

Proteins 2: Sequencing and Secondary Structure

Shoba Ranganathan

Applied Biosciences

T: 02 9850 6262; E: shoba.ranganathan@mq.edu.au

Objectives

- How to determine the order of amino acids in a protein (i.e. sequence or primary structure)
- Where is protein sequence information deposited and shared?
- What are domains?
- Common local (secondary) structures in proteins.

Protein Sequencing Workflow

Current sequencing combines:

1. Sanger's sequencing approach for bovine insulin (1st Nobel Prize) and
2. Mass spectrometry for finding the masses of fragment peptides
 - Masses are then identified with specific sequences.

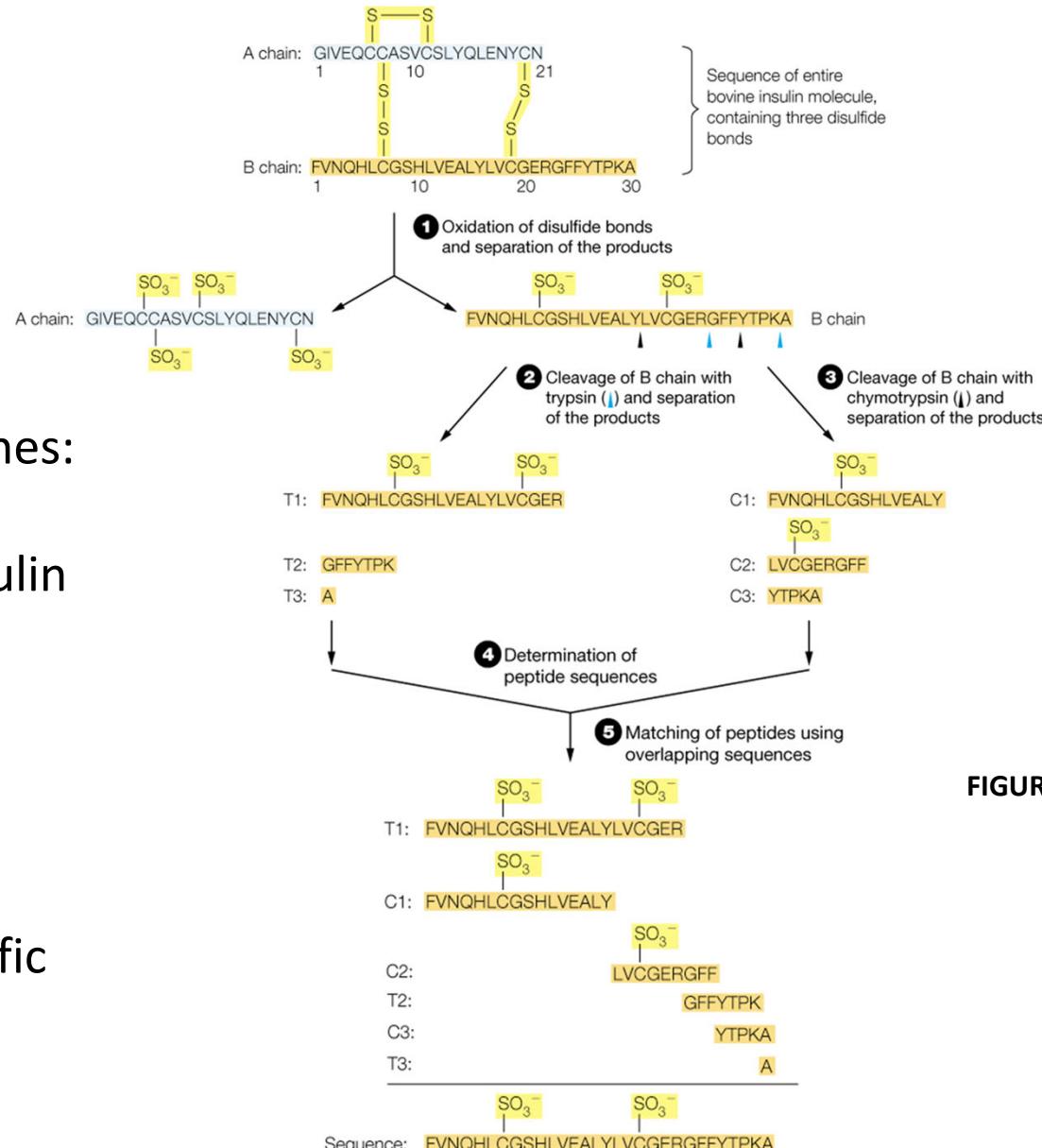
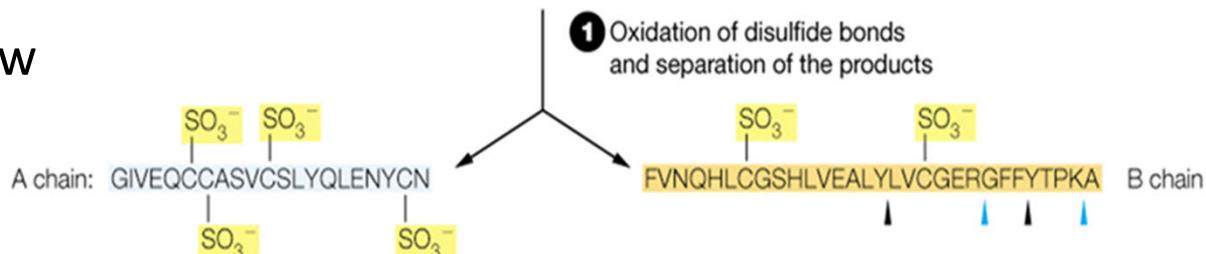
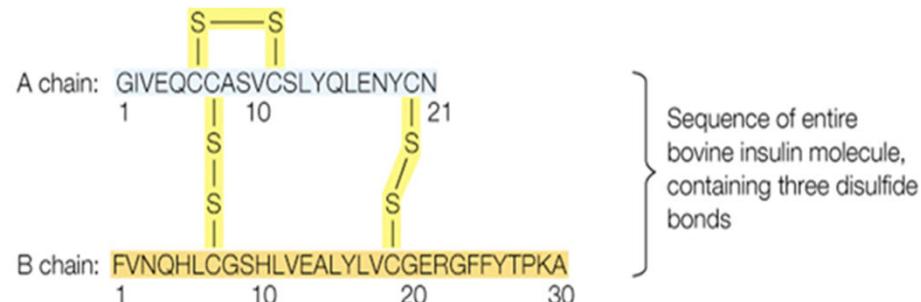
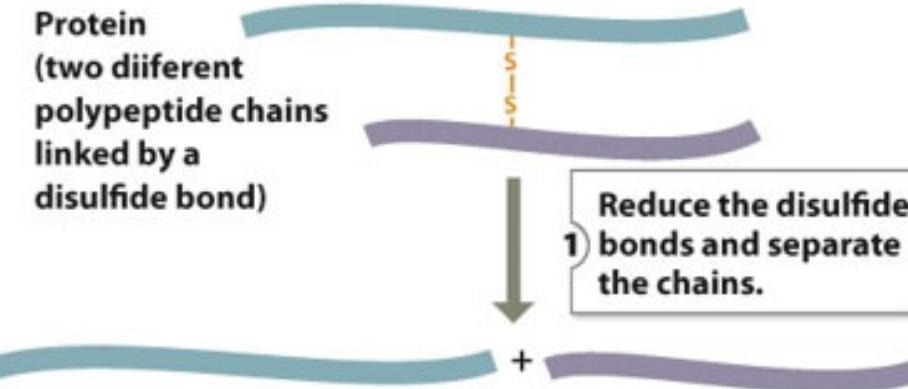


FIGURE 5B.4 Sequencing the

Step 1!

Insulin is made up of two protein chains held together by disulphide bonds

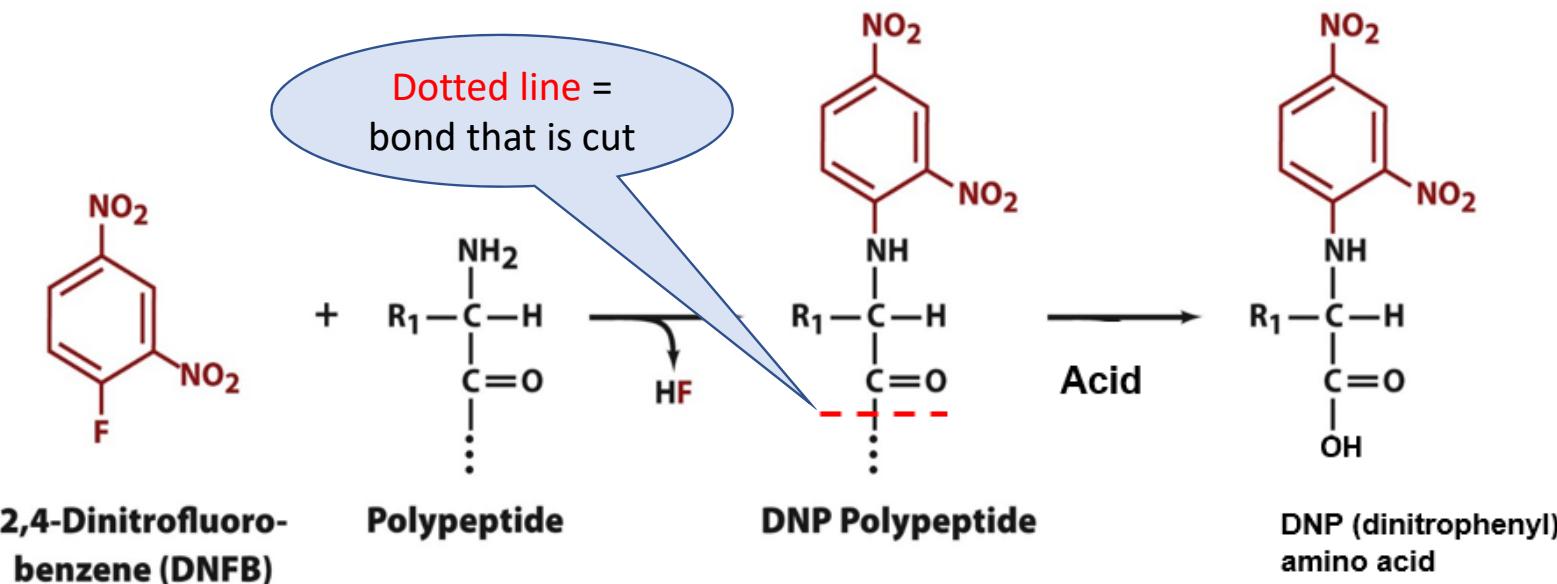
1. Locate the disulphide bonds and
2. Separate the chains by oxidising the disulphide bonds.
3. But we need to know how many chains there are!



How many protein chains are there?

