

## Chapter 01

# Biomolecules and Catalysis

A **biomolecule** is a carbon-based organic compound that is produced by a living organism. More than 25 naturally occurring chemical elements are found in biomolecules, but these biomolecules consist primarily of carbon, hydrogen, nitrogen, oxygen, phosphorus and sulfur. In terms of the percentage of the total number of atoms, four elements such as hydrogen, oxygen, nitrogen and carbon together make up over 99% of the mass of most cells.

Biomolecules include both small as well as large molecules. The **small biomolecules** are low molecular weight (less than 1000) compound which include sugars, fatty acids, amino acids, nucleotides, vitamins, hormones, neurotransmitters, primary and secondary metabolites. Sugars, fatty acids, amino acids and nucleotides constitute the four major families of small biomolecules in cells. Large biomolecules which have high molecular weight are called **macromolecules** and mostly are polymers of small biomolecules. These macromolecules are proteins, carbohydrates and nucleic acids.

major	Small biomolecules	Macromolecules
	Sugars	Polysaccharides
	Amino acids	Polypeptides (proteins)
	Nucleotides	Nucleic acids
	Fatty acids	

minor: - vitamins, hormones, neurotransmitters, Primary & Secondary metabolites  
Nucleic acids and proteins are informational macromolecules. Proteins are polymers of amino acids and constitute the largest fraction (besides water) of cells. The nucleic acids, DNA and RNA, are polymers of nucleotides. They store, transmit, and translate genetic information. The polysaccharides, polymers of simple sugars, have two major functions. They serve as energy-yielding fuel stores and as extracellular structural elements.

### 1.1 Amino acids and Proteins

Amino acids are compounds containing carbon, hydrogen, oxygen and nitrogen. They serve as monomers (building blocks) of **proteins and are composed of an amino group, a carboxyl group, a hydrogen atom, and a distinctive side chain, all bonded to a carbon atom, the  $\alpha$ -carbon.** In an  $\alpha$ -amino acid, the **amino and carboxylate groups are attached to the same carbon atom, which is called the  $\alpha$ -carbon.** The various  $\alpha$ -amino acids differ with respect to the side chain (R group) attached to their  $\alpha$ -carbon. The general structure of an amino acid is:

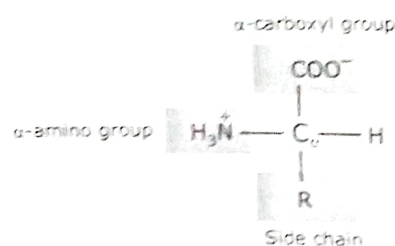
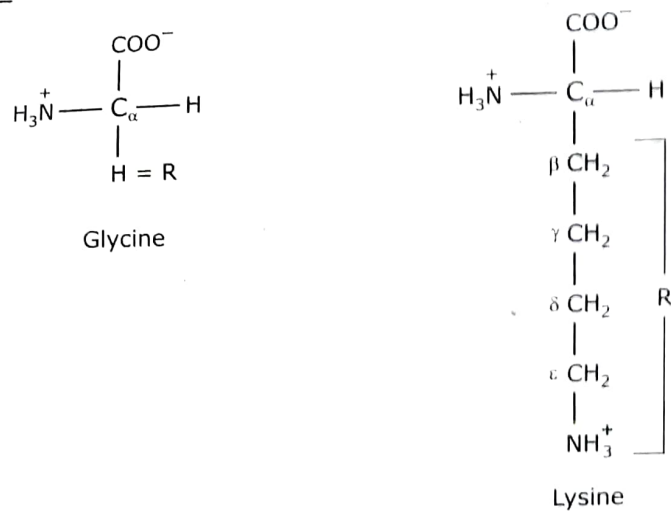


Figure 1.1 General structure of an amino acid.

This structure is common to all except one of the  $\alpha$ -amino acids (*proline* is the exception). The R group or side chain attached to the  $\alpha$ -carbon is different in each amino acid. In the simplest case, the R group is a hydrogen atom and amino acid is glycine.

