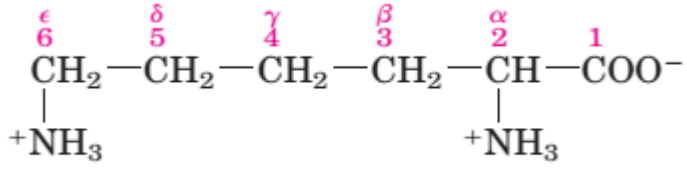
BCH 313 Metabolism of Amino acids and Proteins (3 UNITS)

Proteins are polymers of amino acids, with each amino acid residue joined to its neighbour by a specific type of covalent bond. (The term "residue" reflects the loss of the elements of water when one amino acid is joined to another.) Proteins can be broken down (hydrolysed) to their constituent amino acids by a variety of methods, and the earliest studies of proteins naturally focused on the free amino acids derived from them. Twenty different amino acids are commonly found in proteins. The first to be discovered was asparagine, in 1806. The last of the 20 to be found, threonine, was not identified until 1938. All the amino acids have trivial or common names, in some cases derived from the source from which they were first isolated. Asparagine was first found in asparagus, and glutamate in wheat gluten; tyrosine was first isolated from cheese (its name is derived from the Greek tyros, "cheese"); and glycine (Greek glykos,"sweet") was so named because of its sweet taste.

Amino Acids Share Common Structural Features

All 20 of the common amino acids are α-amino acids. They have a carboxyl group and an amino group bonded to the same carbon atom (the α-carbon). They differ from each other in their side chains, or R groups, which vary in structure, size, and electric charge, and which influence the solubility of the amino acids in water. In addition to these 20 amino acids there are many less common ones. Some are residues modified after a protein has been synthesized; others are amino acids present in living organisms but not as constituents of proteins. The common amino acids of proteins have been assigned three-letter abbreviations and one-letter symbols, which are used as shorthand to indicate the composition and sequence of amino acids polymerized in proteins.

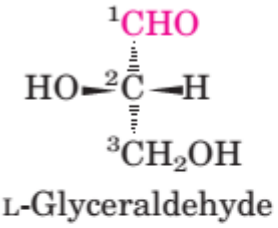
Two conventions are used to identify the carbons in an amino acid—a practice that can be confusing. The additional carbons in an R group are commonly designated β, γ, δ, ε, and so forth, proceeding out from the carbon. For most other organic molecules, carbon atoms are simply numbered from one end, giving highest priority (C-1) to the carbon with the substituent containing the atom of highest atomic number. Within this latter convention, the carboxyl carbon of an amino acid would be C-1 and the α-carbon would be C-2. In some cases, such as amino acids with heterocyclic R groups, the Greek lettering system is ambiguous and the numbering convention is therefore used. For all the common amino acids except glycine, the carbon is bonded to four different groups: a carboxyl group, an amino group, an R group, and a hydrogen atom in glycine, the R group is another hydrogen atom). The α-carbon atom is thus a chiral centre. Because of the tetrahedral arrangement of the bonding orbitals around the α-carbon atom, the four different groups can occupy two unique spatial arrangements, and thus amino acids have two possible stereoisomers. Since they are no superimposable mirror images of each other, the two forms represent a class of stereoisomers called enantiomers. All molecules with a chiral centre are also optically active—that is, they rotate plane-polarized light.



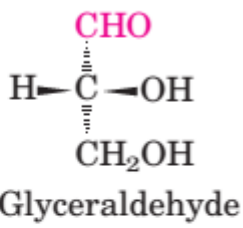
Lysine

For all the common amino acids except glycine, the α-carbon is bonded to four different groups: a carboxyl group, an amino group, an R group, and a hydrogen atom; in glycine, the R group is another hydrogen atom). The α-carbon atom is thus a chiral center. Because of the tetrahedral arrangement of the bonding orbitals around the -carbon atom, the four different groups can occupy two unique spatial arrangements, and thus amino acids have two possible stereoisomers. Since they are no superimposable mirror images of each other, the two forms represent a class of stereoisomers called enantiomers. All molecules with a chiral center are also optically active—that is, they rotate plane-polarized light.

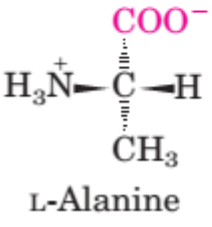
Special nomenclature has been developed to specify the absolute configuration of the four substituents of asymmetric carbon atoms. The absolute configurations of simple sugars and amino acids are specified by the D, L, system, based on the absolute configuration of the three-carbon sugar glyceraldehyde, a convention proposed by Emil Fischer in 1891.



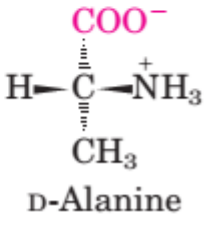
L-Glyceraldehyde



D-Glyceraldehyde



L-Alanine



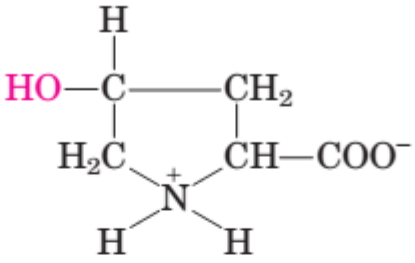
D-Alanine

Nearly all biological compounds with a chiral centre occur naturally in only one stereo isomeric form, either D or L. The amino acid residues in protein molecules are exclusively L stereoisomers. D-Amino acid residues have been found only in a few, generally small peptides, including some peptides of bacterial cell walls and certain peptide antibiotics.

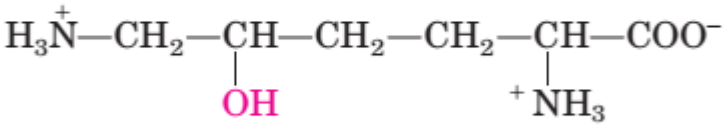
Uncommon Amino Acids Also Have Important Functions

In addition to the 20 common amino acids, proteins may contain residues created by modification of common residues already incorporated into a polypeptide. Among these uncommon amino acids are 4-hydroxyproline, a derivative of proline, and 5-hydroxylysine, derived from lysine. The former is found in plant cell wall proteins, and both are found in collagen, a fibrous protein of connective tissues. 6-NMethyllysineis a constituent of myosin, a contractile protein of muscle. Another important uncommon amino acid is -carboxyglutamate, found in the blood clotting protein prothrombin and in certain other proteins that bind Ca^{2+} as part of their biological function.

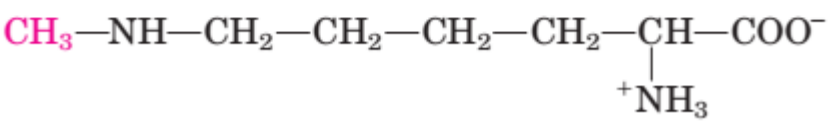
More complex is desmosine,a derivative of four Lys residues, which is found in the fibrous protein elastin. Selenocysteine is a special case. This rare amino acid residue is introduced during protein synthesis rather than created through a post synthetic modification. It contains selenium rather than the sulfur of cysteine. Actually derived from serine, selenocysteine is a constituent of just a few known proteins. Some 300 additional amino acids have been found in cells. They have a variety of functions but are not constituents of proteins. Ornithine and citrulline deserve special note because they are key intermediates (metabolites) in the biosynthesis of arginine and in the urea cycle.



