PROTEINS

LECTURE 3: PEPTIDES AND THE PRIMARY LEVEL OF PROTEIN STRUCTURE

Topic Learning outcomes: At the end of this section, you should be able to:

1. Illustrate peptide structure, properties, function and its synthesis
2. Relate pH to the pI and changes in the structural forms and biological roles of peptides
3. Describe the primary level of protein structure
4. Explain the methods used to determine protein composition and sequence

The formation of the peptide bond

The condensation of the α-amino group of one amino acid with the α-carboxyl group of another amino acid results in the formation of a bond referred to as the peptide bond with the release of a water molecule. This reaction is an important requisite to the polymerization reactions forming peptides and proteins. The reaction is shown in Fig. 4.1.

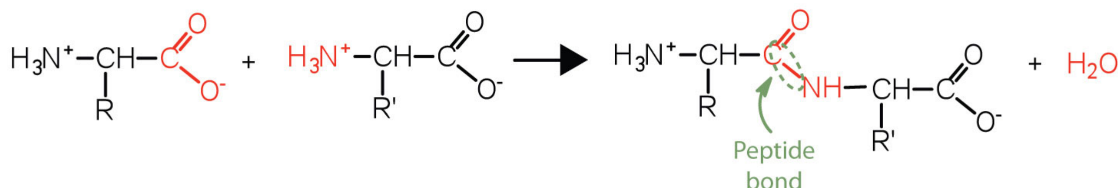


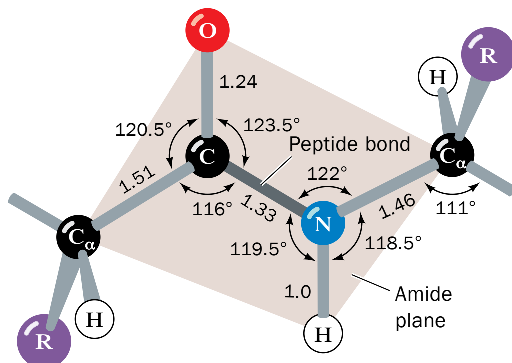
Fig. 4.1. Reaction leading to the formation of the peptide bond
(**https://www.khanacademy.org/science/biology/macromolecules/proteins-and-amino-acids/v/peptide-bond-formation**)

Electron delocalization of the lone pairs of nitrogen towards the δ-positive carbonyl carbon produces a peptide bond that has partial double bond character. This makes the bond very strong. The bond can be visualized having features that are made up of the contributions of all resonance structures due to such electron delocalization. Consequently the peptide bond has the following features:

1. The bond length of the peptide bond is 1.325 which is greater than the bond length of the C=N double bond (1.27 A) but shorter than the bond length of the C-N single bond (1.487).
2. Similarly the bond strength has values between that of the double and

single bonds.

3. The partial double bond character hinders rotation of the peptide bond. Consequently the carbonyl carbon, oxygen and nitrogen atoms lie in a plane with the substituent R groups of a dipeptide perpendicular to this plane and trans to each other. Such arrangement minimizes steric interactions between the R groups.



Source: Marsh, R.E. and Donohue, J. 1967, *Adv. Protein Chem.* **22, 249**

Fig. 4.2. The consensus bond length and bond angles of the peptide bond and adjacent bonds

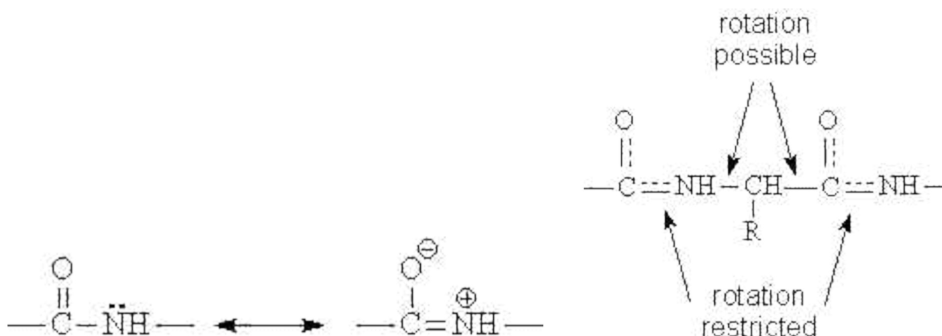


Fig. 4.3. Contributing structures to the peptide bond (A) and the consequent restriction in rotation of the bond

PEPTIDES

https://cdn.kastatic.org/ka-youtube-converted/nv2kfBFkv4s.mp4/nv2kfBFkv4s.mp4#t=0

Peptides are short polymers or oligomers of amino acids linked by peptide bonds. It is composed of usually less than 50 amino acid residues.

