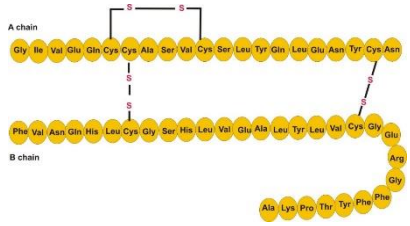


UNIT 6

SEQUENCING AND ANALYSIS OF PROTEIN

Structure

6.1 INTRODUCTION

You are aware and had acquire the knowledge about the amino acids, peptides and proteins in the previous semester. Now, you are going to learn methods of protein sequencing in this unit. The general strategy for determining the amino acid sequence of a protein depends on whether: the protein contains more than one polypeptide chain if yes, then there is a need to separate and purify the chains. It is important to know disulfide bonds in the polypeptide chain must be cleaved. However, it is significant to check the composition of amino acids is determined in each polypeptide chain.

We have studied in the previous units about N-and C-terminals of proteins, hence while performing protein sequencing it is essential to know and identify the amino acids at these terminals. Proteins being macromolecules. We need to generate smaller fragments of peptides, so that the analysis can be performed easily. At the end of the unit, we'll study about protein recombination and overlapping of fragments.

Learners are advised to recall its basic concepts of protein structure and arrangement before proceeding further.

Expected Learning Outcomes

After studying this unit, you should be able to:

- ❖ list the methods of protein sequencing;
- ❖ explain the significance of protein sequencing;
- ❖ describe the significance of disulfide bonds, fragmentation of proteins and overlapping peptides; and
- ❖ illustrate the protein sequencing.

6.2 INTRODUCTION: PROTEIN SEQUENCING

Protein sequencing is a combination of methods or approaches or techniques applied to establish the amino acid sequence of a protein. The amino acid sequence (also called primary structure) of a protein is the order of the amino acids in the protein linear chain.

History of protein sequencing

Sanger's first triumph was to determine the complete amino acid sequence of the two polypeptide chains of bovine insulin, A and B, in 1951, . Prior to this it was widely assumed that proteins were somewhat amorphous. In determining these sequences, Sanger proved that proteins have a defined chemical composition. In 1958, he was awarded a Nobel Prize in Chemistry "for his work on the structure of proteins, especially that of insulin".

Sanger was given the challenge of determining amino acid sequence of insulin, which had never been done before. Using chemistry and chromatography, and by mixing standard techniques with novel ones, he developed a method to read the amino acid sequence of insulin and found that this protein is actually made up of two polypeptide chains linked together by disulfide bonds (Fig. 6.1).

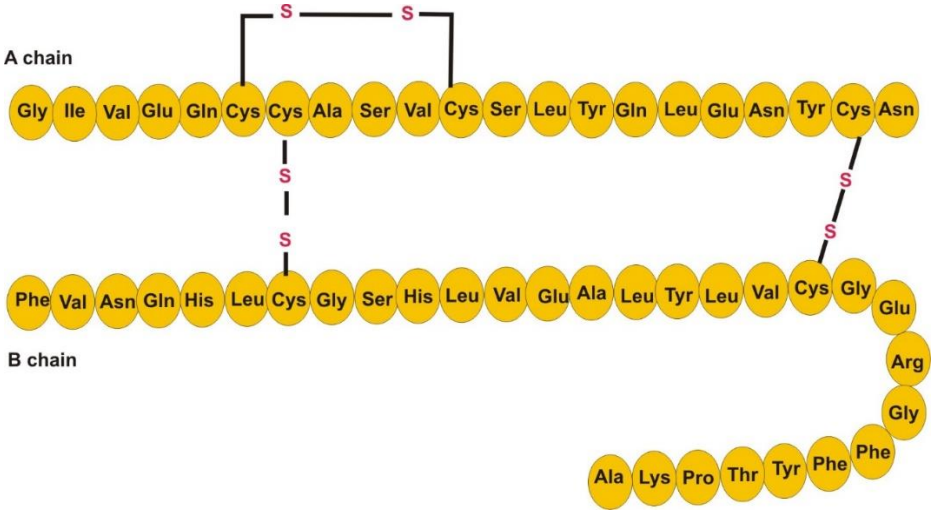


Fig. 6.1: Bovine serum insulin is a protein hormone made of two peptide chains, A (21 amino acids long) and B (30 amino acids long). In each chain, primary structure is indicated by three-letter abbreviations that represent the names of the amino acids in the order they are present (Refer Unit-3 of BBCCT-101). The amino acid cysteine (cys) has a sulfhydryl (SH) group as a side chain. Two sulfhydryl groups can react in the presence of oxygen to form a disulfide (S-S) bond. There are a total of three disulfide bonds, the first two disulfide bonds connect the A and B chains together, and third bond helps the A chain fold into the correct shape.

6.2.1 Protein Sequencing: Sanger Method

Sanger first needed to characterize the free amino groups in insulin. For this he developed a reagent, dinitrofluorobenzene (FDNB or DNFB, also called as Sanger's reagent), which reacted with amino groups present in proteins to form an acid-stable dinitrophrenyl (DNP) derivative (Fig. 6.2).

