BIOMOLECULES
SEM-5, CC-12
PART-4, PPT-24

Part-4: Peptides-I, Properties

CONTENTS

1. *Peptide linkage and its Geometry*
2. *Hydrolysis of Peptide Linkage*
3. *Nomenclature of Peptides*
4. *The Primary Structure of Peptide: Introduction*

Dr. Kalyan Kumar Mandal
Associate Professor
St. Paul's C. M. College
Kolkata

BIOMOLECULES (PART-4, PPT-24)

Peptides-I, Properties

Peptide Linkage and its Geometry

Proteins are hydrolyzed by acids, alkalis, or enzymes to a mixture of amino acids. It is, therefore, suggested that amino acids in proteins are joined in a linear fashion by peptide linkages, i.e., -CONH- group, the carboxyl group of one amino acid molecule forming an amide by combination with the amino group of the next amino acid molecule, etc. Thus, on this basis, a protein molecule may be represented as a linear polymer of amino acid molecules (Figure 1).

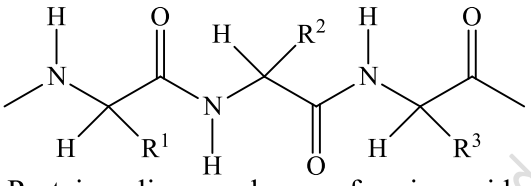

Figure 1: Protein, a linear polymer of amino acid molecules

A peptide bond is, thus, an amide type of covalent chemical bond linking two consecutive α-amino acids from C1 (carbon number 1) of one α-amino acid and N2 (nitrogen on carbon number two) of another, along a peptide or protein chain. The formation of the peptide bond consumes energy, which, in organisms, is derived from ATP. Amino acids are polymerized in living systems by enzymes that form amide linkages from the amino group of one amino acid to the carboxyl group of another.

A peptide bond can be broken by hydrolysis (the addition of water) and the process is known as degradation of peptide bond. In the presence of water, they will break down and release 8-16 kJ/mol (2-4 kcal/mol) of Gibbs energy. This process is extremely slow, with the half-life at 25 °C of between 350 and 600 years per bond. This result indicates the stability of the peptide linkage under normal condition.

In living organisms, the process is normally catalyzed by enzymes known as peptidases or proteases, although there are reports of peptide bond hydrolysis caused by conformational strain as the peptide/protein folds into the native structure.

The IR spectra of amides have been extensively studied and many assignments associated with the -CONH- group have been made. Thus, polypeptides and proteins show bands near 3300 and 3100 cm^{-1}, which are characteristic of the hydrogen bonded N-H group (*stretching*) in secondary amides (RCONHR). They also exhibit absorption bands near 1650 and 1550 cm^{-1} which are characteristic of the C=O (*stretching*) in secondary amides.

UV spectra studies have shown that the peptide bond absorbs in the region 180-220 nm which are characteristic of $n \rightarrow \sigma^*$ transition. This makes it particularly susceptible to UV radiation.

X-ray studies on a number of crystalline peptides provide information on the various bond lengths (in Å) and bond angles in these compounds (Figure 2).

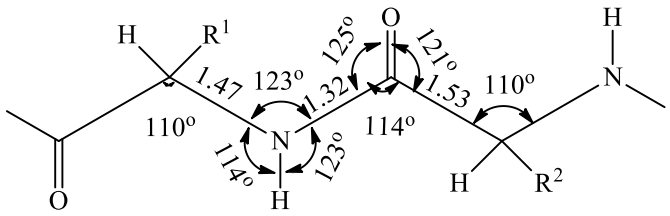

Figure 2: Various bond angles and bond lengths in peptide chain

The conclusion reached from these results were:

1. The atoms in the group, -CO-NH-, are planar and the O and H are *trans*.
2. Since the peptide C-N bond length (i.e., C-N of -CONH-), 1.32 Å is shorter than the usual C-N bond length (~ 1.47 Å), this bond has some double bond character.
3. The partial double bond character of C-N bond in -(O)C-N(H)- as being due to resonance, and hence hindered rotation about this C-N bond permits the possibility of geometrical isomerism, the *trans*-isomer being the more likely because of the much larger steric repulsion operating in the *cis*-isomer (Figure 3).

