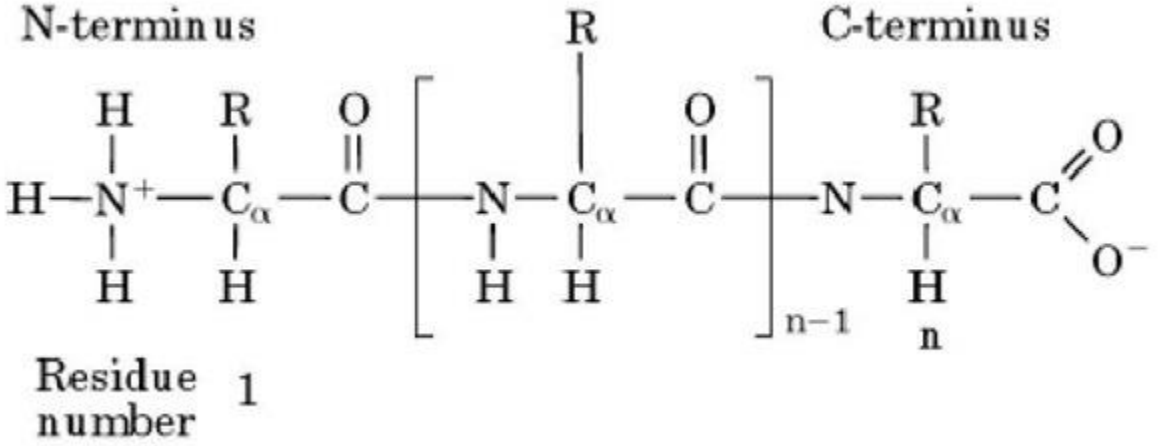
Protein sequencing, structure and peptide synthesis

(semester V, CC12)

Dr. Subhra Mishra
Associate Professor
Department of Chemistry

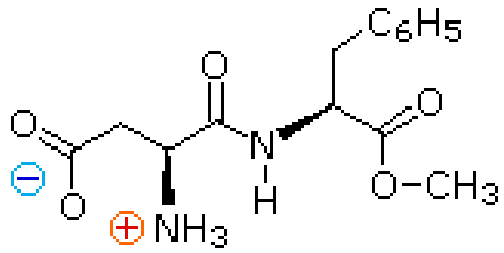
Sequencing of proteins:

- Protein sequencing is a technique to determine amino acids sequence in a protein
- It is a method to understand the structure and function of proteins in a living organism
- Amino acid sequence determines the eventual three dimensional structure of protein



Importance of sequence…….

- Because the N-terminus of a peptide chain is distict from the C-terminus, a small peptide composed of different aminoacids may have a several constitutional isomers (e.g. Asp-Phe or Phe-Asp).
- The methyl ester of the first dipeptide is the artificial sweetner aspartame, which is nearly 200 times sweeter than sucrose. Neither of the component amino acids is sweet (Phe is actually bitter), and derivatives of the other dipeptide (Phe-Asp) are not sweet.



Asp-Phe-OCH_3

Aspartame

- Sequences and composition often (not always) reflect the function of the protein (often proteins of similar function will have similar sequences)
- Homologous proteins from different organisms have homologous sequences

Problems of sequencing……

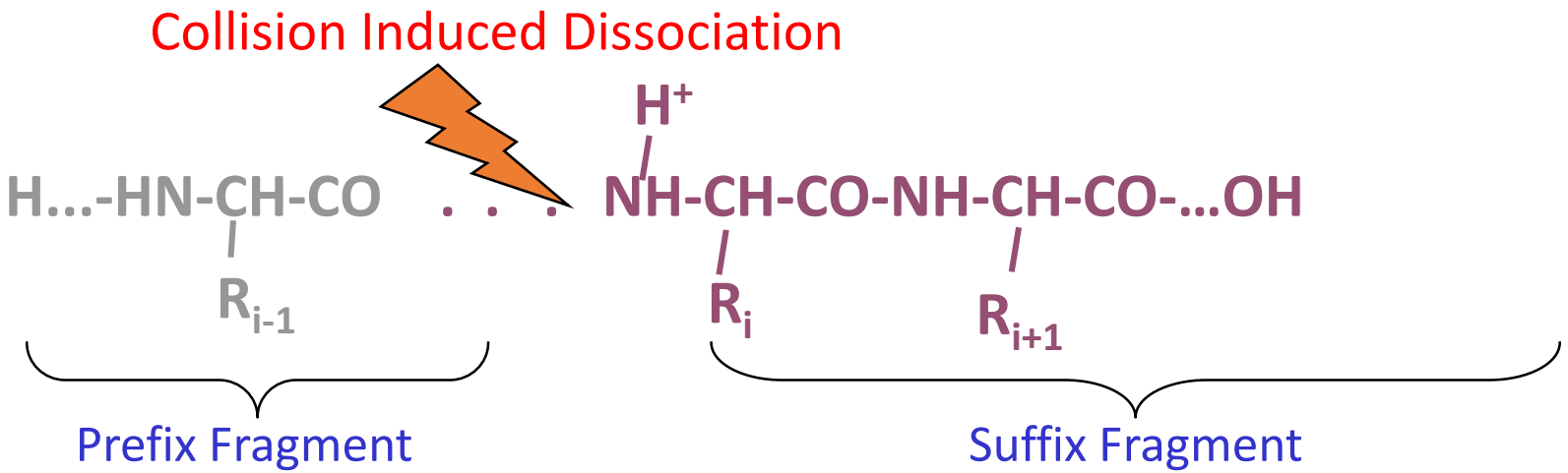
100 amino acid protein has 20^{100} combinations
1953 Frederick Sanger sequenced the two chains of insulin (21 aa)
All of the molecules of a given protein have the same sequence
Proteins can be sequenced in two ways:

- direct amino acid sequencing
- indirect sequencing of the encoding gene (DNA

Conventional Sequencing:

1. If more than one polypeptide chain, separate.
2. Cleave (reduce) disulfide bridges
3. Determine composition of each chain
4. Determine N- and C-terminal residues
5. Cleave each chain into smaller fragments and determine the sequence of each chain
6. Repeat step 5, using a different cleavage procedure to generate a different set of fragments.
7. Reconstruct the sequence of the protein from the sequences of overlapping fragments
8. Determine the positions of the disulfide crosslinks

Peptide Fragmentation



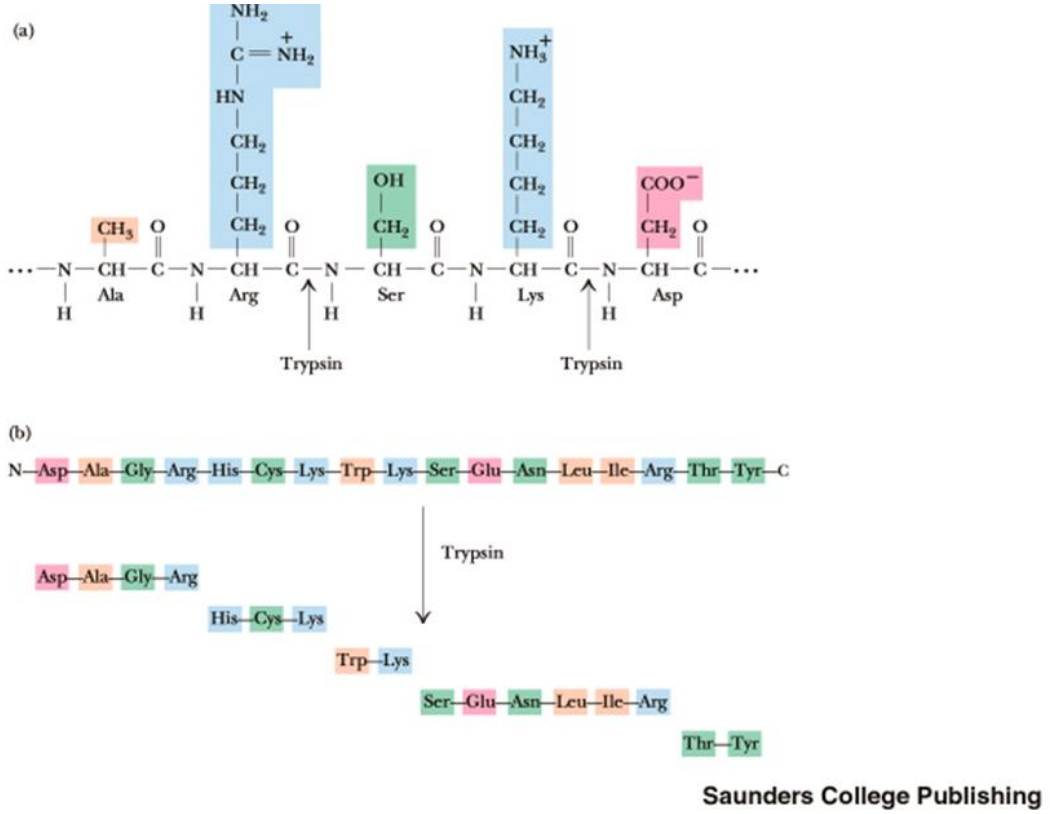
Proteases, e.g. trypsin, break protein into peptides

- Peptides tend to fragment along the backbone.
- Fragments can also loose neutral chemical groups like NH_3 and H_2O.