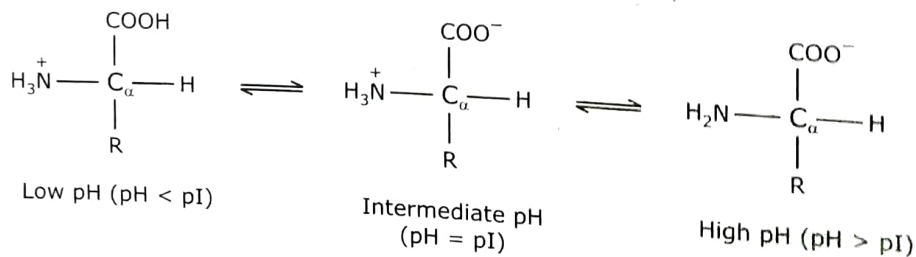


**Figure 1.2** Structure of glycine and lysine.

In  $\alpha$ -amino acids both the amino group and the carboxyl group are attached to the same carbon atom. However, many naturally occurring amino acids not found in protein, have structures that differ from the  $\alpha$ -amino acids. In these compounds the amino group is attached to a carbon atom other than the  $\alpha$ -carbon atom and they are called  $\beta$ ,  $\gamma$ ,  $\delta$ , or  $\epsilon$  amino acids depending upon the location of the C-atom to which amino group is attached.

#### Amino acids can act as acids and bases

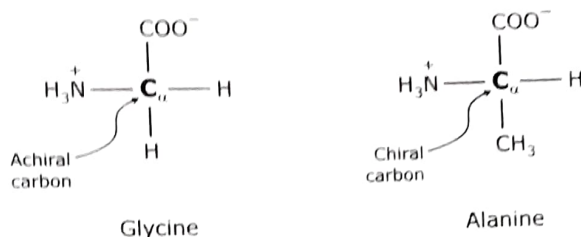
Amino acids contain both an amino ( $-\text{NH}_2$ ) and a carboxyl ( $-\text{COOH}$ ) group. Amino group is basic (*proton acceptor*) and carboxyl group is acidic (*proton donor*). Therefore, amino acids are **amphoteric** in nature. An amphiprotic molecule can either donate or accept a proton, thus acting either as an acid or a base. At high concentrations of protons (low pH), the carboxyl group accepts a proton and becomes uncharged, so that the overall charge on the molecule is positive. Similarly, at low concentrations of protons (high pH), the amino group loses its proton and becomes uncharged; thus the overall charge on the molecule is negative. At specific value of pH called **isoelectric point** (pI), every amino acid exists predominantly as *dipolar ion* or *zwitterion*. A **zwitterion** is a compound with no overall electrical charge, but contains positively and negatively charged groups.



**Figure 1.3** The acid-base behaviour of an amino acid in solution. At low pH, the positively charged species predominates. As the pH increases, the electrically neutral zwitterion becomes predominant. At higher pH, the negatively charged species predominates. *At a pH - ve can*

### 1.1.1 Optical properties

All amino acids except glycine are optically active i.e. they rotate the plane of plane polarized light. Optically active molecules contain *chiral* carbon. A tetrahedral carbon atom with four different constituents are said to be chiral. All amino acids except glycine have chiral carbon and hence they are optically active. An optically active compound can rotate the plane of polarized light either clockwise (to the right) or counterclockwise (to the left). Optically active compounds that rotate the plane of polarized light clockwise are said to be *dextrorotatory*. By convention, this direction is designated by a plus sign (+). Optically active compounds that rotate the plane of polarized light counterclockwise are said to be *levorotatory*. This is designated by a minus sign (-). The + and - forms have also been termed *d-* and *l-*, respectively.



**Figure 1.4** Amino acids showing *achiral* and *chiral* carbon.

Optical activity is measured by *polarimeter*. Optical activity is the ability of an optically active compound to rotate the plane of linearly polarized light. **Optical rotation** is a quantitative measure of the rotation of light caused by the compound. The magnitude of optical rotation indicates the extent to which plane of linearly polarized light is rotated and sign represents the direction of rotation. Optical rotation of an optically active compound depends on the concentration of the compound, temperature, wavelength of light used, solvent used to dissolve the sample and light pathlength. The optical rotation of a solution at a given temperature and wavelength is given by