Nomenclature and convention

Peptides, and proteins for that matter, are written and named according to a convention whereby the first amino acid drawn farthest to the left is the N-

terminal while the last amino acid to the right is the C-terminal residue. The prefixes di-, tri-, tetra-, penta-, etc. are used to indicate the number of amino acid residues linked by peptide bonds. A dipeptide is composed of 2 amino acids; a tripeptide is made up of 3 amino acids, etc. The number of peptide bonds linking the amino acids together is one less than the number of residues. In Fig. 4.4 below is a peptide containing 10 amino acids. It is therefore a decapeptide.

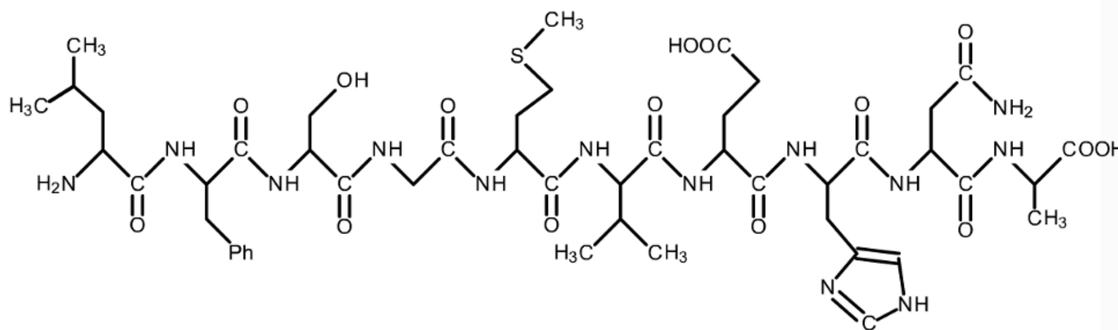


Fig. 4.4. An example of a peptide

The peptide is named starting from the N-terminal such that all these amino acids up to the 9th amino acid are substituents of the 10th amino acid, the C-terminal. Thus the peptide is named: valyl-phenylalanyl-seryl-glycyl-methionyl-valyl-glutamyl-histidyl-asparaginyl-alanine or using the 3-letter abbreviation is simply val-phe-ser-gly-met-val-glu-his-asN-ala. Using the one letter abbreviation, the peptide is VFSGMVEHNA.

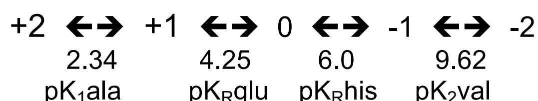
Exercise 4.1.

Acidity and basicity properties of peptides

Peptide bond formation results in the loss of the acidic and basic character of the carboxyl and the amino group of the amino acids that become bonded together. Since the lone pair of nitrogen is partially delocalized towards the carbonyl carbon, the pair of electrons is no longer available for binding to electron-poor species such as electrophiles and protons. Only the N-terminal has a free amino group while the C-terminal amino acid has the only free carboxyl group. Nonetheless, R groups of acidic and basic amino acids or residues as well as R groups of polar amino acids with ionizing protons will contribute to the acidity and basicity of peptides.

For a single amino acid, the ionization of the H^+ of the carboxyl and the protonated amino groups, as well as the ionizable protons in R determines the form, net charges and pI of the amino acids in solution. For peptides, however, the acidity and basicity of the carboxyl and amino groups, respectively, are lost with the formation of the peptide bond. Only the N-terminal amino group and the C-terminal carboxyl group and all R groups with ionizable protons contribute to

the acidity and basicity of peptides. Thus, in determining the forms, net charges and pI of peptides, only ionization of protons from these groups will be considered. For example, for the peptide in Fig. 4.4, at pH lower than the pKa of the carboxyl proton of alanine, i.e. at pH 2.34, the amino group of the N-terminal amino acid, valine and the imidazole ring of histidine will be protonated. At this pH therefore the net charge of the peptide, VFSGMVEHNA is +2. The amide functional group is not basic because the lone pair of nitrogen is delocalized toward the δ-positive carbonyl carbon. That is why, this nitrogen is not basic and protonated. Nonetheless, the peptide is a weak acid of the form H_4A. The protons come from the ionization of the carboxyl, the protonated N-terminal, the protonated imidazole ring of histidine and the carboxyl group in R of glutamic acid, with pKa values of 2.34, 9.62, 6.0 and 4.25, respectively. Ionization occurs in the following manner:



The zwitterion form is the form at which net charge is equal to zero. And this form predominates at pH between 4.25 and 6.0. pI is still the average of only two values, the pKa before and after the zwitterionic form. And again, for the given peptide, the pI is the average of 4.25 and 6.0, i.e. pI = 5.125.