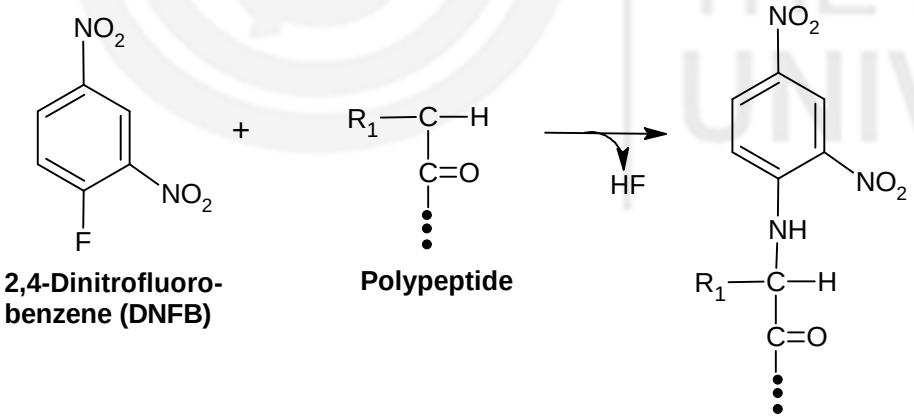


Fig. 6.2: Reaction of 2, 4-dinitrofluorobenzene (DNFB) with polypeptides and the resultant product of DNP polypeptide.

The DNP protein was treated with acid to break the polypeptide backbone, and the free DNP amino acid derivatives were isolated and compared to standards prepared from known amino acids. In this way, Sanger determined that insulin was made up of two peptide chains: one (chain A) with an amino-terminal glycine residue and another (chain B) with an amino-terminal

phenylalanine. Subsequent work revealed that chain A was composed of twenty amino acids and chain B thirty-one. The individual chains were then broken down into smaller components: Acid was used to cleave the polypeptide backbone, ***Performic acid*** was used to break the cysteine disulfide bonds, and proteolytic enzymes were used to hydrolyze the polypeptide at specific sites on the chain. The reaction products were separated from each other and their sequence determined.

Today, there are mode was developed by several refinement and improvements have been in Sanger's method. In this regard, sequencing of most proteins can be performed within a few hours or days using only a small amount or microgram of protein. An overview of protein sequencing is represented by **Fig. 6.3.**

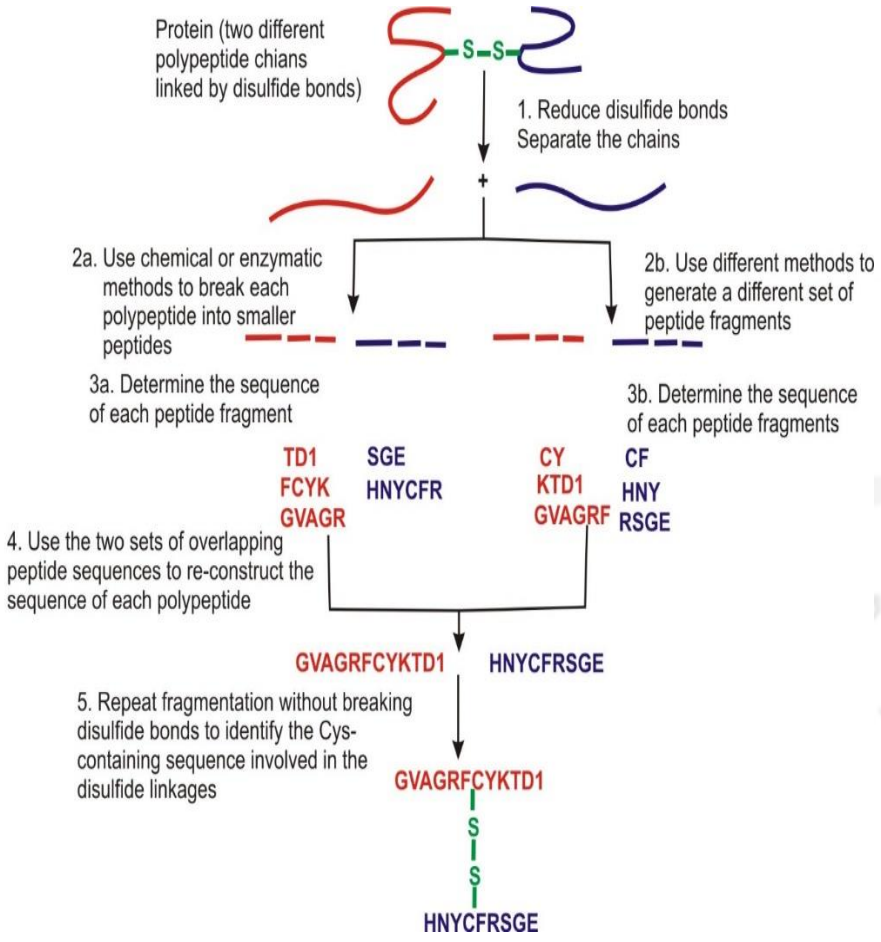


Fig. 6.3: Overview of protein sequencing by Sanger's method.

6.2.2 Sequencing of N-Terminal and C-Terminal Amino Acids

Sequence of amino acids is uncertain because there is an uncertainty to read it left-to-right or right-to-left. The sequence is always read from the N-terminus to the C-terminus of the protein (Fig. 6.4). So, there is a need to know what are the N-terminal and C-terminal amino acids and hence it is considered as amino acid sequence analysis.

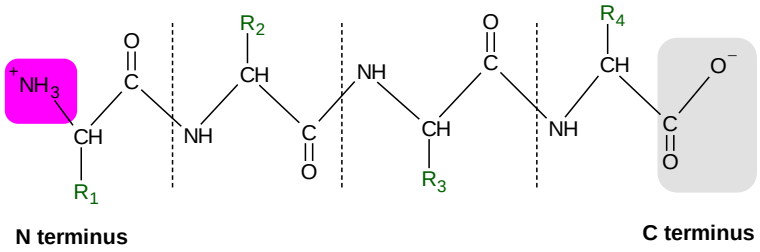


Fig. 6.4: Chain of amino acids in a protein representing the N- terminus and C-terminus.