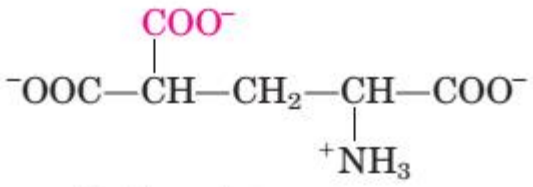
4-Hydroxyproline



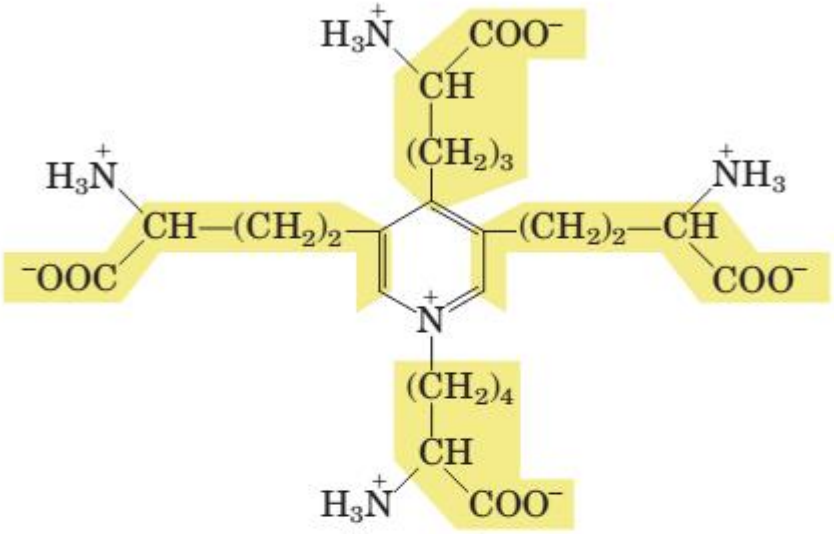
5-Hydroxylysine



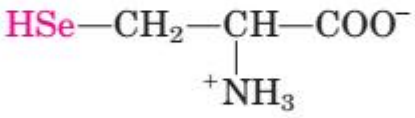
6-*N*-Methyllysine



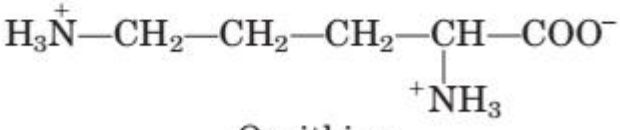
γ-Carboxyglutamate



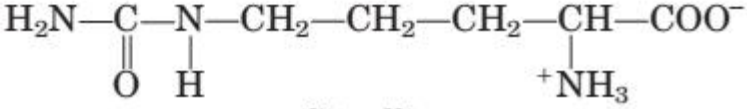
Desmosine



(a) Selenocysteine



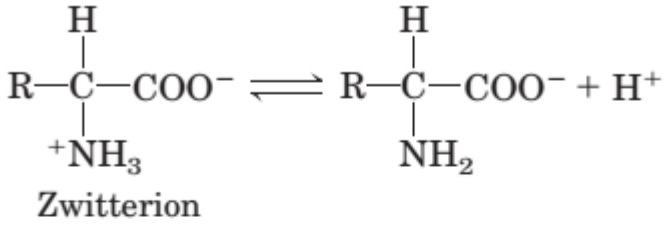
Ornithine



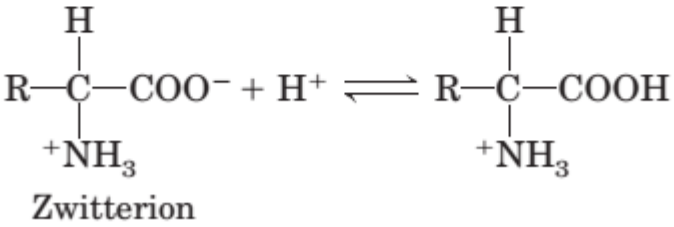
(b) Citrulline

Amino Acids Can Act as Acids and Bases

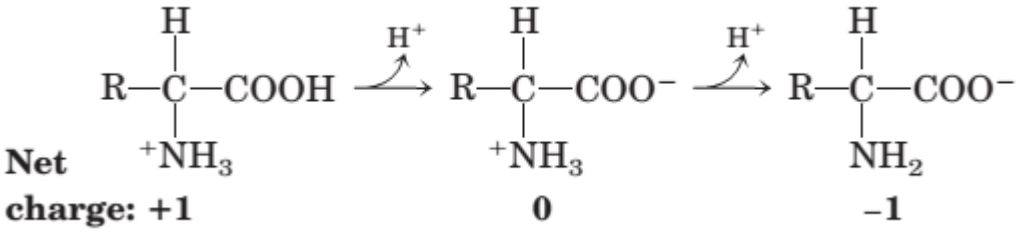
When an amino acid is dissolved in water, it exists in solution as the dipolar ion, or zwitterion (German for "hybrid ion"), A zwitterion can act as either an acid (proton donor):



or a base (proton acceptor):



Substances having this dual nature are amphoteric and are often called ampholytes(from "amphoteric electrolytes"). A simple monoamino monocarboxylic -amino acid, such as alanine, is a diprotic acid when fully protonated—it has two groups, the OCOOH group and the ONH3 group, that can yield protons:



Titration Curves Predict the Electric Charge of Amino Acids

Another important piece of information derived from the titration curve of an amino acid is the relationship between its net electric charge and the pH of the solution. At pH 5.97, the point of inflection between the two stages in its titration curve, glycine is present predominantly as its dipolar form, fully ionized but with no net electric charge. The characteristic pH at which the net electric charge is zero is called the **isoelectric point** or **isoelectric pH**, designated **pI.**

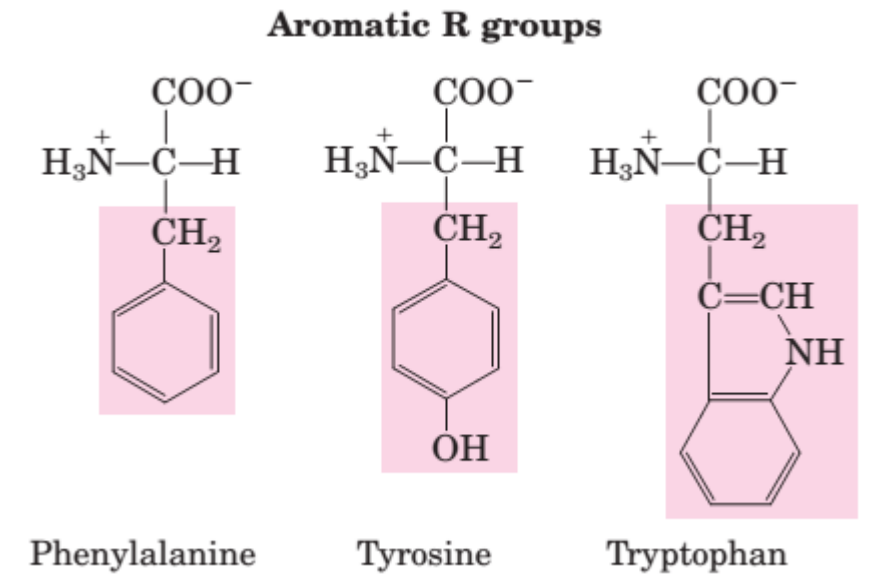
Amino Acids Can Be Classified by R Group

Knowledge of the chemical properties of the common amino acids is central to an understanding of biochemistry. The topic can be simplified by grouping the amino acids into five main classes based on the properties of their R groups (Table 3–1), in particular, their polarity, or tendency to interact with water at biological pH (near pH 7.0). The polarity of the R groups varies widely, from nonpolar and hydrophobic (water-insoluble) to highly polar and hydrophilic (water-soluble) within each class there are gradations of polarity, size, and shape of the R groups. Nonpolar, Aliphatic R Groups the R groups in this class of amino acids are nonpolar and hydrophobic. The side chains of alanine, valine, leucine, and isoleucine tend to cluster together within proteins, stabilizing protein structure by means of hydrophobic interactions. Glycine has the simplest structure. Although it is formally nonpolar, it's very small side chain makes no real contribution to hydrophobic interactions. Methionine, one of the two sulfur-containing amino acids, has a nonpolar thioether group in its side chain. Proline has an aliphatic side chain with a distinctive cyclic structure. The secondary amino (imino) group of proline residues is held in a rigid conformation that reduces the structural flexibility of polypeptide regions containing proline.