Peptide Families

Peptides found in biological systems are usually products of proteolytic cleavage or are derivatives of cleavage products of proteins. The Handbook of Biologically Active Peptides (*Abba J. Kastin, ed. (2013). Handbook of Biologically Active Peptides (2nd ed.). ISBN 978-0-12-385095-9.*) lists families of peptides that includes those derived from plants, bacterial or antibiotic peptides, fungal, invertebrate and marine, amphibian skin, venom, anti-cancer, inflammatory, brain, endocrine, opiates, and other organ-specific peptides. Table 4.1 is a list of some of the bioactive peptides produced by cells.

Table 4.1 Some naturally occurring bioactive peptides

Neme	Composition	Localization	Function
Glutathione	γ-Glu-Cys-Gly (or γECG)	All animal cells	Antioxidant; cell growth regulator
Angiotensin II	Asp-Arg-Val-Tyr-Ile-His-Pro-$PheNH_2$ (or $DRVYIHPFNH_2$)	Brain, kidneys, heart and vascular system	Neuromodulator Regulation of fluid reabsorption
Oxytocin	CYIQNCPLGNH2	Produced by hypothalamus and	Signaling for childbirth,

		secreted into the blood	
Insulin		Produced by the pancreas (beta cells) and secreted when blood glucose level increases	Regulation of blood glucose level by decreasing its concentration outside cells
Glucagon		Produced by the pancreas (alpha cells) and secreted when blood glucose level goes down	Regulation of blood glucose level by increasing its concentration in the blood
Bradykinin		Various animal cells	Promotes inflammation and vasodilation
Gramicidin S	See figure 4.5	Produced by the bacteria, *Bacillus brevis*	antibiotic

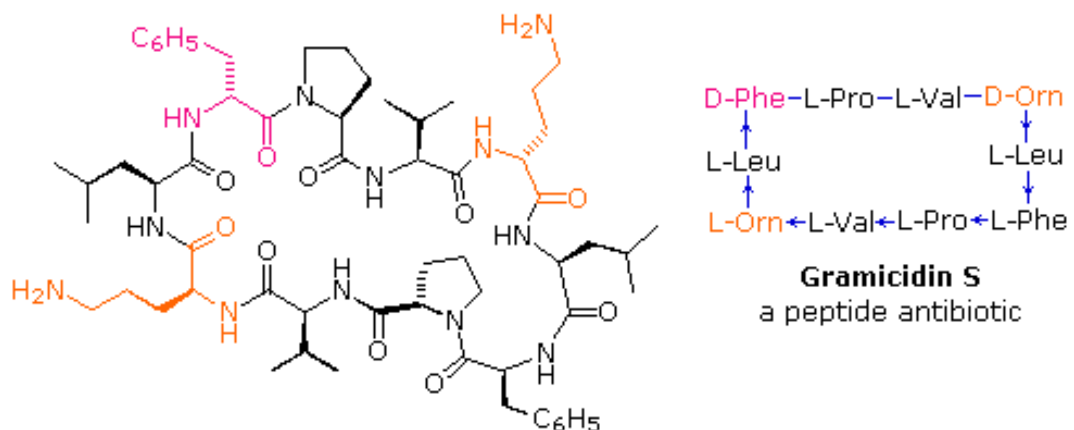


Fig. 4.5. The structure of gramicidin S, a peptide antibiotic produced by Bacillus brevis. Some

non-common amino acids are shown in color such as D-phe, D-ornithine and L-ornithine. When writing the name of the peptide using the 3-letter abbreviation, a single line at the top follows the N→C convention with the vertical and lower lines continuing following the direction of the peptide bond between the amino acids, indicated by the arrows.

Exercise 4.1.
The peptide bradykinin is a nonapeptide. Give the name of the peptide by naming the amino acids from the N-terminal to the C-terminal.

Proteins are complex molecules that are synthesized in the ribosomes of cells. Most peptides however are either degradation or proteolytic cleavage products of larger proteins. Nonetheless, a number of these peptides have bioactivities that are important for the proper functioning of cells such as those involved in immunomodulation, defense, metabolic regulation, etc. Some of these can be synthesized *in vitro* by recombinant technology using expression vectors (e.g. insulin, glucagon and other peptide hormones) but some, particularly those that have been found to be important defense peptides, e.g. antimicrobial peptides, and short peptides as well can also be synthesized in the laboratory. The figure below (Fig. 4.6) shows the synthesis of a peptide by coupling N- and R-protected amino acids.