The N-terminus (also known as the amino-terminus, NH_2-terminus, N-terminal end or amine-terminus) is the start of a protein or polypeptide referring to the free amine group (-NH_2) located at the end of a polypeptide. The C-terminus (also known as the carboxyl-terminus, carboxy-terminus, C-terminal tail, C-terminal end, or COOH-terminus) is the end of an amino acid chain (protein or polypeptide), possessing a free carboxyl group (-COOH). Amino acids sequencing beginning from the N-terminal side of the protein structure is known as N-terminal analysis. Similarly, the sequencing from the C-terminal side of the protein structure is known as C-terminal analysis.

Analysis of N-terminal Amino Acids

The amino acid residue at the N-terminal can be identified by a number of ways. The three widely used N-terminal methods are:

(a) 2**,4-dinitrofluorobenzene (DNFB) method**-This was developed by Sanger in 1945.

(b) **1-dimethylamino naphthalene sulfonyl chloride (Dansyl Chloride; DNS-CL) method**-This method was invented by Gray and Hartley in 1972.

(c) **Phenylisothiocynate (PITC) method or Edman degradation method-** This method was developed by Edman in 1975.

These methods have been in practice as the most important tools for the determination of protein sequence. DNFB and DNS-CL are mostly used only for N-terminal analysis in a qualitative or semi-quantitative manner. These two techniques are required total hydrolysis to release the derivatized amino acids from the N-terminal. Edman degradation method is used for both N-terminal analysis and extended sequence analysis. This method is distinct from both DNFB and DNS-CL, in which release of PITC derivatized N-terminus requires acid hydrolysis instead of total hydrolysis. This reaction can be carried out either in an anhydrous acid solution or in an aqueous acid solution.

N- terminal analysis (Edman degradation)

Determination of the overall amino acid content and the identity of the amino-(N-) terminal residue for each peptide allowed deduction of the sequence of the whole molecule. An alternative approach was described by Edman. This allowed determination of extended sequences of peptides or whole proteins, and has been used widely up to the present day. The method employs a

series of chemical reactions to remove and identify the amino acid residue that is at the N-terminus of the polypeptide chain, i.e. the residue with a free a-amino group. At the same time, the next residue in the sequence is made available and subjected to the same round of chemical reactions. Reappearance of this process reveals the sequence of the polypeptide.

Peptide sequencing by Edman chemistry may be divided into following steps as illustrated in Fig. 6.5.

Description of all the three steps is given below:

STEP-1. Coupling: Phenyl isothiocyanate (PITC) reacts with an amino group (or in the case of prolyl residue with an imino group) at the N-terminal end of the polypeptide chain, to form a phenylthiocarbamyl derivative of the terminal residue. This is the basic condition are required for this reaction.

STEP-2. Cleavage: In the presence of strong acid, cleavage occurs at the first peptide bond, giving the peptide (minus the first residue) and the liberated first residue as the anilinothiazolinone (ATZ) form. Once other reactants and products have been washed away, the shortened polypeptide can be taken through another round of coupling and cleavage to release the second residue, and so on in a cyclical fashion. Currently, trifluoroacetic acid (TFA) is used for this cleavage reaction

STEP-3. Conversion: The ATZ residue is separated from the peptide by extraction in organic solvent (ethyl acetate or chlorobutane), and is then converted to a more stable form to allow better analysis. Conversion to the more stable **phenylthiohydantoin (PTH)** form is done in aqueous acid (25% TFA, v/v in water).

Analysis of PTH residues: The PTH residue generated by each cycle of Edman chemistry is typically identified by chromatography, originally thin-layer chromatography and latterly reversed-phase high-performance liquid chromatography. The PTH amino acid residue derived from each cycle in turn is identified and quantified by comparison with standards, and the sequence is described by the order of residues from the N- to the C-terminus.

Analysis of C-terminal amino acids

Fewer methods of analysis are available for the identification of C-terminal residue of polypeptides. Most widely used methods are described below:

(a) Hydrazinolysis: Proteins are treated with anhydrous hydrazine ($H_2N\text{-}NH_2$) at 100° C. So, all the amino acids of polypeptide chain is reacted with hydrazine and converted into amino acid hydrazides. But, the amino acid attached to carboxyl terminal is not reacted and remain as the free amino acid. This free amino acid can be isolated and identified by chromatographic analysis (refer block I of this course).

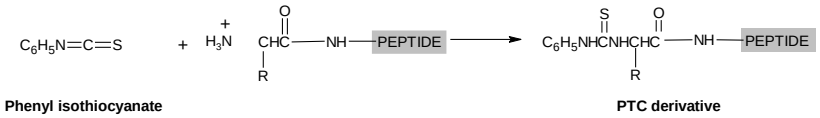
(b) $LiAlH_4$ (Lithium aluminium hydride) based reduction of C-terminus: Proteins are analysed by this method through the process of reduction. Reduction of free α-amino group at the end of a peptide chain into an alcohol is performed by $LiAlH_4$. After completing this reduction process, subsequent acid hydrolysis of protein is carried out and free-amino acids with a single free

alcohol are obtained. This single free alcohol represented the C-terminal residue.

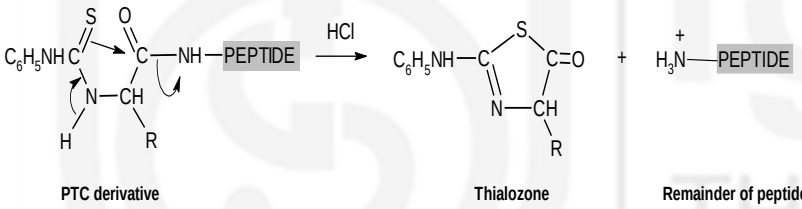
(c) Enzymatic analysis or carboxypeptidase-based analysis: Most recent and popular method of analysis of C-terminus of protein is the enzymatic analysis or carboxypeptidase-based analysis. The enzyme carboxypeptidase cut the amino acid residues from the C-terminus of a protein in a sequential way. Free amino acids from C-terminus of a protein can be isolated and identified by chromatographic analysis.