Aromatic R Groups

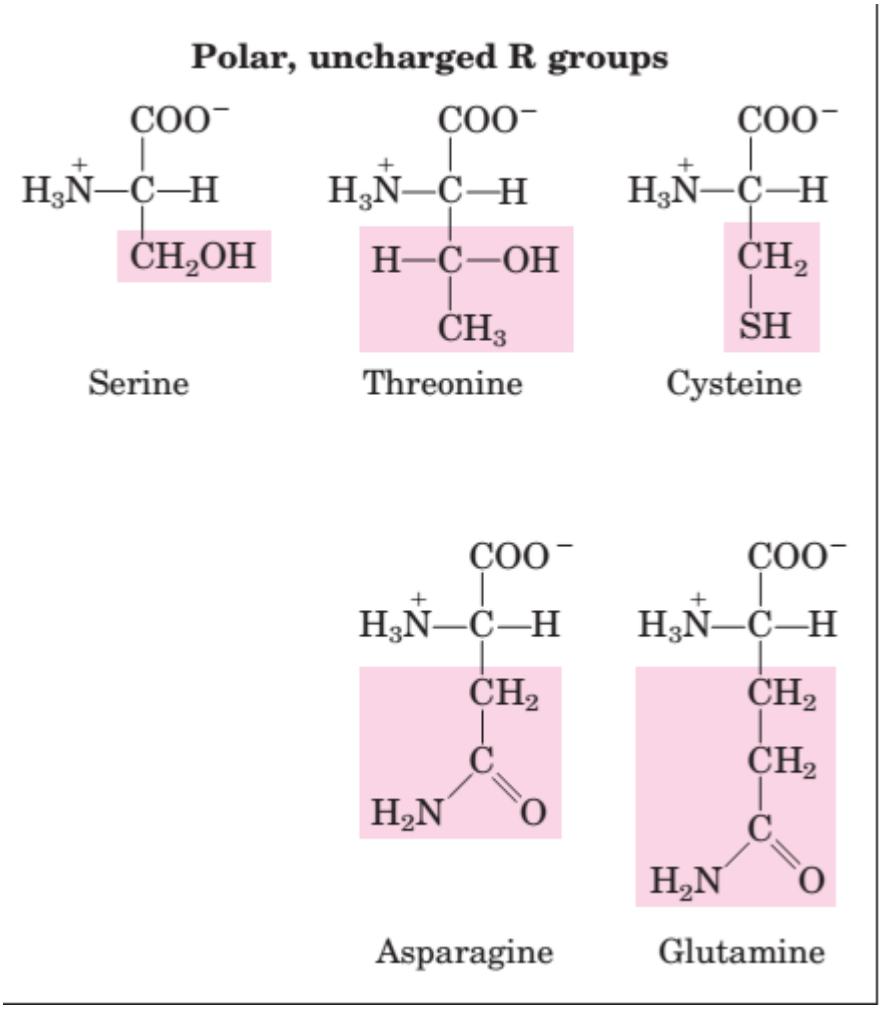
Phenylalanine, tyrosine, and tryptophan, with their aromatic side chains, are relatively nonpolar (hydrophobic). All can participate in hydrophobic interactions. The hydroxyl group of tyrosine can form hydrogen bonds, and it is an important functional group in some enzymes. Tyrosine and tryptophan are significantly more polar than phenylalanine, because of the tyrosine hydroxyl group and the nitrogen of the tryptophan indole ring.

Tryptophan and tyrosine, and to a much lesser extent phenylalanine, absorb ultraviolet light. This accounts for the characteristic strong absorbance of light by most proteins at a wavelength of 280 nm, a property exploited by researchers in the characterization of proteins.



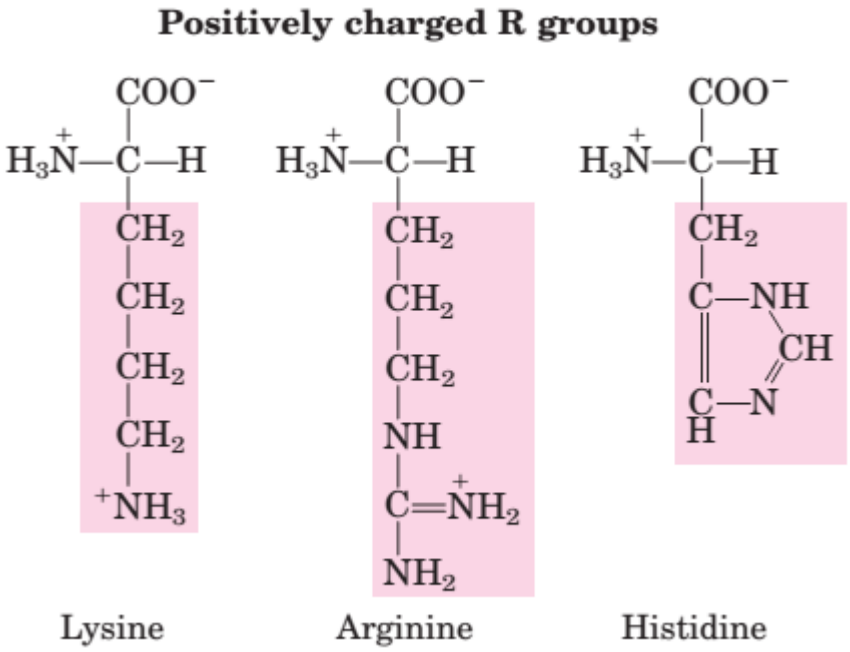
Polar, Uncharged R Groups

The R groups of these amino acids are more soluble in water, or more hydrophilic, than those of the nonpolar amino acids, because they contain functional groups that form hydrogen bonds with water. This class of amino acids includes serine, threonine, cysteine, asparagine, and glutamine. The polarity of serine and threonine is contributed by their hydroxyl groups; that of cysteine by its sulfhydryl group; and that of asparagine and glutamine by their amide groups. Asparagine and glutamine are the amides of two other amino acids also found in proteins, aspartate and glutamate, respectively, to which asparagine and glutamine are easily hydrolyzed by acid or base. Cysteine is readily oxidized to form a covalently linked dimeric amino acid called cystine,in which two cysteine molecules or residues are joined by a disulfide bond . The disulfide-linked residues are strongly hydrophobic (nonpolar). Disulfide bonds play a special role in the structures of many proteins by forming covalent links between parts of a protein molecule or between two different polypeptide chains.