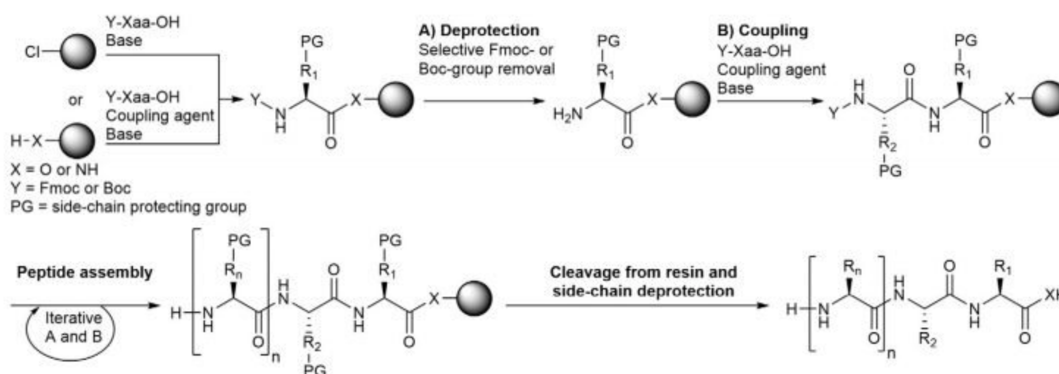


Image by Bédard, F. and Biron, E. (2018) Frontiers in Microbiology 9:1048

Fig. 4.5. Solid-phase synthesis of a peptide using an (amine-functionalized) amide resin. The N-terminal group is protected by either Fmoc or Boc, while the side chain is also protected by chemical modification of its functional group.

PROTEINS: THE PRIMARY STRUCTURE

The term "protein" is derived from the Greek word "Proteios", meaning "of first importance". It is the most abundant cellular component making up approximately 70% of the organic matter in a cell. It is also the first most extensively studied biomolecule owing to the fact that expressed or observable traits called phenotype are attributed to a particular protein component. For example, the color of the skin depends on the level of melanin, a protein pigment

that accumulates in the epithelium of the skin. A type of melanin is also responsible for the color of red hair. Or the curliness of the hair is due to the type of keratin produced by cells in hair follicles. A major change in an organism's genetic trait whether this is due to evolution or mutation is manifested by a change in protein expressed, which in turn changes the organism's characteristics.

Proteins are higher molecular weight-polymers made up at least 50 amino acids. Proteins have specific structures based on1he amino acids making up the protein molecule. All proteins have at least three levels of organization: 1) primary, 2) secondary, and 3) tertiary. A fourth one, quaternary, is also observed for proteins that are made up of more than one polypeptide chain, or is at least a dimer. In this section, you will learn about the primary structure of proteins.

The relative proportion of amino acids present in the polypeptide chain and the arrangement of these amino acids from the N to the C terminal is the primary structure of proteins. The relative proportion is the % composition while the arrangement is referred to as the sequence of the amino acids in the protein molecule. The % composition can be determined by both qualitative and quantitative means.

The first step in the qualitative and quantitative determination of the amino acid content is the partial and/or complete hydrolysis of the protein sample. This is followed by paper or ion-exchange chromatography or the use of automated techniques such as amino acid analyzers. Hydrolysis can be effected by treating the protein sample in acid solution such as 6N HCl or performic acid and heating in a sealed tube in vacuum. Hydrolysis must ensure minimum degradation of amino acids that are easily destroyed such as tryptophan and cysteine, which may require use of more specific techniques. Qualitatively, the presence of specific amino acids can be determined by the visible reaction of particular R groups for certain reagents. For example, the Sakaguchi test will detect presence of the guanidine group of arginine. Substituted aromatic R groups such as the phenol ring of tyrosine and the indole ring of tryptophan can be detected by the Xanthoproteic test. Some chemical reactions are shown in Table 4.2

Table 4.2. Some qualitative tests and the visible result for the detection of amino acids

Reactive group	Reaction/Test	Product/ visible result
α-amino groups	Ninhydrin	Ruhemann's purple, a violet-colored product absorbing at 540 nm.
-NH2 of lysine	Aldehydes	Schiff's base adducts
Guanidine group of arginine	Sakaguchi test (a-naphthol in alkaline solution/NaOH)	Red-orange complex

Substituted benzene	Xanthoproteic test	Yellow precipitate or solution
-SH of cysteine	Pb(OAc)2	Black precipitate, PbS
	Dithionitrobenzoic acid	Yellow product; may be used to quantify –SH groups
Imidazole of his, and phenol of tyrosine	Pauly reagent	Reddish/ yellow product

The quantitative determination of the amino acid composition of the peptide can be determined after complete hydrolysis by separating the products in chromatography. Figure 4.6 is an illustration of the elution profile of peptide hydrolysates to determine the kind and quantity of amino acids present.