1a. Amino terminal identification:

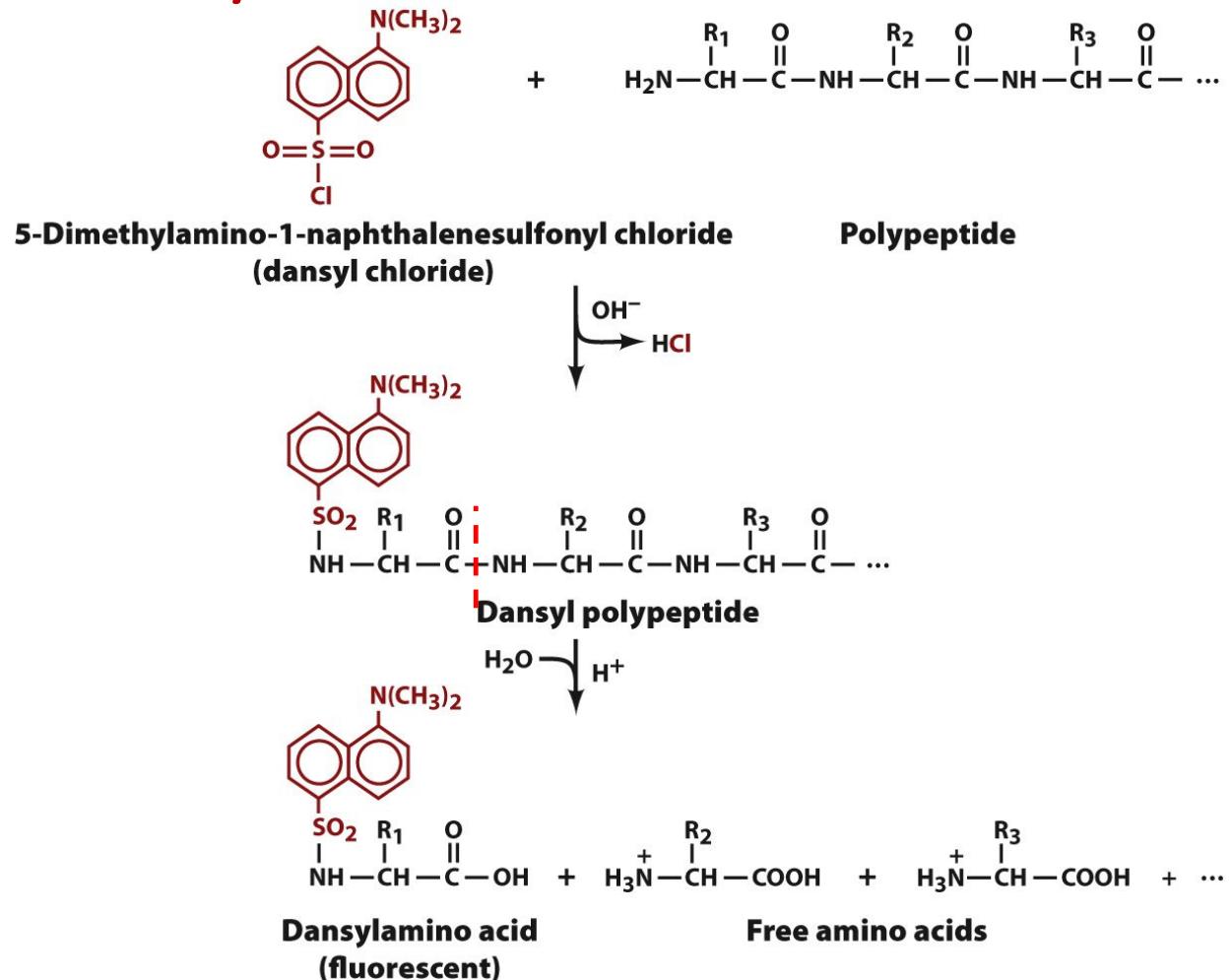
Dinitrofluorobenzene or Sanger method



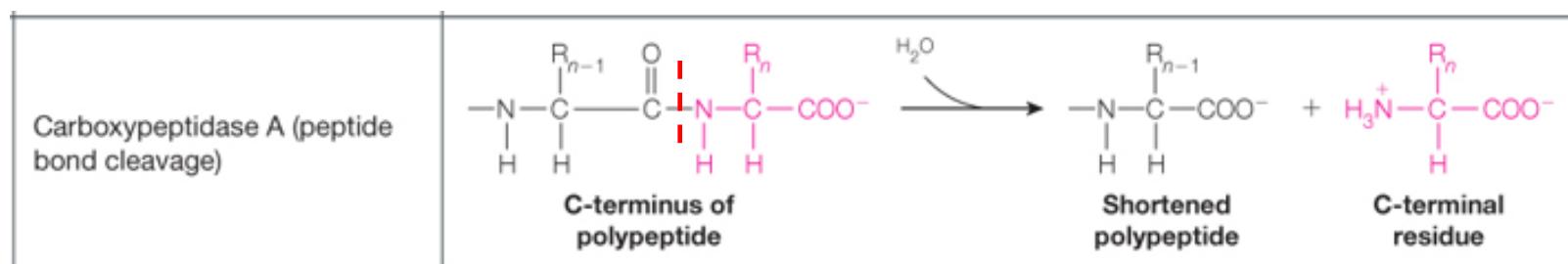
- Took 10 years to sequence insulin - Nobel Prize in 1958
- Currently sequencers take much less time
- Similar method used for sequencing DNA - second Nobel Prize in 1980



1b. Amino terminal identification: Dansyl Chloride Method



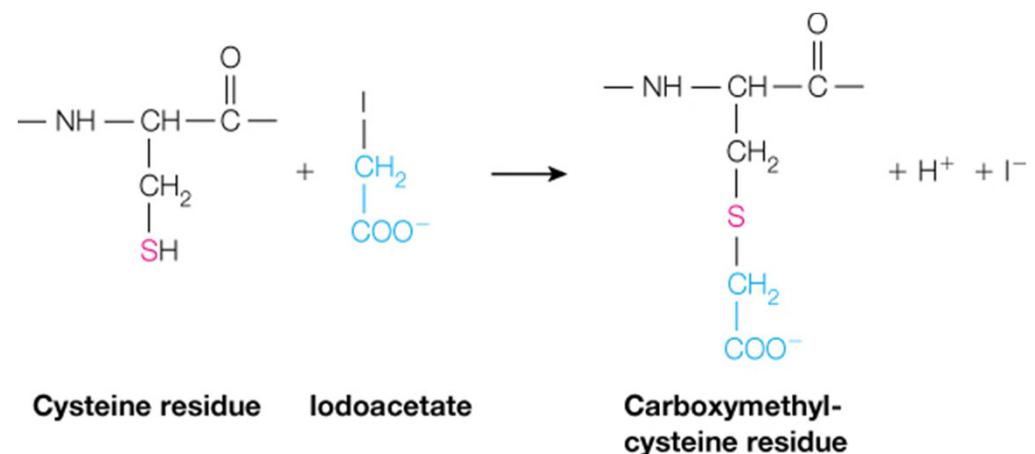
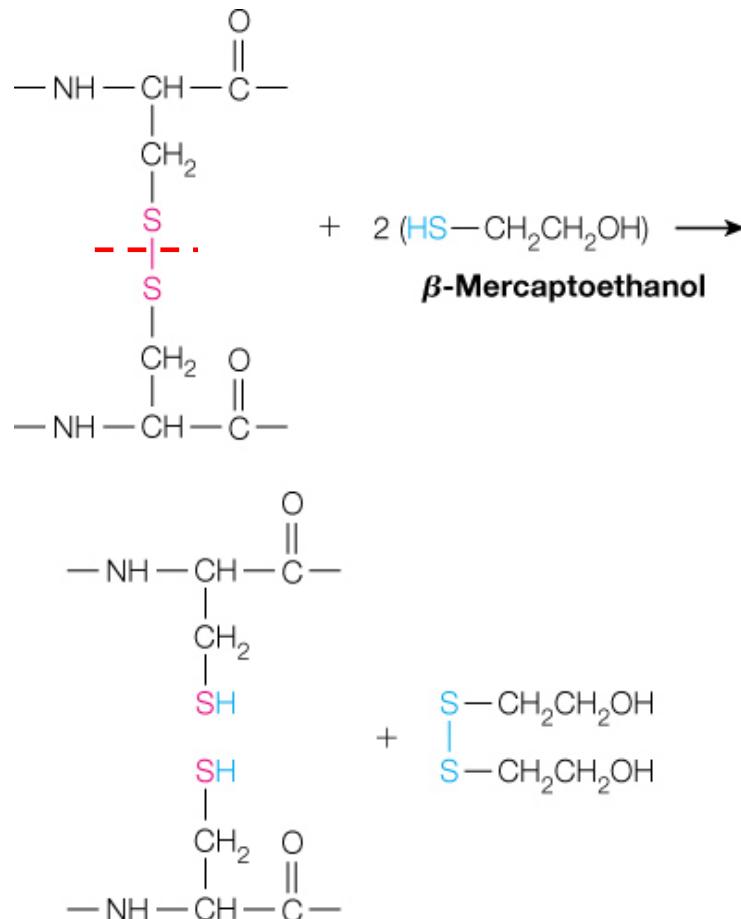
2. Carboxyl terminal identification: Carboxypeptidase



Extracted from Table 8.1

Once we know how many different N- and C- terminal residues there, we can work out if we have more than one protein chain.

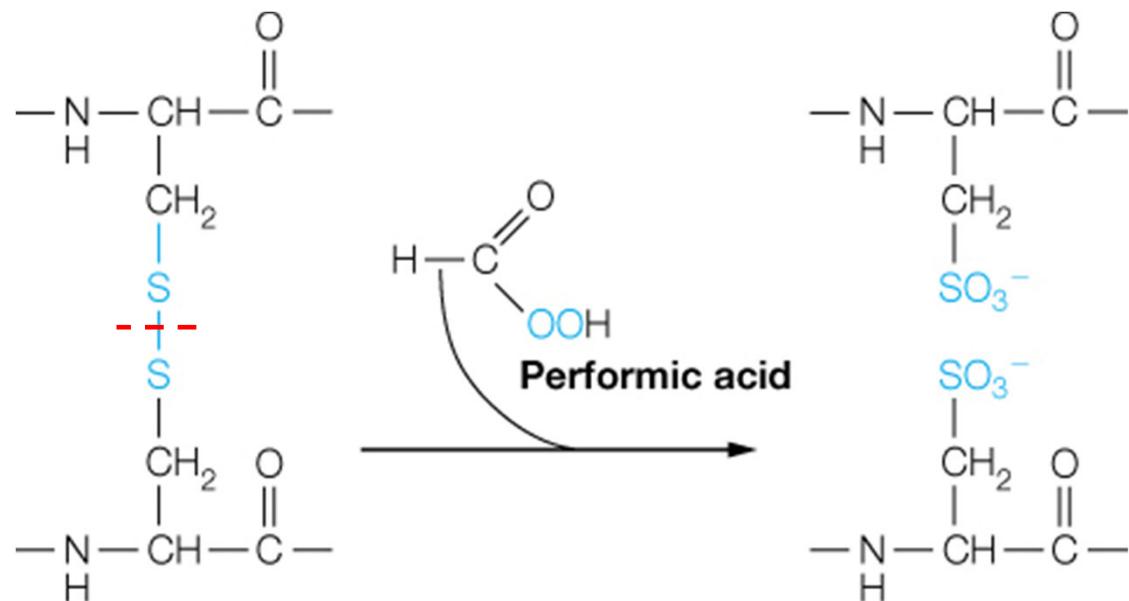
3a. Breaking Disulfide Bonds – 2 steps



Chapter 5, Unnumbered Figures 2 & 3, Page 173

1. Reversible and mild disulphide bond cleavage by **β -mercaptoethanol**
2. The free –SH ends are protected (or blocked) by **iodoacetate** treatment to prevent the free cysteine side chains from forming the same or other disulphide bonds.