Fragmentation of the chains

Enzymatic fragmentation

- Trypsin (R or K)
- Chymotrypsin (F or Y or W)
- Clostripain (R>K) (can be incomplete for K)
- Staphylococcal protease (D or E)
- Chemical fragmentation-cyanogen bromide (Met homoserine lactone)

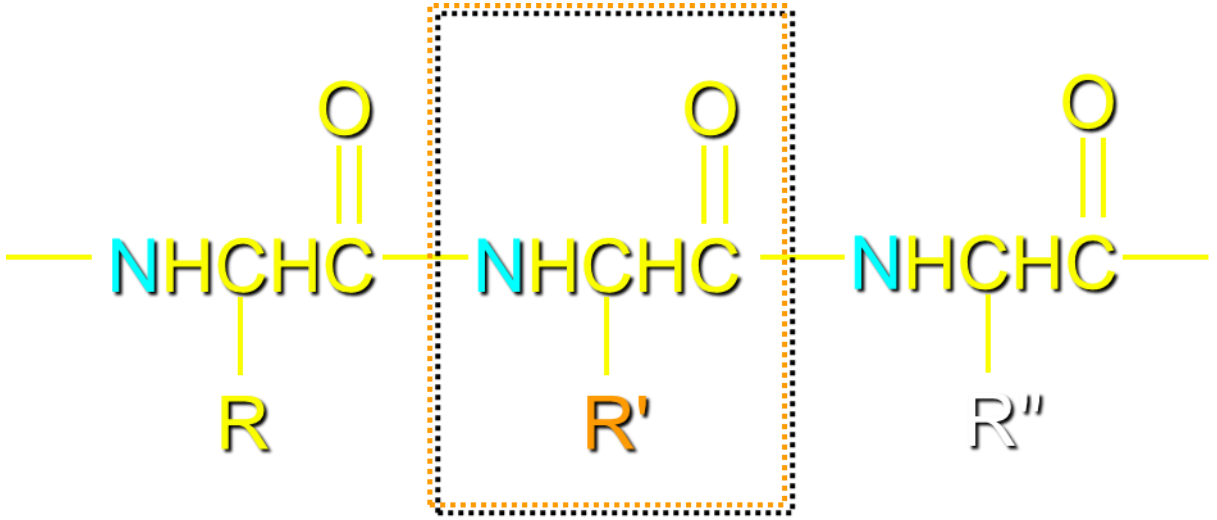


Partial Hydrolysis of Peptides and Proteins:

Acid-hydrolysis of the peptide cleaves all of the peptide bonds.
Cleaving some, but not all, of the peptide bonds gives smaller fragments.
These smaller fragments are then separated and the amino acids present in each fragment determined.
Enzyme-catalyzed cleavage is the preferred method for partial hydrolysis.

Example: Trypsin

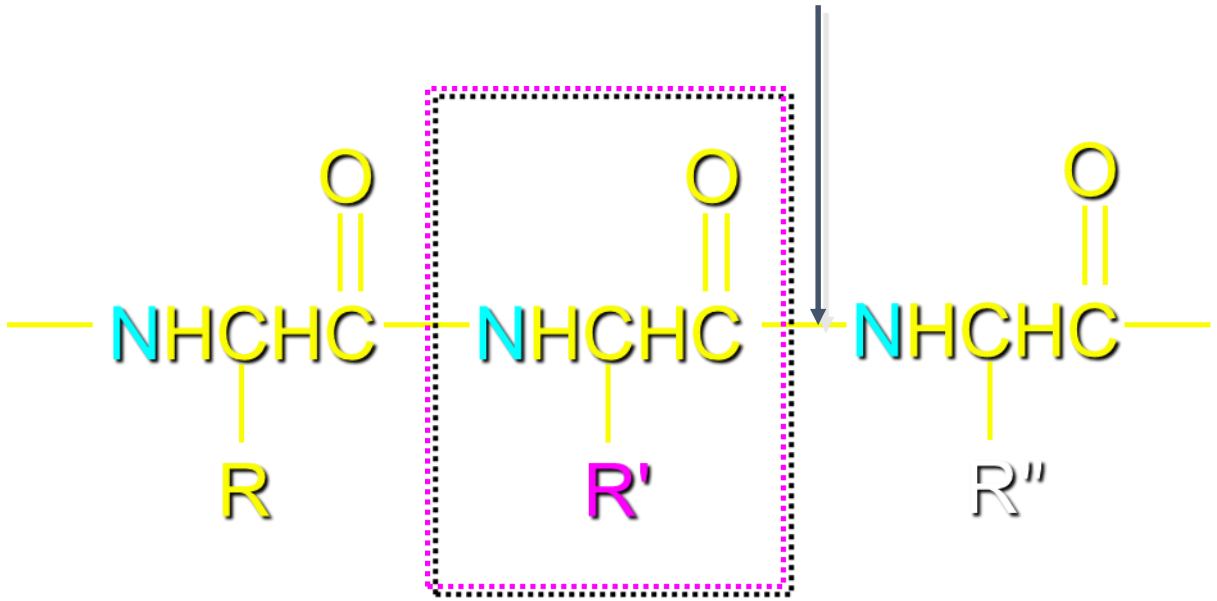
Trypsin is selective for cleaving the peptide bond
to the carboxyl group of lysine or arginine.



Lysine or arginine

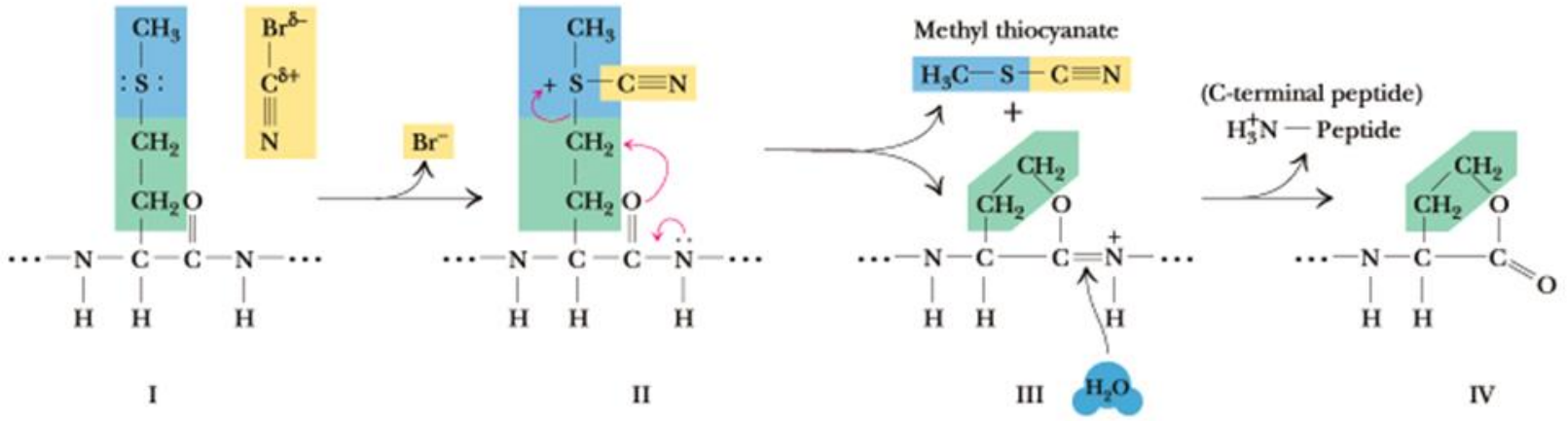
Chymotrypsin

Chymotrypsin is selective for cleaving the peptide bond to the carboxyl group of amino acids with an aromatic side chain.

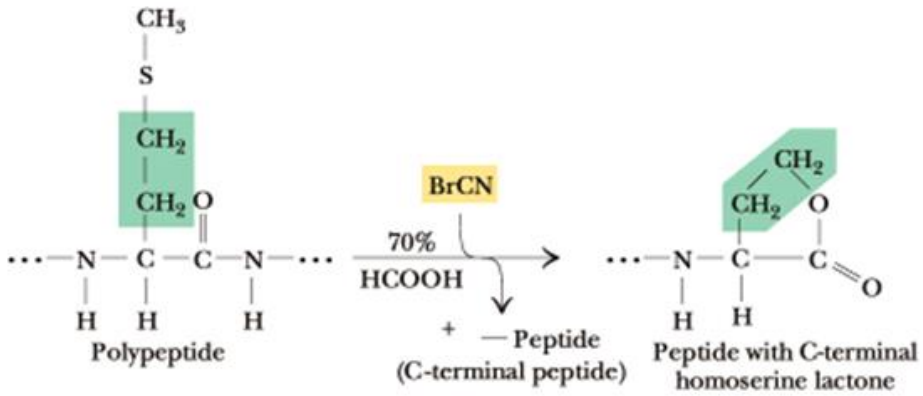


phenylalanine, tyrosine, tryptophan

Chemical fragmentation using cyanogen bromide



OVERALL REACTION:



Saunders College Publishing

End group analysis:

- Amino sequence is ambiguous unless we know whether to read it left-to-right or right-to-left.
- We need to know what the N-terminal and C-terminal amino acids are.
- Several chemical methods have been developed for identifying the N-terminus
- The C-terminal amino acid can be determined by carboxypeptidase-catalyzed hydrolysis.