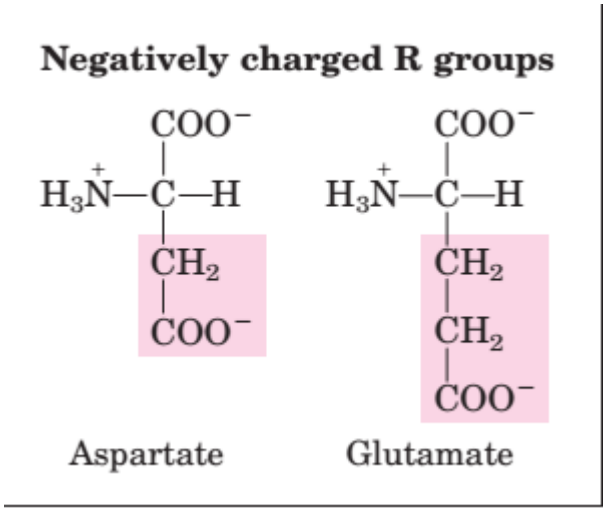
Positively Charged (Basic) R Groups

The most hydrophilic R groups are those that are either positively or negatively charged. The amino acids in which the R groups have significant positive charge at pH 7.0 are lysine, which has a second primary amino group at the position on its aliphatic chain; arginine, which has a positively charged guanidino group; and histidine, which has an imidazole group. Histidine is the only common amino acid having an ionizable side chain with a pKa near neutrality. In many enzyme-catalyzed reactions, a His residue facilitates the reaction by serving as a proton donor/acceptor.



Negatively Charged (Acidic) R Groups

The two amino acids having R groups with a net negative charge at pH 7.0 are aspartate and glutamate, each of which has a second carboxyl group.



Nutritional classification:

1- Essential amino acids: These amino acids can't be formed in the body and so, it is essential to be taken in diet. Their deficiency affects growth, health and protein synthesis. Ten amino acids present in proteins (arginine, histidine, isoleucine, leucine, threonine, lysine, methionine, phenylalanine, tryptophan, valine) are required in the diet of a growing human. Arginine and histidine, although not required in the diets of adults, are required for growth (children and adolescents), because the amounts that can be synthesized are not sufficient to maintain normal growth rates. Larger amounts of phenylalanine are required if the diet is low in tyrosine because tyrosine is synthesized from phenylalanine. Larger amounts of methionine are required if the diet is low in cysteine because the sulfur of methionine is donated for the synthesis of cysteine.

2- **Non-essential amino acids:** These are formed in the body but not in sufficient amount for body requirements especially in children. Twelve amino acids present in proteins are synthesized in the body - eleven (serine, glycine, cysteine, alanine, aspartate, asparagine, glutamate, glutamine, proline, arginine, histidine) are produced from glucose, one (tyrosine) is produced from phenylalanine.

Metabolic classification: according to metabolic or degradation products of amino acids they may be:

1- Ketogenic amino acids: which give ketone bodies. Lysine and Leucine are the only pure ketogenic amino acids.

2- Mixed ketogenic and glucogenic amino acids: which give both ketone bodies and glucose. These are: isoleucine, phenylalanine, tyrosine and tryptophan.

3- Glucogenic amino acids: Which give glucose. They include the rest of amino acids. These amino acids by catabolism yields products that enter in glycogen and glucose formation.

Classification of Amino Acids

An alternative classification scheme.

1. Acidic amino acids and their amides: aspartic acid, asparagine, glutamic acid, glutamine.
2. Basic amino acids: histidine, lysine, arginine.
3. Aromatic amino acids: phenylalanine, tyrosine, tryptophan.
4. Sulphur containing amino acids: cysteine, methionine.
5. Imido acid: proline.
6. Hydrophobic side chains: glycine, alanine, valine, leucine, isoleucine.
7. Hydroxylic amino acids: serine, threonine, (tyrosine).

Nitrogen metabolism

N2

Atmospheric nitrogen N2 is most abundant but is too inert for use in most biochemical processes.

Dietary proteins

Atmospheric nitrogen is acted upon by bacteria (nitrogen fixation) and plants to nitrogen containing compounds. We assimilate these compounds as proteins (amino acids) in our diets.