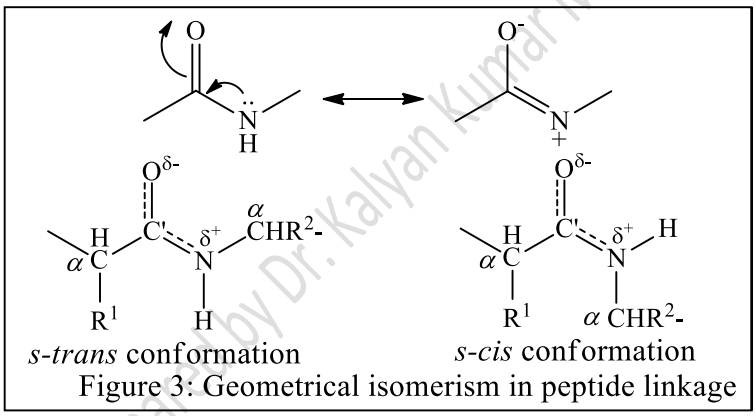

Figure 3: Geometrical isomerism in peptide linkage

On the other hand, rotation can occur about the R^1CH-CO and the R^2CH-NH bond. It is, therefore, possible to describe the conformation of the protein molecule in terms of rotation about the C^{α}-C´ bond and about the N-C^{α} bond. Significant delocalization of the lone pair of electrons on the nitrogen atom with the C=O group gives the -CO-NH- group a partial double bond character, and reduces the C-N bond length than the normal value. The partial double bond renders the amide group planar, occurring in either the *cis* or *trans* isomers.

In the unfolded state of proteins, the peptide groups are free to isomerize and adopt both isomers. However, in the folded state, only a single isomer is adopted at each position. The *trans* form is preferred overwhelmingly in most peptide bonds (roughly 1000:1 ratio in *trans*: *cis* populations).

Hydrolysis of Peptide Linkage

Due to its resonance stabilization, the peptide bond is relatively unreactive under physiological conditions, even less than similar compounds such as esters. Nevertheless, peptide bonds can undergo chemical reactions, usually through an attack of an electronegative atom on the carbonyl carbon, breaking the carbonyl double bond and forming a tetrahedral intermediate. This is the pathway followed in proteolysis.

Proteolysis is the breakdown of proteins into smaller polypeptides or amino acids. Uncatalyzed, the hydrolysis of peptide bonds (Figure 4) is extremely slow, taking hundreds of years. Proteolysis is typically catalyzed by cellular enzymes called proteases.

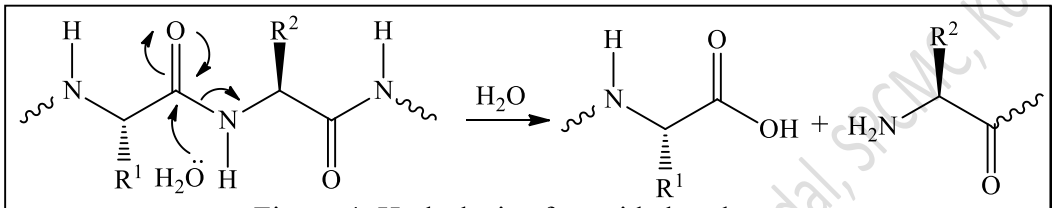

Figure 4: Hydrolysis of peptide bond

Type of Peptides

A molecule formed by joining amino acids together is called a peptide, and the amide linkages (-CO-NH-) in them are called peptide bonds or peptide linkages. Each amino acid in the peptide is called an amino acid residue. '*n*' Number of amino acid residues generate '*n*-1' no of -CO-NH- linkages when combined in a linear fashion (Figure 5). Peptides that contain 2, 3, a few (3-10), or many amino acids are called dipeptides, tripeptides, oligopeptides, and polypeptides, respectively. Proteins are polypeptides consisting of one or more polypeptide chains.