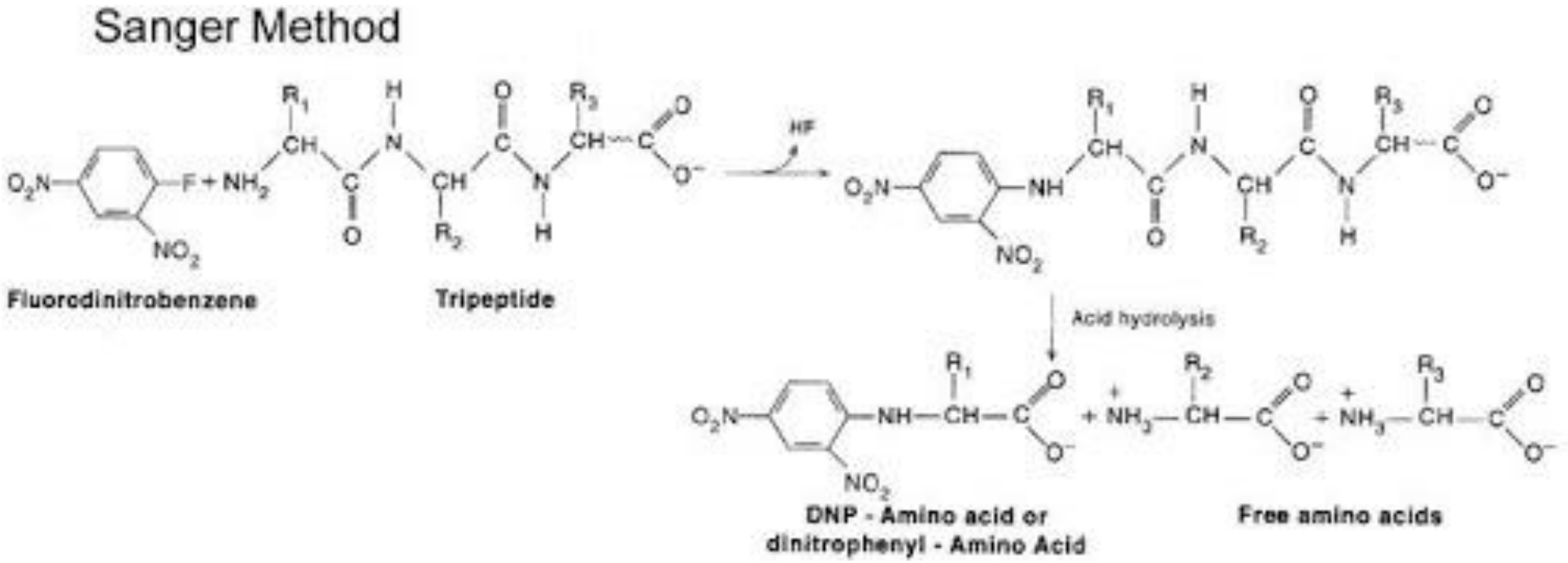
Sanger's Method

Edman degradation

Sanger's Method

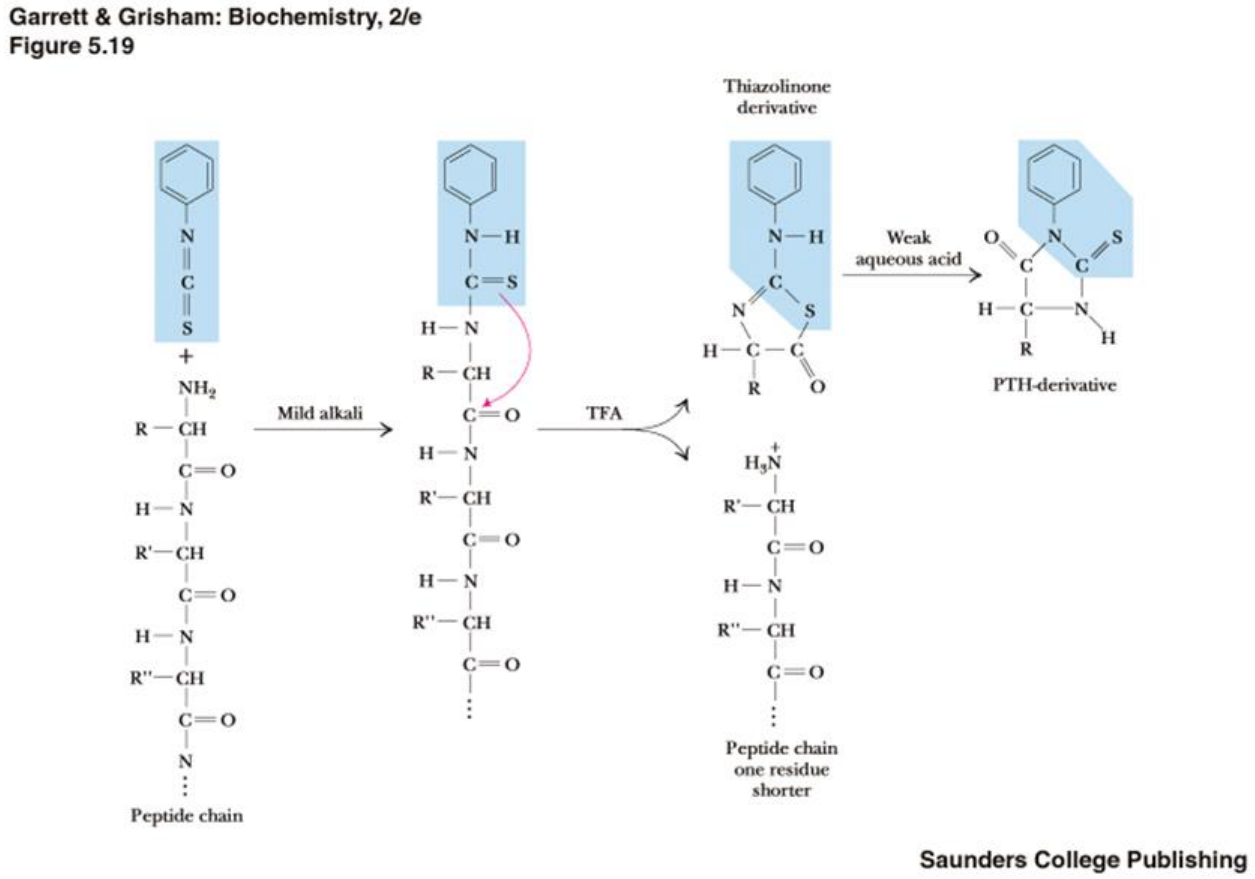
The key reagent in Sanger's method for identifying the N-terminus is 1-fluoro-2,4-dinitrobenzene.

- 1-Fluoro-2,4-dinitrobenzene is very reactive toward nucleophilic aromatic substitution
- 1-Fluoro-2,4-dinitrobenzene reacts with the amino nitrogen of the N-terminal amino acid
- Acid hydrolysis cleaves all of the peptide bonds leaving a mixture of amino acids, only one of which (the N-terminus) bears a 2,4-DNP group.



Edman degradation

- It Can be done sequentially one residue at a time on the same sample.
- Usually one can determine the first 20 or so amino acids from the N-terminus by this method.
- 10^{-10} g of sample is sufficient.
- Has been automated.
- The key reagent in the Edman degradation is phenyl isothiocyanate
- The product is a thiazolone.
- Under the conditions of its formation, the thiazolone rearranges to a phenylthiohydantoin (PTH) derivative.



Edman Degradation

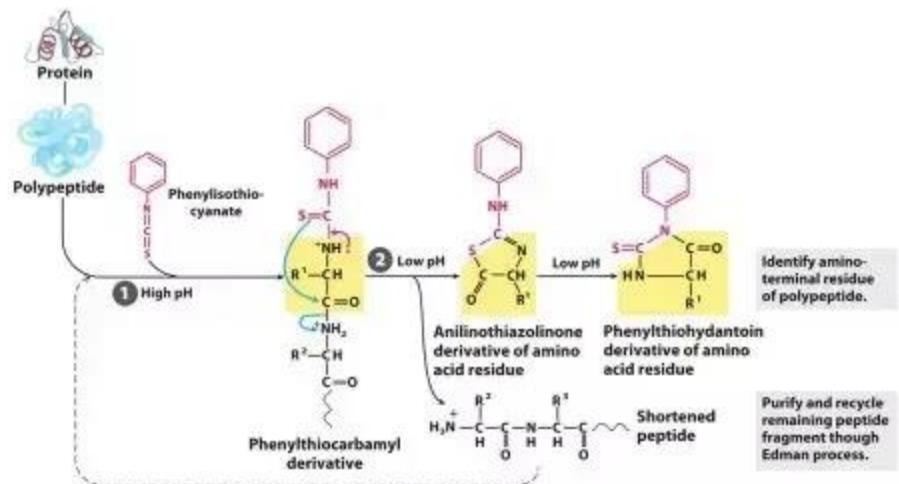


Figure 3-27
Lehninger Principles of Biochemistry, Sixth Edition

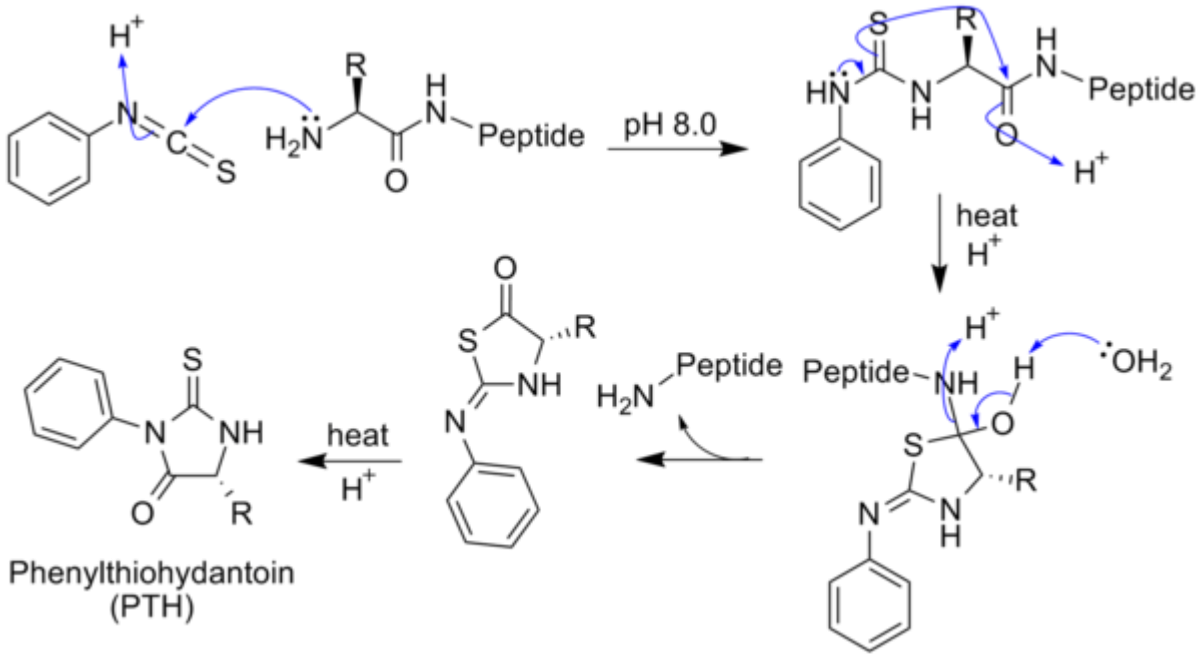
DISADVANTAGE:

•The hydrolysis stage destroys the polypeptide , these procedures cannot be used to sequence a polypeptide beyond its amino-terminal residue .

ADVANTAGE:

•It can help to determine the number of chemically distinct polypeptides in a protein , provided each has a different amino terminal residue.