Amino acids

Conversion of nitrogen into specialized products

Body proteins

α-amino groups

Other nitrogen containing compounds Lecture III

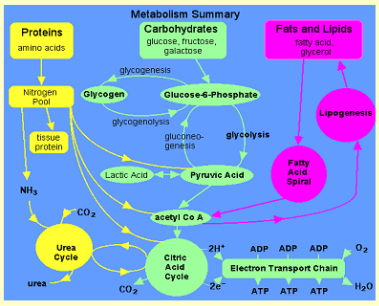
NH4+

Disposal of Nitrogen Lecture I

Urea

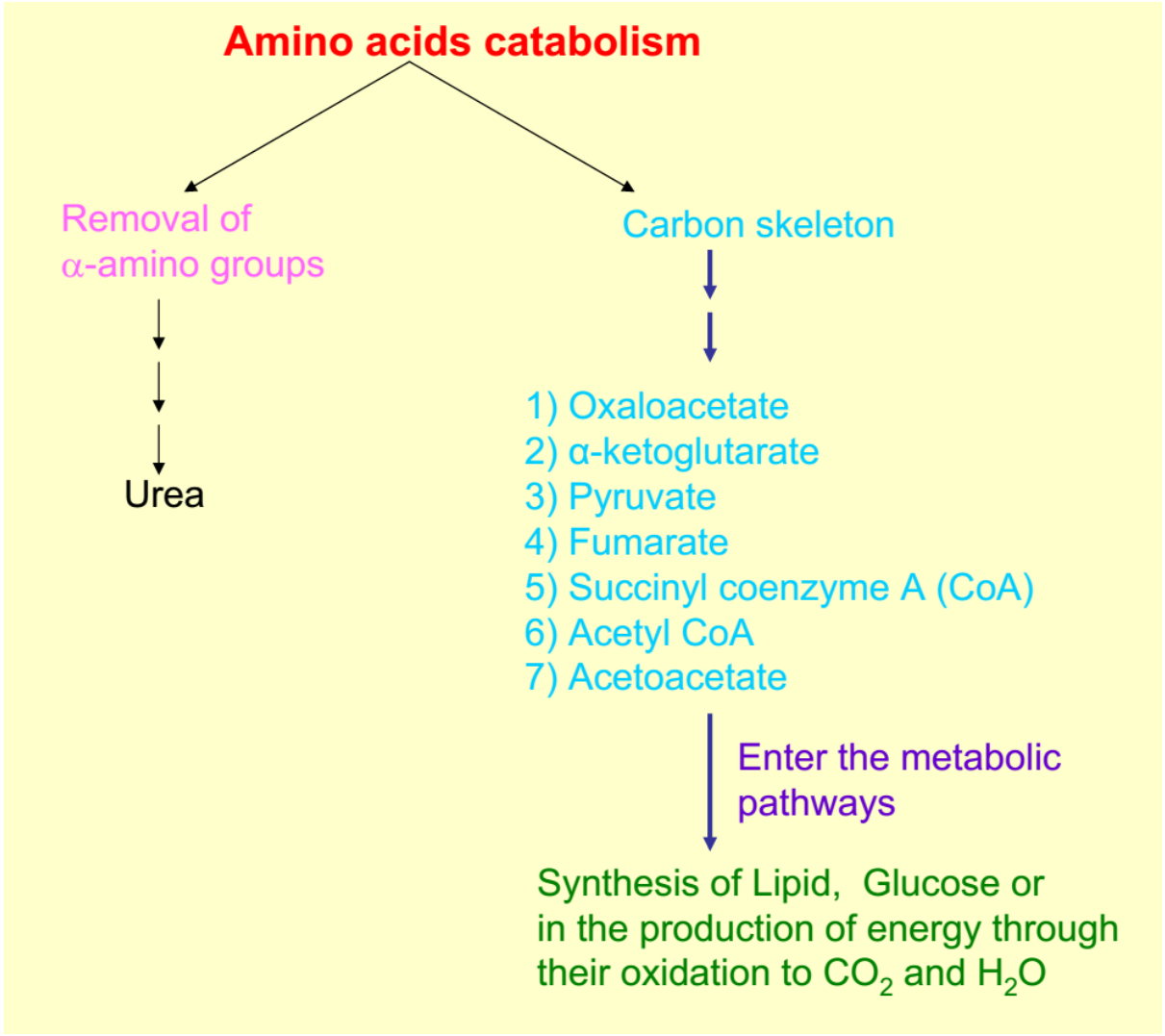
excreted

Carbon skeletons



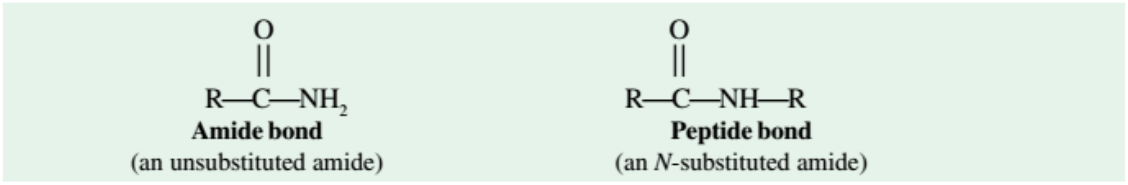
Amino acids synthesis & degradation Lecture II

Enters various metabolic pathways


Amino acids catabolism
Removal of
α-amino groups
Urea
Carbon skeleton
1) Oxaloacetate
2) α-ketoglutarate
3) Pyruvate
4) Fumarate
5) Succinyl coenzyme A (CoA)
6) Acetyl CoA
7) Acetoacetate
Enter the metabolic pathways
Synthesis of Lipid, Glucose or
in the production of energy through
their oxidation to CO_2 and H_2O

Peptides/Peptide Bond

The amino acid units are linked together through the carboxyl and amino groups to produce the primary structure of the protein chain. The bond between two adjacent amino acids is a special type of amide bond, in which the hydrogen atom of amino ($-NH_2$) group is replaced by an R radical. Such a substituted amide bond is known as the peptide bond. And the chain, thus formed, by linking. Two amino acid molecules can be covalently joined through a substituted amide linkage, termed a peptide bond, to yield a dipeptide. Such a linkage is formed by removal of the elements of water (dehydration) from the α-carboxyl group of one amino acid and the α-amino group of another. Peptide bond formation is an example of a condensation reaction, a common class of reactions in living cells. Under standard biochemical conditions, the equilibrium for the reaction favours the amino acids over the dipeptide. To make the reaction thermodynamically more favourable, the carboxyl group must be chemically modified or activated so that the hydroxyl group can be more readily eliminated



together of many amino acid units is called a **peptide chain** (Fig. 9–8).

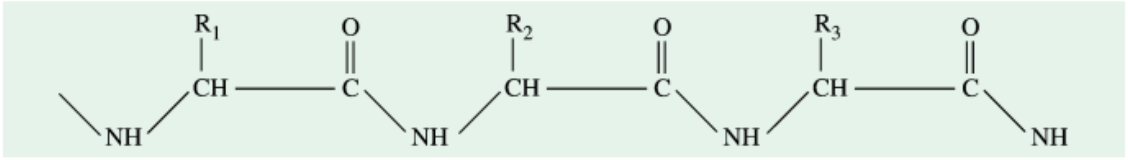


Fig. 9–8. Peptide chain

Each peptide chain is of considerable length and may possess from 50 to millions of amino acid units. Depending on the number of amino acid molecules composing a chain, the peptides may be termed as a dipeptide (containing 2 amino acid units), a tripeptide (containing 3 amino acid units) and so on. If a peptide is made up of not more than 10 amino acids, it is called an oligopeptide ; beyond that it is a polypeptide. Polypeptides when they are made up of over 100 amino acids are, sometimes, called as macropeptides. Strictly speaking, the proteins are polypeptides with more than 100 amino acids. All naturally-occurring important peptides, however, possess a shorter individual name, such as glutathione etc. Proteins differ widely in amino acid content. Various types of proteins in an organism may have varied amounts of a particular amino acid. Some amino acids are in abundance in one protein, may be in meagre amounts in others and may even be lacking in the rest. Tryptophan, for instance, lacks in certain proteins. However, most of the proteins contain all the 20 amino acids. As the number and manner in which the amino acids are grouped is highly variable, the number of proteins approaches almost to infinity.

N- and C-terminals

Each amino acid in the chain is termed a residue. The two ends of the peptide chain are named as amino terminal and carboxyl terminal or simply as an N-terminal and C-terminal respectively. These two terminal groups, one basic and another acidic, are the only ionizable groups of any peptide chain except those present in the side chain. The terminal amino acid with the free amino group is called as the N-terminal amino acid and the one with the free carboxyl group at the other end as C-terminal amino acid.

Naming of Peptide Chain

In naming a polypeptide, the convention is that the N-terminal residue (which is shown at the left hand part of the structure) is written first and the C-terminal residue in the formation of each peptide

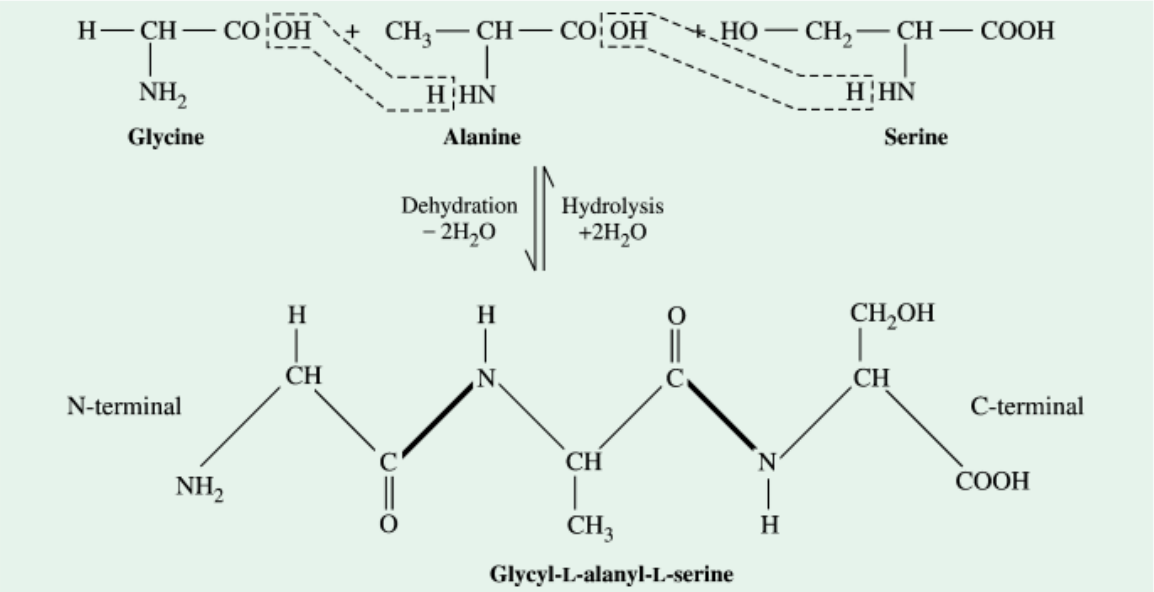


Fig. 9–11. Construction of a tripeptide chain from three different amino acids

Determination of the Amino Acid Sequence of a Polypeptide

This can be explained by taking the example of a dodecapeptide whose composition was found to be Ala, Arg, Glu, Gly, Leu, Lys, Phe, Tyr, Val. It was determined that the N-terminal amino acid of the dodecapeptide was valine and the C-terminal amino acid, leucine. Hydrolysis of the dodecapeptide by trypsin yielded four peptides whose structures were determined and found to be those given in A to D.

