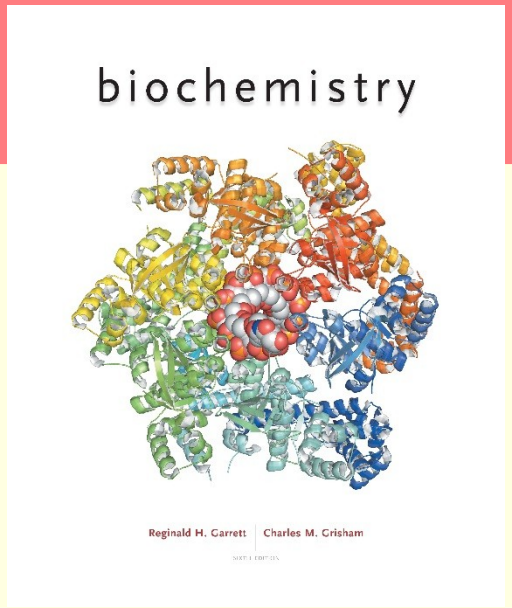
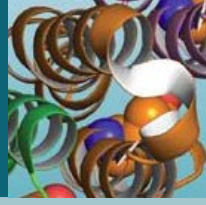


Reginald H. Garrett
Charles M. Grisham

www.cengage.com/chemistry/garrett

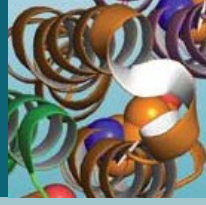
Chapters 4.7 & 5.3-5.6
Proteins: Primary Structure and Sequence Analysis

Outline



- Analysis of amino acid composition
- Determination of amino acid sequences by Edman degradation and mass spectrometry
- Protein evolution
- Solid-phase polypeptide synthesis

How is the amino acid composition of a protein determined?



Protein

heat ↓ 6N HCl

Amino Acids

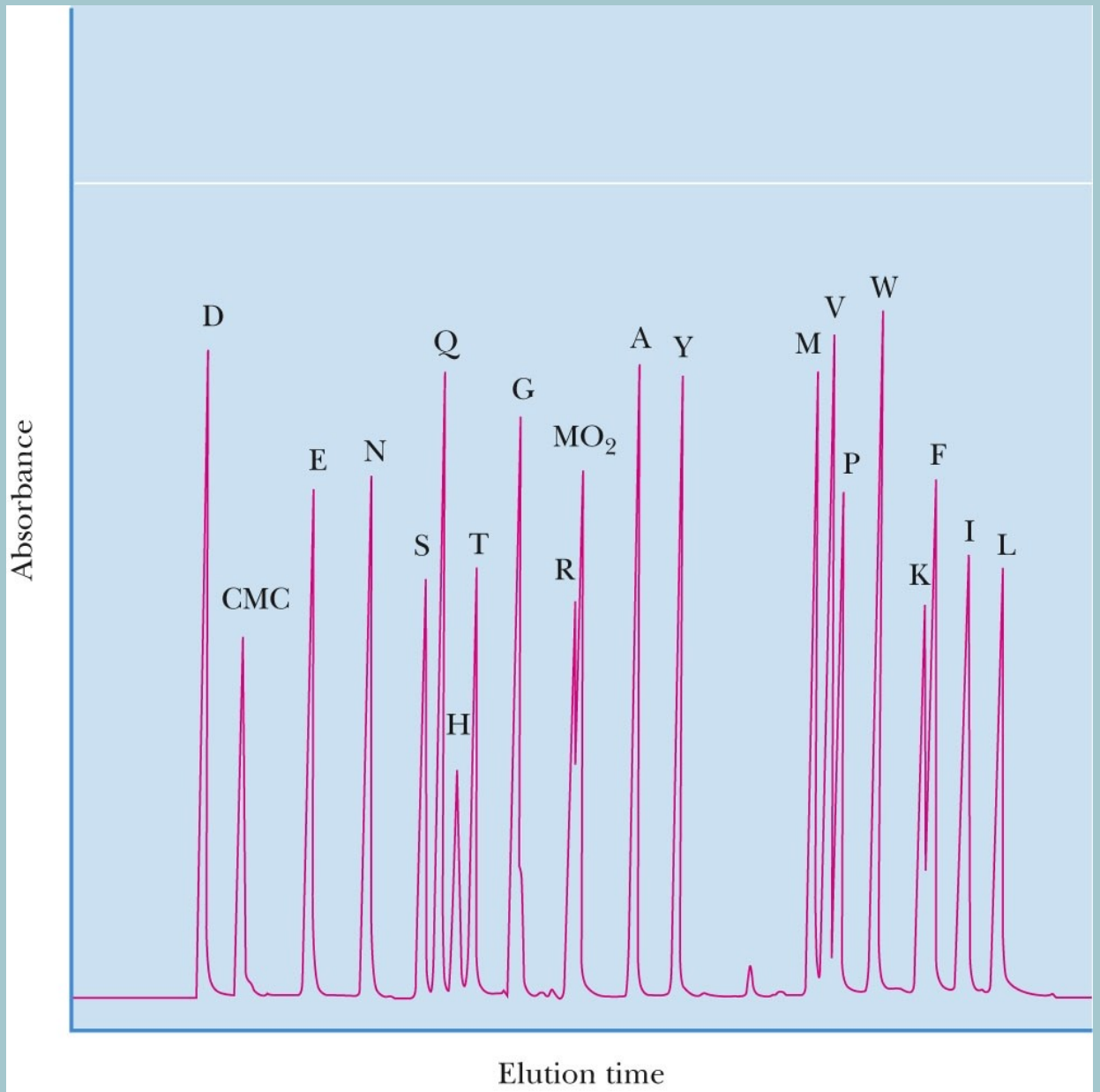
↓ o-phthaldialdehyde
or
↓ phenylisothiocynate
(Edman's Reagent)