EDMAN DEGRADATION

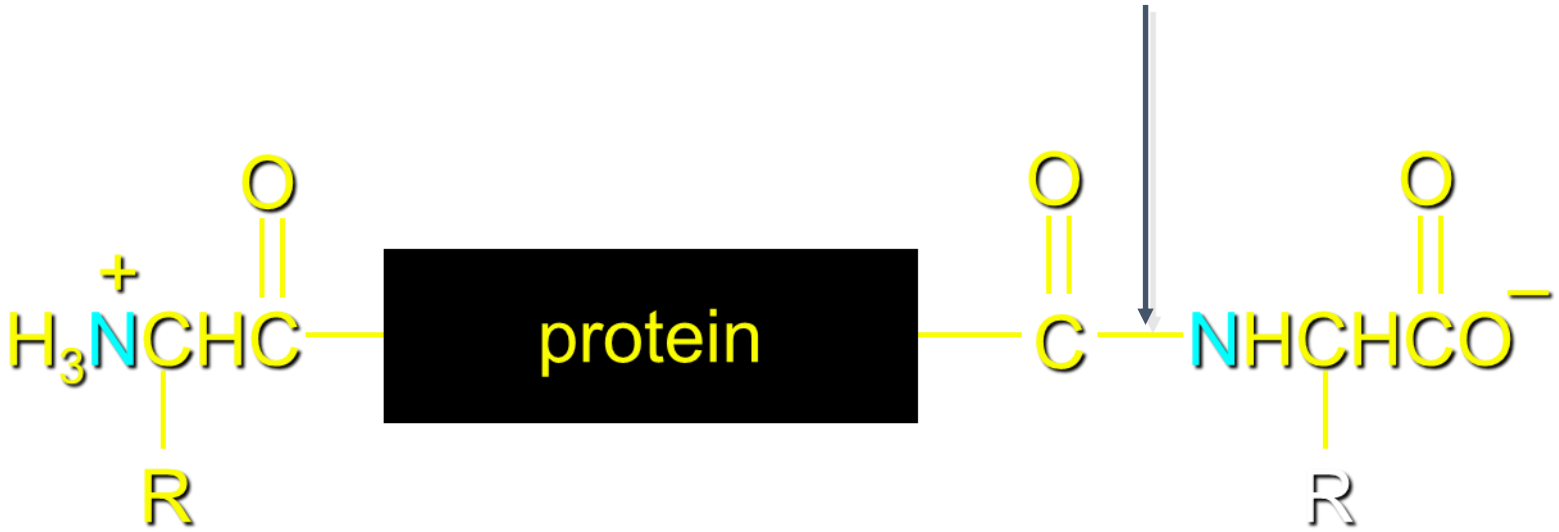


- After removal and identification of the amino terminal residue, the new amino-terminal residue so exposed can be labeled, removed, and identified through the same series of reactions .This procedure is repeated until the entire sequence is determined.
- The Edman degradation is carried out in a machine, called a **sequenator,** that mixes reagents in the proper proportions, separates the products, identifies them, and records the results.
- These methods are extremely sensitive. Often, the complete amino acid sequence can be determined starting with only a few micrograms of protein.

Enzymatic analysis (carboxypeptidase)

- Carboxypeptidase A cleaves any residue except Pro, Arg, and Lys
- Carboxypeptidase B (hog pancreas) only works on Arg and Lys
- Carboxypeptidase C, Y any residue

Carboxypeptidase is selective for cleaving the peptide bond to the C-terminal amino acid.



Hydrolysis (6M HCl, 2M TFA)

Reconstructing the Sequence

- Use two or more fragmentation agents in separate fragmentation experiments
- Sequence all the peptides produced (usually by Edman degradation)
- Compare and align overlapping peptide sequences to learn the sequence of the original polypeptide chain

Overlapping technique……??

Angiotensin II

Amino acid analysis gives Arg, Asp, His, Ile, Phe, Pro, Tyr, Val.

Partial hydrolysis gives four fragments, which are sequenced by several cycles of Edman degradation as:

Asp-Arg-Val
Ile-His-Pro
Pro-Phe
Arg-Val-Tyr
Val-Tyr-Ile

The structure of Angiotensin II is then pieced together by looking for overlaps in these sequences. Fof example, we notice that the Arg-Val sequence occurs in two fragments, so we line those two fragments up with the Arg-Val pair aligned

Asp-Arg-Val
Arg-Val-Tyr

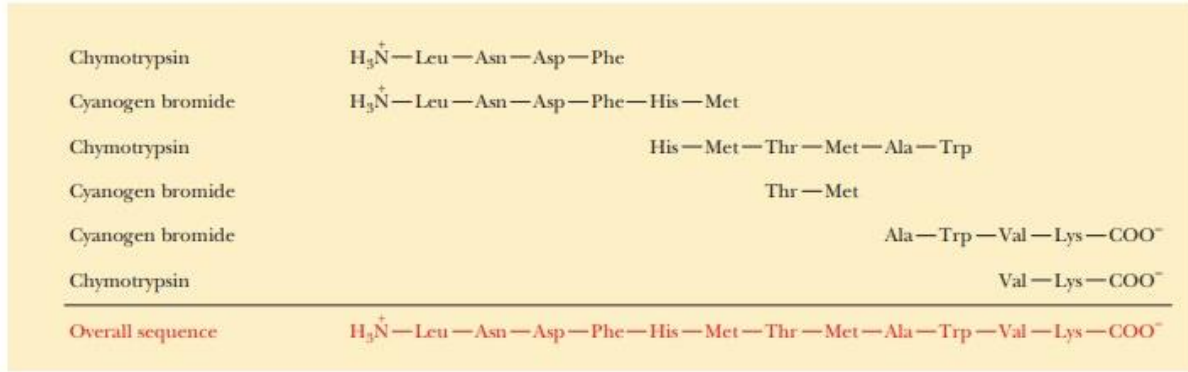
Similarly, the Val-Tyr pair occurs twice, so we add it to the alignment.

Asp-Arg-Val
Arg-Val-Tyr
Val-Tyr-Ile

If we continue to look for these "overlaps," we can assemble the whole sequence.

Asp-Arg-Val
Arg-Val-Tyr
Val-Tyr-Ile
Ile-His-Pro
Pro-Phe

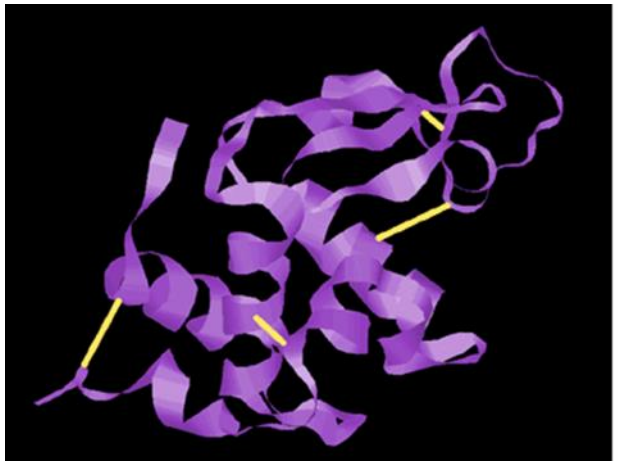
Asp-Arg-Val-Tyr-Ile-His-Pro-Phe

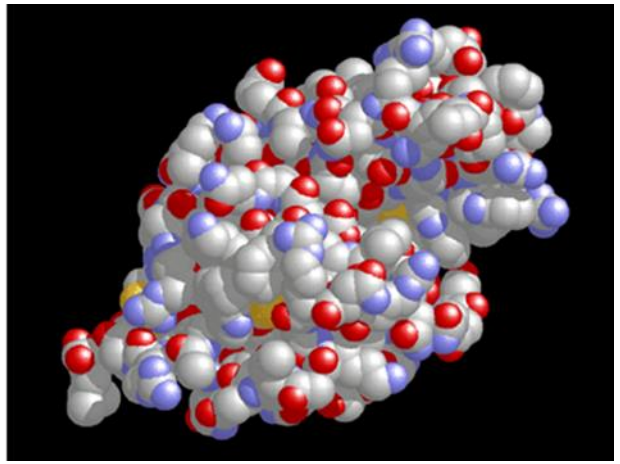


■ **FIGURE 5.19** **Use of overlapping sequences to determine protein sequence.** Partial digestion was effected using chymotrypsin and cyanogen bromide. For clarity, only the original N-terminus and C-terminus of the complete peptide are shown.

High orders of Protein structure

A functional protein is not just a polypeptide chain, but one or more polypeptides precisely twisted, folded and coiled into a molecule of unique shape (conformation). This conformation is essential for some protein function e.g. Enables a protein to recognize and bind specifically to another molecule e.g. hormone/receptor; enzyme/substrate and antibody/antigen.





