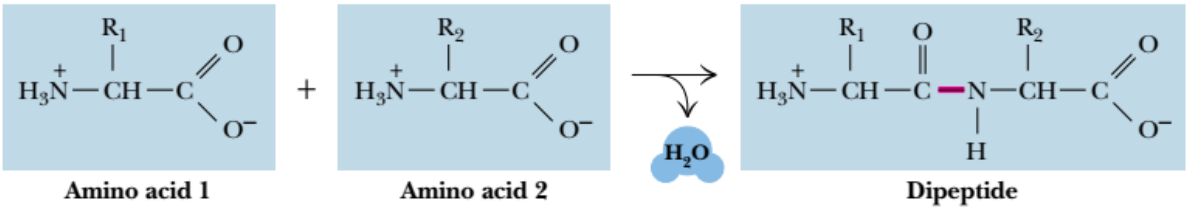
Protein

Proteins are a diverse and abundant class of biomolecules, constituting more than 50% of the dry weight of cells. This diversity and abundance reflect the central role of proteins in virtually all aspects of cell structure and function. Chemically, proteins are unbranched polymers of amino acids linked head to tail, from carboxyl group to amino group, through formation of covalent **peptide bonds,** a type of amide linkage.



Peptide bond formation results in the release of H_2O. The peptide "backbone" of a protein consists of the repeated sequence $-N-C_\alpha-C-$, where the N represents the amide nitrogen, the C is the -carbon atom of an amino acid in the polymer chain, and the final C is the carbonyl carbon of the amino acid, which in turn is linked to the amide N of the next amino acid down the line.

The Peptide Bond Has Partial Double Bond Character

The peptide linkage is usually portrayed by a single bond between the carbonyl carbon and the amide nitrogen Therefore, in principle, rotation may occur about any covalent bond in the polypeptide backbone because all three kinds of bonds $(N-C_\alpha,\ C_\alpha-C_o,$ and the C_o-N peptide bond) are single bonds. In this representation, the C and N atoms of the peptide grouping are both in planar *sp*2 hybridization and the C and O atoms are linked by a bond, leaving the nitrogen with a lone pair of electrons in a 2*p* orbital. However, another resonance form for the peptide bond is feasible in which the C and N atoms participate in a bond, leaving a lone *e* pair on the oxygen (Figure 5.3b). This structure prevents free rotation about the Co- N peptide bond because it becomes a double bond. The real nature of the peptide bond lies somewhere between these extremes; that is, it has partial double bond character, as represented by the intermediate form shown in Figure 5.3c. Peptide bond resonance has several important consequences. First, it restricts free rotation around the peptide bond and leaves the peptide backbone with only two degrees of freedom per amino acid group: rotation around the N-Cα bond and rotation around the Cα-Co bond. Second, the six atoms composing the peptide bond group tend to be coplanar, forming the so-called **amide plane** of the polypeptide backbone Third, the Co-N bond length is 0.133 nm, which is shorter than normal CON bond lengths (for example, the C-N bond of 0.145 nm) but longer than typical C=N bonds (0.125 nm). The peptide bond is estimated to have 40% double-bond character.