Derivitized Amino Acids

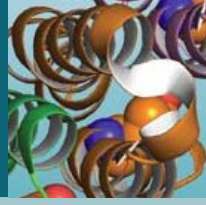
↓ HPLC
to
↓ separate

Separation of Amino Acids for Quantitation

Figure 4.13 Gradient separation of common PTH-amino acids by HPLC (High Performance/ Pressure Liquid Chromatography). Areas under peaks are proportional to the moles of each amino acid.



How is the amino acid sequence of a protein determined?



Protein

↓ Break & Block Disulfide Bonds

↓ cleave into peptides with proteases and sequence-specific reagents

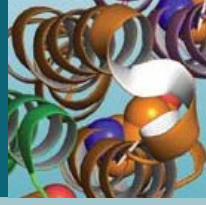
↓ Isolate peptides

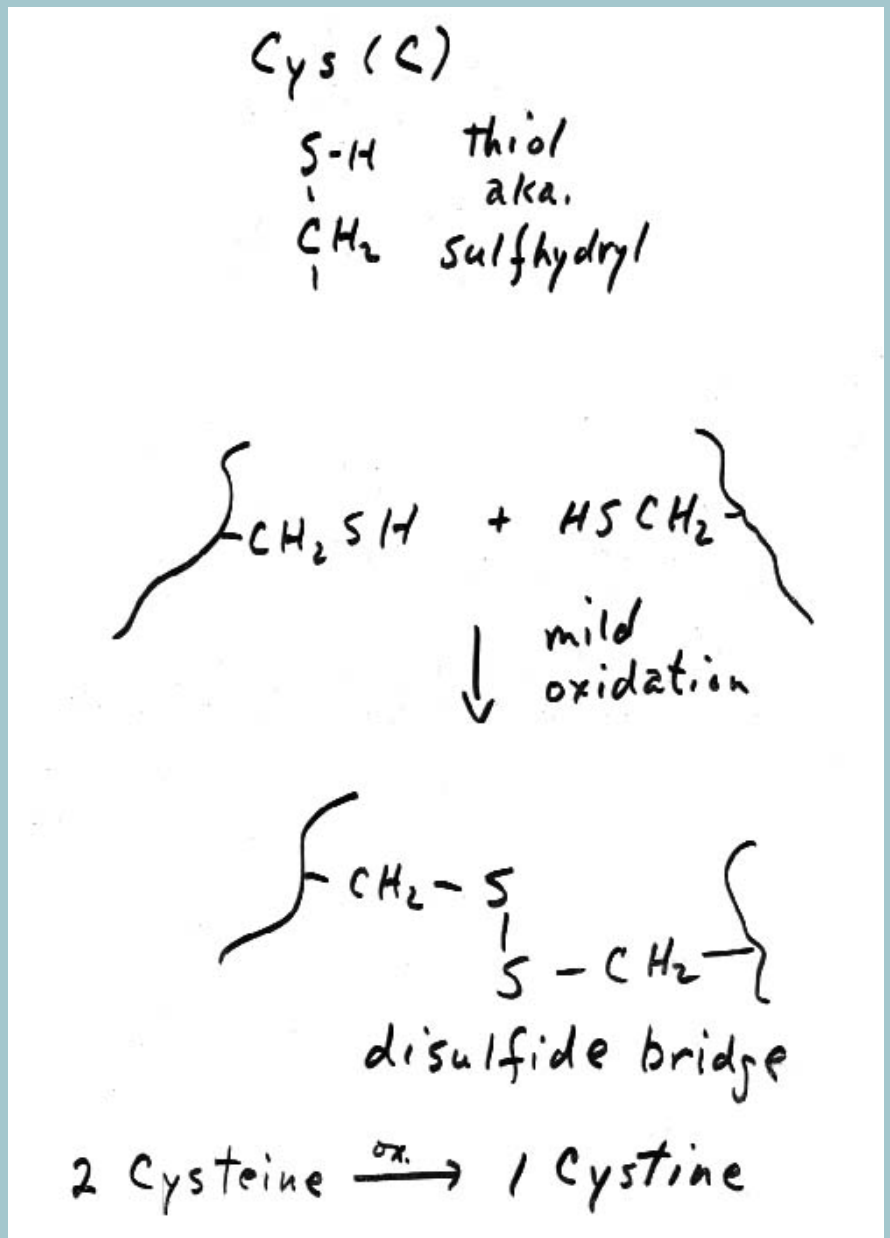
↓ Sequence peptides by Edman degradation

↓ Assemble full Sequence

Alternative: Sequence peptides by tandem Mass Spectrometry

Making and breaking disulfide bonds





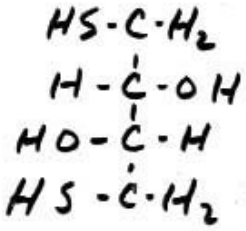
Disulfide bonds can be broken by reduction.

Common reducing agents are

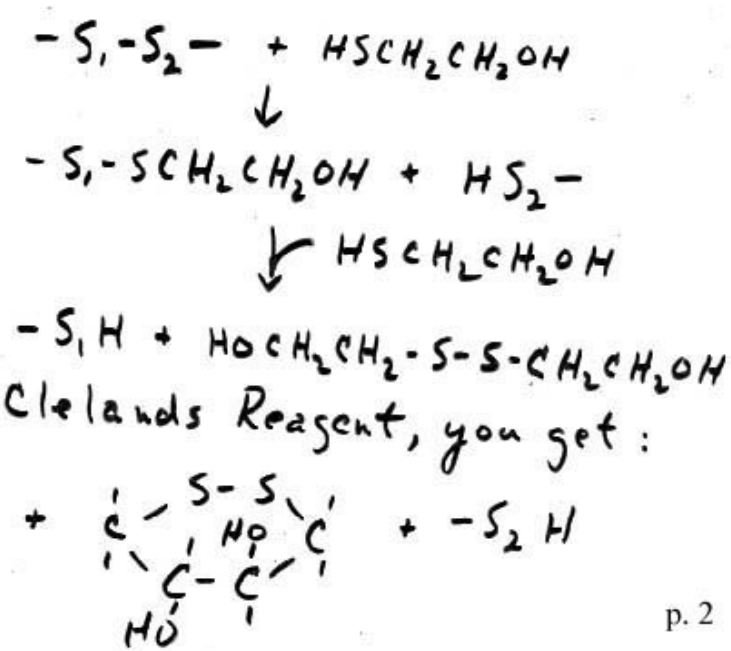
β-mercaptoethanol $HS\text{-}CH_2CH_2\text{-}OH$

aka 2- "

dithiothreitol (Cleland's Reagent)



