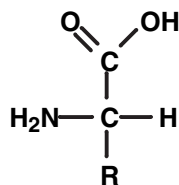# Amino Acids, Peptides, and Proteins
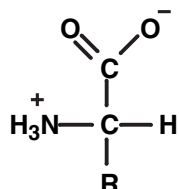
The α Amino Acids

The general structure of the amino acids found in proteins could be depicted as:

As their name implies, they all contain a carboxylic acid group and an amino group. They are called alpha (α) amino acids based upon an archaic system of nomenclature. In this system the carbon adjacent to a carboxyl carbon is designated the α carbon. Since the naturally occurring amino acids have the amino group on the a carbon they are α amino acids.

The carboxylic acid group is a weak acid with a pK_a in the range of 1.7 to 2.6. The carboxylic acid groups of the α amino acids are stronger acids than simple organic acids. The amino group is a weak base having a pK_a in the range of 8.8 to 10.8.

Since the carboxyl group is an acid and the amino group is a base, in solution amino acids self neutralize / autoneutralize. The structure given previously for the naturally occurring amino acids is incorrect, the correct structure for an amino acid in solution is:


This DIPOLAR IONIC form exists when the amino acids are in solution at neutral pH and when they are in the solid state. The dipolar ionic form explains many of the properties of the amino acids. For example their melting points, boiling points, and solubilities more closely resemble those of ionic salts rather than organic acids. The dipolar ionic form of an amino acid is called a ZWITTER ION. Any molecular compound that contains a mixture of positive and negative charges is a zwitter ion.

pH Properties of the Amino Acids

The ionic state of the amino acids depends upon the pH of the solution in which they are dissolved. At low pH (pH ~ 1.0) the amino acid is in the cationic form and if placed in an electric field they will migrate toward the cathode. As the pH is increased the carboxylic acid group ionizes. When the pH is equal to the pK_a the amino acid exists as a 50:50 mixture of the cationic and zwitter ionic forms. Adding more base results in continued ionization of the carboxylic acid group until the zwitter ionic form is the predominant form of the amino acid in solution. By the addition of more base, the pK_a of the amino group is reached and at this point the amino acid exists as a 50:50 mixture of the zwitter ionic form and the anionic form. As the pH is increased further the amino group continues loses its

proton and ultimately, at high pH (pH ~ 12.0), the anionic form is the predominant form in solution. The anionic form will migrate toward the anode if placed in an electric field. This titration curve demonstrates that the a amino acids never exist in an uncharged form.

As the amino acid is titrated, there is a solution pH along the titration curve where the amount of positive charge on the molecule is exactly balanced by the amount of negative charge. At this pH the amino acid in solution has no net charge. The amount of positive charge is exactly balanced by an equal amount of negative charge. The solution pH at which the opposite charges exactly balance one another is called the ISOELECTRIC POINT (pI) of the amino acid.
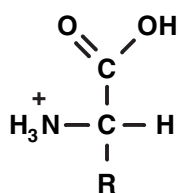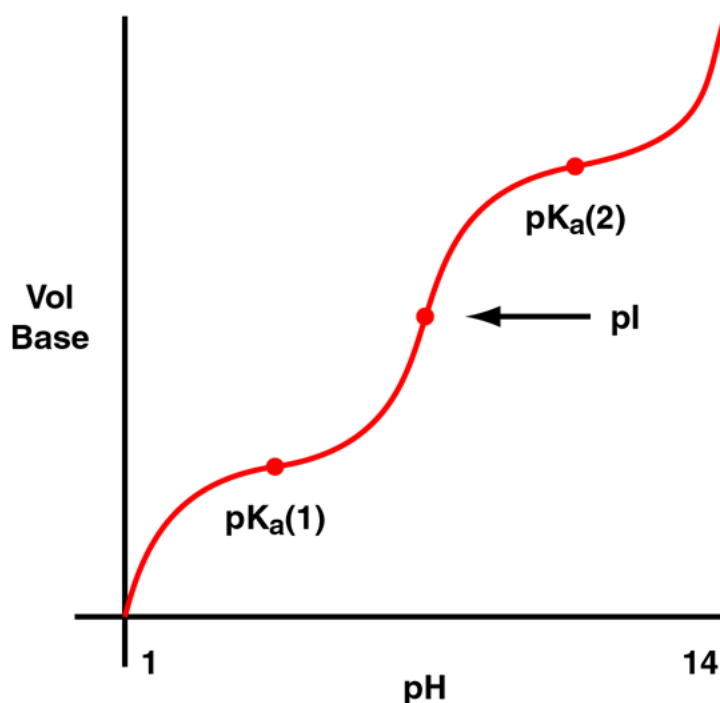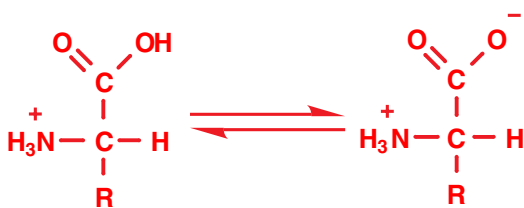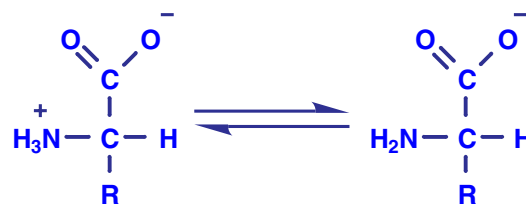