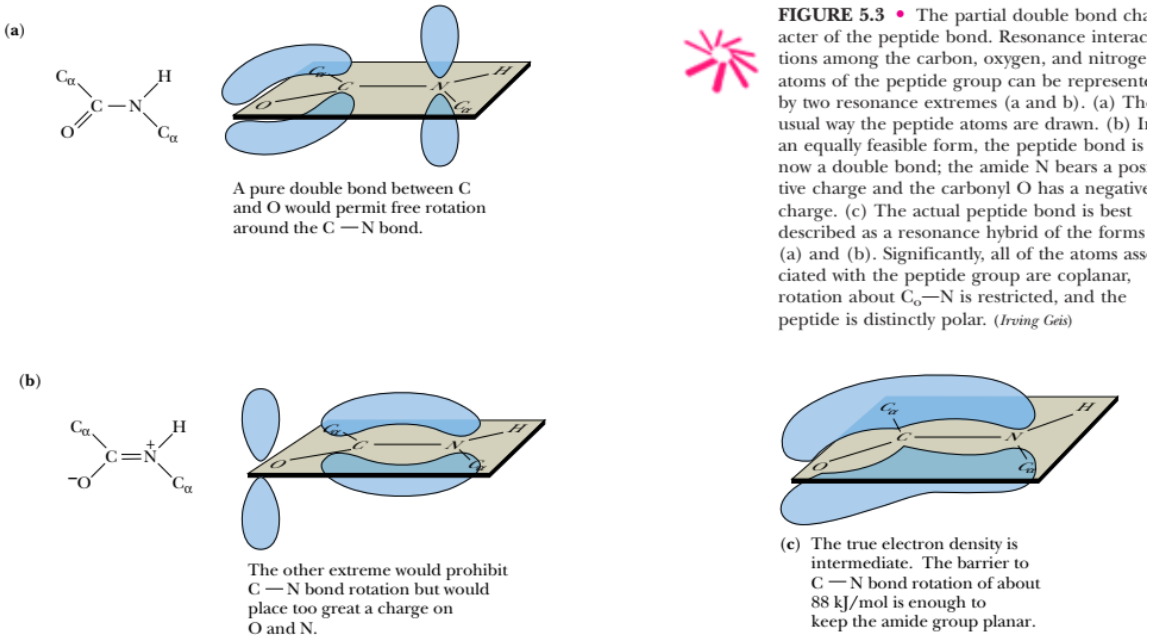


FIGURE 5.3 • The partial double bond character of the peptide bond. Resonance interactions among the carbon, oxygen, and nitrogen atoms of the peptide group can be represented by two resonance extremes (a and b). (a) The usual way the peptide atoms are drawn. (b) In an equally feasible form, the peptide bond is now a double bond; the amide N bears a positive charge and the carbonyl O has a negative charge. (c) The actual peptide bond is best described as a resonance hybrid of the forms (a) and (b). Significantly, all of the atoms associated with the peptide group are coplanar, rotation about C_o—N is restricted, and the peptide is distinctly polar. (*Irving Geis*)

[1]The angle of rotation about the N—C_α bond is designated ϕ, phi, whereas the C_α—C_o angle of rotation is designated ψ, psi.

Peptide is the name assigned to short polymers of amino acids. Peptides are classified by the number of amino acid units in the chain. Each unit is called an **amino acid residue,** the word *residue* denoting what is left after the release of H_2O when an amino acid forms a peptide link upon joining the peptide chain. **Dipeptides** have two amino acid residues; tripeptides have three, tetrapeptides four, and so on. After about 12 residues, this terminology becomes cumbersome, so peptide chains of more than 12 and less than about 20 amino acid residues are usually referred to as **oligopeptides,** and, when the chain exceeds several dozen amino acids in length, the term **polypeptide** is used.

Architecture of Protein Molecules

Protein Shape

1. **Fibrous proteins** tend to have relatively simple, regular linear structures. These proteins often serve structural roles in cells. Typically, they are insoluble in water or in dilute salt solutions.
2. **Globular proteins** are roughly spherical in shape. The polypeptide chain is compactly folded so that hydrophobic amino acid side chains are in the interior of the molecule and the hydrophilic side chains are on the outside exposed to the solvent, water.
3. **Membrane proteins** are found in association with the various membrane systems of cells. For interaction with the nonpolar phase within membranes, membrane proteins have hydrophobic amino acid side chains oriented outward. As such, membrane proteins are insoluble in aqueous solutions but can be solubilized in solutions of detergents.

The Many Biological Functions of Proteins:

1. **Enzymes are** catalysts that accelerate the rates of biological reactions. Each enzyme is very specific in its function and acts only in a particular metabolic reaction.

