Lodish H, Berk A, Zipursky SL, et al. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000.

Section 3.1 Hierarchical Structure of Proteins

Proteins are designed to bind every conceivable molecule—from simple ions to large complex molecules like fats, sugars, nucleic acids, and other proteins. They catalyze an extraordinary range of chemical reactions, provide structural rigidity to the cell, control flow of material through membranes, regulate the concentrations of metabolites, act as sensors and switches, cause motion, and control gene function. The three-dimensional structures of proteins have evolved to carry out these functions efficiently and under precise control. The *spatial* organization of proteins, their shape in three dimensions, is a key to understanding how they work.

One of the major areas of biological research today is how proteins, constructed from only 20 different **amino acids,** carry out the incredible array of diverse tasks that they do. Unlike the intricate branched structure of carbohydrates, proteins are single, unbranched chains of amino acid monomers. The unique shape of proteins arises from noncovalent interactions between regions in the linear sequence of amino acids. Only when a protein is in its correct three-dimensional structure, or **conformation,** is it able to function efficiently. A key concept in understanding how proteins work is that *function is derived from three-dimensional structure, and three-dimensional structure is specified by amino acid sequence.*

The Amino Acids Composing Proteins Differ Only in Their Side Chains

Amino acids are the monomeric building blocks of proteins. The α carbon atom (C_α) of amino acids, which is adjacent to the carboxyl group, is bonded to four different chemical groups: an amino (NH_2) group, a carboxyl (COOH) group, a hydrogen (H) atom, and one variable group, called a *side chain* or *R group* (Figure 3-1). All 20 different amino acids have this same general structure, but their side-chain groups vary in size, shape, charge, hydrophobicity, and reactivity.

Figure 3-1

Amino acids, the monomeric units that link together to form proteins, have a common structure. The α carbon atom (green) of each amino acid is bonded to four different chemical groups and thus is asymmetric. The side chain, or R group (red), is (more...)

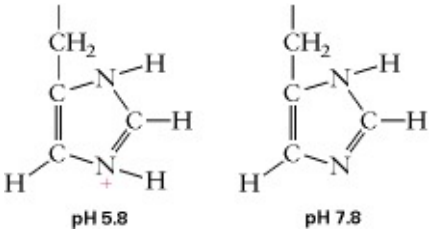
The amino acids can be considered the alphabet in which linear proteins are "written." Students of biology must be familiar with the special properties of each letter of this alphabet, which are determined by the side chain. Amino acids can be classified into a few distinct categories based primarily on their solubility in water, which is influenced by the polarity of their side chains (Figure 3-2). Amino acids with polar side groups tend to be on the surface of proteins; by interacting with water, they make proteins soluble in aqueous solutions. In contrast, amino acids with nonpolar side groups avoid water and aggregate to form the waterinsoluble core of proteins. The polarity of amino acid side chains thus is one of the forces responsible for shaping the final

three-dimensional structure of proteins.

Figure 3-2

The structures of the 20 common amino acids grouped into three categories: hydrophilic, hydrophobic, and special amino acids. The side chain determines the characteristic properties of each amino acid. Shown are the zwitterion forms, which exist at the (more...)

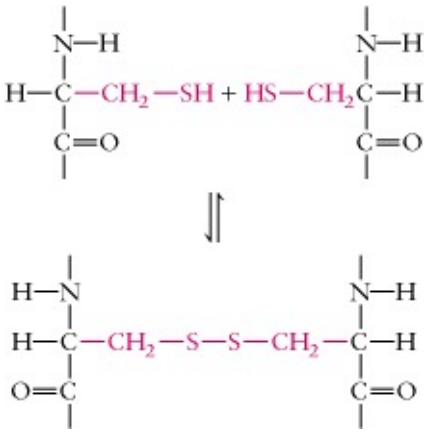
Hydrophilic, or water-soluble, amino acids have ionized or polar side chains. At neutral pH, *arginine* and *lysine* are positively charged; *aspartic acid* and *glutamic acid* are negatively charged and exist as aspartate and glutamate. These four amino acids are the prime contributors to the overall charge of a protein. A fifth amino acid, *histidine,* has an imidazole side chain, which has a pK_a of 6.8, the pH of the cytoplasm. As a result, small shifts of cellular pH will change the charge of histidine side chains:



The activities of many proteins are modulated by pH through protonation of histidine side chains. *Asparagine* and *glutamine* are uncharged but have polar amide groups with extensive hydrogen-bonding capacities. Similarly, *serine* and *threonine* are uncharged but have polar hydroxyl groups, which also participate in hydrogen bonds with other polar molecules. Because the charged and polar amino acids are hydrophilic, they are usually found at the surface of a water-soluble protein, where they not only contribute to the solubility of the protein in water but also form binding sites for charged molecules.