Tyr-Glu-Lys	Phe-Gly-Arg	Val-Lys	Ala-Tyr-Ala-Leu
A	**B**	**C**	**D**

Since valine was the N-terminus and leucine the C-terminus of the dodecapeptide, it is apparent that peptide C must represent the amino acid sequence at the N-terminal end, and peptide D the amino acid sequence at the C-terminal end of the dodecapeptide. To establish the order of the A and B peptides in the interior of the dodecapeptide, another sample of the

dodecapeptide was hydrolyzed by chymotrypsin, the four peptides formed were sequenced, and their structures were found to be those given in E to H.

Ala-Leu	Glu-Lys-Ala-Tyr	Val-Lys-Phe	Gly-Arg-Tyr
E	**F**	**G**	**H**

The sequences Gly-Arg-Tyr in peptide H and Glu-Lys-Ala-Tyr in peptide F clearly establish that peptide B must precede peptide A in the dodecapeptide. Hence, the structure of the dodecapeptide is unambiguously determined to be that as shown belew :

Val-Lys-Phe-Gly-Arg-Tyr-Glu-Lys-Ala-Tyr-Ala-Leu.

Biological Roles of peptides

Peptides participate in a number of biological activities.

1. They serve as intermediates in the formation of proteins.

2. They appear as constituents in a group of compounds called alkaloids. Majority of these have been isolated from fungi, although they are also found in higher plants. Ergotamine is a peptide alkaloid from rye ergot and has pronounced pharmacological properties. The four components of this alkaloid are lysergic acid, alanine, proline and phenylalanine.

3. Many of them possess antibacterial activities and are usually present in fungi and bacteria. Penicillin G with 3 components (valine, cysteine and phenylacetic acid) is a common antibiotics.

4. Certain other peptides serve as growth factors. Folic acid, a water-soluble vitamin, is a noteworthy example of it. Another group of peptides serving as growth factor for a variety of microorganisms is streptogenins.

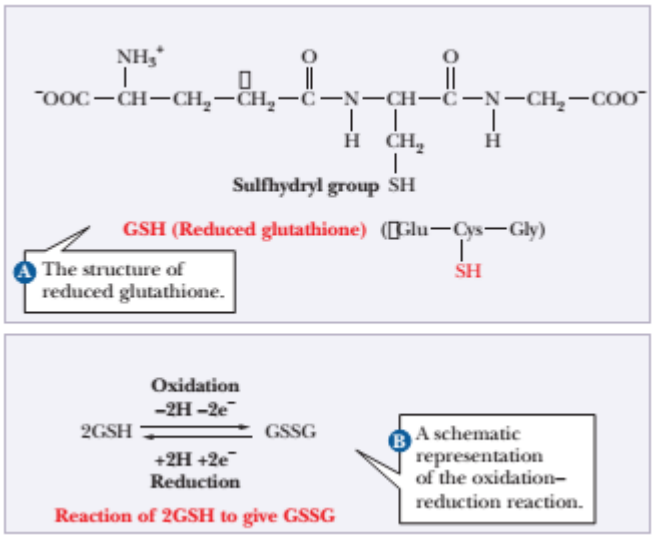
5. Higher animals do synthesize certain peptides serving as hormones.

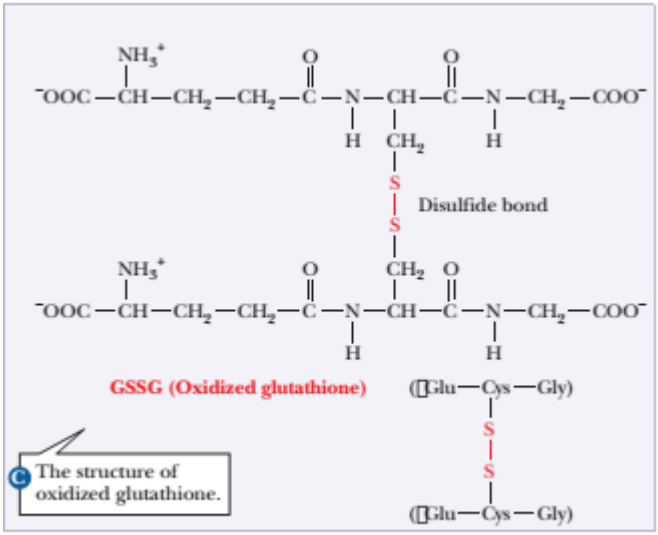
6. Certain peptides like glutathione participate in controlling the oxidation-reduction potential of the cell. This may also serve as a key intermediate in electron-transfer systems.

7. A direct correlation has been found to exist between the amounts of peptides in the urine of the patients and their mental state of disturbance. A group of Norwegian doctors have found an excess of peptides in urine specimens from patients with psychiatric disturbances. The peptides have been shown to induce in animals some of the conditions for the development of psychiatric disorders which lead to mania, depression or schizophrenia. The urine tests can, thus, indicate if a person is suffering from mental illness. The cause of this hyper secretion of peptides is not yet well established. It may exist from birth in organic genetic derangements or may be induced in healthy people by environmental factors.

PHYSIOLOGICAL ACTIVE PEPTIDE

Glutathione is a commonly occurring tripeptide; it has considerable physiological importance because it is a scavenger for oxidizing agents. Recall that oxidation is the loss of electrons; an oxidizing agent causes another substance to lose electrons. (It is thought that some oxidizing agents are harmful to organisms and play a role in the development of cancer.) In terms of its amino acid composition and bonding order, it is g-glutamyll-cysteinylglycine. The letter g(gamma) is the third letter in the Greek alphabet; in this notation, it refers to the third carbon atom in the molecule, counting the one bonded to the amino group as the first. Once again, the N-terminal amino acid is given first. In this case, the g-carboxyl group (the side-chain carboxyl group) of the glutamic acid is involved in the peptide bond; the amino group of the cysteine is bonded to it. The carboxyl group of the cysteine is bonded, in turn, to the amino group of the glycine. The carboxyl group of the glycine forms the other end of the molecule, the C-terminal end. The glutathione molecule shown below is the reduced form. It scavenges oxidizing agents by reacting with them. The oxidized form of glutathione is generated from two molecules of the reduced peptide by forming a disulfi de bond between the !SH groups of the two cysteine residues.





The oxidation and reduction of glutathione. (a) The structure of reduced

Two pentapeptides found in the brain are known as enkephalins, naturally occurring analgesics (pain relievers). For molecules of this size, abbreviations for the amino acids are more convenient than structural formulas. The same notation is used for the amino acid sequence, with the N-terminal amino acid listed first and the C-terminal listed last. The two peptides in question, leucine enkephalin and methionine enkephalin, differ only in their C-terminal amino acids. It is thought that the aromatic side chains of tyrosine and phenylalanine in these peptides play a role in their activities. It is also thought that there are similarities between the three-dimensional structures of opiates, such as morphine, and those of the enkephalins. As a result of these structural similarities, opiates bind to the brain's receptors for the enkephalins and thus produce their physiological activities.

