Today we begin our discussion of the structure and properties of proteins.

Proteins consist in whole or large part of amino acids.

Simple proteins consist only of amino acids.

Conjugated proteins contain a non-amino acid component called the prosthetic group.

The prosthetic group can be an organic or an inorganic component such as a sugar molecule or a metal ion.

Conjugated proteins are classified by the prosthetic group.

Examples are nucleoproteins, lipoproteins and metalloproteins that contain nucleic acids, lipids and metal ions respectively.

There are approximately 20 different amino acids found in proteins.

Most of the 20 common amino acids (a.a.) have the following general structure-

The amino acids are distinguished from each other by the R group often referred to as the side-chain in proteins.

The amino acids have an amino and a carboxylic acid group that are joined to the same carbon atom. This carbon atom is referred to as the α carbon.

In proteins, amino acids are covalently linked to one another by peptide (amide) bonds involving the carboxyl group of one a.a. and the amino group of another a.a. to form linear polymers-

Let's examine the structure and properties of the amino acids. The a.a. are usually grouped according to the polarity and charge of the R group which relates to their role in proteins.

Table 4-1 places a.a. into three groups: a.a. with nonpolar side chains, a.a. with uncharged polar side chains and a.a. with charged polar side chains.

Consider a.a. with nonpolar side chains. This group includes the simplest a.a. glycine whose R group is a H atom.

Most of the amino acids in this group have nonpolar alkyl or aromatic R groups. Alanine, valine, leucine, isoleucine, and phenylalanine have methyl, isopropyl, isobutyl, secondary butyl and benzyl groups.

Methionine has a thioether group that is nonpolar because the electronegativity of sulfur is very similar to carbon.

Included in this group is the amino acid proline that is a secondary α amino acid. The N and α carbon atoms are parts of a five membered ring.

The a.a. tryptophan has a bicyclic ring referred to as indole. Note that while the R group appears to be relatively nonpolar, the N-H group of indole is polar and can H-bond.

Most of the amino acids with uncharged polar side chains have O and/or N atoms.

The side chains are polar because there is a large difference in the electronegativities of these atoms and the atoms to which they are bonded. This difference contributes to an unequal sharing of electrons in the bond that results in a dipole.

Serine and threonine have hydroxyl groups.

Asparagine and glutamine have amide groups as part of their side chains. They differ from one another by a methylene, $-CH_2-$, unit.

Tyrosine has a phenolic group.

Cysteine has a mercapto, -SH, group. The mercapto group is only weakly polar.

The a.a. with charged polar side chains have basic or acidic groups that are partly or completely ionized at pH 7.0.

The basic a.a. include lysine, arginine and histidine. Because they are basic, they will associate with H^+ in solution to form the positively charged groups as shown in Table 4-1 for lysine and arginine.

Lysine has an amino group on the ε (epsilon) carbon atom.

The basic group of arginine is referred to as the guanidino group-

The conjugate acid form shown in Table 4-1 is called the guanidinium group-

The basic group of histidine is an imidazole group.

The imidazole group of histidine is less basic than the amino or guanidino group so that at pH 7.0 it is predominately unprotonated as shown in Table 4-1.

The acid form is called the imidazolium group-

The acidic amino acids include aspartic and glutamic acids that differ from one another by a methylene unit.

Because the carboxylic acid groups are acidic, they will dissociate to give a H^+ and the negatively charged carboxylate group at pH 7.0 as shown in Table 4-1.

Note that these groups are referred to as the β and χ carboxyl groups because they are attached to the β and χ carbon atoms.

Note that each a.a. has a 3 letter and a 1 letter abbreviation. This is useful in representing the sequence of a.a. in a protein.