B.Sc First year Zoology (Subsidiary)


Paper-1

# Structure of Proteins


Increasingly, drug developers are looking to large molecules, particularly proteins, as a therapeutic option. Formulation of a protein drug product can be quite a challenge, and without a good understanding of the nature of protein structure and the conformational characteristics of the specific protein being formulated, the results can be ruinous. This technical brief aims to give the reader a quick overview of protein structure. It will also cover briefly how protein structure can be affected during formulation and some of the analytical methods which can be used both to determine the structure and analyze the stability of the protein.

The term, structure, when used in relation to proteins, takes on a much more complex meaning than it does for small molecules. Proteins are macromolecules and have four different levels of structure – primary, secondary, tertiary and quaternary.

## Primary Structure
There are 20 different standard L-α-amino acids used by cells for protein construction. Amino acids, as their name indicates, contain both a basic amino group and an acidic carboxyl group. This di functionality allows the individual amino acids to join in long chains by forming peptide bonds: amide bonds between the $-NH_2$ of one amino acid and the -COOH of another. Sequences with fewer than 50 amino acids are generally referred to as peptides, while the terms, protein and polypeptide, are used for longer sequences. A protein can be made up of one or more polypeptide molecules. The end of the peptide or protein sequence with a free carboxyl group is called the carboxy-terminus or C-terminus. The terms, amino-terminus and N-terminus, describe the end of the sequence with a free α-amino group.
The amino acids differ in structure by the substituent on their side chains. These side chains confer different chemical, physical, and structural properties to the final peptide or protein. The structures of the 20 amino acids commonly found in proteins are shown in Figure 1. Each amino acid

has both a one-letter and three-letter abbreviation. These abbreviations are commonly used to simplify the written sequence of a peptide or protein.

Figure 1



AMINO ACID STRUCTURES AND ABBREVIATIONS

Lubrizol Life Science
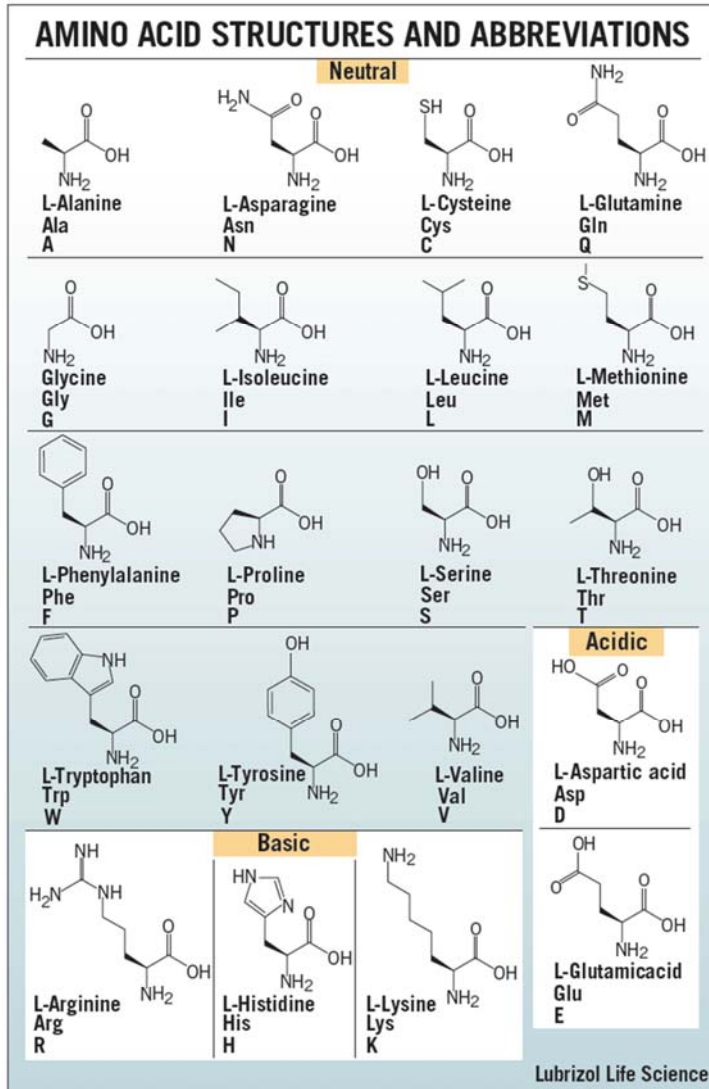
Depending on the side-chain substituent, an amino acid can be classified as being acidic, basic or neutral. Although 20 amino acids are required for synthesis of various proteins found in humans, we can synthesize only ten. The remaining 10 are called essential amino acids and must be obtained in the diet.

