THE SCIENCE WORLD
A Monthly e Magazine
Popular Article
ISSN:2583-2212
December, 2022; 2(12), 2144-2148

A brief focus on Protein Sequencing in Diagnostic Development

Rajeshwar Khandare[1], A.P. Bansod[2], P.M. Khandare[3]
[1]Division of Animal Biotechnology, IVRI, Izatnagar, Bareilly
[2]Division of Animal Nutrition, IVRI, Izatnagar, Bareilly
[3]Division of Plant pathology, VNMKV, Parbhani
https://doi.org/10.5281/zenodo.7442737

Abstract

The building blocks of living system are proteins, carbohydrate, fats, vitamins and minerals. For the proper functioning of living body needs a balanced coordination between these building blocks known as biomolecules. These biomolecules are available in various natural sources as like in fruits, vegetables, meat and now a days in synthetic forms or manufactured as medicinal formulation. Prior to formulate this building blocks one should know the correct sequence of this biomolecules. The proteins are made up of sequence of amino acid, carbohydrates are sequence of saccharides and fats are of glycerol and fatty acid. Among all of these biomolecule's proteins are most important in both in diet for building of muscle and defence mechanism of body.

1. Sanger Sequencing

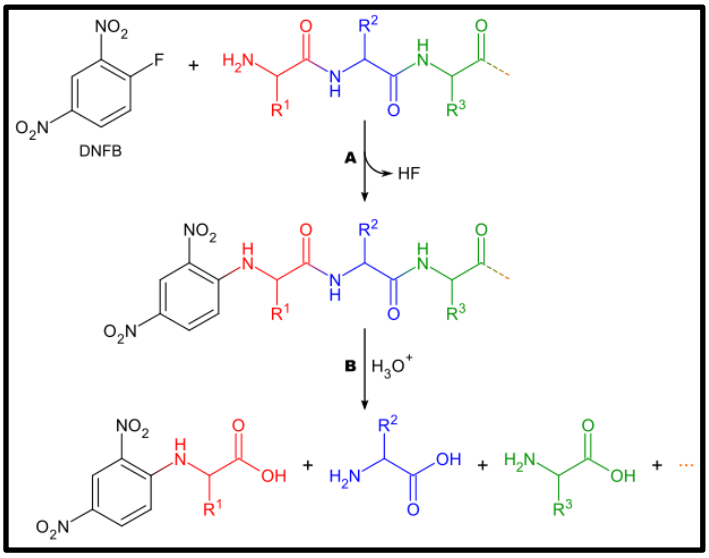
Introduction

Frederick Sanger developed the reagent 1-fluoro-2,4-dinitrobenzene (FDNB) for protein sequencing. Other reagents like Dansyl Chloride and Dabsyl Chloride are also used.

Mechanisms

Addition of FDNB will attack the amino terminal of peptide. Then after labelling of amino terminal residue, the polypeptide is hydrolysed by 6M HCl to its constituent amino acid and labelled amino acid is

identified.



Disadvantages: -

- Once we get the N terminal amino acid, the protein is already hydrolyzed in constituent amino acids. Thus, we cannot repeat the cycle with same sample. For second amino acid sequencing we require new stock of protein sample and the N terminal residue need to be cleaved from the protein using an appropriate protease such as amino peptidase. This makes the process very tedious and complicated.
- These dyes selectively label the amine groups present in the protein and therefore can label the amine groups present in the side chains as well, which may give erroneous results.