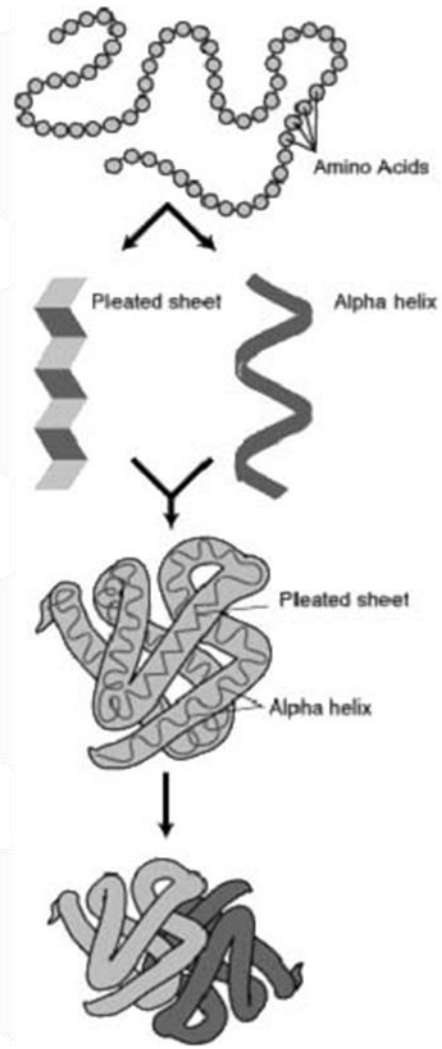
STRUCTURE

Primary

Secondary

Tertiary

Quaternary



Assembly

Folding

Packing

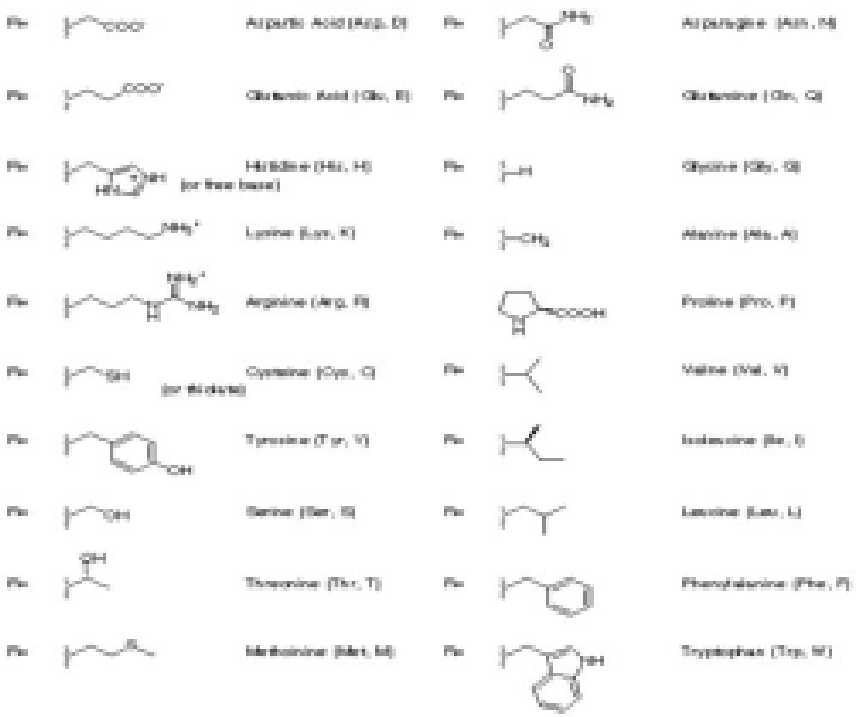
Interaction

Primary Structure

primary structure of human insulin

CHAIN 1: `GIVEQ CCTSI CSLYQ LENYC N`

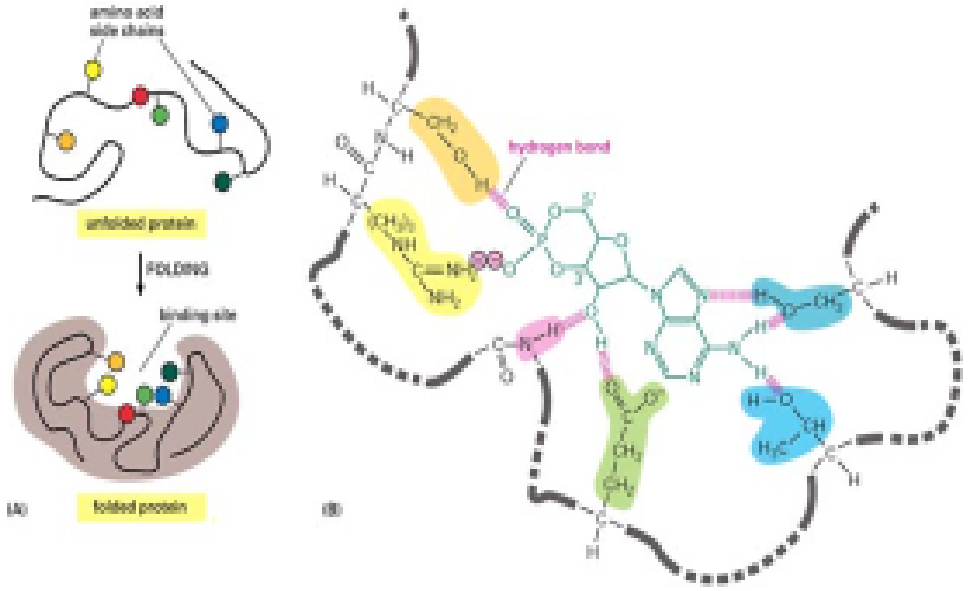
CHAIN 2: `FVNQH LCGSH LVEAL YLVCG ERGFF YTPKT`

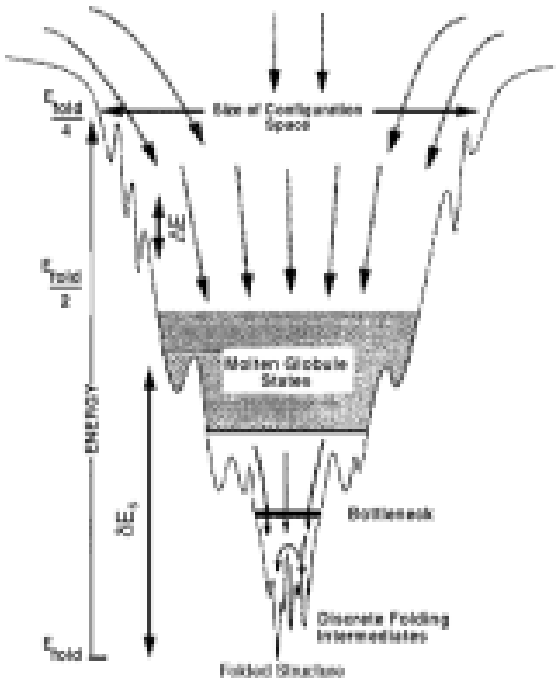


- linear
- ordered
- 1 dimensional
- sequence of amino acid polymer
- by convention, written from amino end to carboxyl end
- a perfectly linear amino acid polymer is neither functional nor energetically favorable → folding!

Protein Folding

- occurs in the cytosol
- involves localized spatial interaction among primary structure elements, i.e. the amino acids
- may or may not involve chaperone proteins
- tumbles towards conformations that reduce ΔE (this process is thermo-dynamically favorable)
- yields secondary structure



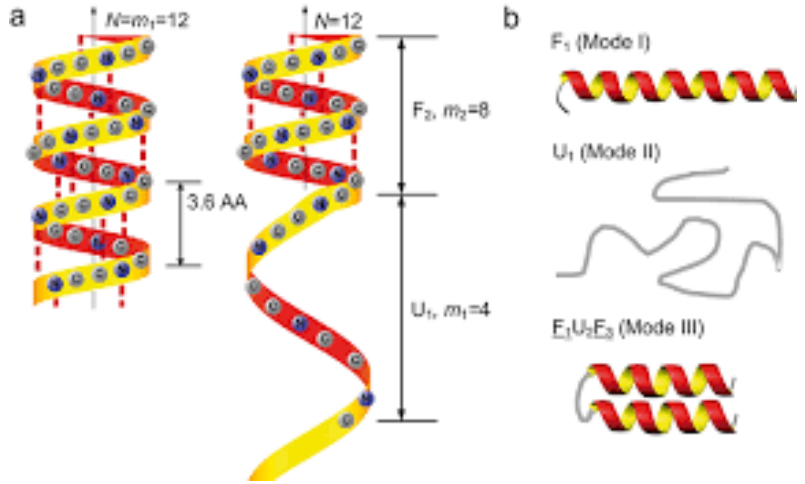


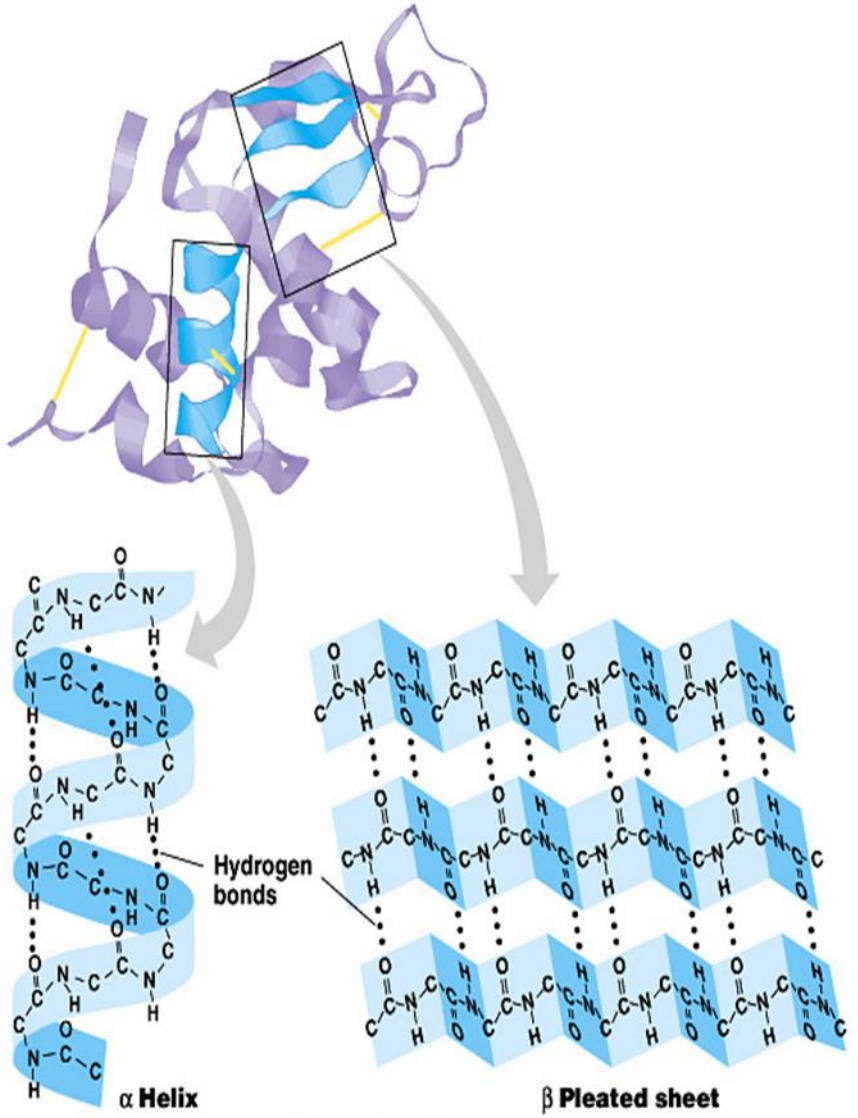
Secondary structure

- It results from hydrogen bond formation between hydrogen of –NH group of peptide bond and the carbonyl oxygen of another peptide bond. According to H-bonding there are two main forms of secondary structure:

α-helix: It is a spiral structure resulting from hydrogen bonding between one peptide bond and the fourth one

β-sheets: is another form of secondary structure in which two or more polypeptides (or segments of the same peptide chain) are linked together by hydrogen bond between H- of NH- of one chain and carbonyl oxygen of adjacent chain (or segment).