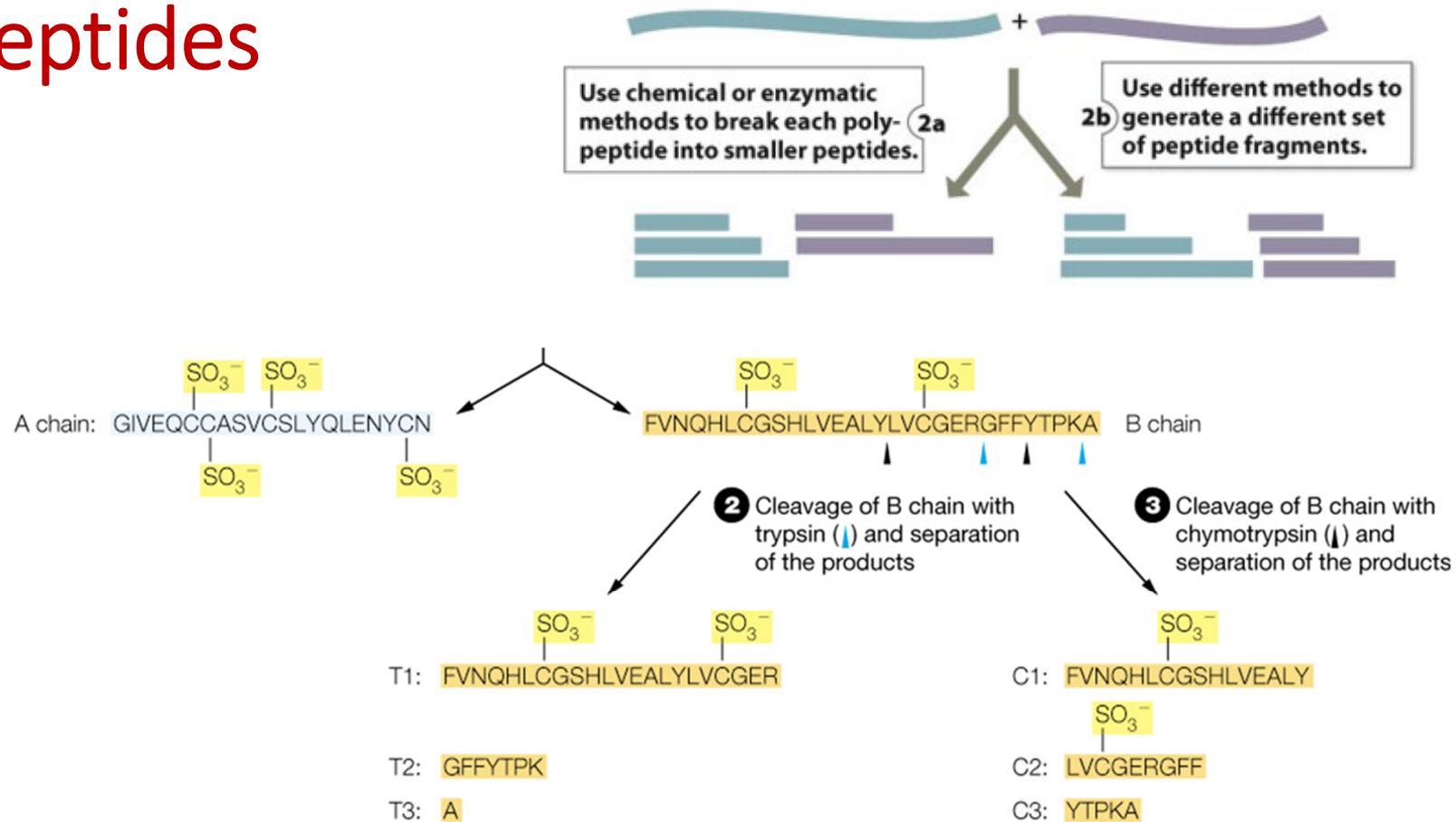
3b. Breaking Disulfide Bonds – 1 step



Chapter 5, Unnumbered Figure 1, Page 173

Performic acid provides irreversible oxidation of disulphide bonds.

Step 2: cutting up the separated protein chains to peptides



Proteins to Peptides - 1: Cyanogen bromide (Met-specific)

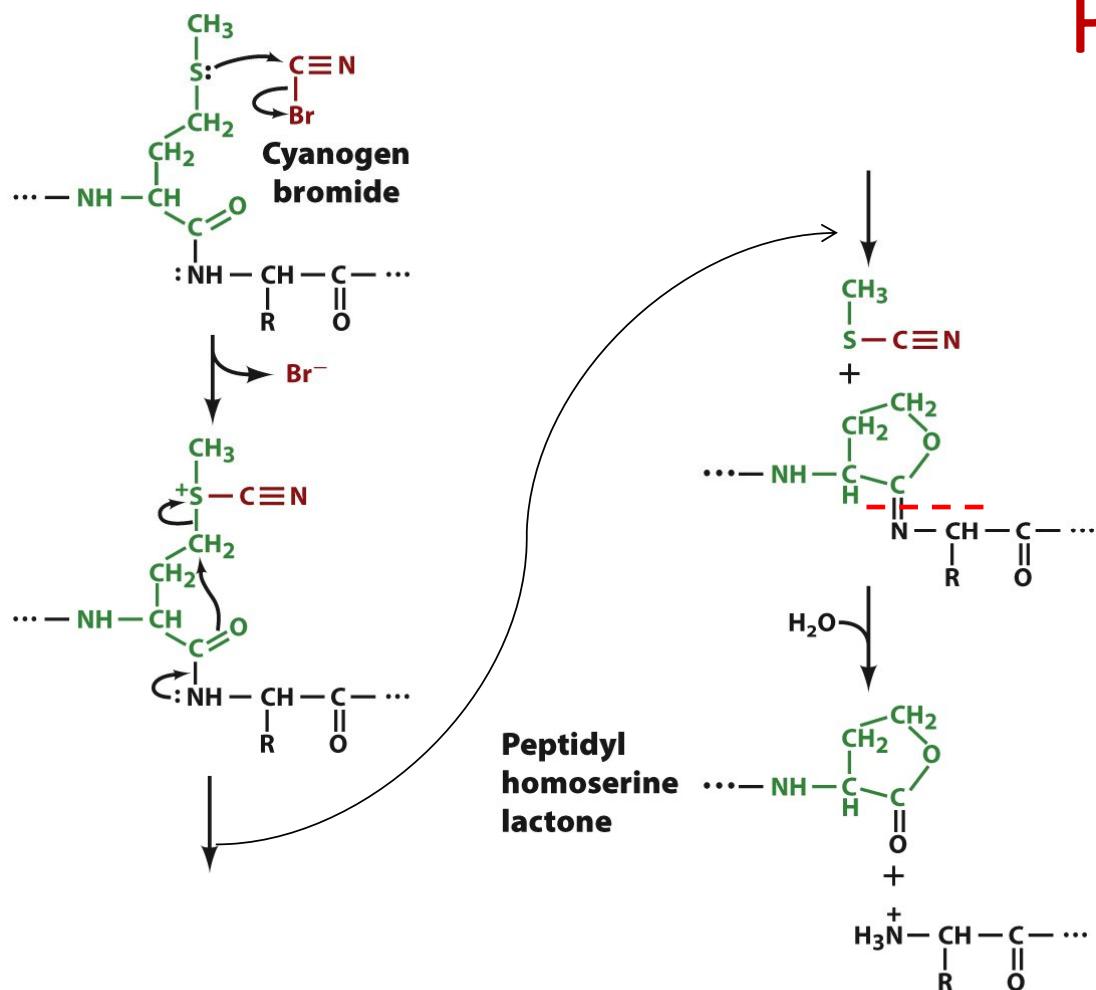
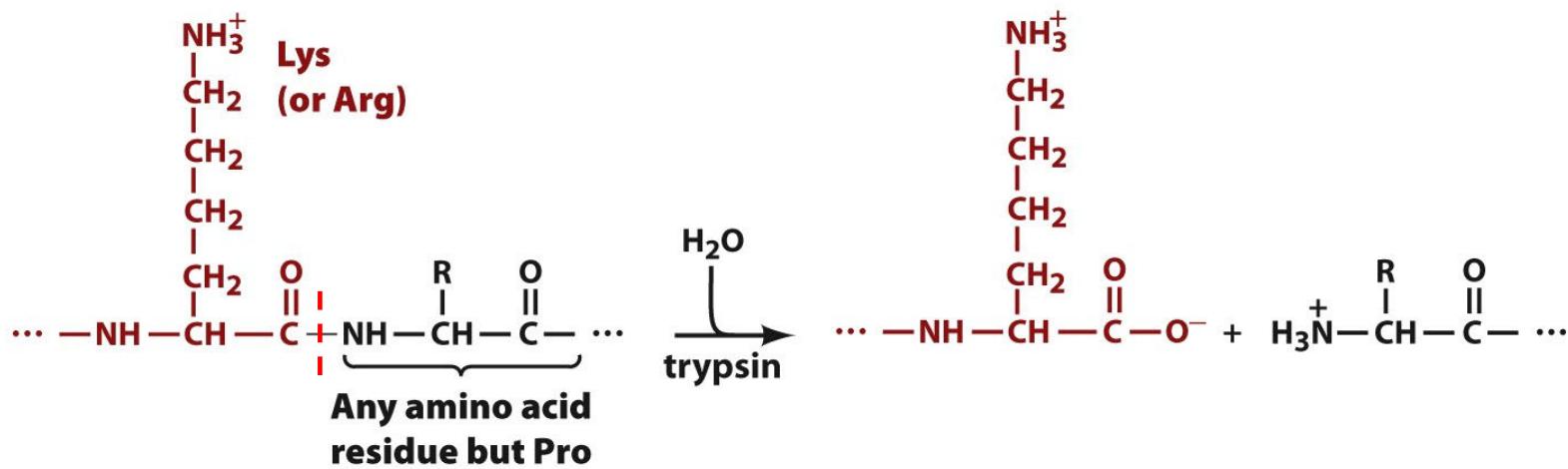


Figure 5-15
© 2013 John Wiley & Sons, Inc. All rights reserved.



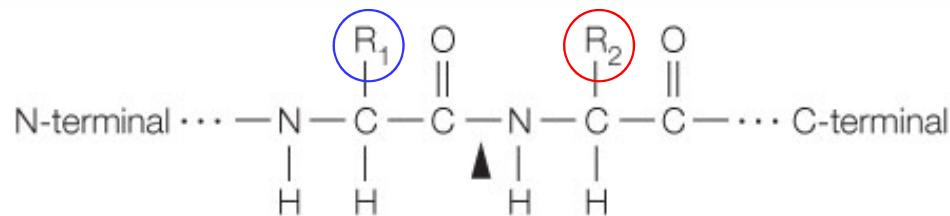
Proteins to Peptides - 2: Specific endopeptidases



Most important method to **fragment proteins**: used in proteomics

TABLE 5.3 The sequence specificities of some proteolytic enzymes and cyanogen bromide

Breaking peptide bonds



Enzyme	Preferred Site ^a	Source
Trypsin	R_2 not Pro $\text{R}_1 = \text{Lys, Arg}$	From digestive systems of animals, many other sources
Chymotrypsin	R_2 not Pro $\text{R}_1 = \text{Tyr, Trp, Phe, Leu}$	Same as trypsin
Thrombin	$\text{R}_1 = \text{Arg}$	From blood; involved in coagulation
V-8 protease	$\text{R}_1 = \text{Asp, Glu}$	From <i>Staphylococcus aureus</i>
Prolyl endopeptidase	$\text{R}_1 = \text{Pro}$ R_2 not Pro	Lamb kidney, other tissues
Subtilisin	Very little specificity	From various bacilli
Carboxypeptidase A	$\text{R}_2 = \text{C-terminal amino acid}$	From digestive systems of animals
Thermolysin	R_1 not Pro $\text{R}_2 = \text{Leu, Val, Ile, Met}$	From <i>Bacillus thermoproteolyticus</i>
Cyanogen bromide	$\text{R}_1 = \text{Met}$	

^aThe residues indicated are those next to which cleavage is most likely. In some cases, preference is determined by the residue on the N-terminal side of the cleaved bond (R_1) and sometimes by the residue to the C-terminal side (R_2). Generally, proteases do not cleave where proline is on the other side of the bond. Even prolyl endopeptidase will not cleave if $\text{R}_2 = \text{Pro}$.



