, this group ionizes and therefore has a negative charge. Lysine, arginine, and histidine are basic amino acids because their side chains have a basic chemical character. At neutral pH, these are positively charged. Serine, threonine, and tyrosine have hydroxyl groups that confer polar character. Asparagine and glutamine have an amide group that is also polar. Nonpolar amino acids include glycine, alanine, valine, leucine, isoleucine, proline, cysteine, methionine, tryptophan, and phenylalanine. Some of these are highly hydrophobic, such as tryptophan, isoleucine, leucine, and phenylalanine.

Proteins are made by the formation of peptide bonds between the amino group of one amino acid and the carboxyl group of another. Because of this bond, the basic character of the amino group is neutralized, and the acidic character of the carboxyl group is neutralized, and the character of the chain of amino acids is determined by the sequence of the side chains.

We describe protein structure on four levels: the primary sequence describes the linear sequence of amino acids along the peptide chain backbone, proceeding from the amino terminus to the carboxy terminus. The protein folds into secondary local structures such as α-helices, β-sheets, and β-turns. The arrangement of these secondary structures in three-dimensional space produces the tertiary structure. Combination of proteins with other structures produces macromolecular complexes with quaternary structure. Some proteins spontaneously fold into their "native" shape, whereas others are assembled on a kind of scaffold that helps them fold up properly into their active form. Denaturation of proteins occurs when the protein loses its normal shape. In some cases this is reversible, but usually loss of the proper folding causes irreversible loss of function. Hydrogen bonding, electrostatic interactions, hydrophobic interactions, and steric hindrance all help stabilize proteins in their secondary structures. Disulfide bonds between cysteine side chains help stabilize higher order structure.

Proteins undergo posttranslational modifications after they are synthesized. These include *N*-glycosylation or *O*-glycosylation, proteolytic cleavage, hydroxylation, methylation, acetylation, γ-carboxylation, covalent attachment of hydrophobic molecules that anchor proteins in membranes, and phosphorylation of specific hydroxyl groups on side chains of serine, threonine, and tyrosine.

Protein function depends on the way their surfaces interact with the surfaces of other materials—substrates, structural elements, or other materials to which the proteins bind. Catalytic activity or structural or regulatory roles of proteins depend on the close match of their surface with the surface of the things they bind to. This also determines the specificity of the protein's action.

In general, activity of proteins in the cell can be regulated by altering the amount of protein or by altering its intrinsic activity. Reversible regulation can be achieved by phosphorylation/dephosphorylation of proteins.

REVIEW QUESTIONS

1. Name the two acidic amino acids. What makes them acidic? Name three basic amino acids. What makes them basic?
2. Name five polar amino acids. What makes them polar? Name the nonpolar amino acids.

3. What does "hydrophobic" mean? What does "hydrophilic" mean? What is the partition coefficient? How does it measure hydrophobicity?
4. What is a peptide bond? Where do you find it?
5. What is the primary structure of a protein?
6. What is an α-helix? β-sheet? What interactions among protein side chains stabilize these structures?
7. What is "posttranslational modification"? Name four different kinds of posttranslational modification.
8. What residues are most often acetylated? What residues are hydroxylated? What is "γ-carboxylation"? What amino acid is γ-carboxylated? What is glycosylation? What is myristoylation or palmitoylation?
9. How can proteins be anchored in hydrophobic membranes?
10. Describe the phosphorylation/dephosphorylation cycle for regulating activity of proteins in the cell.