Hydrophobic amino acids have aliphatic side chains, which are insoluble or only slightly soluble in water. The side chains of *alanine, valine, leucine, isoleucine,* and *methionine* consist entirely of hydrocarbons, except for the sulfur atom in methionine, and all are nonpolar. *Phenylalanine, tyrosine,* and *tryptophan* have large bulky aromatic side groups. As explained in Chapter 2, hydrophobic molecules avoid water by coalescing into an oily or waxy droplet. The same forces cause hydrophobic amino acids to pack in the interior of proteins, away from the aqueous environment. Later in this chapter, we will see in detail how hydrophobic residues line the *surface* of membrane proteins that reside in the hydrophobic environment of the lipid bilayer.

Lastly, *cysteine, glycine,* and *proline* exhibit special roles in proteins because of the unique properties of their side chains. The side chain of cysteine contains a reactive **sulfhydryl group (—SH),** which can oxidize to form a **disulfide bond (—S—S—)** to a second cysteine:



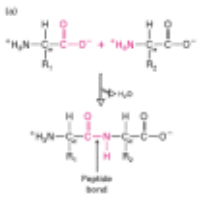
Regions within a protein chain or in separate chains sometimes are cross-linked covalently through disulfide bonds. Although disulfide bonds are rare in intracellular proteins, they are commonly found in extracellular proteins, where they help maintain the native, folded structure. The smallest amino acid, glycine, has a single hydrogen atom as its R group. Its small size allows it to fit into tight spaces. Unlike any of the other common amino acids, proline has a cyclic ring that is produced by formation of a covalent bond between its R group and the amino group on C_α. Proline is very rigid, and its presence creates a fixed kink in a protein chain. Proline and glycine are sometimes found at points on a protein's surface where the chain loops back into the protein.

The 6225 known and predicted proteins encoded by the yeast genome have an average molecular weight (MW) of 52,728 and contain, on average, 466 amino acid residues. Assuming that these average values represent a "typical" eukaryotic protein, then the average molecular weight of amino acids is 113, taking their average relative abundance in proteins into account. This is a useful number to remember, as we can use it to estimate the number of residues from the molecular weight of a protein or vice versa. Some amino acids are more abundant in proteins than other amino acids. Cysteine, tryptophan, and methionine are rare amino acids; together they constitute approximately 5 percent of the amino acids in a protein. Four amino acids—leucine, serine, lysine, and glutamic acid—are the most abundant amino acids, totaling 32 percent of all the amino acid residues in a typical protein. However, the amino acid composition of proteins can vary widely from these values. For example, as discussed in later sections, proteins that reside in the lipid bilayer are enriched in hydrophobic amino acids.

Peptide Bonds Connect Amino Acids into Linear Chains

Nature has evolved a single chemical linkage, the peptide bond, to connect amino acids into a linear, unbranched chain. The peptide bond is formed by a condensation reaction between the amino group of one amino acid and the carboxyl group of another (Figure 3-3a). The repeated amide N, C_α, and carbonyl C atoms of each amino acid residue form the backbone of a protein molecule from which the various side-chain groups project. As a consequence of the peptide linkage, the backbone has polarity, since all the amino groups lie to the same side of the C_α atoms. This leaves at opposite ends of the chain a free (unlinked) amino group (the N-terminus) and a free carboxyl group (the C-terminus). A protein chain is conventionally depicted with its N-terminal amino acid on the left and its C-terminal amino acid on the right (Figure 3-3b).

Figure 3-3



The peptide bond. (a) A condensation reaction between two ar the peptide bond, which links all the adjacent residues in a pro Side-chain groups (R) extend from the backbone of a protein c the amino N, α (more...)

Many terms are used to denote the chains formed by polymerization of amino acids. A short chain of amino acids linked by peptide bonds and having a defined sequence is a peptide; longer peptides are referred to as polypeptides. Peptides generally contain fewer than 20–30 amino acid residues, whereas polypeptides contain as many as 4000 residues. We reserve the term protein for a polypeptide (or a complex of polypeptides) that has a threedimensional structure. It is implied that proteins and peptides represent natural products of a cell.