Edman Degradation for N-terminal sequencing of peptides

- Repeated proteolysis
- For step-wise N-terminal identification: gives 1st aa, 2nd aa, 3rd aa etc.

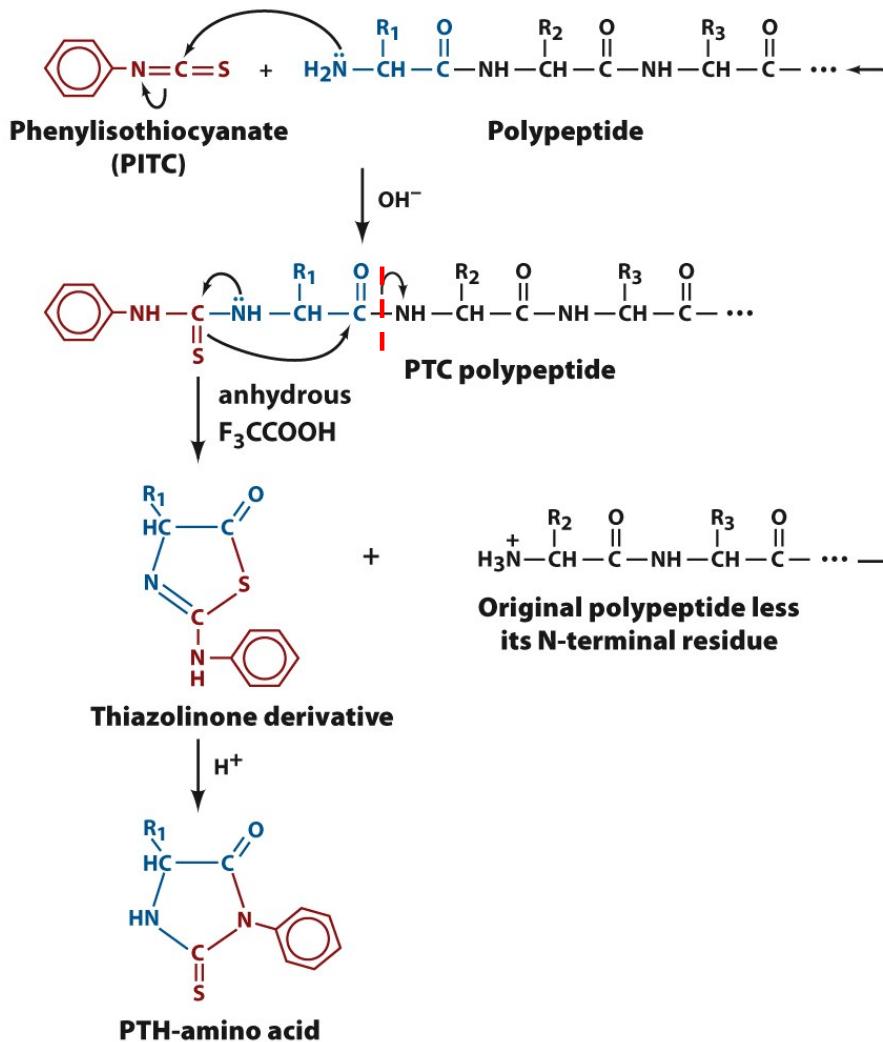
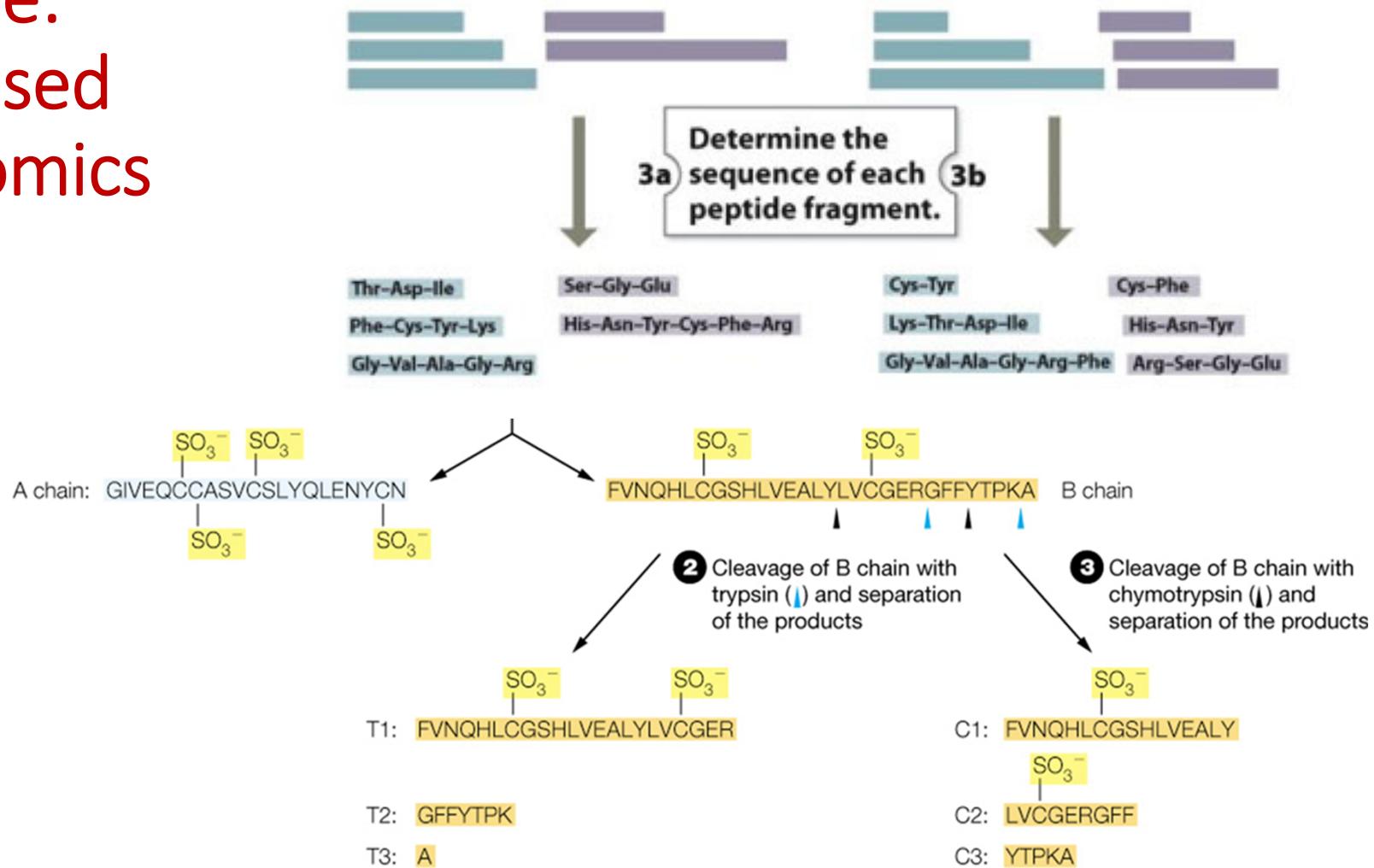


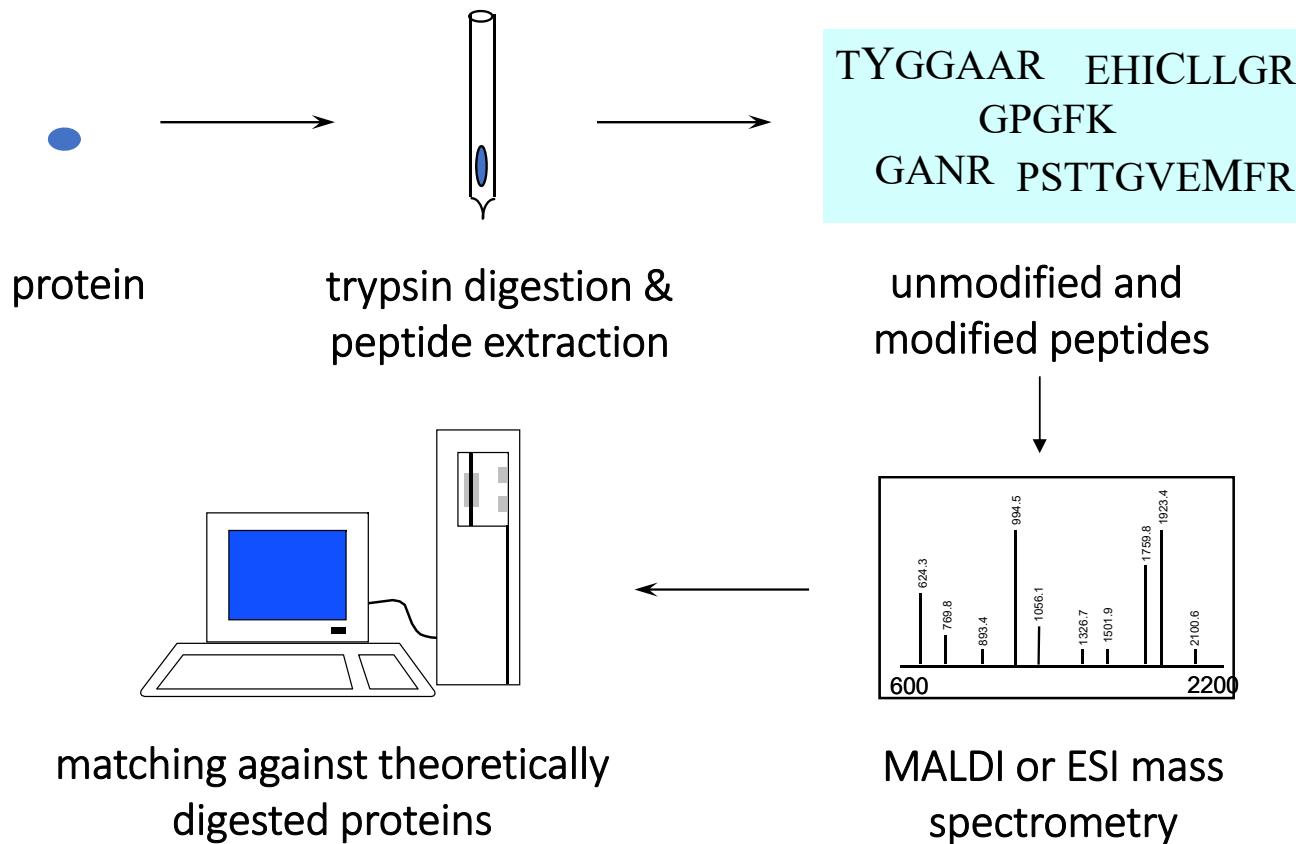
Figure 5-16
© 2013 John Wiley & Sons, Inc. All rights reserved.