The Edman Degradation

Step 1: A peptide is treated with phenyl isothiocyanate to give a phenylthiocaramoyl (PTC) derivative.



Step 2: On reaction with hydrogen chloride in an anhydrous solvent, the thiocarbonyl sulfur of the PTC derivative attacks the carbonyl carbon of the N-terminal amino acid. The N-terminal amino acid is cleaved as a thiazolone derivative from the remainder of the peptide.



Step 3: Once formed, the thialozone derivative isomerizes to a more stable phenylthiohydantoin (PTH) derivative, which is isolated and characterized, thereby providing identification of the N-terminal amino acid. The remainder of the peptide (formed in step 2) can be isolated and subjected to a second Edman degradation.

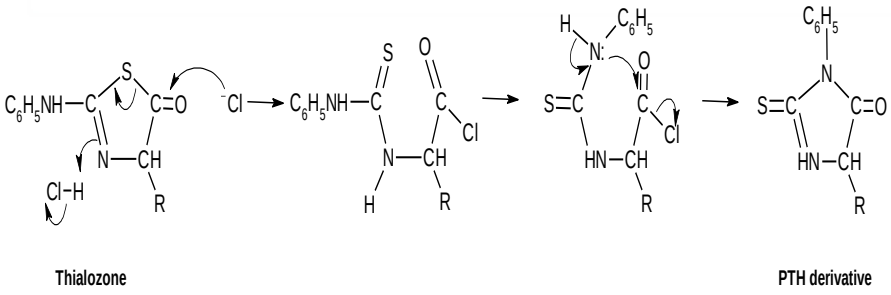


Fig. 6.5: Illustration of the all the steps of Edman degradation.

C- terminal analysis (Carboxypeptidase based method)

Carboxypeptidases are enzymes that cleave amino acid residues from the C-termini of polypeptides in a successive fashion. Four carboxypeptidases (A, B, C, and Y) are in general use. Carboxypeptidase A (from bovine pancreas) works well in hydrolyzing the C-terminal peptide bond of all residues except proline, arginine, and lysine. Carboxypeptidase B is effective only when Arg or Lys is situated on the C-terminal residues. Thus, a mixture of carboxypeptidases A and B liberates any C-terminal amino acid except proline. Carboxypeptidase C from citrus leaves and carboxypeptidase Y from yeast act on any C-terminal residue (Table-6.1).

Table 6.1: Source and specificity of carboxypeptidases

Enzyme	Source	Specificity
Carboxypeptidase A	Bovine pancreas	$R_n \neq$ Arg, Lys, Pro; $R_{n-1} \neq$ Pro
Carboxypeptidase B	Bovine pancreas	$R_n \neq$ Arg, Lys, $R_{n-1} \neq$ Pro
Carboxypeptidase C	Citrus leaves	All free C-terminal residues; pH optimum=3.5
Carboxypeptidase Y	Yeast	All free C-terminal residues, but slowly with R_n=Gly

Enzymatic C-terminal amino acid cleavage by one of several carboxypeptidase enzymes is a fast and convenient method of analysis. Because the shortened peptide product is also subject to enzymatic cleavage, care must be taken to control the conditions of reaction so that the products of successive cleavages are properly monitored. The following example illustrates this feature. A peptide having a C-terminal sequence: ~Gly-Ser-Leu is subjected to carboxypeptidase cleavage, and the free amino acids cleaved in this reaction are analyzed at increasing time intervals. The leucine is cleaved first, the serine second, and the glycine third, as demonstrated by the sequential analysis.

Significance

Protein sequencing (the whole N-terminal and C-terminal analysis) gives the most valuable information . Examples of Such information given below:

- Information of the sequence of a protein is very essential in explaining its mechanism of action.
- The three-dimensional structure of proteins is explained by the information of amino acid sequence.
- Amino acid sequence provides a relation between the genetic message in RNA and three-dimensional structure that performs a proteins biological function.
- Amino acid sequence determination is an essential part of molecular pathology. Alterations in amino acid sequence can produce abnormal function of protein and resulting in disease development, such as sickle cell anemia & cystic fibrosis.
- Evolutionary history is also determined by protein sequencing. Protein resembles with another amino acid sequence only if they have

a common ancestor. Therefore, molecular events in evolution can be traced from amino acid sequences.

SAQ 1

1. What is protein sequencing?

2. Who is the first scientist to invent the protein sequencing?

3. What is the meaning of N-terminal and C-terminal analysis?

4. Indicate whether the following statements are true or false:

i) Proteins are treated with anhydrous hydrazine (H_2N-NH_2) at 100° C. So, all the amino acids of polypeptide chain are reacted with hydrazine and converted into amino acid hydrazides. ( )

ii) Reduction of free α-amino group at the end of a peptide chain into an alcohol is performed by LiAlH4. ( )

iii) Most recent and popular method of analysis of C-terminus of protein is the enzymatic analysis or carboxypeptidase-based analysis. ( )

iv) Carboxypeptidase A (from bovine pancreas) works well in hydrolyzing the C-terminal peptide bond of all residues except proline, arginine, and lysine. ( )

6.3 GENERATION OF FRAGMENTED PROTEIN OR POLYPEPTIDE CHAIN

Up to now we have studied about Sangers, Edman and Carboxy peptidase based protein sequencing. In this section of unit, you'll to studying about various methods used for generation of fragmented proteins.

6.3.1 Introduction

During the analysis of proteins, there are numerous circumstances come which strength to cleave a protein into peptide fragments. Proteins or polypeptides, which contain more than 40 to 100 amino acids, cannot be sequenced in straight forward manner. There is a need of cleavage of proteins or polypeptides, such cleavage generates the fragments of protein or polypeptides, which contain a smaller number of amino acids. The number of amino acids is always smaller in these fragments as compared to the protein or polypeptide, from which these fragments are generated. The whole process is known as fragmentation of protein or generation of the fragment of proteins or polypeptides.