The size of a protein or a polypeptide is reported as its mass in daltons (a dalton is 1 atomic mass unit) or as its molecular weight (a dimensionless number). For example, a 10,000-MW protein has a mass of 10,000 daltons (Da), or 10 kilodaltons (kDa). In the last section of this chapter, we will discuss different methods for measuring the sizes and other physical characteristics of proteins.

Four Levels of Structure Determine the Shape of Proteins

The structure of proteins commonly is described in terms of four hierarchical levels of organization. These levels are illustrated in Figure 3-4, which depicts the structure of hemagglutinin, a surface protein on the influenza virus. This protein binds to the surface of animal cells, including human cells, and is responsible for the infectivity of the flu virus.

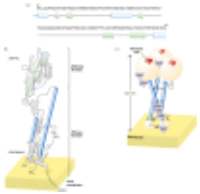


Figure 3-4

Four levels of structure in hemagglutinin, which is a long multimeric molecule whose three identical subunits are each composed of two chains, HA_1 and HA_2. (a) Primary structure is illustrated by the amino acid sequence of residues 68 –195 (more...)

The primary structure of a protein is the linear arrangement, or *sequence,* of amino acid residues that constitute the polypeptide chain.

Secondary structure refers to the localized organization of parts of a polypeptide chain, which can assume several different spatial arrangements. A single polypeptide may exhibit all types of secondary structure. Without any stabilizing interactions, a polypeptide assumes a *random-coil* structure. However, when stabilizing hydrogen bonds form between certain residues, the backbone folds periodically into one of two geometric arrangements: an **α helix,** which is a spiral, rodlike structure, or a **β sheet,** a planar structure composed of alignments of two or more β *strands,* which are relatively short, fully extended segments of the backbone. Finally, U-shaped four-residue segments stabilized by hydrogen bonds between their arms are called *turns.* They are located at the surfaces of proteins and redirect the polypeptide chain toward the interior. (These structures will be discussed in greater detail later.)

polypeptide chain, that is, the three-dimensional arrangement of all the amino acids residues. In contrast to secondary structure, which is stabilized by hydrogen bonds, tertiary structure is stabilized by hydrophobic interactions between the nonpolar side chains and, in some proteins, by disulfide bonds. These stabilizing forces hold the α helices, β strands, turns, and random coils in a compact internal scaffold. Thus, a protein's size and shape is dependent not only on its sequence but also on the number, size, and arrangement of its secondary structures. For proteins that consist of a single polypeptide chain, monomeric proteins, tertiary structure is the highest level of organization.

Multimeric proteins contain two or more polypeptide chains, or *subunits,* held together by noncovalent bonds. Quaternary structure describes the number (stoichiometry) and relative positions of the subunits in a multimeric protein. Hemagglutinin is a trimer of three identical subunits; other multimeric proteins can be composed of any number of identical or different subunits.

In a fashion similar to the hierarchy of structures that make up a protein, proteins themselves are part of a hierarchy of cellular structures. Proteins can associate into larger structures termed *macromolecular assemblies.* Examples of such macromolecular assemblies include the protein coat of a virus, a bundle of actin filaments, the nuclear pore complex, and other large submicroscopic objects. Macromolecular assemblies in turn combine with other cell biopolymers like lipids, carbohydrates, and nucleic acids to form complex cell organelles.

Graphic Representations of Proteins Highlight Different Features

Different ways of depicting proteins convey different types of information. The simplest way to represent three-dimensional structure is to trace the course of the backbone atoms with a solid line (Figure 3-5a); the most complex model shows the location of every atom (Figure 3-5b; see also Figure 2-1a). The former shows the overall organization of the polypeptide chain without consideration of the amino acid side chains; the latter details the interactions among atoms that form the backbone and that stabilize the protein's conformation. Even though both views are useful, the elements of secondary structure are not easily discerned in them.

Figure 3-5

Various graphic representations of the structure of Ras, a guanine nucleotide–binding protein. Guanosine diphosphate, the substrate that is bound, is shown as a blue space-filling figure in parts (a)–(d). (a) The C_α trace of Ras, (more...)

Another type of representation uses common shorthand symbols for depicting secondary structure, cylinders for α helices, arrows for β strands, and a flexible stringlike form for parts of the backbone without any regular structure (Figure 3-5c). This type of representation emphasizes the organization of the secondary structure of a protein, and various combinations of secondary structures are easily seen.

However, none of these three ways of representing protein structure conveys much information about the protein surface, which is of interest because this is where other molecules bind to a protein. Computer analysis in which a water molecule is rolled around the surface of a protein

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.