Determination of amino acid composition

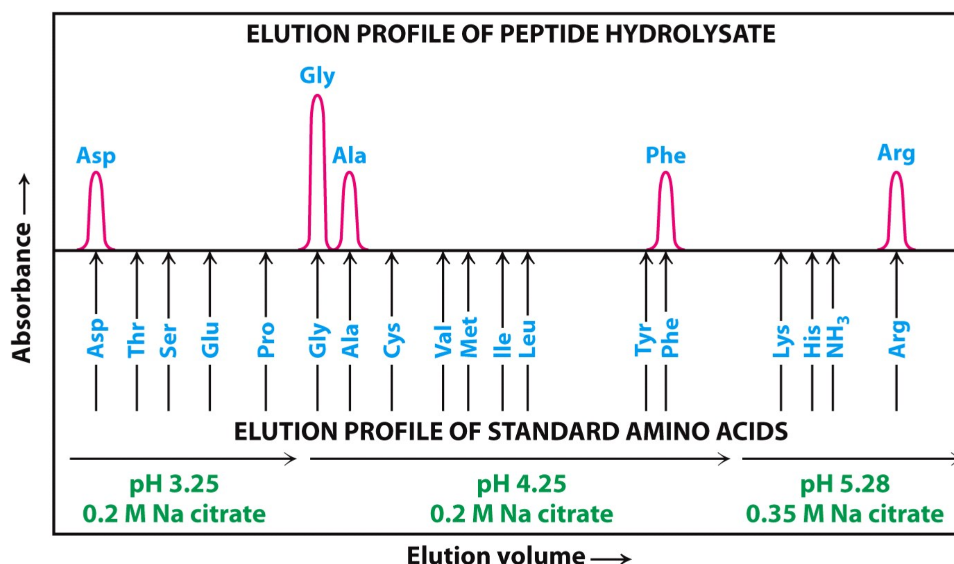


Fig. 4.6. The elution profile of amino acids in a hydrolysate versus standard amino acids.

Amino acid sequence can be determined by several techniques including the use of automated instruments such as amino acid sequencers. Prior to the use of sequencers, scientists used only a combination of chemical and enzymatic methods to determine protein sequences. The revolutionary work of Frederick Sanger revolutionized the study of proteins, which previously could not be sequenced owing to its large size. Sanger determined the sequence of insulin, a 51 amino acid protein, and was awarded the 1958 Nobel Prize in chemistry for such work. He simply used acids to hydrolyze the protein into smaller sized products, which he separated by using electrophoresis and chromatography.

Although it took him years to determine the amino acid sequence of insulin, he was also able to correctly determine that insulin is made up of 2 polypeptide chains, A and B linked by a disulfide bond. https://www.researchgate.net/publication/11064896_The_first_sequence_Fred_Sanger_and_insulin

The methods available if protein sequencers are not available include the use of chemical and enzymatic techniques. Some of these methods are listed in Table 4.3.

Table 4.3. Chemical and enzymatic method used for determining protein sequence without the use of sequencers.

Chemical Analyses		
Method/ Reagent	**Specificity**	**Additional use**
DNFB (2, 4-dinitrofluorobenzene) or the Sanger reagent	Determination of the N-terminal amino acid producing a dinitrophenol (DNP)-amino acid	Detection of N-terminal residues only; results in complete hydrolysis of the protein
Phenylisothiocyanate (Edman reagent)	Determination of N-terminal amino acid producing a phenylthiohydantoin (PTH)-amino acid and cleaving this from the parent peptide	This reagent is currently used in protein sequencers as it cleaves only the N-terminal amino acid while keeping the rest of the peptide intact
Hydrazine	Determination of the C-terminal amino acid	
Cyanogen bromide (CNBr)	Cleaves peptide bonds in which the carboxyl groups are contributed by the amino acid, methionine	At least one hydrolytic product will have methionine for its C-terminal amno acid
Hydroxylamine HCl	Cleaves peptide bonds joining asN and gly resides only.	If the cleavage products are two shorter peptides, one of these will have asN as its C terminal while the other will have gly as its C-terminal.
Iodosobenzoate	Cleaves at the carboxyl side of tryptophan residues	If there are 2 hydrolytic products, one will have trp as its C-terminal residue
2-nitro-5-thiocyanobenzoate	Cleaves at the amino side of cysteine	If there are 2 hydrolytic products, the N-terminal amino acid of one will be