2. **Regulatory Proteins** A number of proteins do not perform any obvious chemical transformation but nevertheless can regulate the ability of other proteins to carry out their physiological functions. Such proteins are referred to as **regulatory proteins.** well-known example is *insulin,* the hormone regulating glucose metabolism in animals. Insulin is a relatively small protein (5.7 kD). Other hormones that are also proteins include pituitary *somatotropin* (21 kD) and *thyrotropin* (28 kD), which stimulates the thyroid gland.
3. **Transport Proteins** These proteins function to transport specific substances from one place to another. One type of transport is exemplified by the transport of oxygen from the lungs to the tissues by *hemoglobin* These *membrane transport proteins* take up metabolite molecules on one side of a membrane, transport them across the membrane, and release them on the other side. Examples include the transport proteins responsible for the uptake of essential nutrients into the cell, such as glucose or amino acids.
4. **Storage Proteins** Proteins whose biological function is to provide a reservoir of an essential nutrient are called **storage proteins.** Because proteins are amino acid polymers and because nitrogen is commonly a limiting nutrient for growth, organisms have exploited proteins as a means to provide sufficient nitrogen in times of need. For example, *ovalbumin,* the protein of egg white, provides the developing bird embryo with a source of nitrogen during its isolation within the egg. *Casein* is the most abundant protein of milk and thus the major nitrogen source for mammalian infants.
5. **Contractile and Motile Proteins** Certain proteins endow cells with unique capabilities for movement. Cell division, muscle contraction, and cell motility represent some of the ways in which cells execute motion. The **contractile** and **motile proteins** underlying these motions share a common property: they are filamentous or polymerize to form filaments. Examples include *actin* and *myosin,* the filamentous proteins forming the contractile systems of cells, and *tubulin,* the major component of microtubules (the filaments involved in the mitotic spindle of cell division as well as in flagella and cilia).
6. **Structural Proteins** An apparently passive but very important role of proteins is their function in creating and maintaining biological structures. **Structural proteins** provide strength and protection to cells and tissues. Monomeric units of structural proteins typically polymerize to generate long fibers (as in hair) or protective sheets of fibrous arrays, as in cowhide (leather). -*Keratins* are insoluble fibrous proteins making up hair, horns, and fingernails. *Collagen,* another insoluble fibrous protein, is found in bone, connective tissue, tendons, cartilage, and hide, where it forms inelastic fibrils of great strength.
7. **Scaffold Proteins (Adapter Proteins)** Some proteins play a recently discovered role in the complex pathways of cellular response to hormones and growth factors. These proteins, the **scaffold** or **adapter proteins,** have a modular organization in which specific parts (**modules**) of the protein's structure recognize and bind certain structural elements in other proteins through **protein–protein interactions.**
8. **Protective and Exploitive Proteins** In contrast to the passive protective nature of some structural proteins, another group can be more aptly classified as **protective** or **exploitive proteins** because of their biologically active role in cell defense, protection, or exploitation. Prominent among the protective proteins are the *immunoglobulins* or *antibodies* produced by the lymphocytes of vertebrates.
9. **Exotic Proteins** Some proteins display rather exotic functions that do not quite fit the previous classifications. *Monellin,* a protein found in an African plant, has a very sweet

taste and is being considered as an artificial sweetener for human consumption. *Resilin,* a protein having exceptional elastic properties, is found in the hinges of insect wings.

10. **Some Proteins Have Chemical Groups Other Than Amino Acids**
    - **A) GLYCOPROTEINS.** Glycoproteins are proteins that contain carbohydrate. Proteins destined for an extracellular location are characteristically glycoproteins. For example, fibronectin and proteoglycans are important components of the extracellular matrix that surrounds the cells of most tissues in animals.
    - **B) LIPOPROTEINS.** Blood plasma lipoproteins are prominent examples of the class of proteins conjugated with lipid. The plasma lipoproteins function primarily in the transport of lipids to sites of active membrane synthesis. Serum levels of *low density lipoproteins*.
    - **C) NUCLEOPROTEINS.** Nucleoprotein conjugates have many roles in the storage and transmission of genetic information. Ribosomes are the sites of protein synthesis.
    - **D) PHOSPHOPROTEINS.** These proteins have phosphate groups esterified to the hydroxyls of serine, threonine, or tyrosine residues. Casein, the major protein of milk, contains many phosphates and serves to bring essential phosphorus to the growing infant.
    - **E) METALLOPROTEINS.** Metalloproteins are either metal storage forms, as in the case of ferritin, or enzymes in which the metal atom participates in a catalytically important manner.
    - **F) HEMOPROTEINS.** These proteins are actually a subclass of metalloproteins because their prosthetic group is **heme,** the name given to iron protoporphyrin IX.
    - **G) FLAVOPROTEINS.** *Flavin* is an essential substance for the activity of a number of important oxidoreductases. We discuss the chemistry of flavin and its derivatives, FMN and FAD, in the chapter on electron transport and oxidative phosphorylation.