Dominant form
at pH ≤ 1
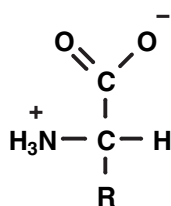
Dominant form
at pI
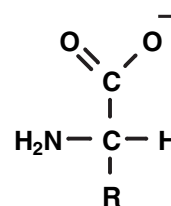
Dominant form
at pH ≥ 12

Forms present at pK$_a$(1)

Forms present at pK$_a$(2)

At the isoelectric point an amino acid will not migrate in an electric field since it has no net charge. At pH's below the isoelectric point, the amino acid has a net positive charge and it will migrate toward the cathode; above the isoelectric point the molecule has a net negative charge and it will migrate toward the anode.

The R Groups - The Amino Acid Side Chains

The different R groups, the different amino acid side chains give each of the 20 amino acids different chemical and physical properties. Different combinations of the amino acids within a protein give the protein its unique chemical and physical properties. The unique sequence of amino acids defines the structure and function of the protein.

Its time to examine the structure of the side chains of the 20 naturally occurring amino acids

The amino acids can be divided into two major groups: Nonpolar (Hydrophobic) Amino Acids and Polar (Hydrophilic) Amino Acids. The Polar Amino Acids can be divided into three subgroups: Polar But Not Charged Amino Acids, Amino Acids with Negatively Charged Side Chains (Acidic Amino Acids), and Amino Acids with Positively Charged Side Chains (Basic Amino Acids).
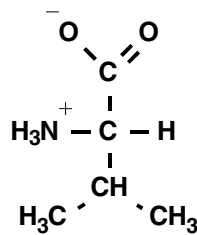
The Hydrophobic (Nonpolar) Amino Acids are
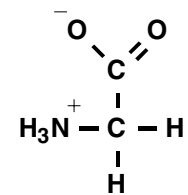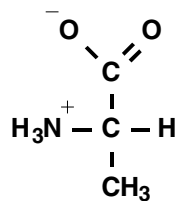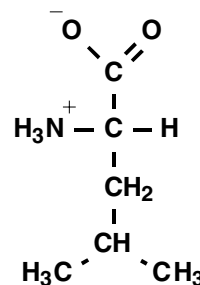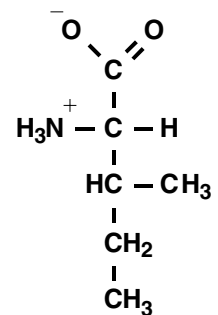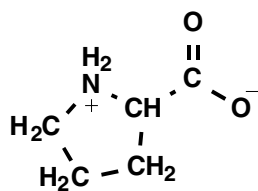