Fragment of proteins or polypeptides can be produced by enzymatic cleavage or chemical of the native protein. Enzymatic methods have a tendency to cleave adjacent to the more common amino acids. In this regard, these methods produce 50 or more peptides from a protein. Chemical methods are

likely to produce large fragments. These methods perform the cleavage at the less common amino acids and produce the two or three large peptides.

6.3.2 Enzyme Based Generation of Proteins or Polypeptide Fragments

Several specific enzymes are used for the generation of the fragment of proteins or polypeptides. These enzymes are proteases, which are further divided into endopeptidases and exopeptidases. Endopeptidases are those enzymes which catalyze the hydrolysis of internal peptide bond. Exopeptidases are those enzymes which catalyse the hydrolysis of N or C-terminal amino acids residues (towards either ends).

Exopeptidases and endopeptidases are very specific for their action. They act on the bond that is to be cleaved and flanking with specific amino acid residue. This bond is known as scissile peptide bond (Fig. 6.6).

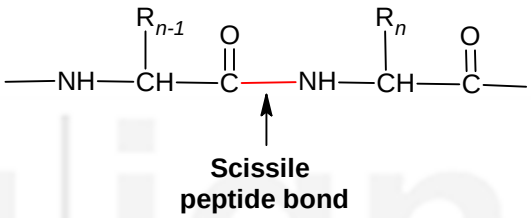


Fig. 6.6: The scissile peptide bond of a protein.

Here, the meaning of specificity is in the context of amino acid residues which are recognized by proteases (endopeptidases and exopeptidases) and their presence is necessary in protein or polypeptides. Without the presence of these amino acid residues in the protein or polypeptides, no action is performed by the endopeptidases and exopeptidases.

Endopeptidases

A number of endopeptidases used for the generation of the fragments of proteins or polypeptides. There are trypsin, chymotrypsin, elastase, pepsin, thermolysine and endopeptidase. Refer Table 6.2 to know more about these enzymes.

Table 6.2: Source and specificity of endopeptidases

Enzyme	Source	Specificity	Comments
Trypsin	Bovine pancreas	R_{n-1}=positively charged residues: Arg, Lys; $R_n \neq$ Pro	Highly specific
Chymotrypsin	Bovine pancreas	R_{n-1}= bulky hydrophobic residues: Phe, Trp, Tyr; $R_n \neq$ Pro	Cleaves more slowly for R_{n-1}=Asn, His, Met, Leu
Elastase	Bovine pancreas	R_{n-1}= small neutral residues:	

		Ala, Gly, Ser, Val; $R_n \neq$ Pro	
Thermolysin	*Bacilllus thermoproteolyticus*	R_n=IIe, Met, Phe, Trp, Tyr, Val; $R_{n-1} \neq$ Pro	Occasionally cleaves at R_n=Ala, Asp, His, Thr; heat stable
Pepsin	Bovine gastric mucosa	R_n=Leu, Phe, Trp, Tyr; $R_{n-1} \neq$ Pro	Also, others; quite nonspecific pH optum=2
Endopeptidase V8	*Staphylococcus aureus*	R_{n-1}= Glu	

Exopeptidases

Exopeptidases are also used for the generation of the fragment of proteins or polypeptides. Leucine amino peptidase cleaves the N-terminal amino acid leucine and does not cleave the proline situated at N-terminal. ***Amino peptidase*** *M* performs the cleavage of all N-terminal residues. ***Carboxypeptidase-A*** performs the cleavage of all N-terminal residues except arginine, lysine and proline. ***Carboxypeptidase-B*** cleaves C-terminal arginine and lysine in the condition of the absence of proline as a next residue. ***Carboxypeptidase-C*** cleaves C-terminal residues.

6.3.3 Chemical Compounds or Reagents-Based Generation of Proteins or Polypeptide Fragments

Several highly specific chemical methods of proteolysis are available, the most widely used being **cyanogen bromide** (CNBr) cleavage. CNBr acts upon methionine residues. A number of other chemical methods give specific fragmentation of polypeptides, including cleavage at asparagine–glycine bonds by hydroxylamine (NH_2OH) at pH 9 and selective hydrolysis at aspartyl–prolyl bonds under mildly acidic conditions. Table 6.3 summarizes the various procedures described here for polypeptide cleavage.

Table 6.3: Chemical compounds and their respective cleavage sites in proteins

Chemical compounds	**Cleavage sites**
BNPS skatole	C-terminal side of tryptophan residues
N- Bromosuccinimide	
o-Iodosobenzoate	
Cyanogen bromide	C-terminal side of methionine residues
Hydroxylamine	Asparagine-glycine bonds
2-Nitro-5-thiocyanobenzoate	N-terminal side of cysteine residues

6.3.4 Overlapping Peptides

In the determination of primary structure of protein, several methods (enzymatic or chemical) are simultaneously employed. This results in the formation of overlapping peptides. This is due to the specific action of different agents on different sites in the polypeptide. Overlapping peptides are very useful in determining the amino acid sequence.

Two sets of overlapping peptidases or protein fragments are made by using trypsin to cleave the polypeptide its Arg or Lys residues and, in a separate reaction, using CNBr to leave all its met residues (Fig. 6.7).