All biological processes involve the specialized functions of one or more protein molecules. Proteins function to produce other proteins, control all aspects of cellular metabolism, regulate the movement of various molecular and ionic species across membranes, convert and store cellular energy, and carry out many other activities. Essentially all of the information required to initiate, conduct, and regulate each of these functions must be contained in the structure of the protein itself.

Forces Influencing Protein Structure:

Hydrogen Bonds Hydrogen bonds are generally made wherever possible within a given protein structure. In most protein structures that have been examined to date, component atoms of the peptide backbone tend to form hydrogen bonds with one another. For example, in -helices, the C-O and N-H groups of every residue participate in H bonds. The importance of H bonds in protein structure cannot be overstated.

Hydrophobic Interactions Hydrophobic "bonds," or, more accurately, *interactions,* form because nonpolar side chains of amino acids and other nonpolar solutes prefer to cluster in a nonpolar environment rather than to intercalate in a polar solvent such as water. The forming of hydrophobic bonds minimizes the interaction of nonpolar residues with water and is therefore highly favorable.

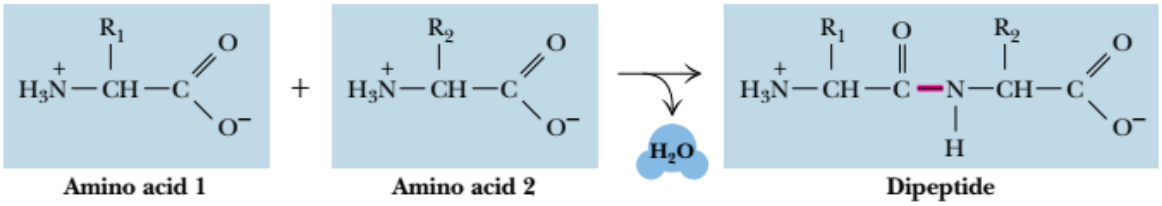
Electrostatic Interactions

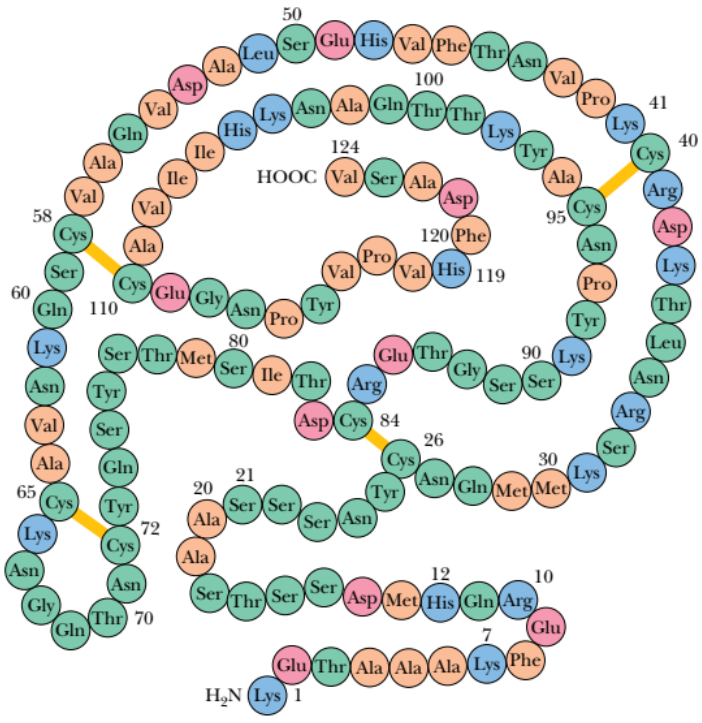
Ionic interactions arise either as electrostatic attractions between opposite charges or repulsions between like charges. Amino acid side chains can carry positive charges, as in the case of lysine, arginine, and histidine, or negative charges, as in aspartate and glutamate. In addition, the NH2-terminal and COOH-terminal residues of a protein or peptide chain usually exist in ionized states and carry positive or negative charges, respectively.

Van der Waals Interactions Both attractive forces and repulsive forces are included in van der Waals interactions. The attractive forces are due primarily to instantaneous dipole-induced dipole interactions that arise because of fluctuations in the electron charge distributions of adjacent non bonded atoms.

The Levels of Protein Structure

Primary Structure The amino acid sequence is the primary (1°) structure of a protein.