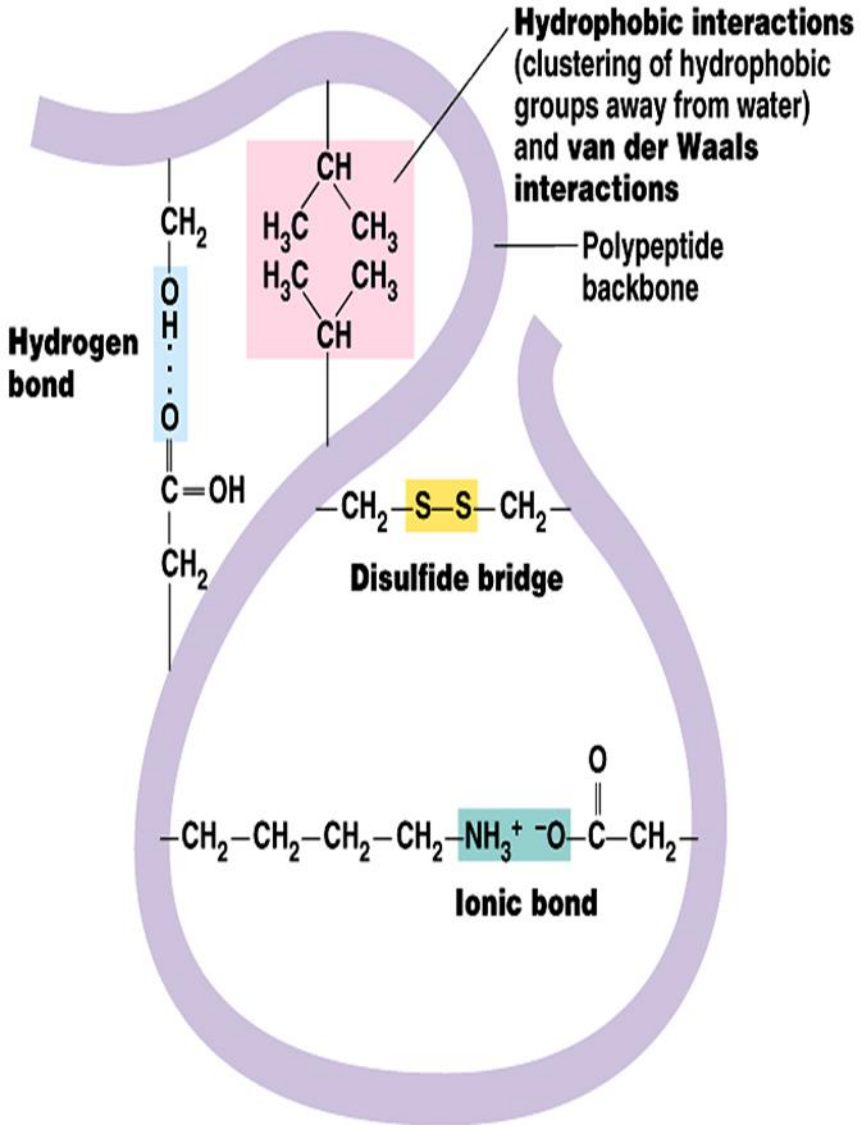




Tertiary Structure

- It is non-linear, 3 dimensional
- global but restricted to the amino acid polymer
- It is formed and stabilized by hydrogen bonding, covalent (e.g. disulfide) bonding, hydrophobic packing toward core and hydrophilic exposure to solvent
- A globular amino acid polymer folded and compacted is somewhat functional (catalytic) and energetically favorable

Tertiary structure is determined by a variety of interactions (bond formation) among R groups and between R groups and the polypeptide backbone.

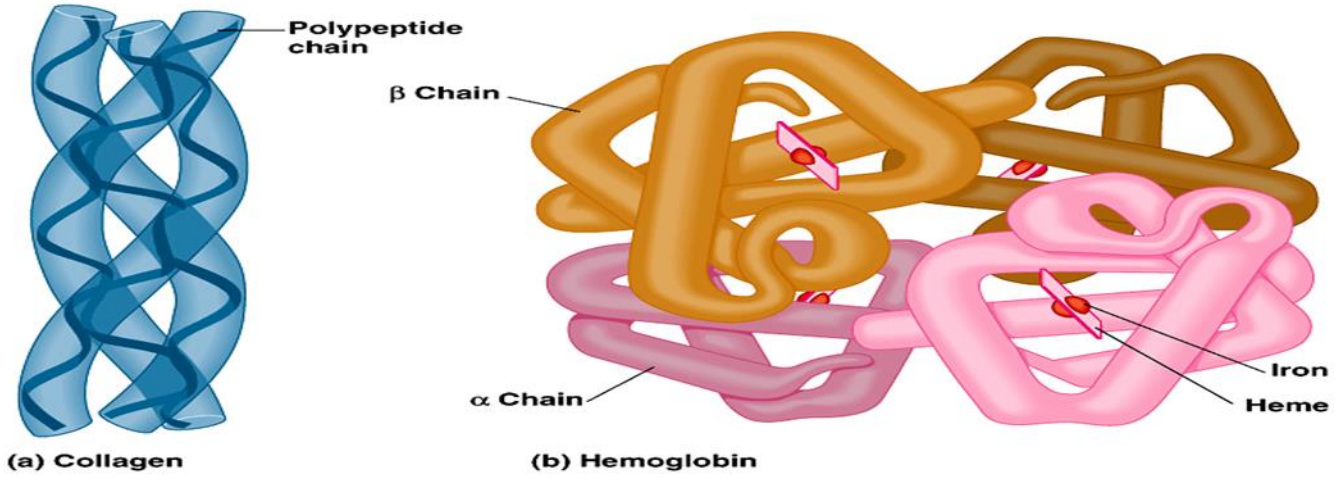


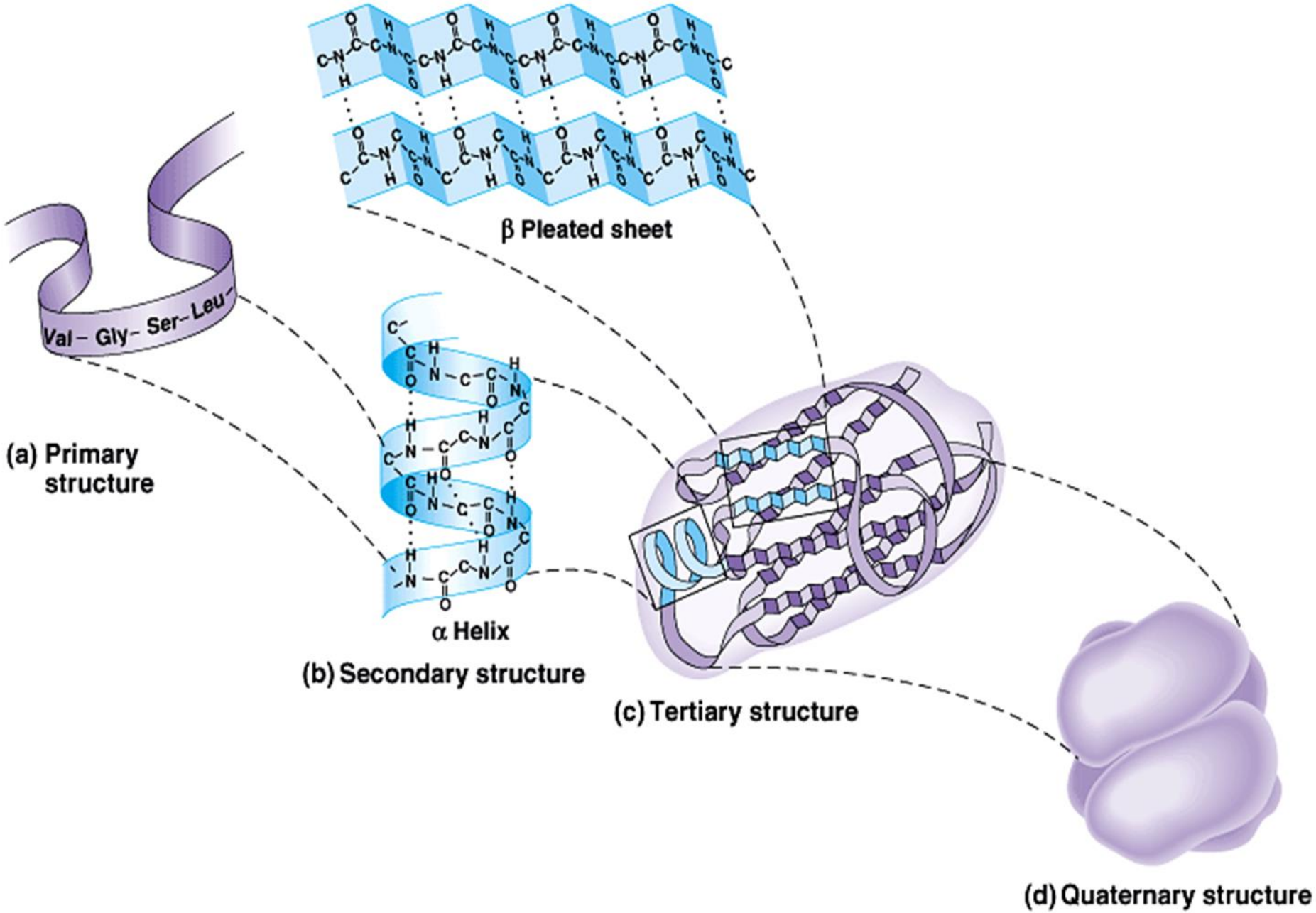
Quaternary structure

- It results from the aggregation (combination) of two or more polypeptide subunits held together by non-covalent interaction like H-bonds, ionic or hydrophobic interactions.
- Examples on protein having quaternary structure:
- Collagen is a fibrous protein of three polypeptides (trimeric) that are supercoiled like a rope.
- This provides the structural strength for their role in connective tissue.

Hemoglobin is a globular protein with four polypeptide chains (t

Insulin : two polypeptide chains (dimeric)