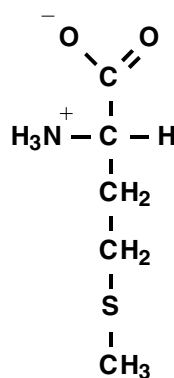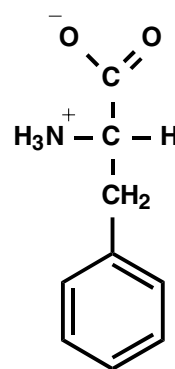Glycine (Gly, G)   Alanine (Ala, A)   Valine (Val, V)   Leucine (Leu, L)   Isoleucine (Ile, I)
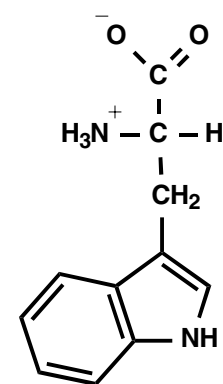
Proline (Pro, P)   Methionine (Met, M)   Phenylalanine (Phe, F)   Tryptophan (Trp, W)

Note 1: The amino acids Tryptophan (Trp) and Glycine (Gly) are sometimes included in the polar but not charged group of amino acids. Tryptophan is sometimes included in this group because the

nitrogen in the ring system makes it slightly polar. However, the complex ring system is fairly non-polar and overall Trp behaves most often like a nonpolar amino acid. Glycine is included because its side chain is the small H and because its solubility in water resembles that of the other polar amino acids.

The Polar But Not Charged Amino Acids are:
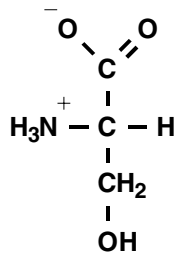
Serine(Ser, S)
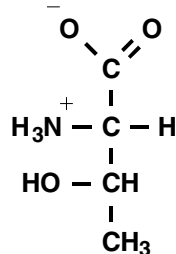
Threonine (Thr, T)
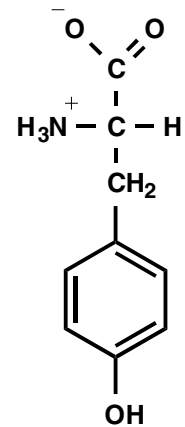
Tyrosine (Tyr, Y)
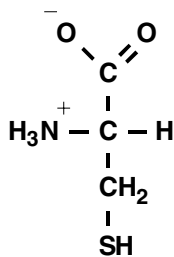
Cysteine (Cys, C)
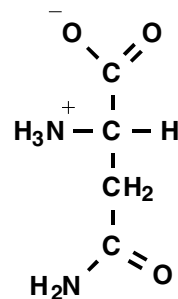
Asparagine (Asn, N)
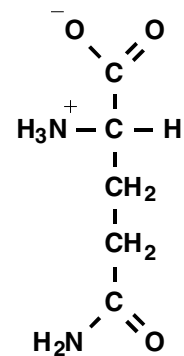
Glutamine (Gln, Q)

Note 2: The side chains of cysteine and tyrosine have some acid / base properties. The -SH group on cysteine has a pK_a of 8.3 and the -OH group on the tyrosine side chain has a pK_a of 10.1. These pK_a's can be measured *in vitro* (in a test tube). At physiological pH, pH 7.4, these side chains are for the most part unionized.

Note 3: Phenylalanine, Tyrosine and Tryptophan are called the aromatic amino acids because they all contain an aromatic ring as part of their side chains. The aromatic groups on the side chains of Phenylalanine, Tyrosine and Tryptophan absorb light at the ultraviolet region of the spectrum, at 280 nm. This property is employed by biochemists to detect and quantify proteins in solutions and / or biological fluids.