		cysteine
Enzymatic hydrolysis		
Trypsin	Cleaves peptide bonds in which the carbonyl group is contributed by basic residues, lys and arg (see Fig. 4.7)	All hydrolytic products except the product towards the C-terminal will have lys or arg as its C-terminal.
Chymotrypsin	Cleaves peptide bonds at the carboxyl side of aromatic amino acids, tyr, phe and trp	All hydrolytic products except the product towards the C-terminal will have tyr or phe or trp as its C-terminal.
Carboxypeptidase A	Cleaves all C-terminal amino acids except proline	Determination of C-terminal amino acid
Carboxypeptidase B	Cleaves C-terminal amino acids lys, and arg only	Determination of C terminal amino acid
Thermolysin	Cleaves amino side of residues where R is contributed by leu and val	All hydrolytic products except N-terminal will have leu or val as N-terminal
Streptococcal protease	Cleaves peptide bonds at the carboxyl side of acidic amino acids	All hydrolytic products except the product towards the C-terminal will have glu or asp as its C-terminal residue.
Thrombin	Cleaves peptide bonds at the carboxyl side of arginine	If there are 2 hydrolytic products, one will have arg as its C-terminal amino acid
Elastase	Cleaves peptide bonds at the carboxyl side of ala	If there are 2 hydrolytic products, one will have ala as its C-terminal amino acid

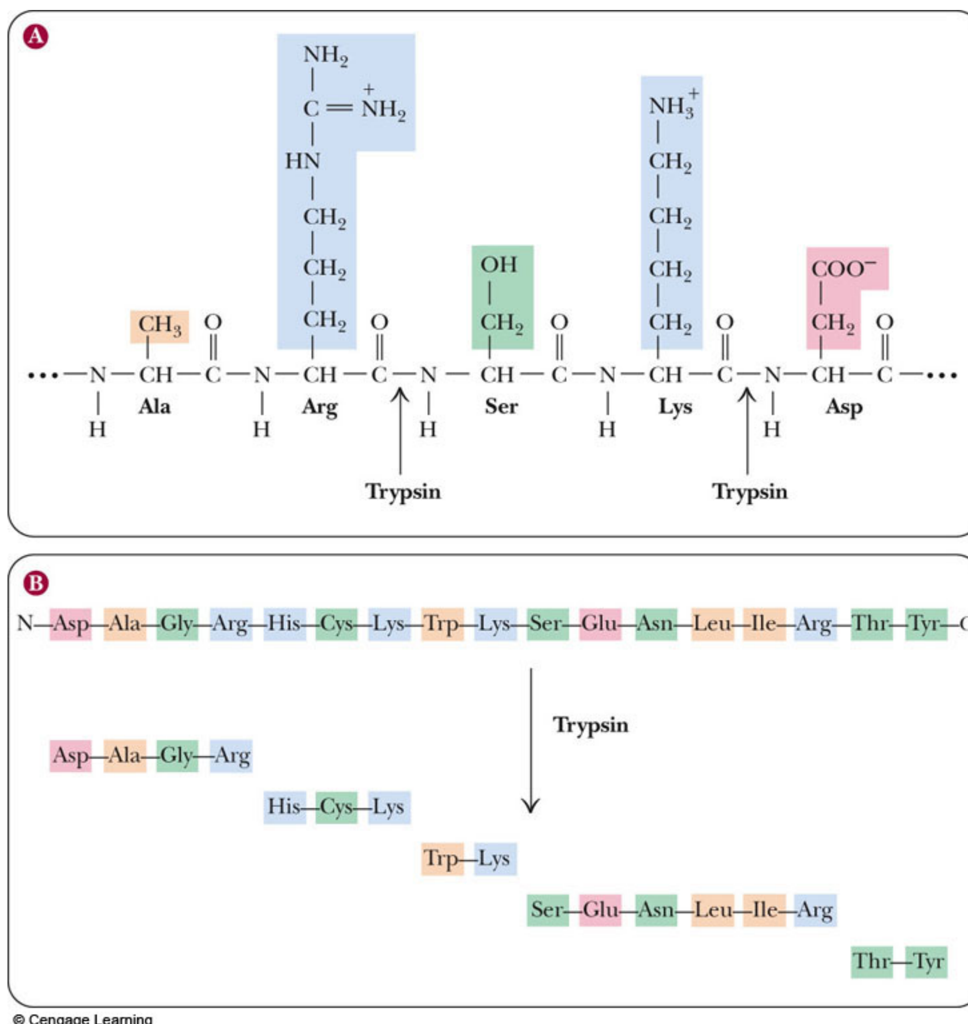


Fig. 4.7. Specificity of enzyme (trypsin) cleavage of the peptide bond of a protein

Prior to sequencing, large proteins are cleaved into smaller peptide products for ease of sequencing. The chemical and enzymatic methods described in Table 4.2 are useful for these purposes.