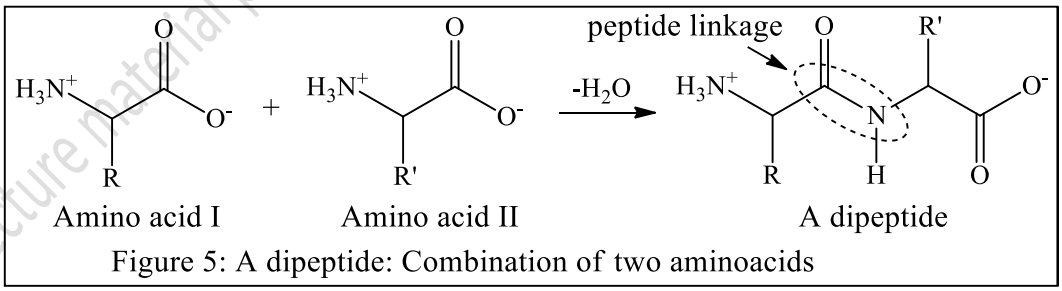

Figure 5: A dipeptide: Combination of two aminoacids

Polypeptides are linear polymers. One end of a polypeptide chain terminates in an amino acid residue that has a free $-NH_3^+$ group and the other terminates in an amino acid residue with a free $-CO_2^-$ group. These two groups are called the *N*-terminal and the *C*-terminal residues, respectively and these are shown in Figure 6.

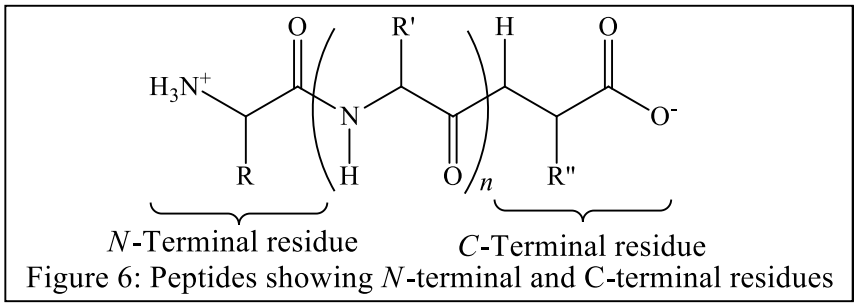

Figure 6: Peptides showing *N*-terminal and C-terminal residues

Conventionally peptide and protein structures are written with the *N*-terminal amino acid residue on the left and the *C*-terminal residue on the right. Two dipeptides are shown in Figure 7 for examples.

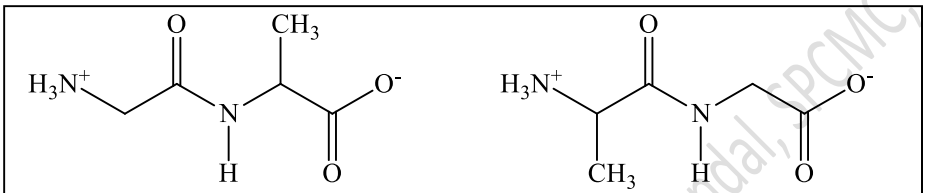

Figure 7: Dipeptides of glycine and alanine

The tripeptide glycylalanylphenylalanine has the following structural formula (Figure 8):

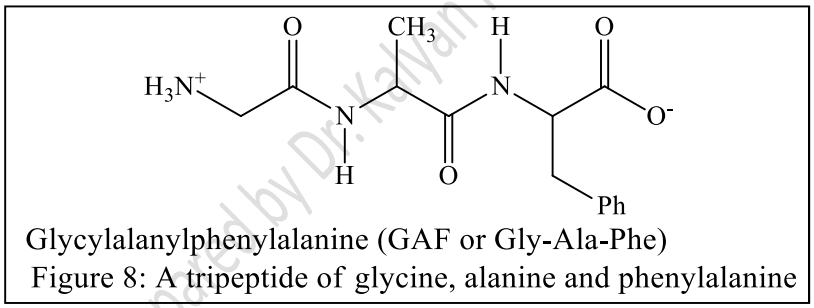

Figure 8: A tripeptide of glycine, alanine and phenylalanine

Nomenclature of Peptides

The terminology and nomenclature associated with peptides are best illustrated by an example. Let us consider the following peptide formed from the three amino acids alanine, valine, and tyrosine (Figure 9).