The amino acid sequence of a protein is encoded in DNA. Proteins are synthesized by a series of steps called transcription (the use of a DNA strand to make a complimentary messenger RNA strand – mRNA) and

translation (the mRNA sequence is used as a template to guide the synthesis of the chain of amino acids which make up the protein). Often, post-translational modifications, such as glycosylation or phosphorylation, occur which are necessary for the biological function of the protein. While the amino acid sequence makes up the primary structure of the protein, the chemical/biological properties of the protein are very much dependent on the three-dimensional or tertiary structure.

## Secondary Structure

Stretches or strands of proteins or peptides have distinct, characteristic local structural conformations, or secondary structure, dependent on hydrogen bonding. The two main types of secondary structure are the α-helix and the ß-sheet.

The *α-helix* is a right-handed coiled strand. The side-chain substituents of the amino acid groups in an α-helix extend to the outside. Hydrogen bonds form between the oxygen of each C=O bond in the strand and the hydrogen of each N-H group four amino acids below it in the helix. The hydrogen bonds make this structure especially stable. The side-chain substituents of the amino acids fit in beside the N-H groups.

The hydrogen bonding in a *ß-sheet* is between strands (inter-strand) rather than within strands (intra-strand). The sheet conformation consists of pairs of strands lying side-by-side.  The carbonyl oxygens in one strand bonds with the amino hydrogens of the adjacent strand. The two strands can be either parallel or anti-parallel depending on whether the strand directions (N-terminus to C-terminus) are the same or opposite. The anti-parallel ß-sheet is more stable due to the more well-aligned hydrogen bonds.

## Tertiary Structure

The overall three-dimensional shape of a protein molecule is the tertiary structure. The protein molecule will bend and twist in such a way as to achieve maximum stability or lowest energy state. Although the three-dimensional shape of a protein may seem irregular and random, it is fashioned by many stabilizing forces due to bonding interactions between the side-chain groups of the amino acids.

Under physiologic conditions, the hydrophobic side-chains of neutral, non-polar amino acids such as phenylalanine or isoleucine tend to be buried on the interior of the protein molecule, thereby shielding them from the aqueous medium. The alkyl groups of alanine, valine, leucine and isoleucine often form hydrophobic interactions between one another, while aromatic groups such as those of phenylalanine and tyrosine often stack together. Acidic or basic amino acid side-chains will generally be exposed on the surface of the protein as they are hydrophilic.

The formation of disulfide bridges by oxidation of the sulfhydryl groups on cysteine is an important aspect of the stabilization of protein tertiary structure, allowing different parts of the protein chain to be held together covalently. Additionally, hydrogen bonds may form between different side-chain groups. As with *disulfide bridges*, these hydrogen bonds can bring together two parts of a chain that are some distance away in terms of sequence. Salt bridges, ionic inter- actions between positively and negatively charged sites on amino acid side chains, also help to stabilize the tertiary structure of a protein.

## Quaternary Structure

Many proteins are made up of multiple polypeptide chains, often referred to as protein subunits. These subunits may be the same, as in a homodimer, or different, as in a heterodimer. The quaternary structure refers to how these protein subunits interact with each other and arrange themselves to form a larger aggregate protein complex. The final shape of the protein complex is once again stabilized by various interactions, including hydrogen-bonding, disulfide-bridges and salt bridges. The four levels of

protein structure are shown in Figure 2.

LEVELS OF PROTEIN STRUCTURE

Primary Structure

Secondary Structure

β-Sheet

α-Helix

Tertiary Structure

Quaternary Structure

Lubrizol Life Science