Step 3: Determining the primary sequence of each peptide: MS-based Proteomics

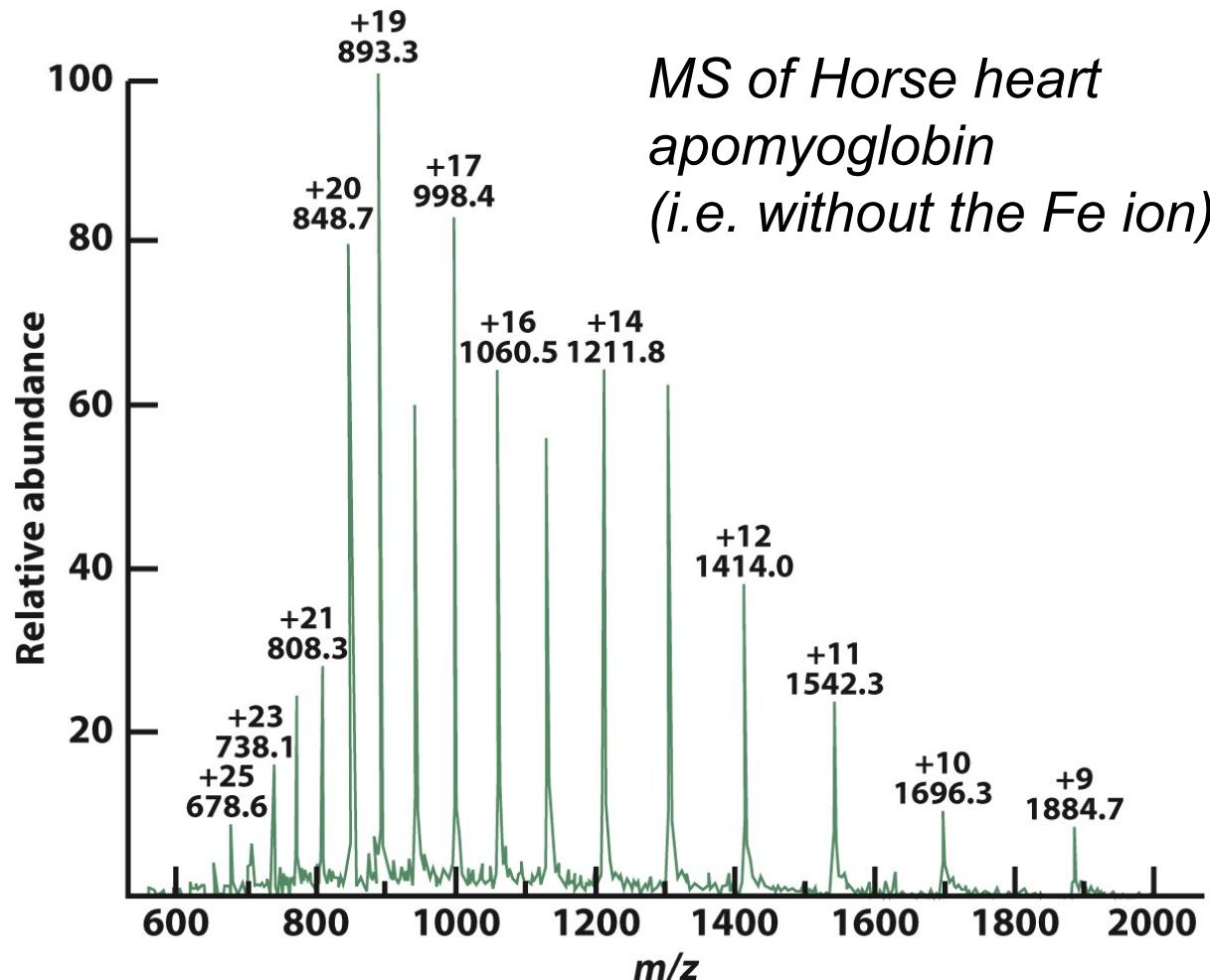


Peptide Masses from Mass Spectrometry

Mass spectrometers accurately measure peptide masses

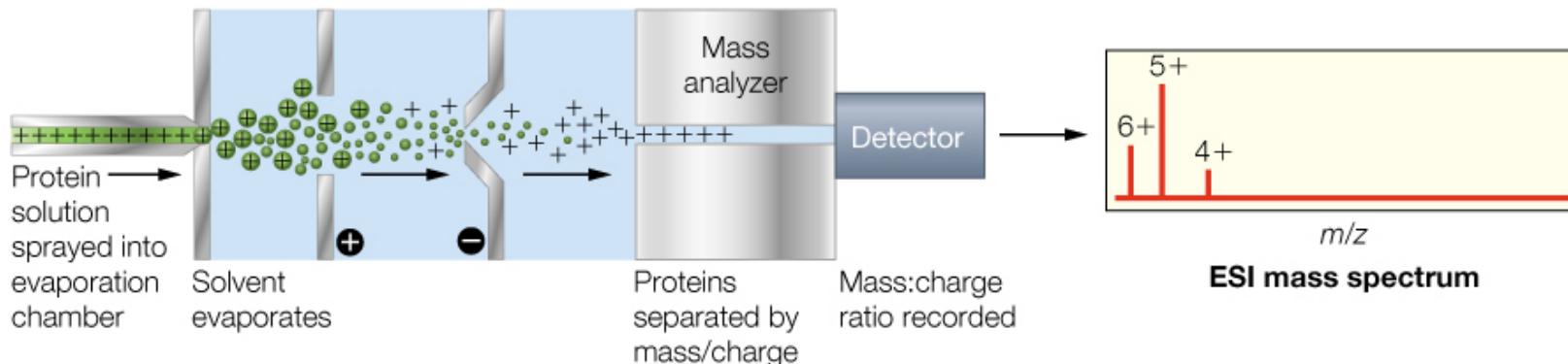


Mass Spectrum gives m/z of peptides



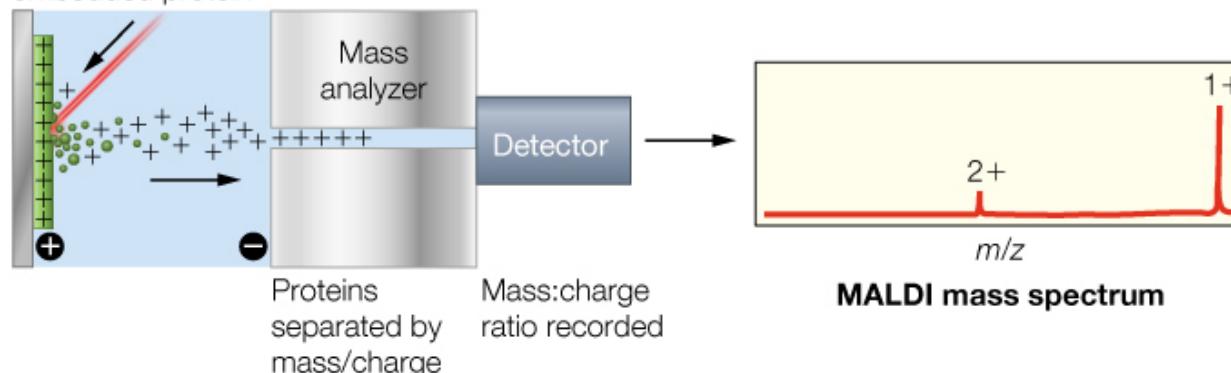
MS methods

Electrospray Ionization (ESI)



Matrix-Assisted Laser Desorption/Ionization (MALDI)

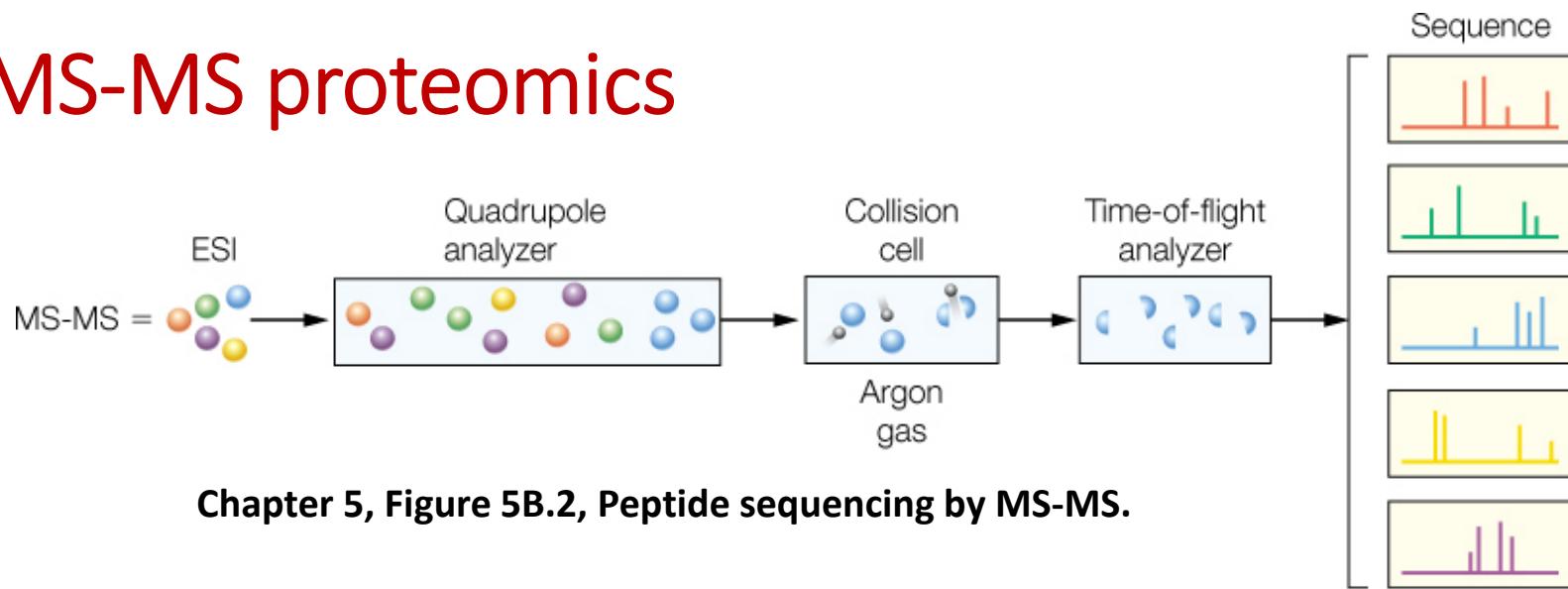
Laser pulse (orange) vaporizes matrix containing embedded protein



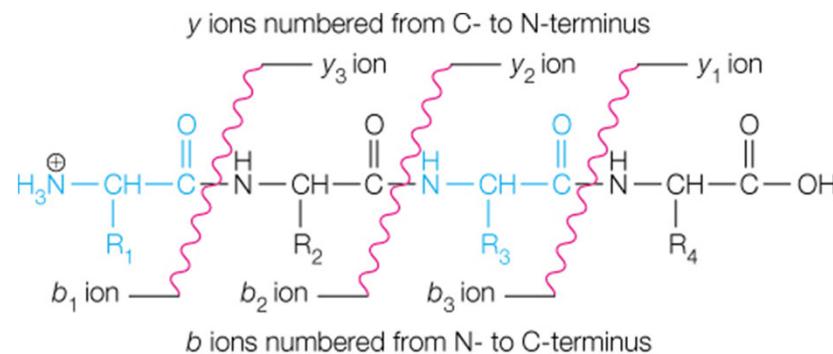
Chapter 5, Figure 5B.1, Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) mass spectrometry techniques.



MS-MS proteomics



Chapter 5, Figure 5B.2, Peptide sequencing by MS-MS.



Chapter 5, Figure 5B.3, Principal ions generated by low-energy collision-induced fragmentation.