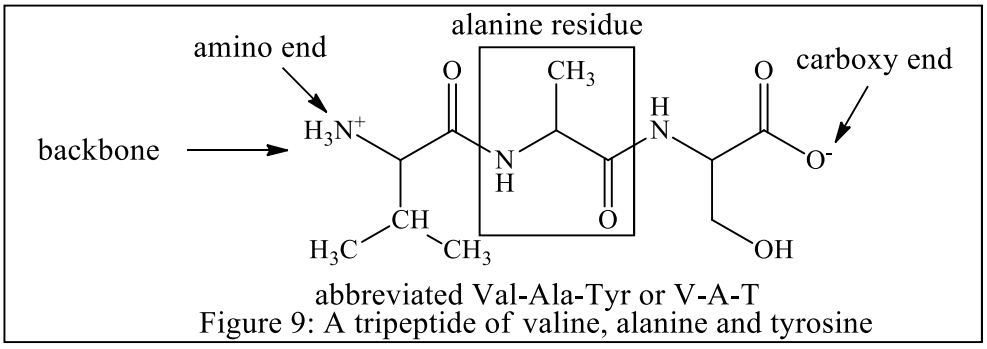

Figure 9: A tripeptide of valine, alanine and tyrosine

The peptide backbone is the repeating sequence of nitrogen, α-carbon, and carbonyl groups obtained by combining of amino acids shown in green colour in the structure in Figure 9. The characteristic amino acid side chains are attached to the peptide backbone at the respective α-carbon atoms. Each amino acid unit in the peptide is called a residue. For example, the part of the peptide derived from *alanine*, the *alanine residue*, is outlined in red. The ends of a peptide are labeled as the *amino end* or *amino terminus* and the *carboxy end* or *carboxy terminus*.

A peptide can be characterized by the number of residues it contains. For example, the preceding peptide is a tripeptide because it contains three amino acid residues. A peptide containing two, three, or five amino acids would be called a dipeptide, tripeptide, or pentapeptide, respectively. A relatively short peptide of unspecified length containing a few amino acids is sometimes referred to as an oligopeptide (from a Greek root meaning "scant" or "few").

A peptide is conventionally named by giving successively the names of the amino acid residues, *starting at the amino end*. The names of all but the carboxy-terminal residue are formed by dropping the final ending (*ine, ic,* or *an*) and replacing it with *yl*. Thus, the foregoing peptide is named valylalanyltyrosine. The peptides then named as acylated derivatives of the terminal amino acid residue on the right-hand side.

In practice, this type of nomenclature is cumbersome for all but the smallest peptides. A simpler way of naming peptides is to connect with hyphens the three-letter (or one-letter) abbreviations of the component amino acid residues beginning with the amino-terminal residue. Thus, the preceding peptide is also written as Val-Ala-Tyr or V-A-T.

It becomes a significant task to write a full structural formula for a polypeptide chain that contains any more than a few amino acid residues. In this situation, use of the one-letter abbreviations is the norm for showing the sequence of amino acids. Very short peptide sequences are sometimes still represented with the three-letter abbreviations. Both type of the abbreviations is shown in Tables 1-3, Part-1, PPT-21.

The Primary Structure of Peptide: Introduction

The primary structure of a peptide (or protein) is the sequence of the amino acid residues in the molecule. First, let us consider a dipeptide composed of two different amino acids, H_2N-A-CO_2H and H_2N-B-CO_2H. These may be combined in two different ways:

1. H_2N-A-CONH-B-CO_2H (I)
2. HO_2C-A-NHCO-B-NH_2 (II) or H_2N-B-CONH-A-CO_2H (IIa)

Inspection of either (I) or (II) shows that the two ends of each molecule are different. The 'amino-end' is said to be *N*-terminal and the 'carboxyl-end' is said to be *C*-terminal.

The general method of writing the sequence of amino acids in a peptide (polypeptide or protein) is with the terminal amino-group on the left. Therefore, the dipeptide (I) is in

accordance with this convention, but not the dipeptide (II), which should therefore be written as (IIa). The peptides are, therefore, named as the acylated derivatives of the terminal amino acid residue on the right-hand side. When the sequence of the amino acids is not known, the symbols are enclosed in brackets and are separated by commas. When the sequence is known, the units are separated by dots or dashes, or by arrows which indicate the direction of linkage from carboxyl to amino.