$$\hat{A} = [\alpha]_D^T \times C \times l$$

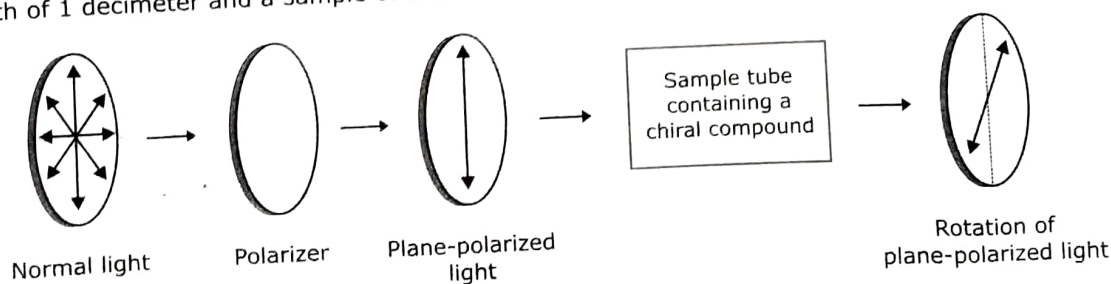
where,  $\hat{A}$  = observed rotation in degrees

$C$  = concentration of the solution in g/ml

$l$  = light path length in decimeters (dm)

$[\alpha]_D^T$  = the specific rotation of compound at temperature,  $T$  (in degrees Celsius) and wavelength,  $\lambda$  (in nm).  
If the wavelength of the light used is 589 nm, the symbol 'D' is used,  $[\alpha]_D^T$ .

**Specific rotation** is the reference value of optical rotation for a given concentration of compound at a given temperature and fixed wavelength. At a given temperature and for a given wavelength of light, the *specific rotation* is defined as the observed value of optical rotation when plane polarized light is passed through a sample with a path length of 1 decimeter and a sample concentration of 1g per milliliter.

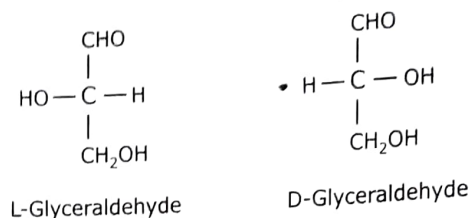


**Figure 1.5** When plane polarized light is passed through a solution that contains an optically active compound, there is net rotation of the plane polarized light. The light is rotated either clockwise (*dextrorotatory*) or counterclockwise (*levorotatory*) by an angle that depends on the molecular structure and concentration of the compound, the path length and the wavelength of the light.

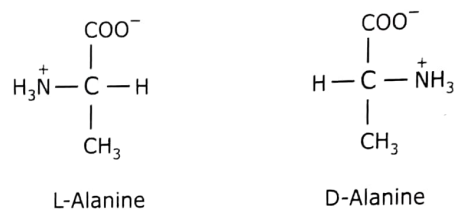


### 1.1.2 Absolute configuration

An amino acid with a chiral carbon can exist in two configurations that are non-superimposable mirror images of each other. These two configurations are called **enantiomers**. An enantiomer is identified by its absolute configuration. For example, glyceraldehyde has two absolute configurations. When the hydroxyl group is on the right, the configuration is D; when the hydroxyl group is on the left, the configuration is L.

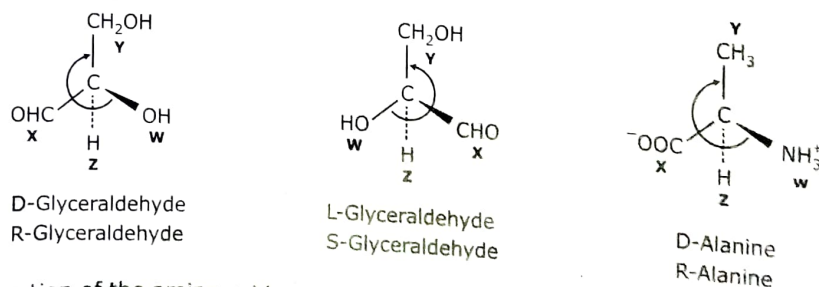


In the above figure, prefixes D- and L- refer to absolute configuration of glyceraldehyde. Similarly, absolute configuration of amino acids are specified by the D- and L- system. The designation of D or L to an amino acid refers to its absolute configuration relative to the structure of D- or L-glyceraldehyde, respectively.



All amino acids except glycine exist in these two different enantiomeric forms. However, all the amino acids which ribosomally incorporated into proteins exhibit L-configuration. Therefore, they are all L- $\alpha$ -amino acids. The basis for preference for L-amino acids is not known. D-form of amino acids are not found in ribosomally synthesized proteins, although they exist in nature. D-form of amino acids are found in some peptide antibiotics and peptidoglycan cell wall of eubacteria.