The Amino Acids with Negatively Charged Side Chains (Acidic Amino Acids) are:

Aspartate (Asp, D)
[Aspartic Acid]

Glutamate (Glu, E)
[Glutamic Acid]
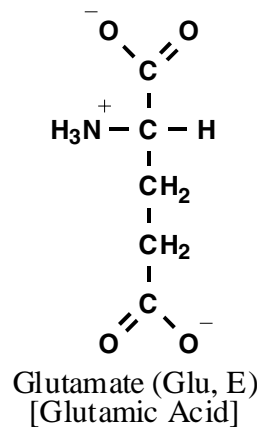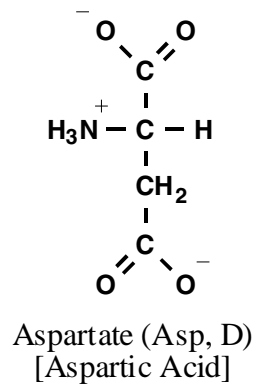
The side chain carboxyl group of aspartate (aspartic acid) has a pK_a of 3.9 and the side chain carboxyl group of glutamate (glutamic acid) has a pK_a of 4.3. These values are closer the pK_a's of simple organic acids.

The Basic Amino Acids are:

Lysine (Lys, K)

Histidine (His, H)

Arginine (Arg, R)
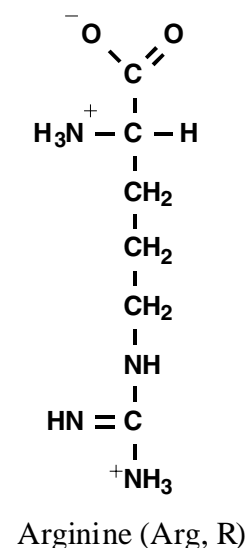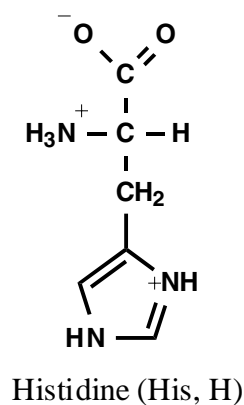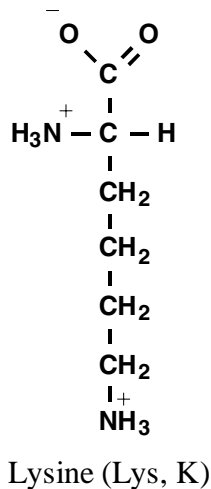
The imidazole ring of histidine has a pK_a of 6.0, the side chain amino group of lysine has a pK_a of 10.5, and the guanidinium group of arginine has a pK_a of 12.5.

Chiral Carbon

When the R group (side chain) of an amino acid is an organic group other than a hydrogen (H) atom, the α carbon of the molecule is an asymmetric carbon; a chiral carbon (a tetrahedral stereocenter). A chiral carbon in the molecule means that the molecule can exists as a pair of stereoisomers. These stereoisomers have mirror image configurations; they exist as a pair of enantiomers. The pair of enantiomers for an α amino acid have the following configurations:
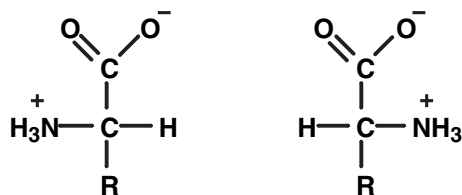
L Form   D Form

In nature, when two enantiomers exist, usually only one is synthesized and used by cells. In the case of the amino acids, only the L configuration exists in nature and is used by cells. Some bacteria and fungi synthesize and use D configuration amino acids. The molecules synthesized with these D-form amino acids are often toxic to other living organisms. With these few minor exceptions all of the naturally occurring amino acids are in the L-form.

Calculation of Amino Acid Isoelectric Point

For the nonpolar and polar amino acids with two pK_a's, the isoelectric point is calculated by taking the numerical average of the carboxyl group pK_a and the α-amino group pK_a.