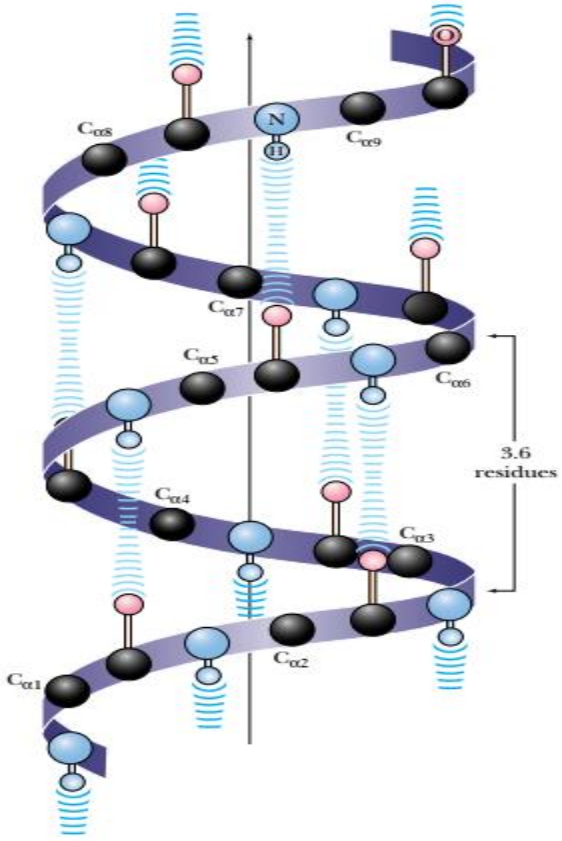
Secondary Structure in Proteins

The Alpha-Helix

These structures tend to form in cooperative fashion and involve substantial portions of the peptide chain. Structures resulting from these interactions constitute **secondary structure** for proteins. When a number of hydrogen bonds form between portions of the peptide chain in this manner, two basic types of structures can result: α-*helices* and β-*pleated sheets.* In 1951, Linus Pauling, Robert Corey, and their colleagues at the California Institute of Technology summarized a large volume of crystallographic data in a set of dimensions for polypeptide chains. With these data in hand, Pauling, Corey, and

their colleagues proposed a new model for a helical structure in proteins, which they called the α**-helix.**

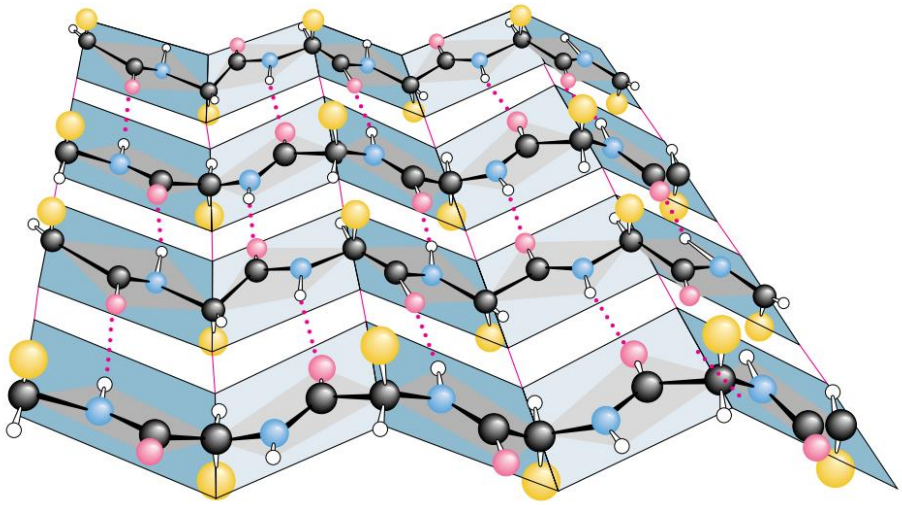
1. One turn of the helix represents 3.6 amino acid residues. (A single turn of the -helix involves 13 atoms from the O to the H of the H bond. For this reason, the -helix is sometimes referred to as the 3.613 helix.)
2. Each amino acid residue extends **1.5 Å (0.15 nm)** along the helix axis. With **3.6 residues per turn,** this amounts to 3.6 1.5 Å or **5.4 Å (0.54 nm)** of travel along the helix axis per turn.
3. On average, there are about 10 residues per helix. Myoglobin, one of the first proteins in which - helices were observed, has eight stretches of -helix that form a box to contain the heme prosthetic group.