Sequencing of the B Chain of Insulin Using Proteases and MS-MS: locating disulphide bonds in insulin

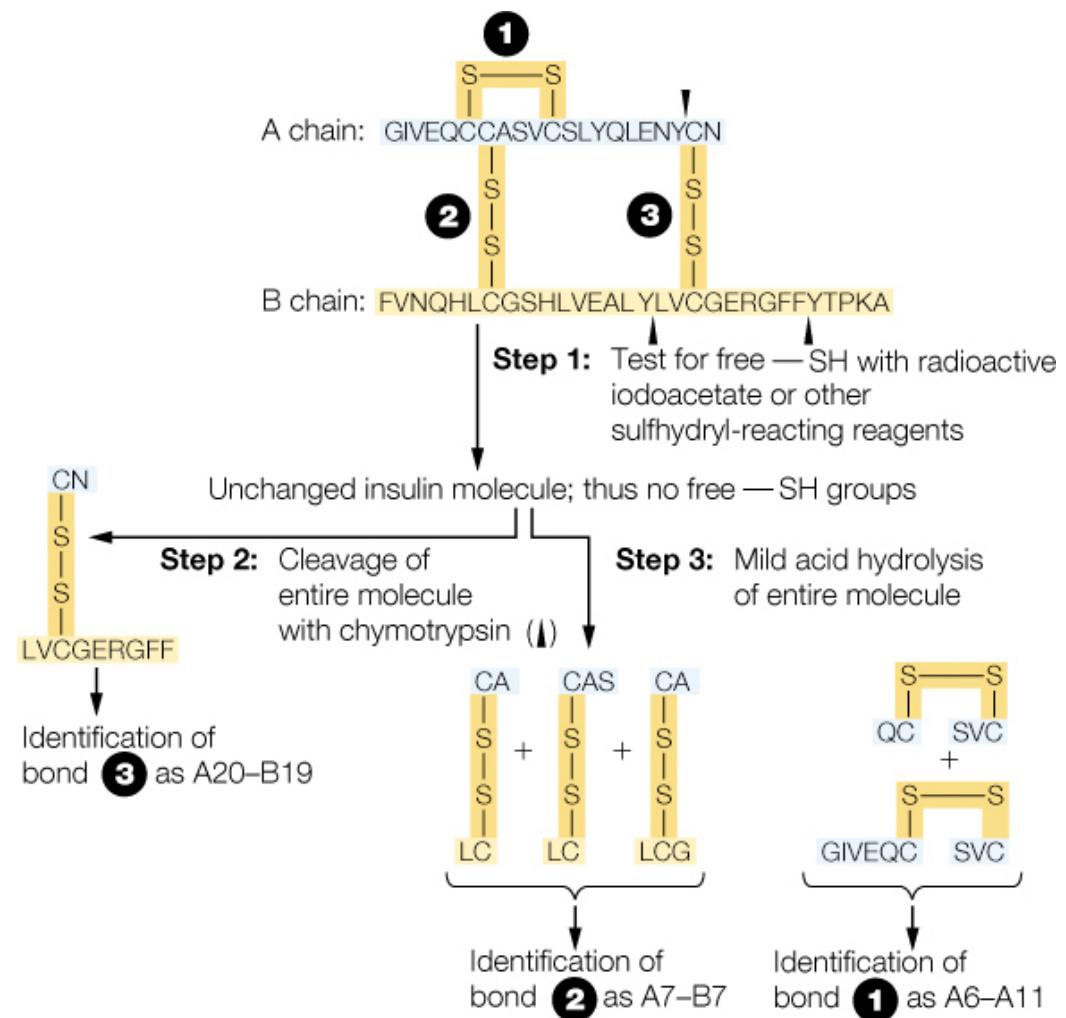
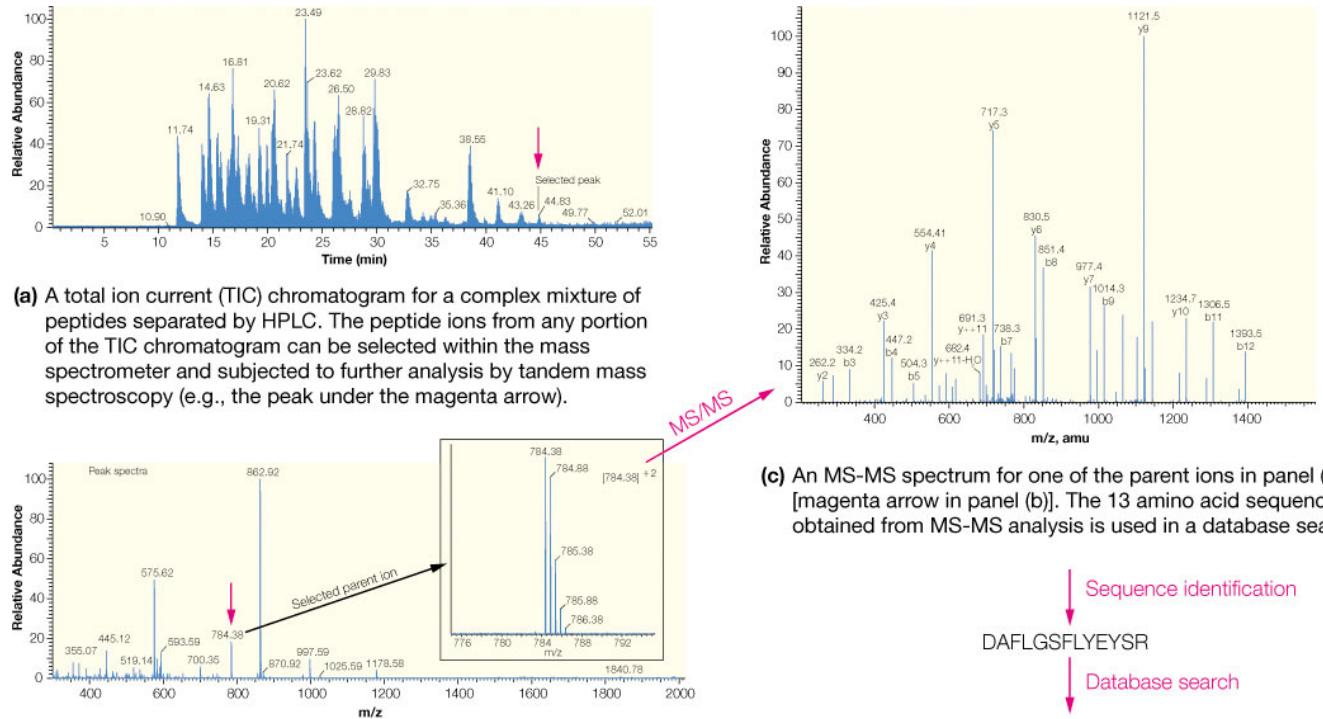


FIGURE 5B.5 Locating the disulfide bonds in insulin.

Finding possible proteins from proteomics



Chapter 5, Figure 5B.6, Identification of a protein of interest using proteomics methods.

(c) An MS-MS spectrum for one of the parent ions in panel (b) [magenta arrow in panel (b)]. The 13 amino acid sequence obtained from MS-MS analysis is used in a database search.

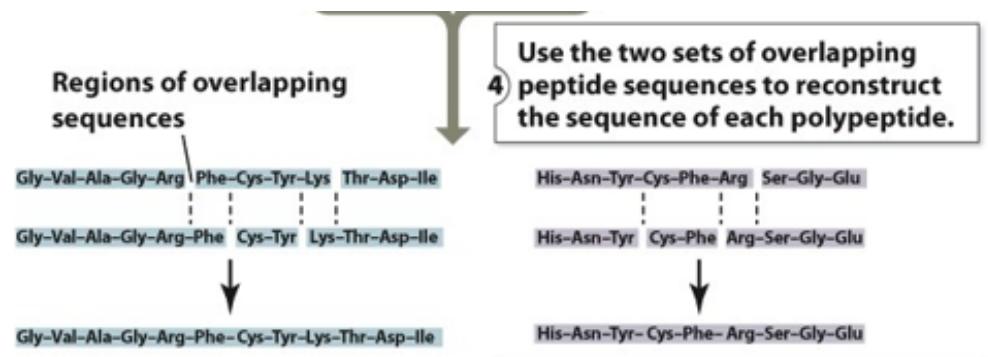
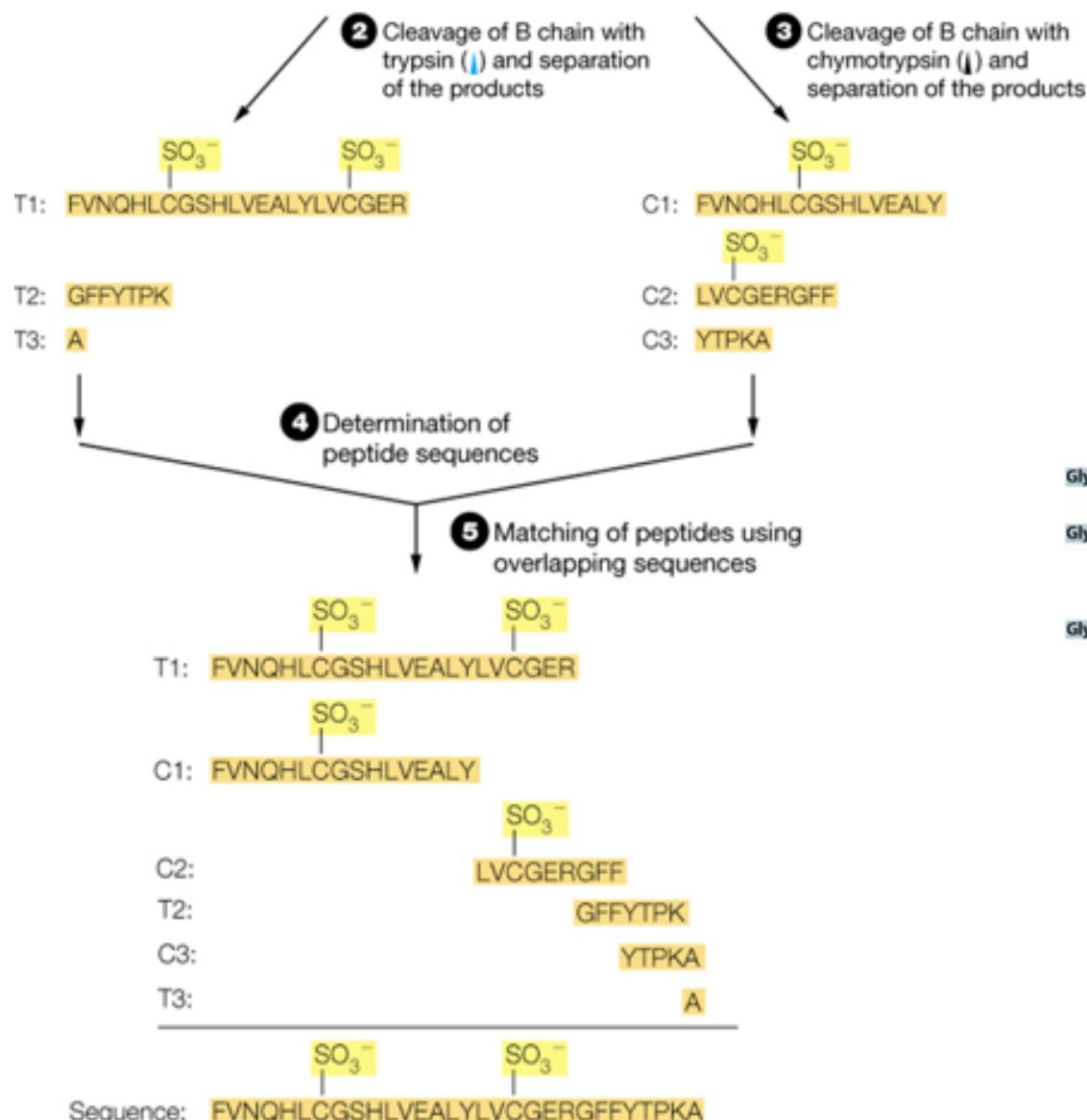
Sequence identification
 DAFLGSFLYYSR
 Database search

P02769 Serum albumin precursor (Allergen Bos d 6) (BSA)