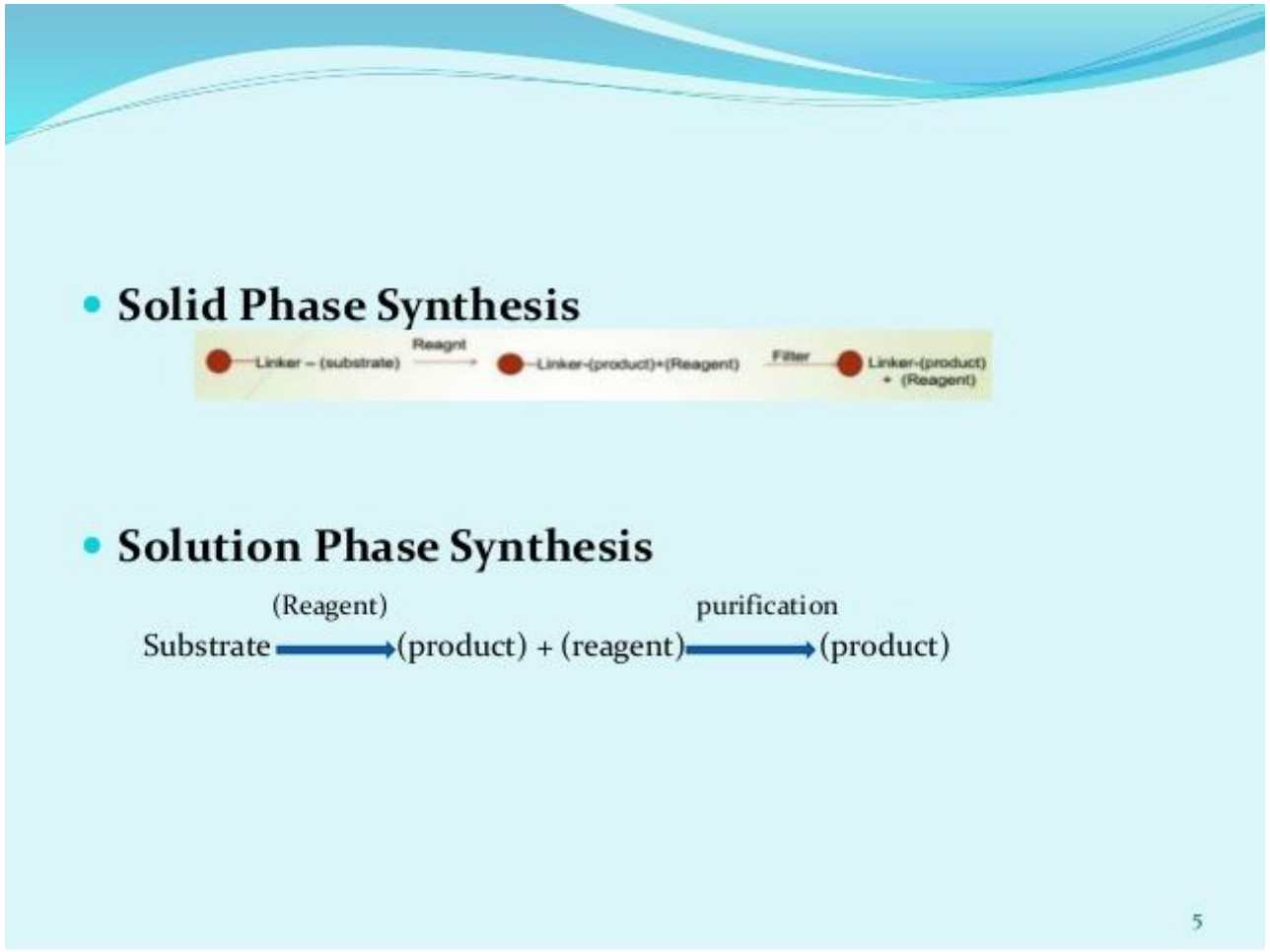

β Pleated sheet
Val – Gly– Ser–Leu–
(a) Primary structure
α Helix
(b) Secondary structure
(c) Tertiary structure
(d) Quaternary structure

INTRODUCTION

- Peptides are the sequence of amino acids which are formed by the condensation of two amino acids .
- In which a peptide bond is formed between a carbonyl group of one amino acid and the amino group of another amino acid .
- Peptides are synthesized in two ways they are

  1.solid phase synthesis

  2.solution phase synthesis

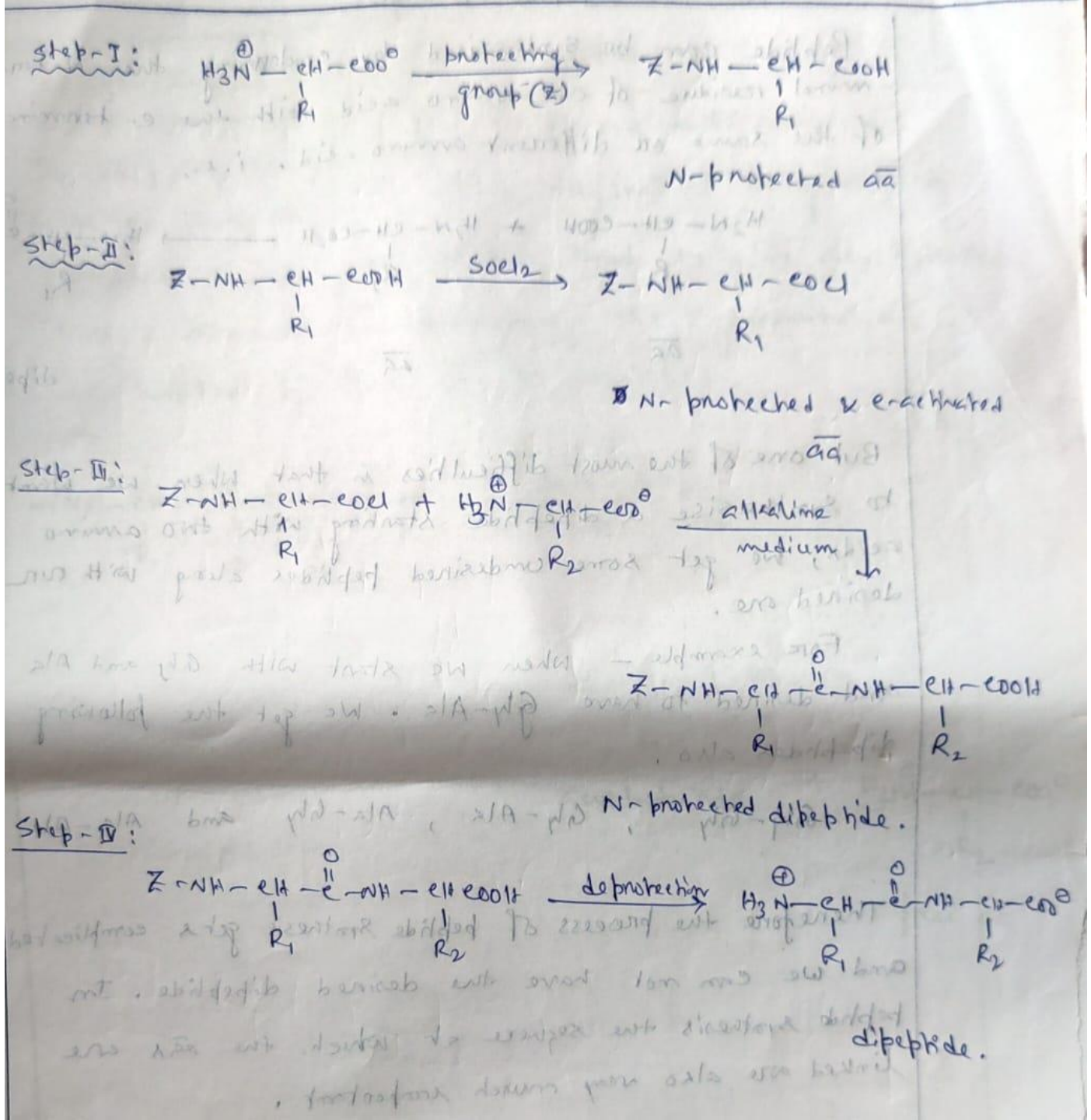


Difficulties in conventional peptide synthesis

- Making peptide bonds between amino acids is not difficult.
- The challenge is connecting amino acids in the correct sequence.
- Random peptide bond formation in a mixture of phenylalanine and glycine, for example, will give four dipeptides.

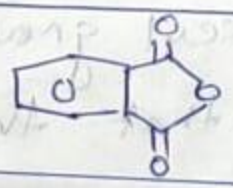
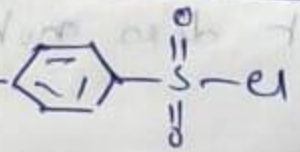
Phe—PheGly—Gly Phe—Gly Gly—Phe

Steps in Solution phase peptide synthesis:



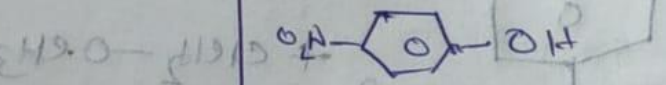
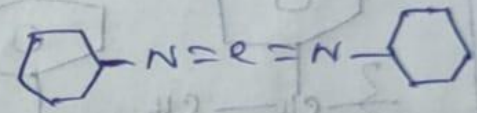
Protection & Deprotection of N-terminal :-

The N-terminal of an aa can be protected using many compd. They are -

Terminal	Protecting group	Str	Deprotection agent.
N-terminal	i) Benzyl chloro formate. (cbz)	$PhCH_2O-\overset{\overset{O}{\|\|}}{C}-Cl$	a) H_2/Pd on C, (b) HBr in cold AcOH
	ii) t-Butoxy carbonyl group (Boc)	$\left((CH_3)_3C-O-\overset{\overset{O}{\|\|}}{C}\right)_2O$	a) CF_3COOH on C, (b) HCl in HOAc
	iii) Pthalic anhydride		Hydrazine (H_2N-NH_2)
	iv) ~~Tosyl~~ Tosyl chloride (TsCl)		Na / liq. NH_3

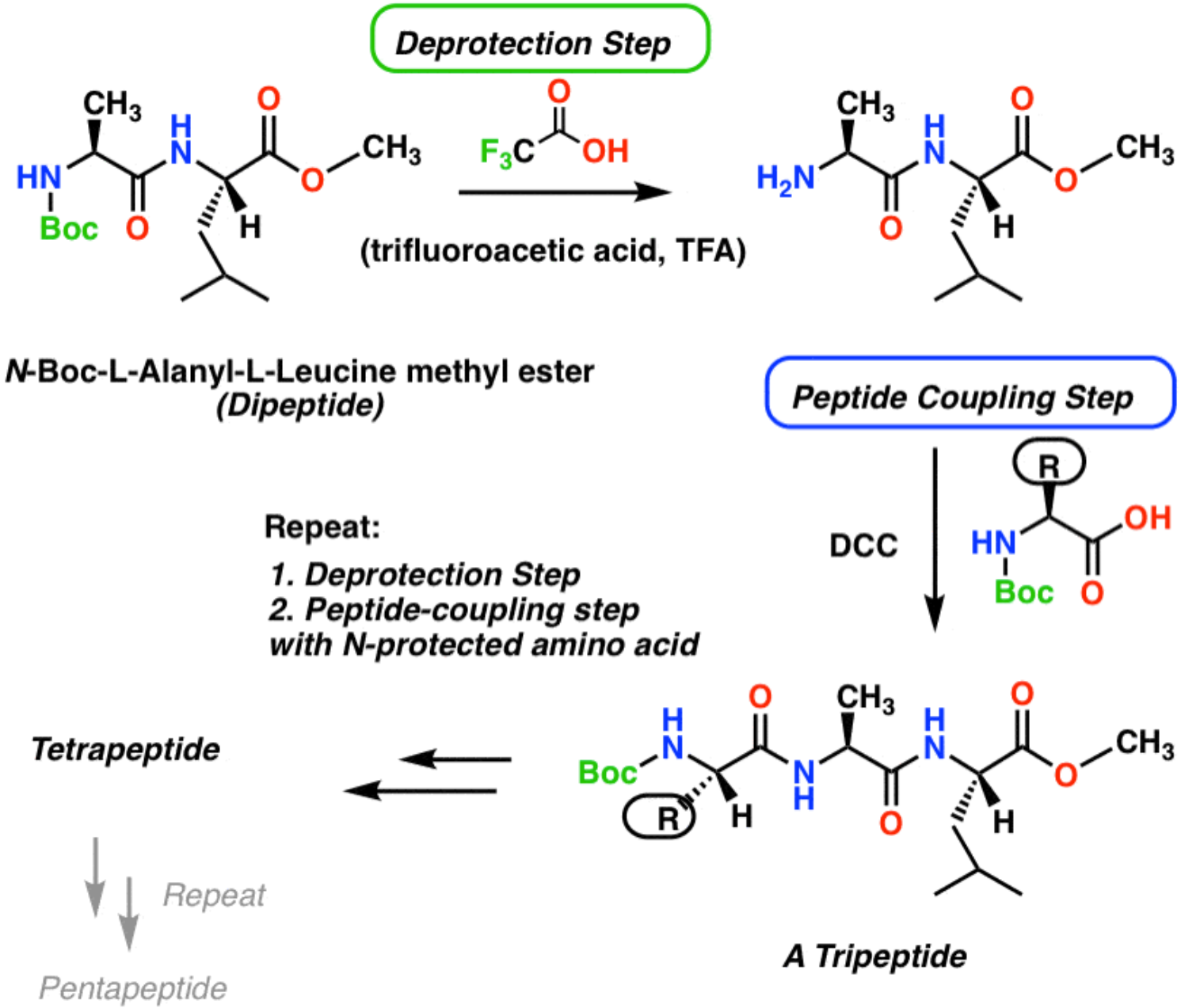
Activation of acid group (c-terminal) :-