The titration curves for seven of the amino acids (Cys, Try, Glu, Asp, His, Lys, & Arg) demonstrate three inflection points, three pK_a's. One for the α-amino group, one for the carboxyl group (carbon 1), and one for the ionizable side chain. The isoelectric point for these amino acids in calculated by taking the numerical average of the pK_a's of the groups with like charge when ionized. For example to calculate the isoelectric point of Glu, the pK_a's of the two carboxyl groups are averaged. To calculate the isoelectric point of Arg, the pK_a's of the α-amino group and the guanidinium group are averaged. Remember, at the isoelectric point the amino acid has no net charge. For the amino acids with three ionizable groups, the total charge on the groups with like charge must equal one (1) so that it can be balanced by the one (1) opposite charge present on the molecule.

Proteins: Their Primary Structure and Biological Functions

The word protein comes from the Greek PROTEIOS which means first or primary. Proteins are of primary importance to cells and organisms. Proteins can be compared to words of the English language. The words that are read, written, and spoken are composed of the 26 letters of the alphabet. The sequence of letters is the primary structure of the word. From the 26 letters of the alphabet an infinite number of "words" can be made. Some of these words have meaning, others are just gibberish. For proteins the alphabet is the 20 amino acids. The sequence of amino acids is the primary structure of the polypeptide. From these 20 amino acids an infinite number of polypeptides can be made, each with a different primary structure. Many of these polypeptides are just gibberish. The cell only synthesizes proteins that have meaning.
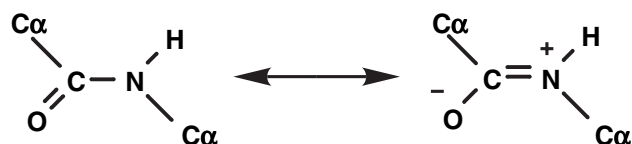
Peptide Bonds
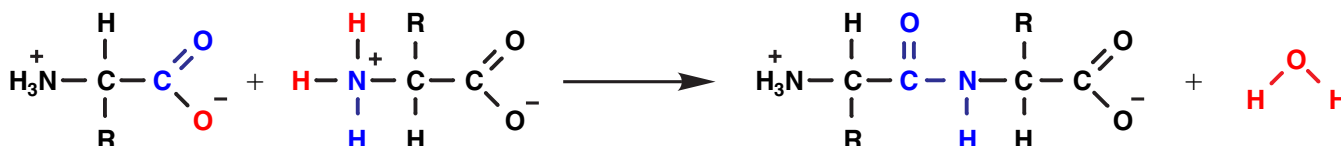
Amino acids are linked in proteins in a head to tail manner by a condensation reaction between the carboxyl group of one amino acid and the α amino group of the second. The "back bone" of a protein consists of the repeating sequence -N-Cα-C-. The amino acid side chains of the amino acids project perpendicularly from the back bone of the molecule.

The bond between amino acids in a protein is an amide bond. Since this amide bond holds peptides and proteins together it is called a PEPTIDE BOND. This amide linkage, the peptide bond, has no acid or base properties. It will neither donate nor accept a proton. However, the peptide bond is very polar.

Physical chemical studies have shown that the peptide bond exists in two resonance forms:

The real nature of the peptide bond lies somewhere between these two extremes, it has partial double bond character. The partial double bond character of the peptide bond restricts free rotation about this bond limiting the possible number of conformations that the peptide or protein can assume. It also places the six atoms of the peptide bond in the same plane.
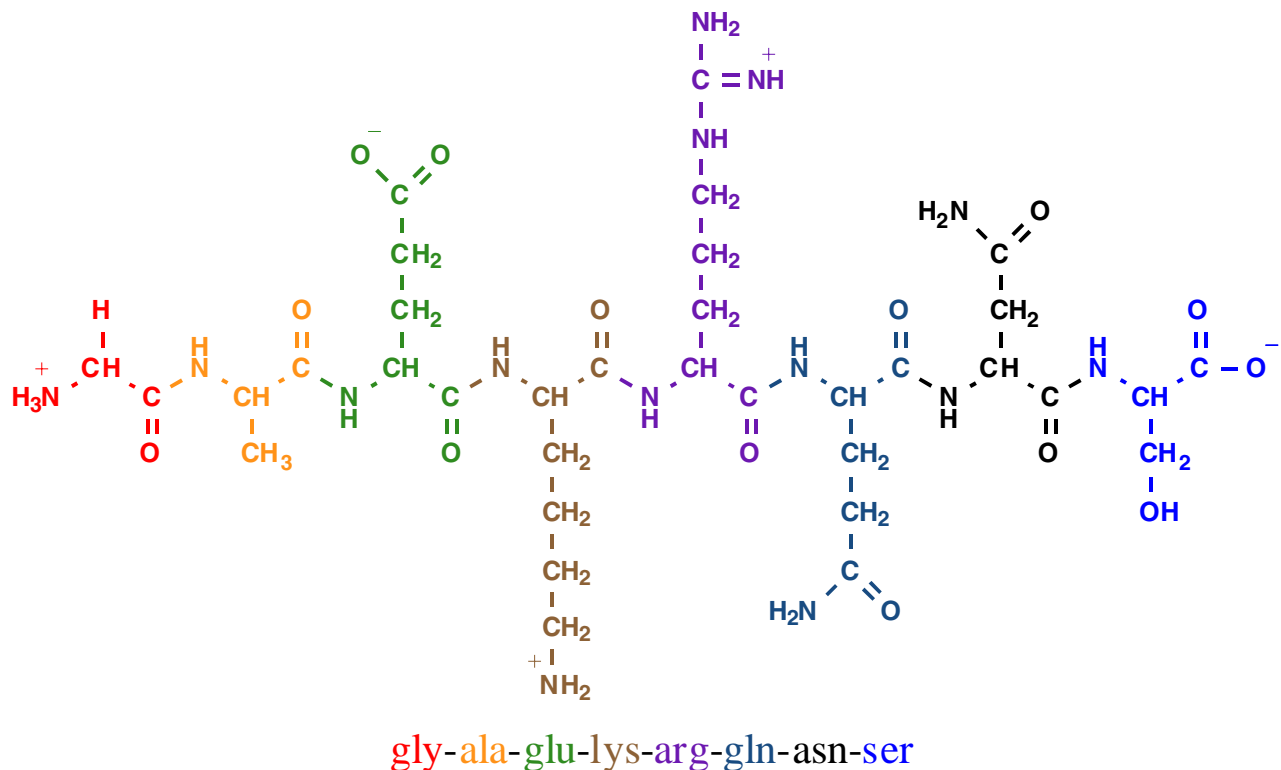
Peptide / Protein Terminology

The unique sequence of amino acids in a peptide or protein is termed the Primary (1°) Structure of the Protein. A gene contains the information necessary for the synthesis of a protein, for the assembly of the primary structure. The primary structure contains the information necessary for the protein to fold into its final three dimensional conformation and once correctly folded to assume its cellular function. If the 1° structure of a protein is changed very often the final shape changes resulting in a nonfunctional polypeptide.

A molecule containing two amino acids joined by a peptide bond is a Dipeptide; one with three amino acids held together by two peptide bonds is a Tripeptide; four is a Tetrapeptide; etc. In general, Peptides contain 12 or fewer amino acid residues. An Oligopeptide contains between 12 and 20 amino acids and a Polypeptide contains greater than 20 amino acids. Twenty appears to be a magic number with respect to peptide/protein structure. Oligopeptides with 20 or fewer amino acids do not fold into, do assume a single low energy conformation, rather they exist in numerous random shapes. Molecules with greater than 20 amino acids very often fold into a single stable conformation, a single low energy conformation. The terms polypeptide and protein are often used interchangeably. However, the term protein is often reserved for molecules that perform some cellular function.

Monomeric Proteins contain a single polypeptide chain. Some proteins are supramolecular complexes composed of more than one polypeptide chain. These proteins are called Multimeric Proteins. Homomultimeric Proteins are composed of several polypeptides all with the same sequence of amino acids, all with the same primary structure. Heteromultimeric Proteins are composed of several different polypeptides; polypeptides with different primary structures.