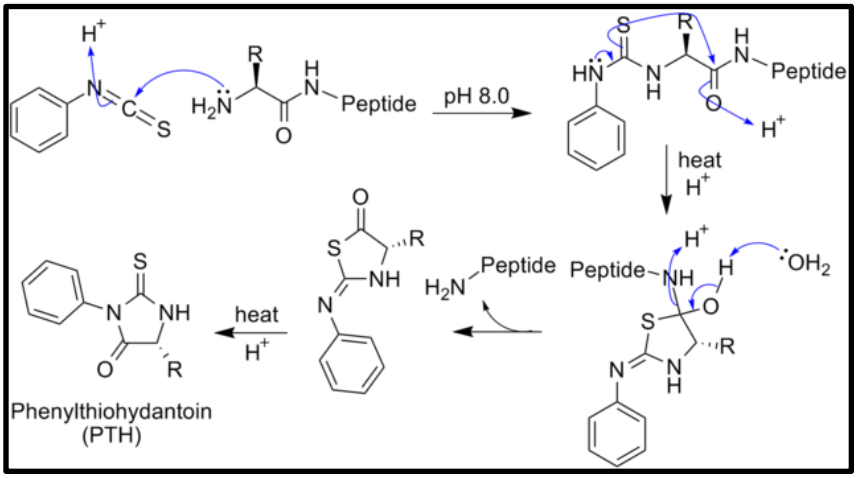
2. Edman Degradation

Introduction

- Edman degradation, developed by **Pehr Edman**, is a method of sequencing amino acids in a peptide. In this method, the amino-terminal residue is labeled and cleaved from the peptide without disrupting the peptide bonds between other amino acid residues.
- The Edman degradation reaction was automated in 1967 by Edman and Beggs to speed up the process.

Mechanism

- Phenyl isothiocyanate is reacted with an uncharged N-terminal amino group, under mildly alkaline conditions, to form a cyclical phenyl thiocarbamoyl derivative. Then, under acidic conditions, this derivative of the terminal amino acid is cleaved as a thiazolinone derivative. The thiazolinone amino acid is then selectively extracted into an organic solvent and treated with acid to form the more stable phenylthiohydantoin (PTH)- amino acid derivative that can be identified by using chromatography or electrophoresis.
- This procedure can then be repeated again to identify the next amino acid.



Disadvantages: -

- A major drawback of this technique is that the peptides being sequenced in this manner cannot have more than 50 to 60 residues (and in practice, fewer than 30).
- The peptide length is limited due to the cyclical derivatization not always going to completion. The derivatization problem can be resolved by cleaving large peptides into smaller peptides before proceeding with the reaction. It is able to accurately sequence up to 30 amino acids with modern machines capable of over 99% efficiency per amino acid.

Advantages: -

- An advantage of the Edman degradation is that it only uses 10 - 100 pico-moles of peptide for the sequencing process.