A second absolute configuration notation using the symbols *R* (from *rectus*, Latin for right) and *S* (from *sinister*, Latin for left) can also be used. In this approach, the substituents on an asymmetric carbon (a chiral carbon with four different substituents) are prioritized by decreasing the atomic number. Atoms of higher atomic number bonded to a chiral centre are ranked above those of lower atomic number. For example, the oxygen atom of a  $-\text{OH}$  group takes precedence over the carbon atom of the  $-\text{CH}_3$  group that is bonded to the same chiral carbon atom. If any of the first substituent atoms are of the same element, the priority of these groups is established from the atomic number of the second, third etc, atoms outward from the chiral carbon atom. Hence, a  $\text{CH}_2\text{OH}$  group takes precedence over a  $\text{CH}_3$  group. The prioritized groups are assigned the letters W, X, Y and Z with the order of priority rating is  $W > X > Y > Z$ . Configuration is assigned by looking down the bond to the lowest priority substituent, Z. If the order of the group  $W \rightarrow X \rightarrow Y$  is clockwise, then the configuration of the chiral centre is designated R. If the order of  $W \rightarrow X \rightarrow Y$  is counterclockwise, the chiral centre is designated S.



The absolute configuration of the amino acids at the  $\alpha$ -carbon is typically described by D-L system rather than the more modern R-S system. According to the R-S system, all the L-amino acids from proteins are S-amino acids, with the exception of L-cysteine, which is R-cysteine.

### 1.1.3 Standard and non-standard amino acids

There are hundreds of different amino acids present in the living organisms; however, only 22 different amino acids participate in protein synthesis which are incorporated ribosomically into proteins. Such amino acids are called *standard* or *proteinogenic amino acids*. Some amino acids are more abundant in proteins than other amino acids. Four amino acids – leucine, serine, lysine and glutamic acid – are the most abundant amino acid residues in a typical protein. Tryptophan and methionine are rare amino acids in a protein. Standard L- $\alpha$ -amino acids are specified by simple three letter codons.  $\alpha$ -amino acids can be classified on the properties of their side chain (or R group), in particular, their polarity, or tendency to interact with water at physiological pH (near pH 7). The polarity of the side chain varies widely, from nonpolar and hydrophobic to highly polar and hydrophilic.

#### 1. *Amino acids with nonpolar side chain* / hydrophobic

Among standard amino acids, nine amino acids contain nonpolar side chain. These are glycine, alanine, valine, leucine, isoleucine, proline, methionine, phenylalanine and tryptophan. *Proline* differs from other members in having its side chain bonded to both the nitrogen and the  $\alpha$ -carbon atoms. *Phenylalanine* and *tryptophan* have aromatic side chains. The side chain of phenylalanine contains a phenyl ring whereas tryptophan has an indole ring.

#### 2. *Amino acids with uncharged polar side chain* / hydrophilic

Six amino acids contain uncharged polar side chain – serine, threonine, cysteine, asparagine, glutamine and tyrosine. Three amino acids, serine, threonine and tyrosine contain *hydroxyl groups* attached to a side chain. Cysteine is structurally similar to serine but contains a *sulfhydryl*, or *thiol group* (-SH) in place of the hydroxyl group.

#### 3. *Amino acids with charged polar side chain*

*Positively charged side chain:* *Lysine* and *arginine* have side chains that contain *positively charged groups* at neutral pH or physiological pH. Lysine has an amino group whereas arginine contains a guanidinium group. *Histidine* contains an imidazole aromatic ring (a planar five member heterocyclic ring). The imidazole ring can be uncharged or positively charged near neutral pH, depending on its local environment.

*Negatively charged side chain:* Amino acids *aspartate* and *glutamate* contain *acidic side chains* that contain *negatively charged carboxyl groups* at physiological pH.

**Table 1.1** Classification of amino acids based on chemical nature of R group

Chemical nature of R group	Example
Aliphatic	Gly, Ala, Val, Leu
Aromatic	Phe, Tyr, Trp
Hydroxyl	Ser, Thr
Carboxylic	Asp, Glu
Sulphur containing	Cys, Met
Imino	Pro
Amino	Lys, Arg
Amide	Asn, Gln

The structures of the 20 standard amino acids are shown in the figure and some of their properties are listed in table 1.2. Properties of selenocysteine (Sec) and pyrrolysine (Pyl), termed as 21<sup>st</sup> and 22<sup>nd</sup> amino acids respectively are mentioned in the next section.