Since the conventional way of writing peptide formulae has the terminal amino group on the left, shown as (I) and (IIa), the arrow will point from left-to-right. A terminal amino group may be indicated by H and a terminal carboxyl group by OH. Finally, carbonamide groups may be indicated by the addition of NH_2 to the symbol, e.g., asparagine and glutamine (these have proposed alternatives): $AspNH_2$ (Asn) and $GluNH_2$ (Gln). Also, a propped alternative for tryptophan is Trp. The following formulae illustrate the conventions:

(Ala, Gly, Tyr) Thr. (Val. Arg.) Ala. Glu. Phe or

H.Ala, Glu, Phe.OH or Ala → Glu → Phe or Ala - Glu - Phe

In connection to the problem of amino acid sequence, it can be shown that, in general, for *n* different amino acids, *n*! different combinations are possible. Furthermore, had not the common amino acids been all L-acids, the total number of possible combinations would have been very much larger. As a simple example let us consider a tripeptide. The first thing to do in this case is to determine the nature of the amino acid residues. This is usually carried out by acid hydrolysis and chromatography. Suppose the amino acids are shown to be A, B and C. These can be written in six (3!) different combinations. These are:

(i) A. B. C; (ii) A. C. B; (iii) B. A. C;

(iv) B. C. A; (v) C. A. B and (vi) C. B. A

The problem then is to ascertain which is the actual combination of the tripeptide under investigation. Inspection of the formulae shows that if determination of *N*-terminal amino acid (*N*-T-AA) residue of the tripeptide is possible, then six possibilities can be reduced into three pairs. Again, in this situation it is possible to determine the C-terminal amino acid (*C*-T-AA) residue of the tripeptide. Thus:

N-terminal: (*a*) A. B. C and A. C. B

(*b*) B. A. C and B. C. A

(*c*) C. A. B and C. B. A

C-terminal: (*d*) B. C. A and C. B. A

(*e*) A. C. B and C. A. B

(*f*) A. B. C and B. A. C

Since this results in different pairing, the determination of both *N*- and *C*-terminal groups will give the amino acid sequence of the tripeptide, e.g., if the *N*-T-AA determination showed that the tripeptide was in group (*b*) and the *C*-T-AA determination showed that the tripeptide was

in group (d), the tripeptide is, therefore, B. C. A. Now let us assume that the *N*-and *C*-T-AA methods are such that their application results in the removal of the respective terminal amino acid. In these circumstances it would leave with different fragments according to the order of application (of the methods), e.g., for B. C. A:

(i) *N*-T-AA first: residual fragment C. A.
(ii) *C*-T-AA first: residual fragment B. C.

By repeating either of these determinations, the amino acid sequence is solved. Thus, the sequence may be determined by use of one method twice or by use of each method once. Now let consider the tetrapeptide whose acids have been shown to be (A, B, C, D). There are 24 (4!) possible combinations as the following.

A. B. C. D	B. A. C. D	C. A. B. D	D. A. B. C
A. B. D. C	B. A. D. C	C. A. D. B	D. A. C. B
A. C. B. D	B. C. A. D	C. B. A. D	D. B. A. C
A. C. D. B	B. C. D. A	C. B. D. A	D. B. C. A
A. D. B. C	B. D. A. C	C. D. A. B	D. C. A. B
A. D. C. B	B. D. C. A	C. D. B. A	D. C. B. A

Suppose the *N*- and *C*-T-AA determinations showed respectively B and D. The tetrapeptide is therefore B. A. C. D or B. C. A. D. As for the tripeptide, the fragments obtained will depend on the order of application:

1. *N*-T-AA first: residual fragments A. C. D or C. A. D.
2. *C*-T-AA first: residual fragments B. A. C or B. C. A.
3. Both *N*- and *C*-T-AA (irrespective of order): residual fragments A. C. or C. A.
4. Hence, the amino acid sequence of the tetrapeptide may be determined by using either the *N*-T-AA or the *C*-T-AA method three times, or by a combination of these methods (also three operations).
5. The general method of amino acid sequence analysis, however, does not use both end-group analyses on the original peptide. Only one end-group is determined and this is then followed by fragmentation of the peptide in at least two different ways.
6. The smaller peptides are then subjected to amino acid sequence determination by end-group analysis. In this way, the various small peptides ‘overlap’ and so it becomes possible to deduce the complete sequence of the amino acids in the original peptide.