The Edman degradation is the method used in the sequencing of amino acids by a sequencer. The process involves an initial protection of the side chain carboxyl and hydroxyl groups converting it to either amides or esters, while the C-terminal carboxyl group is converted into an anhydride. Reaction with an isothiocyanate causes the cyclization of the N-terminal aminto acid to a hydantoin ring which is cleaved from the peptide. The resulting product is a thiohydantoin derivative and a peptide shorter by one amino acid. The Edman degradation cycle is repeated until about 10-20 amino acids from the N-terminal have been identified and the sequence similarity with known proteins searched using databases. Fig. 4.8 shows the Edman degradation process.

Edman degradation to sequence proteins is most precise for smaller

peptides. Sequencing large proteins may require combining both the initial chemical and enzymatic degradation, and the Edman degradation. An example of the application of using the data for overlapping peptides from hydrolysis to guide sequence analyses is shown in Fig. 4.9.

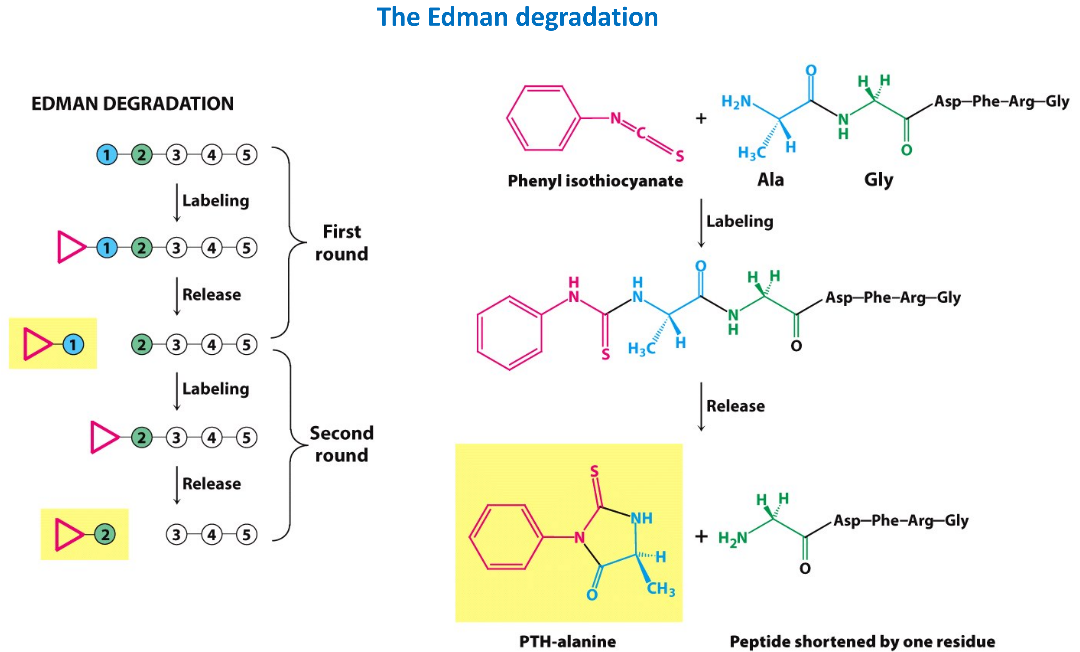


Fig. 4.8. The Edman degradation cycle

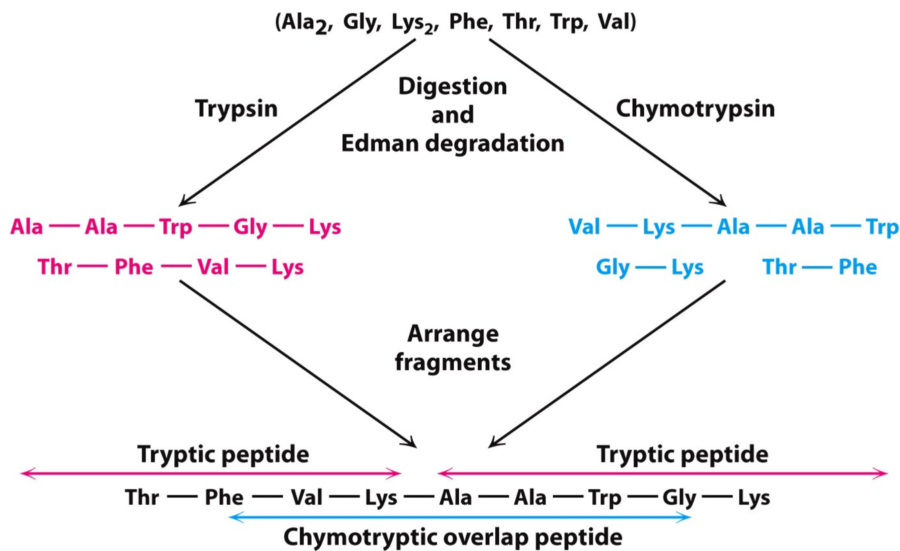
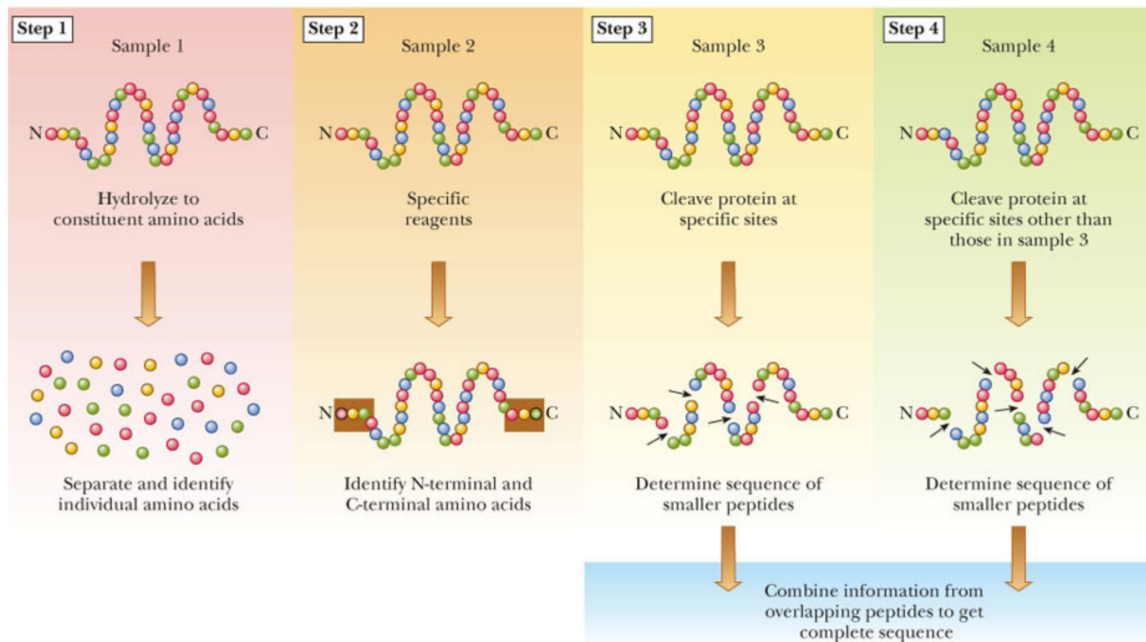


Fig. 4.9. Use of a combination digestion and Edman degradation for sequence analysis

To sum up the sequencing procedure, the following general steps are followed as are shown in Fig. 4.10.



© Cengage Learning

Fig. 4.10. General steps to sequence proteins

The primary level will give both a qualitative and quantitative description of a protein molecule. Additionally it will help provide data to infer the following:

1. Structure and function of the protein molecule by comparing with existing protein databases
2. It could give a clue to evolutionary relationships of the protein and the species where the protein can be found. For example, the relationship of the different coronavirus strains that have evolved can be compared based on sequence analysis.
3. Structural modifications of the protein and the presence of internal repeats, motifs and domains.
4. Sequence similarities with internal structures that regulate its function and binding with ligands or interacting partners
5. Basis for diseased states
6. Gene localization by using cDNA sequences derived from the protein sequences

References:

https://www.open.edu/openlearn/science-maths-technology/science/biology/proteins/content-section-1.2

INTERESTING FACTIS

https://www.nature.com/scitable/topicpage/frederick-sanger-method-man-problem-solver-6537485/

https://www.genetics.org/content/162/2/527