Tyr—Gly—Gly—Phe—Leu (three-letter abbreviations)

Y—G—G—F—L (one-letter abbreviations)

Leucine enkephalin

Tyr—Gly—Gly—Phe—Met

Y—G—G—F—M

Methionine enkephalin

Some important peptides have cyclic structures. Two well-known examples with many structural features in common are oxytocin and vasopressin

PROTEIN STRUCTURE

The covalent backbone of a typical protein contains hundreds of individual bonds. Because free rotation is possible around many of these bonds, the protein can assume a very large number of conformations. However, each protein has a specific chemical or structural function, strongly suggesting that each has a unique three-dimensional structure. Given that, generally, the ordered array of molecules in a crystal can form only if the molecular units are identical, the finding that many proteins could be crystallized was evidence that even very large proteins are discrete chemical entities with unique structures. This conclusion revolutionized thinking about proteins and their functions This section of the note will examine how sequence of amino acids in a polypeptide chain is translated into a discrete, three-dimensional protein structure. This note will emphasize five themes. First, the three-dimensional structure of a protein is determined by its amino acid sequence. Second, the function of a protein depends on its structure. Third, an isolated protein usually exists in one or a small number of stable structural forms. Fourth, the most important forces stabilizing the specific structures maintained by a given protein are non-covalent interactions. Finally, amid the huge number of unique protein structures, we can recognize some common structural patterns that help to organize our understanding of protein architecture.

The spatial arrangement of atoms in a protein is called its conformation. The possible conformations of a protein include any structural state it can achieve without breaking covalent bonds. A change in conformation could occur, for example, by rotation about single bonds. Of the many conformations that are theoretically possible in a protein containing hundreds of single bonds, one or (more commonly) a few generally predominate under biological conditions. The need for multiple stable conformations reflects the changes that must take place in most proteins as they bind to other molecules or catalyze reactions. The conformations existing under a given set of conditions are usually the ones that are thermodynamically the most stable-that is, having the lowest Gibbs free energy (G). Proteins in any of their functional, folded conformations are called native proteins.

Hydrophobic interactions are clearly important in stabilizing conformation; the interior of a protein is generally a densely packed core of hydrophobic amino acid side chains. It is also important that any polar or charged groups in the protein interior have suitable partners for hydrogen bonding or ionic interactions. One hydrogen bond seems to contribute little to the stability of a native structure, but the presence of hydrogen bonding groups without partners in the hydrophobic core of a protein can be so destabilizing that conformations containing these groups are often thermodynamically untenable. The favorable free-energy change resulting from the combination of several such groups with partners in the surrounding solution can be greater than the free-energy difference between the folded and unfolded states. In addition, hydrogen bonds between groups in a protein form cooperatively (formation of one makes the next one more likely) in repeating secondary structures that optimize hydrogen bonding. Hydrogen

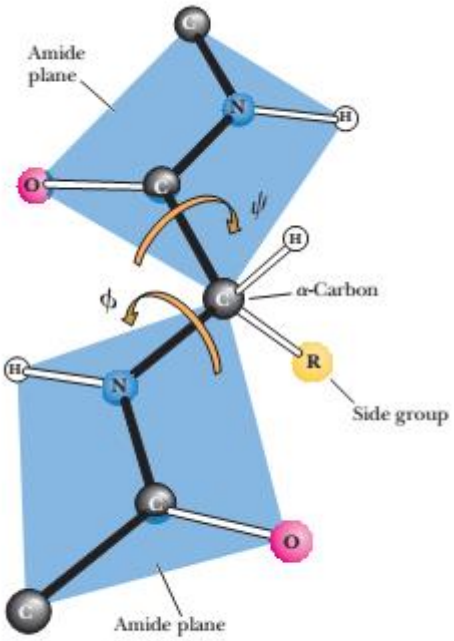
bonds often have an important role in guiding the protein-folding process. The interaction of oppositely charged groups that form an ion pair, or salt bridge, can have either a stabilizing or destabilizing effect on protein structure. As in the case of hydrogen bonds, charged amino acid side chains interact with water and salts when the protein is unfolded, and the loss of those interactions must be considered when evaluating the effect of a salt bridge on the overall stability of a folded protein. However, the strength of a salt bridge increases as it moves to an environment of lower dielectric constant.

Primary structure is the order in which the amino acids are covalently linked together. The peptide Leu-Gly-Thr-Val-Arg-Asp-His (recall that the N-terminal amino acid is listed first) has a different primary structure from the peptide Val-His-Asp-Leu-Gly-Arg-Thr, even though both have the same number and kinds of amino acids. Note that the order of amino acids can be written on one line. The primary structure is the one-dimensional first step in specifying the three-dimensional structure of a protein. Some biochemists define primary structure to include all covalent interactions, including the disulfide bonds that can be formed by cysteines; however, we shall consider the disulfide bonds to be part of the tertiary structure. One of the most striking demonstrations of the importance of primary structure is found in the hemoglobin associated with sickle-cell anemia. In this genetic disease, red blood cells cannot bind oxygen efficiently. The red blood cells also assume a characteristic sickle shape, giving the disease its name. The sickled cells tend to become trapped in small blood vessels, cutting off circulation and thereby causing organ damage. These drastic consequences stem from a change in one amino acid residue in the sequence of the primary structure.

Secondary Structure of Proteins

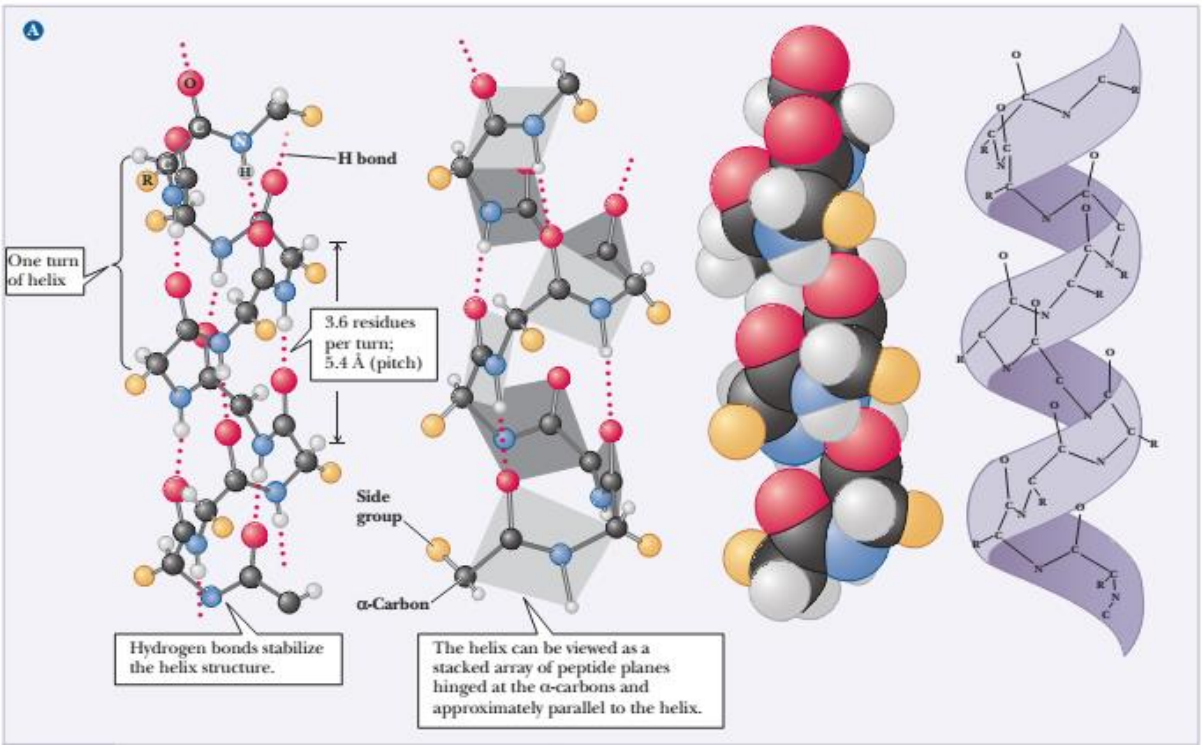
The secondary structure of proteins is the hydrogen-bonded arrangement of the backbone of the protein, the polypeptide chain. The nature of the bonds in the peptide backbone plays an important role here. Within each amino acid residue are two bonds with reasonably free rotation: (1) the bond between the α-carbon and the amino nitrogen of that residue and (2) the bond between the α-carbon and the carboxyl carbon of that residue. The combination of the planar peptide group and the two freely rotating bonds has important implications for the three-dimensional conformations of peptides and proteins. A peptide-chain backbone can be visualized as a series of playing cards, each card representing a planar peptide group. The cards are linked at opposite corners by swivels, representing the bonds about which there is considerable freedom of rotation (Figure 4.1). The side chains also play a vital role in determining the three-dimensional shape of a protein, but only the backbone is considered in the secondary structure. The angles f(phi) and c(psi), frequently called Ramachandran angles (after their originator, G. N. Ramachandran), are used to designate rotations around the C!N and C!C bonds, respectively. The conformation of a protein backbone can be described by specifying the values of fand cfor each residue (2180° to 180°). Two kinds of secondary structures that occur frequently in proteins are the repeating a-helixand b-pleated sheet (or b-sheet) hydrogen-bonded structures. The fand cangles repeat themselves in contiguous amino acids in