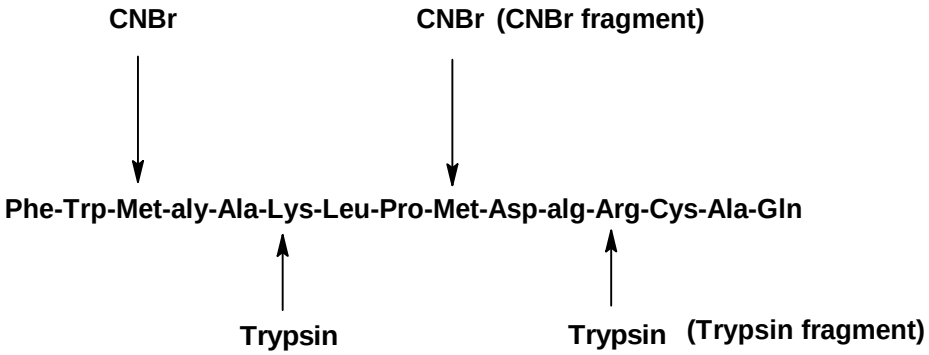


Fig. 6.7: Production of overlapping peptides by CNBr and trypsin treated proteins or polypeptides.

SAQ 2

1. Define the exopeptidases and endopeptidases?

2. Give the three examples of endopeptidases?

3. What is scissile bond?

4. Indicate whether the following statements are true or false:

i) Trypsin, chymotrypsin and pepsin are digestive enzymes. ( )

ii) Exopeptidases and endopeptidases are not specific for their action. ( )

iii) Proteins or polypeptides which contain more than 40 to 100 amino acids cannot be sequenced in a straight forward manner. ( )

iv) Endopeptidase V8 is produced from the bovine pancrease. ( )

v) Pepsin is produced from the bovine gastric mucosa. ( )

6.4 DISULFIDE BONDS AND THEIR LOCATION

Up to now we have explored enzymatic and chemical methods of protein fragmentation. In this section we'll learn about identifying disulfide bond and the need to break it with respect to protein sequencing.

6.4.1 Disulfide Bond

Disulfide bond is a covalent bond. A disulfide (-S-S-) is produced by the sulphhydryl groups (-SH) of two cysteine residues. After formation of disulfide bond in protein, the two cysteine residues are converted into cystine (Fig. 6.8). The disulfide bond is formed in a single peptide chain as well as in between different polypeptide chains. Disulfide bond is responsible for the structural conformation and stability of proteins (Refer Unit-4 of BBCCT-101).

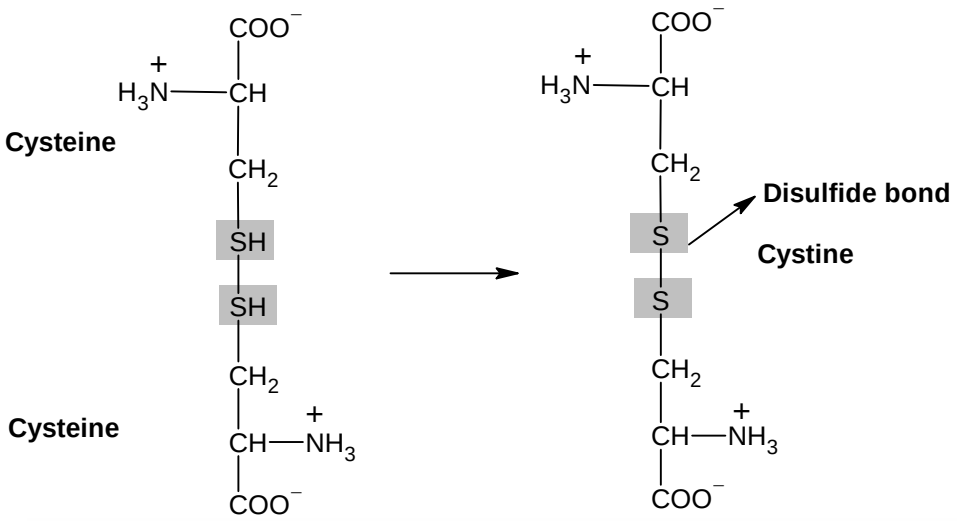


Fig. 6.8: Two cysteine residues join together through the disulfide bond and form a cystine molecule.

6.4.2 Disulfide Bond of Formation

Disulfide bonds are formed by the oxidation of thiol groups (-SH) between two cysteine residues. The disulfide bond formation in protein is a type of covalent modification. This is a reversible process. Oxidation is required for the formation of disulfide bond and the reduction reaction is responsible to disrupt this bond (Fig. 6.9). Oxidation reaction is carried out within the same protein (intra-molecular) or within different proteins (inter-molecular). Intra-molecular disulfide bonded proteins are ribonuclease, insulin, lysozyme etc whereas inter-molecular disulfide bonded proteins are light and heavy chains of antibody. Extracellular proteins are generally disulfide-bonded because the conditions inside the cytosol are reducing and this is a necessary condition for the existence of cystine. Some examples of extracellular proteins are alkaline phosphatase, hGH (Human growth hormone), glucose-1 phosphatase, IGF-I (Interferon like growth factor I), BPTI (Bovine pancreatic trypsin inhibitor and tissue plasminogen activator. The proper and error free disulfide bond formation is assisted by the cellular enzymes such as protein disulfide isomerases. Much more abundant cellular proteins are disulfide-bonded.

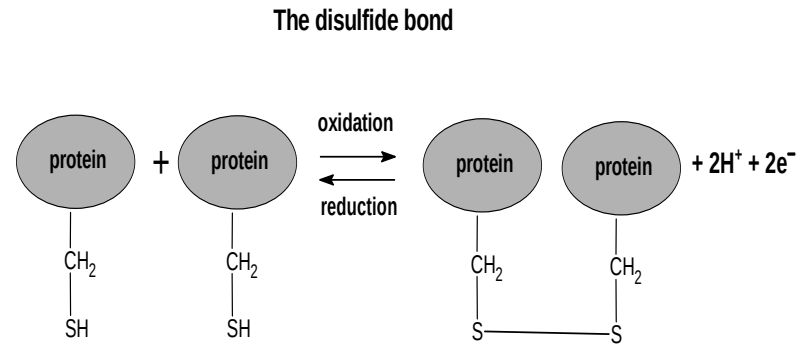


Fig. 6.9: The disulfide bond formation is occurred through oxidation process.