Sequence Convention

gly-ala-glu-lys-arg-gln-asn-ser

A protein is a long linear sequence of amino acids joined by peptide bonds. One end of the molecule has a free amino group, an amino group not involved in a peptide bond. This is the Amino Terminus or N-Terminus of the protein. The opposite end of the molecule has a free carboxyl group, a carboxyl group not involved in a peptide bond. This end is the Carboxy Terminus, Carboxyl Terminus or C-Terminus of the protein.

When biochemists write the sequence of a protein they use either the three letter or the one letter abbreviation for the amino acids, usually separated by hyphens. The amino terminus is always the Left most amino acid and the carboxyl terminus is always the Right most amino acid.

Biological Functions of Proteins

Within the cell / organism proteins serve a wide range of important biological functions.

Enzymes are the biological catalysts of the chemical reactions that occur within the cell. Enzymes will be examined in detail shortly.

Regulatory Proteins regulate the activities of the cell and the ability of other proteins to carry out their cellular function. The *peptide and protein hormones* are regulatory proteins. They play a role in regulating overall metabolism, growth, development, and maintenance of the organism. *Allosteric enzymes* can be considered regulatory proteins since they control key cellular reactions. *Gene inducers* and *gene repressors* are also considered regulatory proteins. Gene inducers stimulate gene expression, they turn genes "on". Gene repressors inhibit gene expression, they turn genes "off".

Transport Proteins carry specific substances from one place to another. *Membrane transporters* carry polar molecules across cell membranes. *Hemoglobin* transports oxygen from the lungs to the tissues; *serum albumin* carries a wide variety of drugs and metabolites through out the body.

Storage Proteins provide a reservoir of an essential nutrient. *Myoglobin* stores oxygen in skeletal muscle tissue; *ferritin* stores iron in the liver and bone marrow; the protein *ovalbumin* serves as a storage depot of amino acids for developing birds; and *casein*, the major protein in mammalian milk stores amino acids for growing infants.

Structural proteins provide strength, support, and form to cells, tissues, and organisms. *Collagen* is the major protein of bone, tendons, and cartilage. *a-Keratin* is the protein of hair, horns, hooves, and fingernails. There are cytoskeletal protein fibers adjacent to the cell membrane that give the cell shape and support. *Tubulin, actin,* and *spectrin* are some of the proteins that make-up the cytoskeleton.

Contractile and Motile Proteins provide the cell / organism with motion. Examples include *actin* and *myosin* of muscle cells and *tubulin* of cilia and flagella. The proteins *dynein* and *kinesin* drive the movement of vesicles and organelles along cytoskeletal tracks within the cell. A single protein can serve several functions within a cell. For example actin and tubulin can be classified as structural proteins or motile proteins depending upon their location within the cell.

Scaffold Proteins act as bridges by binding to and localizing specific proteins to specific sites within the cell. They act as a site upon which supramolecular complexes are formed.

Protective Proteins play an active role in cell defense or in defense of the organism. *Antibodies* of the immune system protect the organism from foreign invaders. The *hemostasis cascade* of proteins protects the organism from excessive blood loss. *Toxins*, such as snake venoms, also fall into this class of proteins.

Exotic Proteins display functions that do not fit the other classifications. One example is the *glue protein* secreted by mussels. This protein allows the mussels to anchor to hard surfaces.

Conjugated Proteins

Conjugated proteins contain nonprotein components integral to their structure and necessary for their function. The nonprotein part is called a Prosthetic Group. Conjugated proteins include:

Glycoproteins - the prosthetic groups are carbohydrates.

Lipoproteins - the prosthetic groups are lipids.

Nucleoproteins - the prosthetic groups are nucleotides or nucleic acids.

Phosphoproteins - have phosphoryl (phosphate) groups in ester linkage to hydroxyl groups on the protein.

Metalloproteins - contain metal ions attached either by ionic interactions or by coordinate covalent bonds.