regular secondary structures. The a-helix and b-pleated sheet are not the only possible secondary structures, but they are by far the most important.

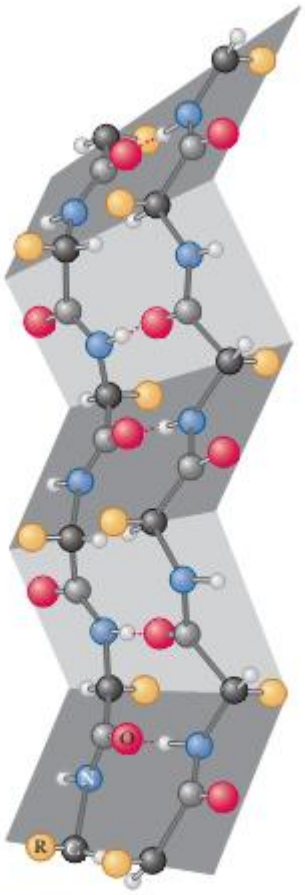


■ **FIGURE 4.1 Phi and psi angles of the peptide backbone.** Definition of the angles that

The a-helix is stabilized by hydrogen bonds parallel to the helix axis within the backbone of a single polypeptide chain. Counting from the N-terminal end, the C-O group of each amino acid residue is hydrogen bonded to the N-H group of the amino acid four residues away from it in the covalently bonded sequence. The helical conformation allows a linear arrangement of the atoms involved in the hydrogen bonds, which gives the bonds maximum strength and thus makes the helical conformation very stable. Proteins have varying amounts of a-helical structures, varying from a few percent to nearly 100% e.g. the helices are represented by the regularly coiled sections of the ribbon diagram. Myohemerythrin is an oxygen-carrying protein in invertebrates and collagen.



The arrangement of atoms in the β-pleated sheet conformation differs markedly from that in the a-helix. The peptide backbone in the b-sheet is almost completely extended. Hydrogen bonds can be formed between different parts of a single chain that is doubled back on itself (intra-chain bonds) or between different chains (inter-chain bonds). If the peptide chains run in the same direction (i.e., if they are all aligned in terms of their N-terminal and C-terminal ends), a parallel pleated sheet is formed. When alternating chains run in opposite directions, an antiparallel pleated sheet is formed. The hydrogen bonding between peptide chains in the b-pleated sheet gives rise to a repeated zigzag structure; hence, the name “pleated sheet”. Note that the hydrogen bonds are perpendicular to the direction of the protein chain, not parallel to it as in the a-helix.



■ **FIGURE 4.5 The three-dimensional form of the antiparallel *β*-pleated sheet arrangement.** The chains do not fold back on each other but are in a fully extended conformation. *(Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)*

Protein Tertiary and Quaternary Structures

Protein Architecture- The overall three-dimensional arrangement of all atoms in a protein is referred to as the protein's tertiary structure. Whereas the term "secondary structure" refers to the spatial arrangement of amino acid residues that are adjacent in a segment of a polypeptide, tertiary structure includes longer-range aspects of amino acid sequence. Amino acids that are far apart in the polypeptide sequence and are in different types of secondary structure may interact within the completely folded structure of a protein. The location of bends (including β- turns) in the polypeptide chain and the direction and angle of these bends are determined by the number and Iocation of specific bend-producing residues, such as Pro, Thr, Ser, and GIy. Interacting segments of polypeptide chains are held in their characteristic tertiary positions by several kinds of weak interactions (and sometimes by covalent bonds such as disulflde crosslinks) between the segments. Some proteins contain two or more separate polypeptide chains, or subunits, which may be identical or different. The arrangement of these protein subunits in three-dimensional complexes constitutes quaternary structure. In considering these higher levels of structure, it is useful to classify proteins into two major groups: fibrous proteins, with polypeptide chains arranged in long strands or sheets, and globular

proteins, with polypeptide chains folded into a spherical or globular shape. The two groups are structurally distinct. Fibrous proteins usually consist largely of a single type of secondary structure, and their tertiary structure is relatively simple. Globular proteins often contain several types of secondary structure. The two groups also differ functionally: the structures that provide support, shape, and external protection to vertebrates are made of flbrous proteins, whereas most enzymes and regulatory proteins are globular proteins.

In a globular protein, different segments of the polypeptide chain (or multiple polypeptide chains) fold back on each other, generating a more compact shape than is seen in the fibrous proteins (Fig. 4-14). The folding also provides the structural diversity necessary for proteins to carry out a wide array of biological functions. Globular proteins include enzymes, transport proteins, motor proteins, regulatory proteins, immunoglobulins, and proteins with many other functions. The first breakthrough in understanding the three-dimensional structure of a globular protein came from x-ray diffraction studies of myoglobin carried out by John Kendrew and his colleagues in the1950s. Myoglobin is a relatively small (M, 16,700),oxygen-binding protein of muscle cells. It functions both to store oxygen and to facilitate oxygen diffusion in rapidly contracting muscle tissue. Myoglobin contains a single polypeptide chain of 153 amino acid residues of known sequence and a single iron protoporphyrin, or heme, group. The same heme group that is found in myoglobin is found in hemoglobin, the oxygen-binding protein of erythrocytes, and is responsible for the deep red-brown color of both myoglobin and hemoglobin.

α-Keratin, collagen, and silk flbroin nicely illustrate the relationship between protein structure and biological function (Table 4-2). Fibrous proteins share properties that give strength and./or flexibility to the structures in which they occur. In each case, the fundamental structural unit is a simple repeating element of secondary structure. All fibrous proteins are insoluble in water, a property conferred by a high concentration of hydrophobic amino acid residues both in the interior of the protein and on its surface. These hydrophobic surfaces are largely buried as many similar polypeptide chains are packed together to form elaborate supramolecular complexes. α-Keratin The α-keratins have evolved for strength. Found only in mammals, these proteins constitute almost the entire dry weight of hair, wool, nails, claws, quills, horns, hooves, and much of the outer layer of skin. The a-keratins are part of a broader family of proteins called intermediate f,lament (IF) proteins. Other IF proteins are found in the cytoskeletons of animal cells. All IF proteins have a structural function and share the structural features exemplified by the a-keratins.