6.4.3 Identification of Disulfide Bonds in Proteins

In protein sequence analysis, it is important to determine the location and numbers of disulfide bonds. The native protein (i.e., with disulfide bridges intact) is cleaved with a proteolytic enzyme (e.g., trypsin) to produce a number of small peptides. After these procedural steps, it is necessary to perform the breaking of disulfide bonds for proper sequencing of proteins. This breaking procedure is carried out by performic acid, 2-mercaptoethanol and DTT (dithiothreitol). The resulting free-sulfhydryl groups then alkylated usually by treatment with iodoacetate (ICH_2COOH) to prevent the reformation of disulfide bonds through oxidation by oxygen. The whole procedure is depicted in Fig. 6.10 Now, the separation and sequencing of polypeptides are performed.

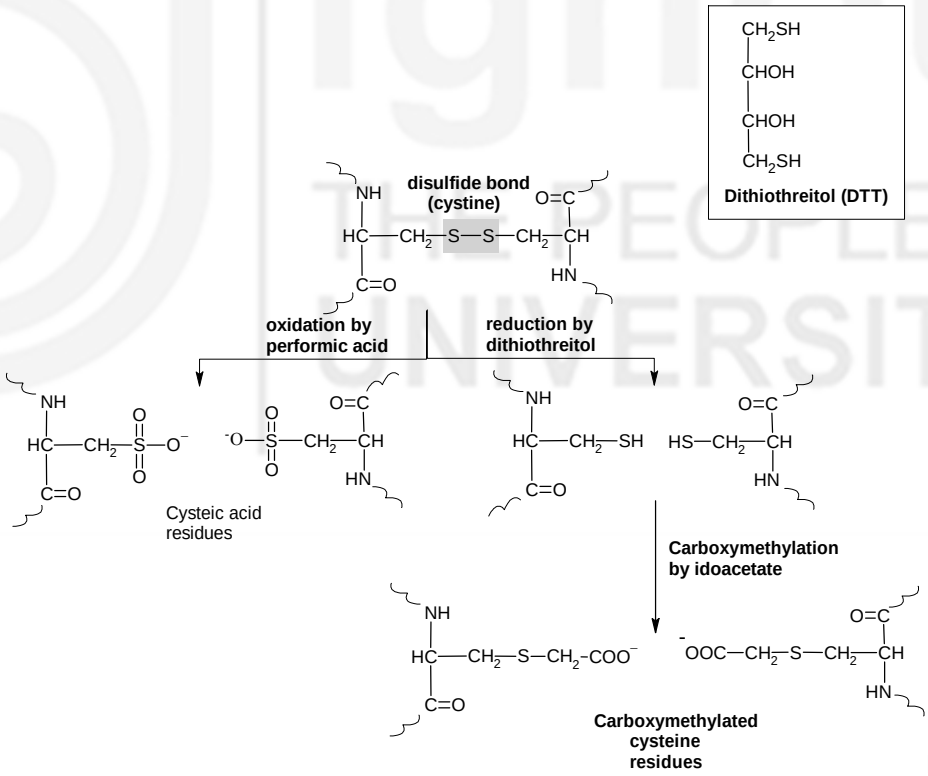


Fig. 6.10: The disulfide bond formation is occurred through oxidation process.

Significance of disulfide bond

Stabilization of the native structures of proteins and peptides is carried out by several kinds of interactions. These interactions are responsible to state directly or indirectly the folding, conformation, and flexibility of the molecule. The majority of these interactions, for example hydrogen bonds and

hydrophobic interactions are noncovalent and relatively breakable. In other side, the covalent interactions are usually stronger. These interactions are accountable to apply a larger influence on stability and conformation, predominantly in peptides, which have a tendency to have fewer and weaker hydrophobic interactions because of their smaller size. The most widespread illustration of a covalent interaction is the disulfide bond. There are various significant roles of disulfide bonds in proteins. Some of these are described below:

(1) Under usual consideration, disulfide bonds are accountable to increase the thermodynamic stability of proteins. This thermodynamic stability affords them less vulnerable to denaturation and degradation in the extracellular environment as well as provides more resistance against to extremes of temperature and pH.

(2) The disulfides also participate in the catalytic role of enzymes. This role is appeared in the enzyme such as thioredoxin. In thioredoxin, it performs as a cellular redox sensor through the oxidation status of its tail groups.

(3) Several disulfides in proteins show to include no direct functional role, somewhat their major target is to preserve the conformation of the protein. In definite cases, conformational changes are linked with the reduction and oxidation of these bonds may permit a protein to switch between different functions.

SAQ 3

1. Define disulfide bonds in protein or polypeptides?
2. Which amino acid is responsible to form disulfide bonds?
3. Give the name of two chemical compounds which are used to reduce the disulfide in the protein or polypeptides?
4. Indicate whether the following statements are true or false:

i) The cleavage of disulfide bonds can be performed by 2-mercaptoethanol. ( )

ii) Iodoacetate is used to prevent the reformation of disulfide bonds. ( )

iii) Methionine is responsible to produce the disulfide bonds in protein or polypeptide. ( )

iv) All the sulfur containing amino acids are responsible to create the disulfide bond in protein or peptides. ( )

6.5 SUMMARY

- Protein sequencing is a method or approaches or techniques applied to establish the amino acid sequence of a protein. The amino acid sequence (also called primary structure) of a protein is the order of the amino acids in the protein chain.