Four N-H groups at The N-terminal end of an -helix and four C=O groups at the C-terminal end cannot participate in hydrogen bonding. The formation of H-bonds with other nearby donor and acceptor groups is referred to as **helix capping.**

The Beta-Pleated Sheet

Another type of structure commonly observed in proteins also forms because of local, cooperative formation of hydrogen bonds. That is the pleated sheet, or -structure, often called the β**-pleated sheet.** This structure was also first postulated by Pauling and Corey in 1951 and has now been observed in many natural proteins. A -pleated sheet can be visualized by laying thin, pleated strips of paper side by side to make a “pleated sheet” of paper.



A “β-pleated sheet” of paper with an antiparallel -sheet drawn on it.

In the **parallel β-pleated sheet,** adjacent chains run in the same direction (N - C or C- N). In the **antiparallel β–pleated sheet,** adjacent strands run in opposite directions. Each single strand of the -sheet structure can be pictured as a twofold helix, that is, a helix with two residues per turn. The arrangement of successive amide planes has a pleated appearance due to the tetrahedral nature of the C atom. It is important to note that the hydrogen bonds in this structure are essentially *inter*strand rather than *intra*strand. The distance between residues is 0.347 nm for the antiparallel pleated sheet, but only 0.325 nm for the parallel pleated sheet.

Protein Folding and Tertiary Structure

The folding of a single polypeptide chain in three-dimensional space is referred to as its **tertiary structure.** First, secondary structures—helices and sheets—form whenever possible as a consequence of the formation of large numbers of hydrogen bonds. Second, -helices and -sheets often associate and pack close together in the protein. No protein is stable as a single-layer structure, for reasons that become apparent later. There are a few common methods for such packing to occur. Third, because the peptide segments between secondary structures in the protein tend to be short and direct, the peptide does not execute complicated twists and knots as it moves from one region of a secondary structure to another.

Fibrous Proteins

fibrous proteins, globular proteins, and *membrane proteins.* Fibrous proteins contain polypeptide chains organized approximately parallel along a single axis, producing long fibers or large sheets.

α-Keratin the structure of the **α-keratins** is dominated by -helical segments of polypeptide. The amino acid sequence of -keratin subunits is composed of central -helix–rich rod domains about 311 to 314 residues in length, flanked by non helical N- and C-terminal domains of varying size and composition.

The **fibroin** proteins found in silk fibers represent another type of fibrous protein. These are composed of stacked antiparallel β-sheets.

Collagen: A Triple Helix **Collagen** is a rigid, inextensible fibrous protein that is a principal constituent of connective tissue in animals, including tendons, cartilage, bones, teeth, skin,

and blood vessels. The high tensile strength of collagen fibers in these structures makes possible the various animal activities such as running and jumping that put severe stresses on joints and skeleton. Broken bones and tendon and cartilage injuries to knees, elbows, and other joints involve tears or hyperextensions of the collagen matrix in these tissues.

Subunit Interactions and Quaternary Structure

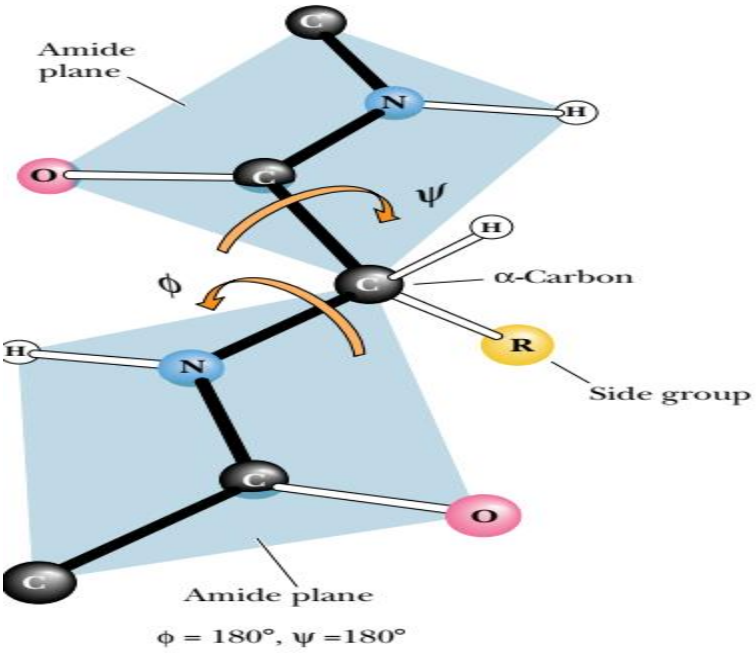
Many proteins exist in nature as **oligomers,** complexes composed of (often symmetric) noncovalent assemblies of two or more monomer subunits. In fact, subunit association is a common feature of macromolecular organization in biology. Most intracellular enzymes are oligomeric and may be composed eitherof a single type of monomer subunit (*homomultimers*) or of several different kinds of subunits (*heteromultimers*).