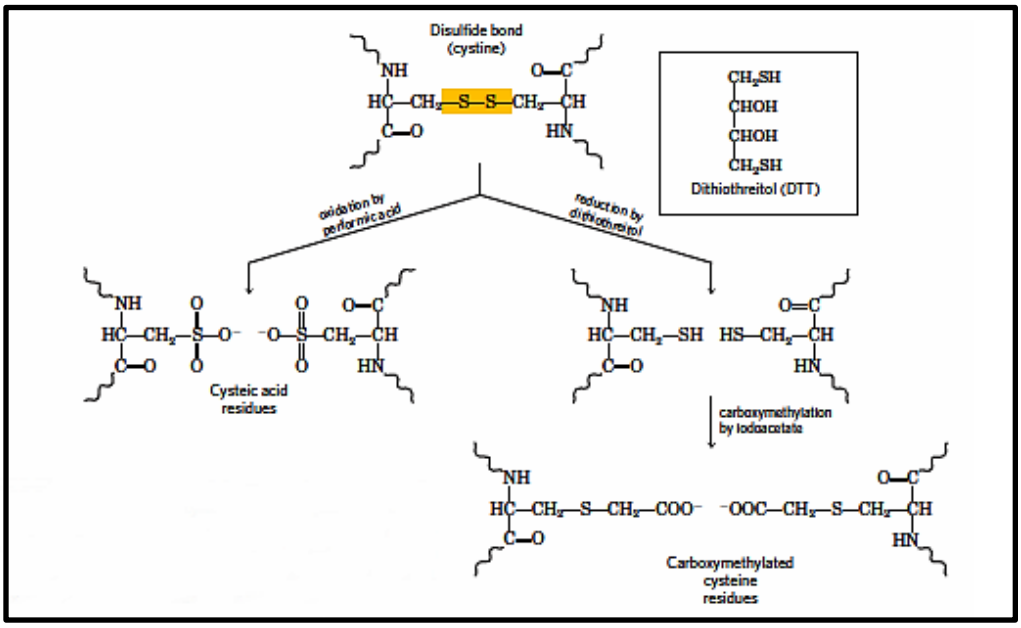
Breaking of Disulfide Bond: -

- Oxidation by performic acid produces two cysteic acid residues
- Reduction by Dithiothreitol or β-mercaptoethanol to form Cys residues which must be followed by further modification of the reactive –SH group to prevent reformation of the disulphide bond. So Acetylation by Iodoacetate serves this purpose.

- **Methylation** of N-terminal amino groups is rare; different methylases do modify specific proteins, including ribosomal proteins, methylation of ribosomes affects their function
- **Amidation of peptides** (*e.g.*, hormones) sometimes occurs at the C-terminus

❖ **Phosphorylation**

- phosphorylation can affect the activity and structure of proteins perhaps as many as 1 in 8 proteins are phosphorylated



Chemical Modification of Protein

- too many examples to list: e.g., HSF activity is modulated by phosphorylation; cell-signalling molecules are best characterized

❖ **Glycosylation**

- Glycosylation takes place in the ER, golgi by a variety of enzymes glycosylated proteins often found on the surface of cells or are secreted folding/assembly of glycosylated proteins requires ER molecular chaperones addition of GlcNAc (beta-O-linked N-acetylglucosamine) residues occurs in the cytoplasm and nucleus modifications are carried out by O-linked GlcNac transferases (OGTs) proteins modified by O-GlcNAc include: cytoskeletal proteins, hormone receptors, kinases & other signalling molecules, nuclear pore proteins, oncogenes, transcription factors, tumor suppresors, transcriptional & translational machinery, viral proteins

❖ **Prenylation, fatty acid acylation**

- proteins without major hydrophobic (transmembrane) domains can be directed to membranes by prenylation of their C-terminal cysteine residue

❖ **Hydroxylation and oxidation, carboxylation**

- A variety of derivatives are known; *e.g.*,hydroxyamino acids (hydroxyproline) and are very common in collagen

❖ **Selenocysteine/selenomethionine modification**

- Essentially all selenium in cells occurs as selenocysteine

- Selenomethionine is a useful for protein crystallography: can grow cells in the presence of the modified amino acid and produce protein containing se-Met; can deduce 'phase' of protein this way

❖ **Pegylation**: Chemically attaching single- or branched-chain polyethylene glycol (PEG) groups to proteins is a form of labelling or modification that is primarily used to confer water-solubility and/or inert molecular mass to proteins. Forms of PEG that have been synthesized to contain reactive chemical groups comprise ready-to-use, activated reagents for pegylation.

❖ **Block sulfhydryls**: Protein sulfhydryls (side chain of cysteine) are important regulators of protein structure and function. Certain reagents are capable of reacting permanently or reversibly with sulfhydryl groups (e.g., NEM or MMTS, respectively). These reagents add a very small "cap" on the native sulfhydryl, enabling the activity of certain enzymes to be controlled for specific assay purposes.

❖ **Convert amines to sulfhydryls**: SATA (N-succinimidyl-S-acetyl-thioacetate) and related reagents contain an amine-reactive group and a protected sulfhydryl group. By reacting the compound to a purified protein, the side chain of lysine residues can be modified to contain a sulfhydryl group for targeting with sulfhydryl-specific crosslinkers or immobilization chemistries. The method does not actually convert the amine into a sulfhydryl; rather it attaches a sulfhydryl-containing group to the primary amine. The effect is also to extend the length of the side chain by several angstroms.

Reference

1. Smith, Bryan John, ed. *Protein sequencing protocols*. Vol. 211. Springer Science & Business Media, 2002.
2. Fraenkel, D. G., & Vinopal, R. T. (1973). Carbohydrate metabolism in bacteria. *Annual Reviews in Microbiology*, *27*(1), 69-100.
3. O'brien, R. D. (2008). *Fats and oils: formulating and processing for applications*. CRC press.