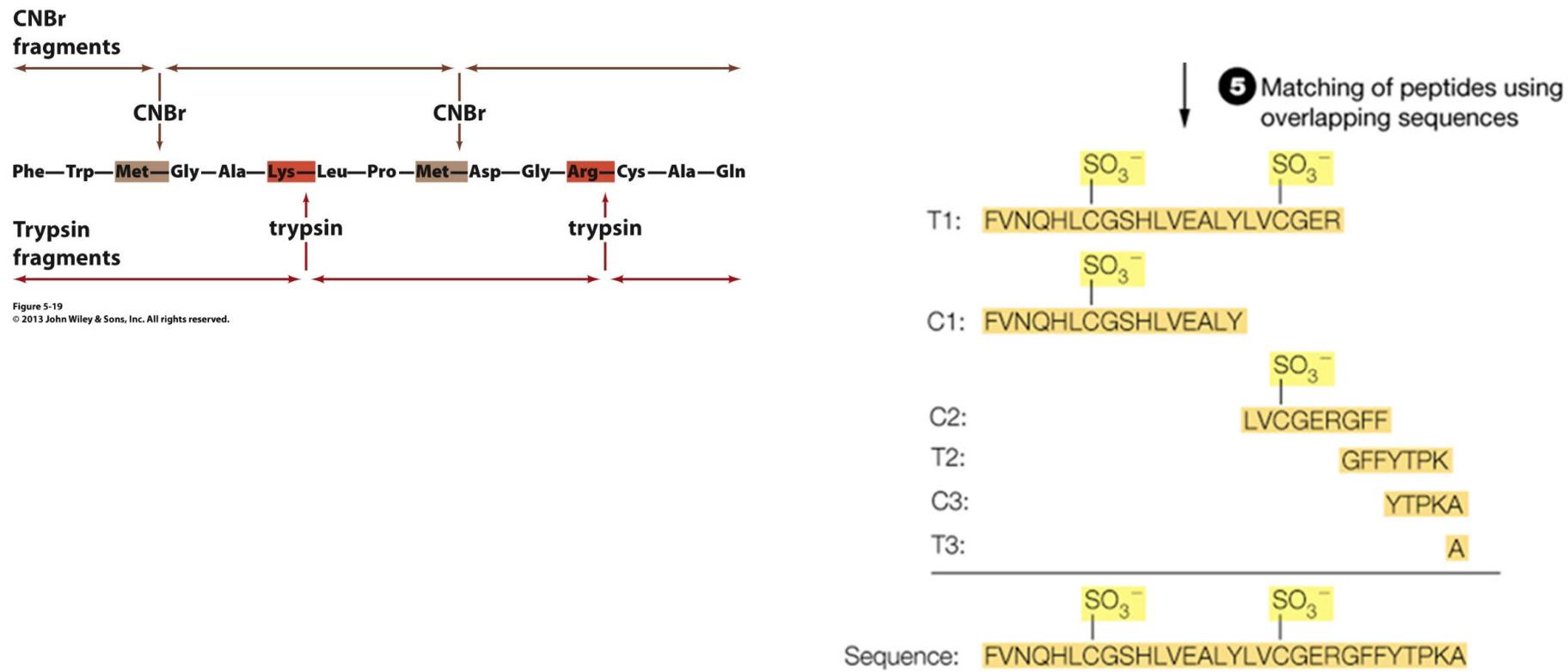
MKWVTFISLLLLFSSAYSRSRGVFRRDTHKSEIAHRFKDLGEEHFKGVLIAFSQYLQQCPFDEHVKLVNELEFAKTCVADESHAGECEKSLHTLFGDELCKVASLRETYGDMADCCEKQEPERNECFLSHKDDSPDLPKLPDPNTLCDEFKADEKKFWGKYLYEIARRHPFYAPELLYYAN KYNGVFQECCQAEDKGACLLPKIETMREKVLIASSARQRLRCASIQ KFGERALKAWSVARLSQKFPKAEFVEVTKLVDLTKVHKECCHGD LLECADDRADLAKYICDNQDTISSKLKECCDKPLLEKSHCIAEVK DAIPENLPPPLTADFAEDKDVKNCYEAQDAFLGSFLYYSR RHEPV AVSVLLRLAKEYEATLEECCAKDDPHACYSTVFDKLKHLVDEPQLNIKQNCDQFEKLGEYGFQNALIVRYTRKVPQVSTPTLVEVSRSLGKV GTRCCTKPESERMPCTEDYLSLILNRLCWLHEKTPVSEKVKCCTE SLVNRRPCFSALTPDETYVPKAFDEKLFTFHADICTLPDTEKQIKKQ TALVELLKHKPKATEEQLKTVMENFVAFVDKCCAADDKEACFAVEG PKLVVSTQTALA

(d) A database search using the sequence DAFLGSFLYYSR (magenta) shows that the peptide is part of the amino acid sequence of bovine serum albumin precursor protein.

Step 4: Figuring out how to fit the peptide sequences together



Overlapping Fragments Allow Determination of Amino Acid Sequence



Protein & DNA Sequence Data Banks for identifying peptides from proteomics

TABLE 5-5 Internet Addresses for the Major Protein and DNA Sequence Data Banks

Data Banks Containing Protein Sequences

ExPASy Proteomics Server: <http://expasy.org/>

Protein Information Resource (PIR): <http://pir.georgetown.edu/>

UniProt: <http://www.uniprot.org/>

Data Banks Containing Gene Sequences

GenBank: <http://www.ncbi.nlm.nih.gov/genbank/>

European Bioinformatics Institute (EBI): <http://www.ebi.ac.uk>

GenomeNet: [http://www.genome.jp/](http://www.genome.jp)

Table 5-5
© 2013 John Wiley & Sons, Inc. All rights reserved.



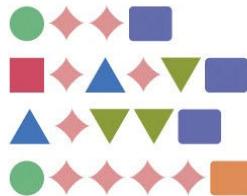
Protein sequences reveal shared parts called domains

(a) Fibronectin



(b) Blood clotting proteins

Factors VII, IX, X, and protein C



Factor XII

Tissue-type plasminogen activator

Protein S

Key

- ▲ Fibronectin domain 1
- Fibronectin domain 2
- Fibronectin domain 3
- γ-Carboxyglutamate domain
- ◆ Epidermal growth factor domain
- Serine protease domain
- ▼ Kringle domain
- Unique domain

1. Parts that are common to different proteins are called **domains**.
2. A single domain usually has the same function, even in different proteins.
3. In combination with other domains, proteins have different overall functionality.

Protein Sequencing summary

- To be sequenced, a protein must be separated into individual polypeptides that can be cleaved into sets of overlapping fragments.
- The amino acid sequence can be determined by Edman degradation, a procedure for removing N-terminal residues one at a time – very slow.
- Mass spectrometry can quickly identify amino acid sequences from the mass-to-charge ratio of gas-phase protein fragments.
- Protein sequence data are deposited in online databases.

Points to think about

- Summarize the steps involved in sequencing a protein.
- Why is it important to identify the N-terminal residue(s) of a protein?
- What are some advantages of sequencing peptides by mass spectrometry rather than by Edman degradation?
- Explain why long polypeptides must be broken into at least two different sets of peptide fragments for sequencing.
- What types of information can be retrieved from a protein sequence database?

Protein structure – definition of terms

- Zeroth level of protein structure: aa composition (historical)
- Primary (1°) structure
 - Amino acid sequence
- Secondary (2°) structure
 - Structural elements in proteins that are primarily formed through peptide backbone hydrogen bonds
- Tertiary (3°) structure
 - The three dimensional (3D) structure of a protein
- Quaternary (4°) Structure
 - Arrangement of subunits within a multisubunit protein

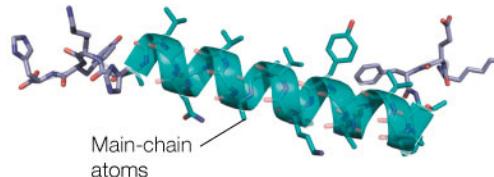
1° , 2° , 3° & 4° Structure: haemoglobin

...KEFTPPVQAAYQKVAGVANALAHKYH...

(a) Primary structure (amino acid sequence):

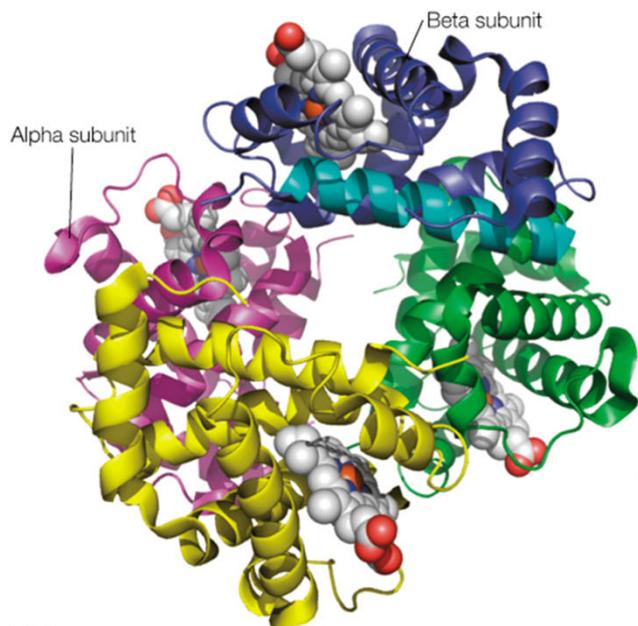
A portion of the amino acid sequence of human beta globin is shown. The sequence highlighted in cyan adopts a helical conformation, and is shown in the same orientation in parts (b-d).

Some parts of the primary sequence adopt a local regular repeating structure (" 2° structure")



(b) Secondary structure:

A stick representation of the amino acid sequence from part (a) is shown. Superimposed on the stick structure is a cartoon rendering of

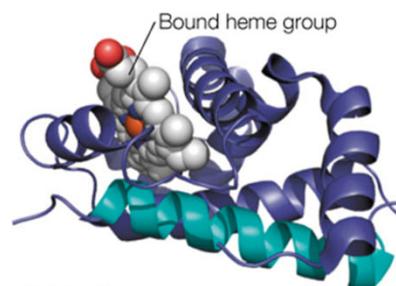


(d) Quaternary structure:

Four separate protein subunits, two alpha subunits (magenta and green) and two beta subunits (yellow and blue/cyan) associate to form the fully assembled hemoglobin protein. The four subunits are shown in cartoon rendering with hemes in space-filling display (PDB ID: 2hhb).

Quaternary structure (" 4° structure") arises when two or more proteins folded into tertiary structures interact to form well-defined multisubunit complexes.

Several 2° structure elements associate along their hydrophobic surfaces to give a stably folded structure (" 3° structure")



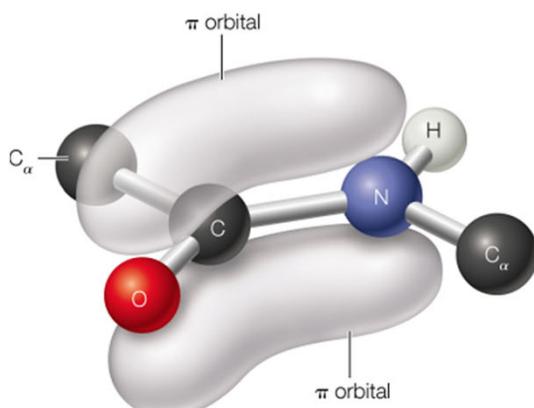
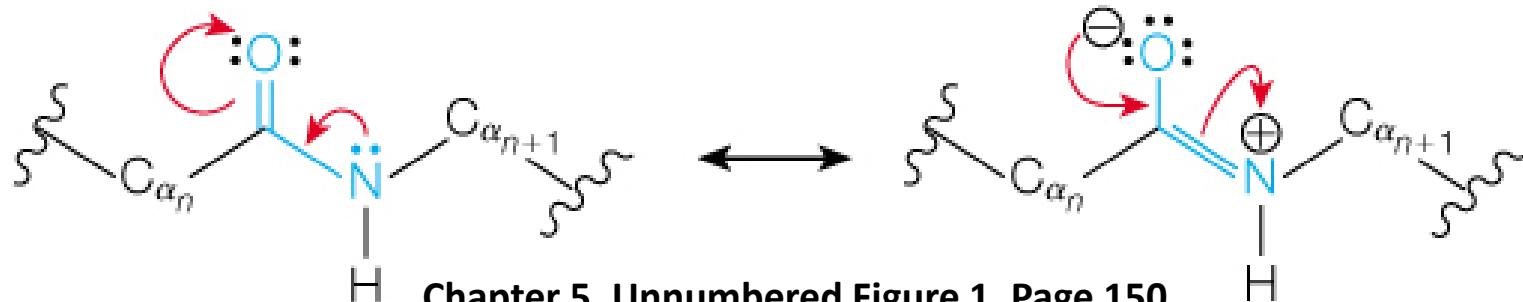
(c) Tertiary structure:

The entire beta globin chain is shown in its well-defined folded structure. As in myoglobin the helical regions interact to define the folded structure, which binds a heme (shown in space-filling display).