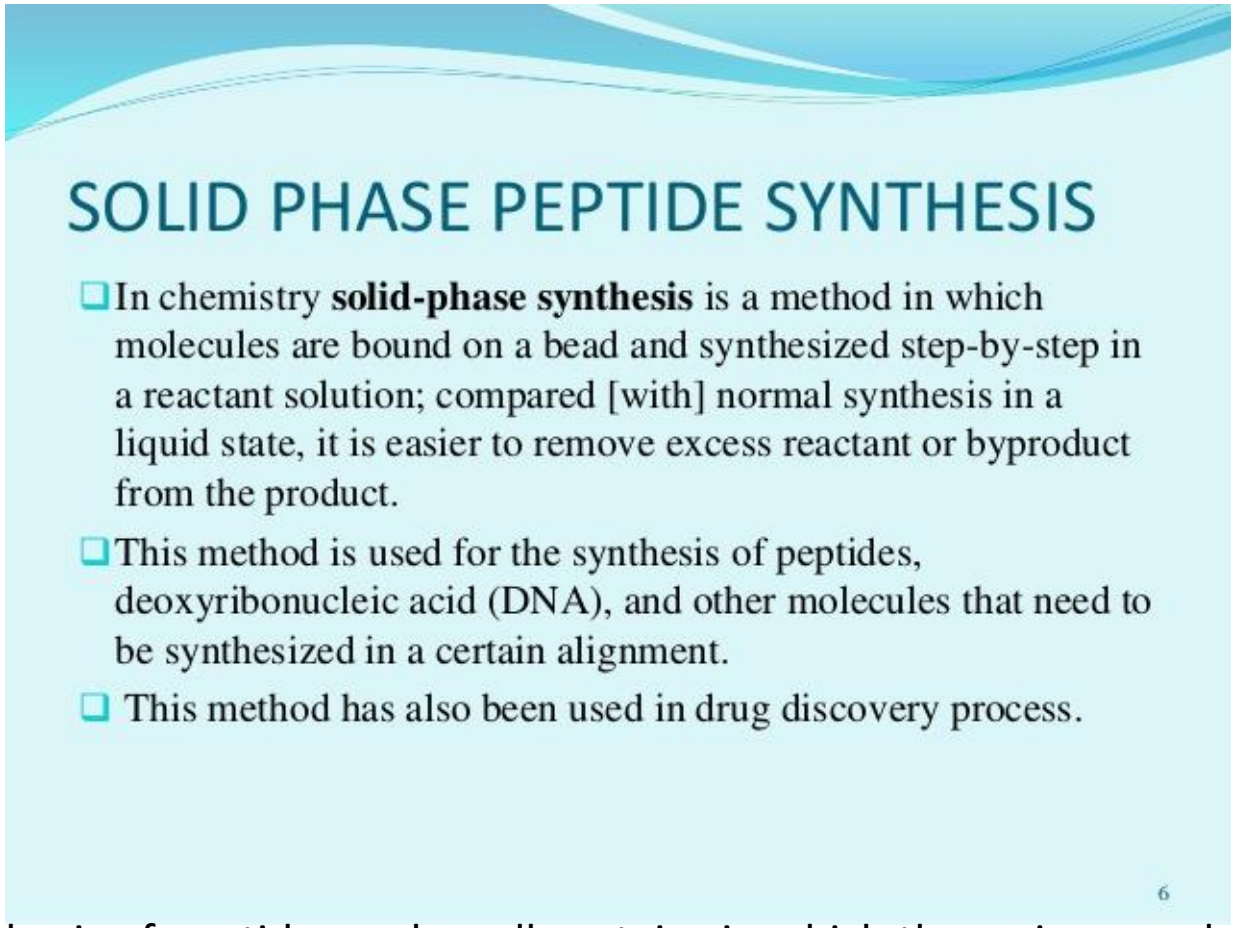
The c-terminal of an aa can be activated by converting it into some other groups having a better leaving group than $-OH^{\ominus}$.

Terminal	-COOH group to be converted to-	Reagent	Auxilary reagent.
c-terminal.	i) acyl chloride	$SOCl_2$	—
	ii) anhydride.	Ethyl chloroformate ($Cl-\overset{\overset{O}{\|\|}}{C}-OEt$)	Et_3N
	iii) p-nitrophenyl ester	p-nitrophenol 	DCC 
	iv) acyl azide	Hydrazine. (H_2N-NH_2)	i) CH_3OH, ii) HNO_2

To Tripeptides... and Beyond !

Example:



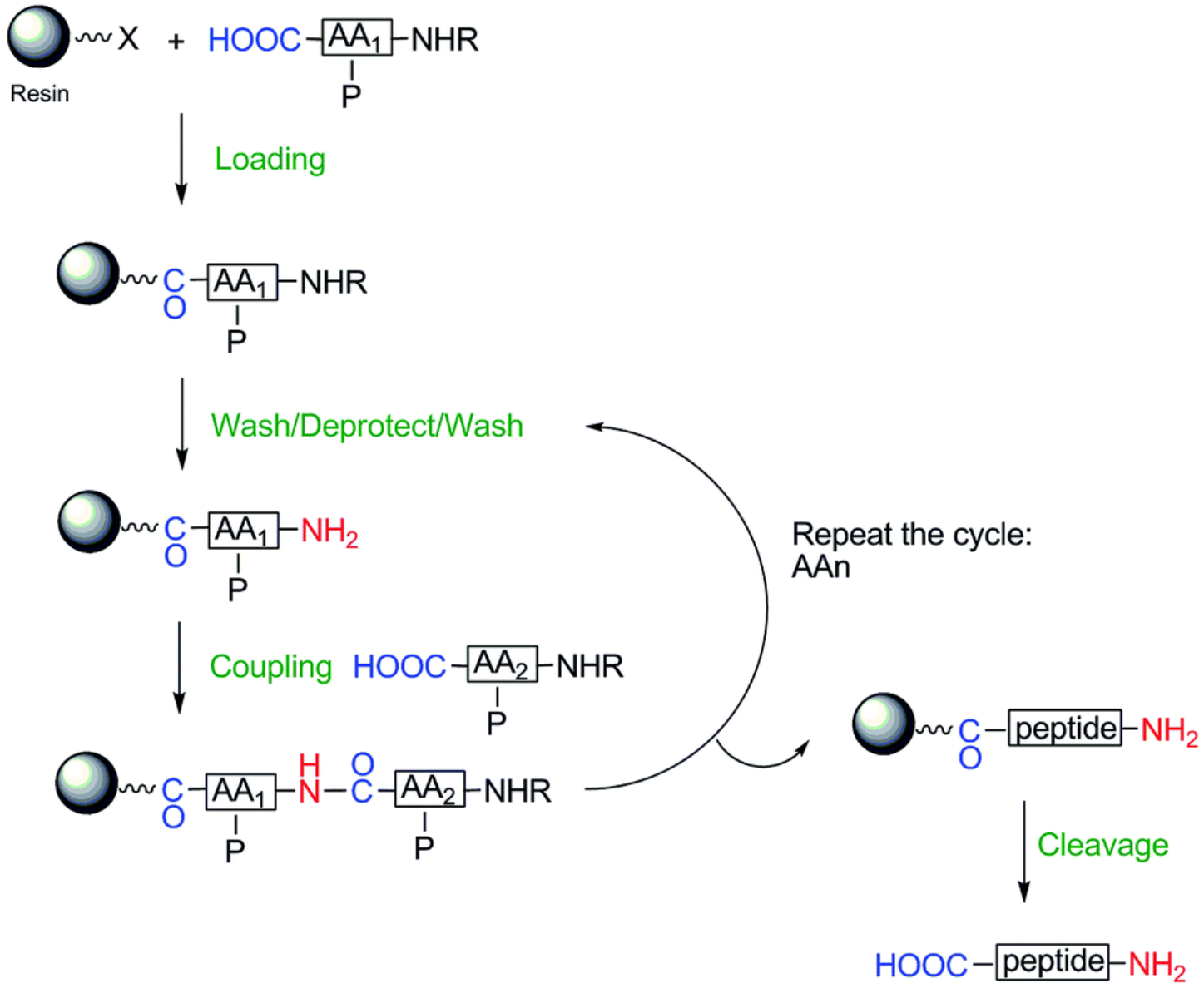


The synthesis of peptides and small proteins in which the resinous polymer supported amino acid and succeeding peptide repeatedly reacts with N-protected amino acids followed by deprotection until the desired peptide or protein is assembled is generally referred to as the Merrifield solid phase peptide synthesis

Merrifield synthesized a nonapeptide (bradykinin) in 1962 in 8 days in 68% yield. He synthesized ribonuclease (124 amino acids) in 1969 using 369 reactions; 11,391 steps
He got Nobel Prize in chemistry: 1984

Solid phase peptide synthesis:

In this peptide synthesis, an amino-protected amino acid is bound to a solid phase material or resin (most commonly, low cross-linked polystyrene beads), forming a covalent bond between the carbonyl group and the resin, most often an amido or an ester bond.



Advantages of Solid phase peptide synthesis:

General advantages of solid phase synthesis are easy purification, rapid generation of linear peptide intermediates, and precedent in the synthesis of large peptides. Given the ease with which peptide linear precursors are produced on solid phase, we opted to employ this approach in addition to a solution phase method.