Ramachandran Plot

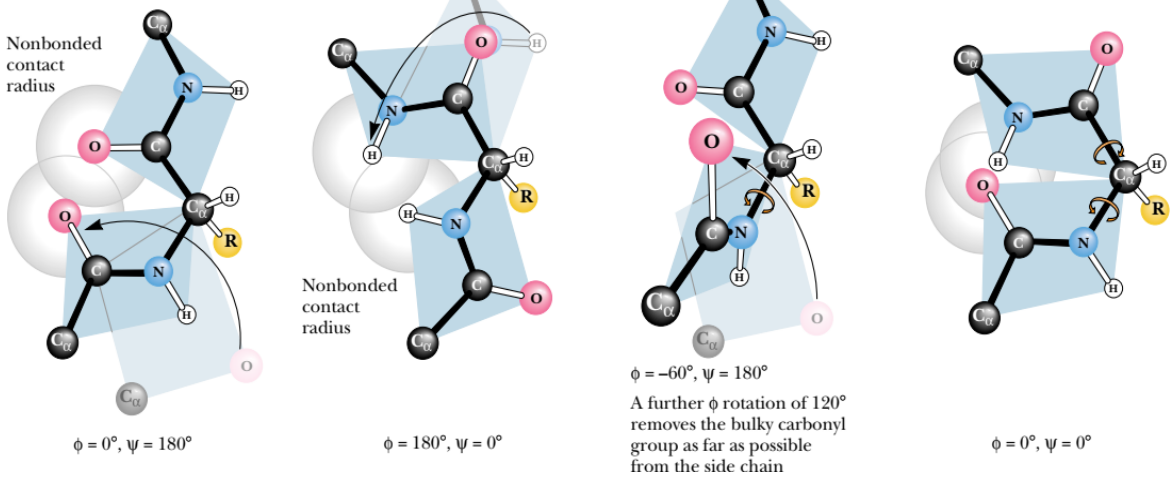
G. N. Ramachandran and his coworkers in Madras, India, first showed thatit was convenient to plot values against values to show the distribution ofallowed values in a protein or in a family of proteins. A typical **Ramachandran plot** is shown in Figure 6.4. Note the clustering of and values in a few regions of the plot. Most combinations of and are sterically forbidden, and the corresponding regions of the Ramachandran plot are sparsely populated. The combinations that are sterically allowed represent the subclasses of structure

The planarity of the peptide bond means that there are only two degrees of freedom per residue for the peptide chain. Rotation is allowed about the bond linking the -carbon and the carbon of the peptide bond and also about the bond linking the nitrogen of the peptide bond and the adjacent α-carbon.

α-carbon is the joining point for two planes defined by peptide bonds. The angle about the C_α—N bond is denoted by the Greek letter (phi) and that about the C_α—Co is denoted by (psi). For either of these bond angles, a value of 0° corresponds to an orientation with the amide plane bisecting the H—C_α—R (sidechain) plane and a *cis* configuration of the main chain around the rotating bond.



The amide or peptide bond planes are joined by the tetrahedral bonds of the α-carbon. The rotation parameters are ϕ and ψ . The conformation shown corresponds to ϕ = 180° and ψ= 180°. Note that positive values of and correspond to clockwise rotation as viewed from C . Starting from 0°, a rotation of 180° in the clockwise direction (180°) is equivalent to a rotation of 180° in the counterclockwise direction (180°).



ϕ = 0°, ψ = 180°

ϕ = 180°, ψ = 0°

ϕ = –60°, ψ = 180°

A further ϕ rotation of 120° removes the bulky carbonyl group as far as possible from the side chain

ϕ = 0°, ψ = 0°