- Sanger's first triumph was to determine the complete sequence of the two polypeptide chains of bovine insulin, A and B, in 1952 and 1951, respectively. Prior to this it was widely assumed that proteins were somewhat amorphous.
- In determining these sequences, Sanger proved that proteins have a defined chemical composition. In 1958, he was awarded a Nobel Prize in Chemistry "for his work on the structure of proteins, especially that of insulin". Amino acids sequencing from the N-terminal side of the protein structure is known as N-terminal analysis. Similarly, the sequencing from the C-terminal side of the protein structure is considered as C-terminal analysis.
- The amino acid residue at the N-terminal can be identified by a number of ways. The three widely used N-terminal methodologies are: 2,4-dinitrofluorobenzene (DNFB) method, 1-dimethylamino naphthalene sulfonyl chloride (Dansyl Chloride; DNS-CL) method and Phenylisothiocynate (PITC) method or Edman degradation method.
- Fewer methods of analysis are available for the identification of C-terminal residue of polypeptides. Most widely used methods are hydrazinolysis and $LiAlH_4$ (Lithium aluminium hydride) based reduction of C-terminus. Determination of the overall amino acid content and the identity of the amino- (N-) terminal residue for each peptide allowed deduction of the sequence of the whole molecule.
- An alternative approach was that described by Edman. This allowed determination of extended sequences of peptides or whole proteins, and has been used widely up to the present day.
- The method employs a series of chemical reactions to remove and identify the amino acid residue that is at the N-terminus of the polypeptide chain, i.e., the residue with a free a-amino group. At the same time, the next residue in the sequence is made available and subjected to the same round of chemical reactions.
- Carboxypeptidases are enzymes that cleave amino acid residues from the C-termini of polypeptides in a successive fashion. Four carboxypeptidases are in general use: A, B, C, and Y. Carboxypeptidase A (from bovine pancreas) works well in hydrolyzing the C-terminal peptide bond of all residues except proline, arginine, and lysine. Fragment of proteins or polypeptides can be produced by chemical or enzymatic cleavage of the native protein.
- Chemical methods are likely to produce large fragments. These methods perform the cleavage at the less common amino acids and produce the two or three large peptides.
- Enzymatic methods have a tendency to cleave adjacent to the more common amino acids. In this regard, these methods produce 50 or more peptides from a protein. Disulfide bond is a covalent bond. A disulfide (-S-S-) is produced by the sulphhydryl groups (-SH) of two cysteine

residues. After formation of disulfide bond in protein, the two cysteine residues are converted into cystine.

- The disulfide bond is formed in a single peptide chain as well as in between different polypeptide chains. Disulfide bond is responsible for the structural conformation and stability of proteins. In protein sequence analysis, it has to determine the location and numbers of disulfide bonds. The native protein (i.e., with disulfide bridges intact) is cleaved with a proteolytic enzyme (e.g., trypsin) to produce a number of small peptides.
- After these procedural steps, it is necessary to perform the breaking of disulfide bonds for proper sequencing of proteins. This breaking procedure is carried out by performic acid, 2-mercaptoethanol and DTT (dithiothreitol). The resulting free-sulfhydryl groups then alkylated usually by treatment with iodoacetate (ICH_2COOH) to prevent the reformation of disulfide bonds through oxidation by oxygen.

6.6 TERMINAL QUESTIONS

1. Discuss and elaborate the protein sequencing and clarify the N-terminal and C-terminal analysis.
2. Discuss all the steps of the Edman degradation methods.
3. Illustrate the steps involved in carboxypeptidase-based C-terminal analysis of the protein.
4. Write the significance of protein sequencing.
5. Describe the enzyme based and chemical reagent-based cleavage of protein or polypeptide.
6. Discuss the cleavage of disulfide bonds.

6.7 ANSWERS

SAQ 1

1. Protein sequencing is a methods or approaches or techniques applied to establish the amino acid sequence of a protein. The amino acid sequence (also called primary structure) of a protein is the order of the amino acids in the protein chain.
2. Protein sequencing is invented by Sanger.
3. Amino acids sequencing from the N-terminal side of the protein structure is known as N-terminal analysis. Similarly, the sequencing from the C-terminal side of the protein structure is considered as C-terminal analysis.
4. i) True ii) True iii) True iv) True.

SAQ 2

1. Endopeptidases are those enzymes which catalyze the hydrolysis of internal peptide bond. Exopeptidases are those enzymes which catalyse the hydrolysis of N or C-terminal amino acids residues.
2. Trypsin, chymotrypsin and elastase.
3. Exopeptidases and endopeptidases are very specific for their action. They act on the bond that is to be cleaved and flanking with specific amino acid residue. This bond is known as scissile peptide bond
4. i) True ii) False iii) False iv) True v) True.

SAQ 3

1. Disulfide bond is a covalent bond. A disulfide (-S-S-) is produced by the sulphhydryl groups (-SH) of two cysteine residues in protein or polypeptides
2. Cysteine amino acid is responsible to form disulfide bonds
3. 2-mercaptoethanol and DTT (dithiothreitol) are used to reduce the disulfide in the protein or polypeptides
4. (i) True, (ii) True, (iii) False, (iv) False.

Terminal Questions

1. Refer to section 6.2.
2. Refer to section 6.2.2.
3. Refer to section 6.2.3.
4. Refer to section 6.2.4.
5. Refer to section 6.3.
6. Refer to section 6.4.

6.8 FURTHER READINGS

1. Creighton, T. E., 1983. Proteins: Structure and Molecular Properties. San Francisco:W.H. Freeman and Co., 515 pp.
2. Creighton, T. E., ed., 1997. Protein Function—A Practical Approach, 2nd ed. Oxford: IRL Press at Oxford University Press.
3. Dayhoff, M. O., 1972–1978. The Atlas of Protein Sequence and Structure, Vols. 1–5. Washington, DC: National Medical Research Foundation.
4. Deutscher, M. P., ed., 1990. Guide to Protein Purification. Vol. 182, Methods in Enzymology. San Diego: Academic Press, 894 pp.
5. Goodsell, D. S., and Olson, A. J., 1993. Soluble proteins: Size, shape and function. Trends in Biochemical Sciences 18**:**65–68.