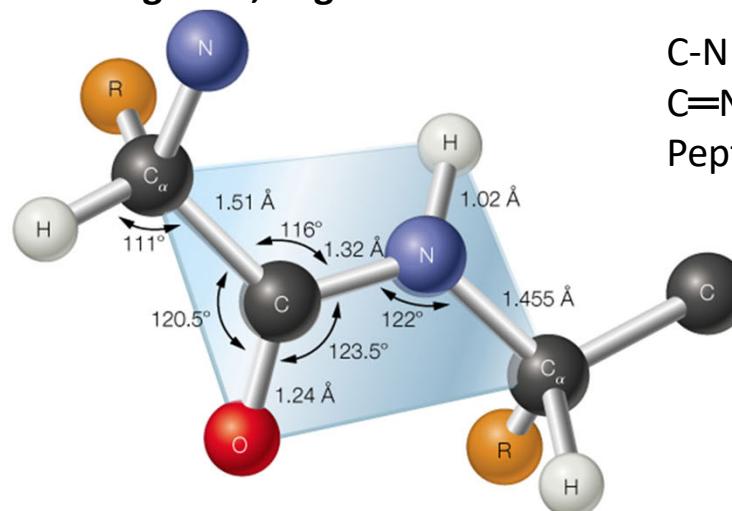
FIGURE 6.2 The four levels of structural organization in proteins.



Revisiting the peptide bond: neither single nor double



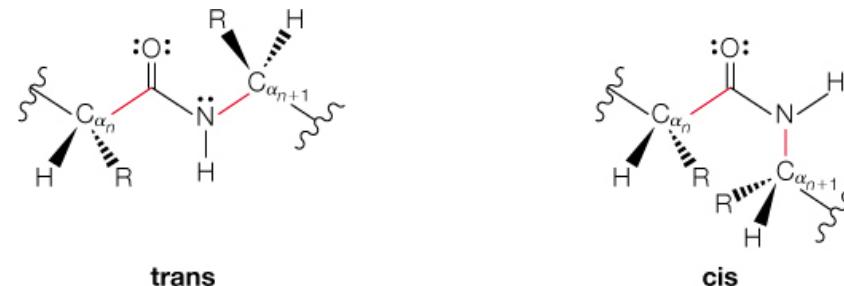
(a) Delocalization of the π -electron orbitals over the three atoms O—C—N accounts for the partial double-bond character of the C—N bond.



(b) The presently accepted values for bond angles and bond lengths (in Angstrom, \AA or m^{-10}) are given here. The six atoms shown in (a) and those in the blue rectangle in (b) are nearly coplanar.

C—N single bond: 1.49 \AA
 C=N double bond: 1.27 \AA
 Peptide bond: 1.32 \AA

Secondary structure: Peptide bonds impose planarity on the 6 atoms involved



Chapter 5, Unnumbered Figure 2, Page 150

Extended Conformation of Polypeptide

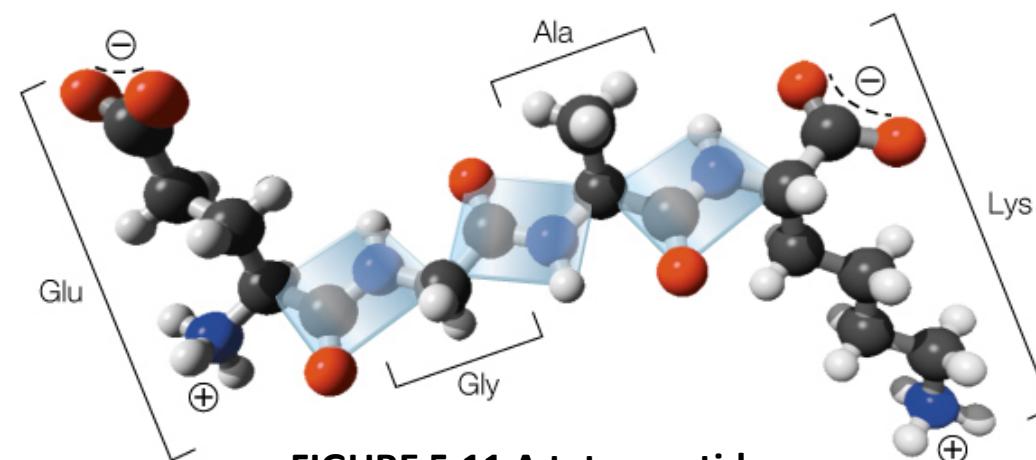


FIGURE 5.11 A tetrapeptide.



Rotations around the N-C_α and the C_α-C bonds of the protein backbone

- Free rotation is only allowed about the C_α carbons.
 - However, this rotation is restricted by steric interactions.
- Such rotation is described by two torsional angles termed ϕ (between N and C_α) and ψ (between C_α and the carbonyl C).
- Torsions refer to two different residues.
- Each residue maintains its peptide bond planarity of the atoms involved in the peptide bond.
- Steric interference of adjacent R groups leads to only some allowed Φ (phi) and Ψ (psi) values.

φ (phi), ψ (psi)

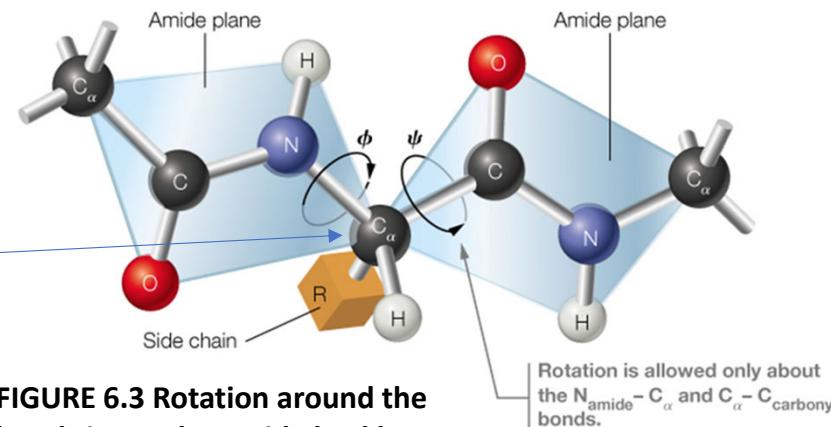
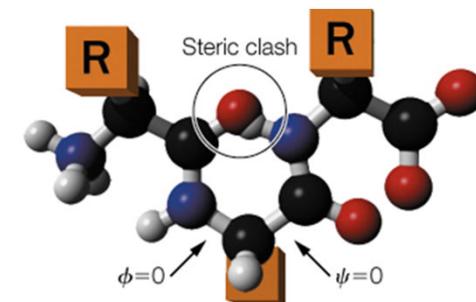


FIGURE 6.3 Rotation around the bonds in a polypeptide backbone.



(a) A sterically nonallowed conformation. The conformation $\phi = 0^\circ$, $\psi = 0^\circ$ is not allowed in any polypeptide chain because of the steric crowding between main-chain atoms. A tripeptide is shown, where the central amino acid has $\phi = 0^\circ$ and $\psi = 0^\circ$. Notice that the carbonyl oxygen of residue #1 (on the left) would clash with the amide hydrogen of residue #3 (on the right).

Figure 6.8, Steric interactions determine peptide conformation.

Naturally occurring secondary structures in proteins

Figure 6.4

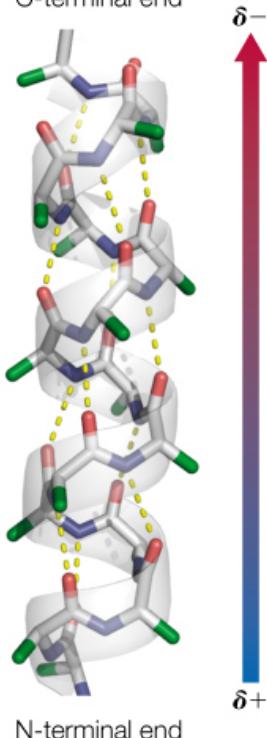
The right-handed α helix,

β sheet,

and

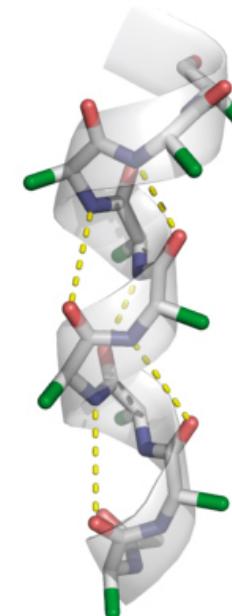
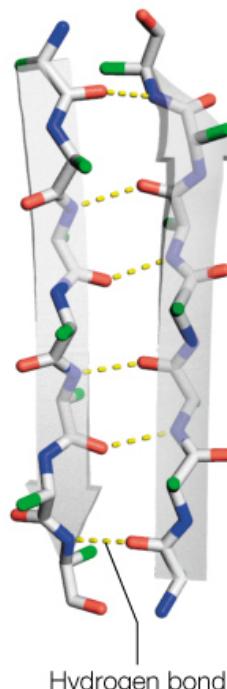
3_{10} helix.

C-terminal end



N-terminal end

- (a) In the α helix, the hydrogen bonds are within a contiguous stretch of amino acids and are almost parallel to the helix axis. This orientation of the amide bonds in the helix gives rise to a helical macrodipole moment shown by the arrow (see Figure 2.5). The N-terminal end of the helix has partial (+) charge character, and the C-terminal end has partial (−) charge character.
- (b) In the β sheet, the hydrogen bonds are between adjacent strands (only two strands are shown here), which are not necessarily contiguous in the primary sequence. In this structure, the hydrogen bonds are nearly perpendicular to the chains. Note that in the cartoon rendering a strand is shown by a flat arrow, where the head of the arrow points to the C-terminus of the strand.



- (c) The 3_{10} helix is found in proteins but is less common than the α helix. Note that, compared to the α helix, the 3_{10} helix forms a tighter spiral.



Summary

- The peptide bond imposes planarity on the 6 backbone atoms of each peptide
 - Free rotation is only allowed about the C_α carbons.
- Steric interactions from side chains limit the types of secondary structures that can form in 3D space.
 - Secondary structures are mainly stabilized by backbone hydrogen bonds.
 - There are two main types of structures that are able to form these bonds
 - ❖ the **α -helix** and
 - ❖ the **β -sheet**, made up of β -strands.

Spots test on textbook website

- Access textbook via iLearn
- Look for L5 Spot Test
- 2 questions; 5 mins; one attempt