Rx:

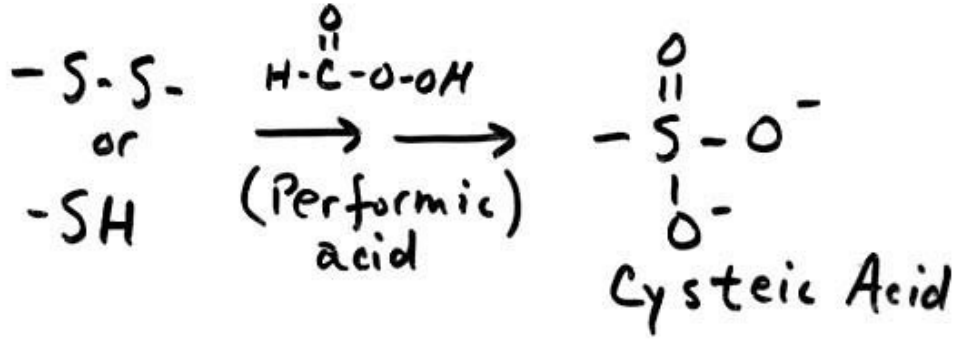


with Clelands Reagent, you get:

Also,
Tris (2-carboxyethyl) phosphine (TCEP)

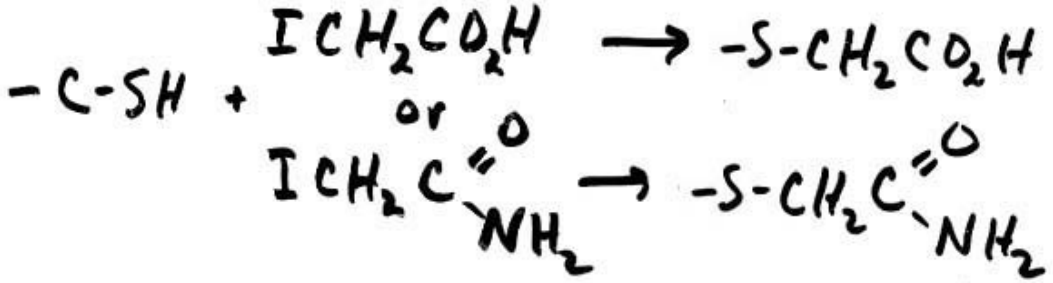
$RSSR + TCEP + H_2O \rightarrow 2RSH + TCEP{=}O$

Disulfide bonds can also be broken, permanently, by strong oxidation:

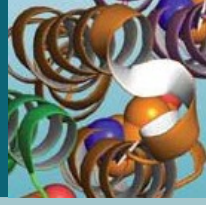


Cysteic Acid

When preparing proteins for sequencing -SH groups must be blocked, typically with iodoacetic acid or iodoacetamide



Ways to cleave a protein into oligopeptides



A.Enzymatic Cleavage of denatured protein

1. Trypsin - after Arg, Lys, or aminoethyl-Cys, produced by reaction with ethyleneimine; maleylation of Lys by reaction with maleic anhydride blocks cleavage by trypsin

2. V8 Protease (Staphylococcal) - after Glu in NH_4acetate, pH 4, or NH_4HCO_3, pH7.8, but also after Asp in PO_4 buffers

3. Chymotrypsin - after Phe, Trp, Tyr, Leu

4. Pepsin - after (and sometimes before) Phe, Tyr, Trp, and Leu, but it is less specific than chymotrypsin. (Good for recovery of original disulfide bridges, -S-S-, since it is able to digest proteins at pH2.0)

5. Thermolysin - before Leu, Ile, Val & other hydrophobic amino acids.

Usually none of the above cleave bonds when proline follows the amino acid

B. Chemical Cleavage

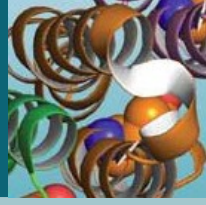
1. Cyanogen bromide, N= C-BR - after Met

2. Hydroxylamine, NH_20H - between Asn & Gly

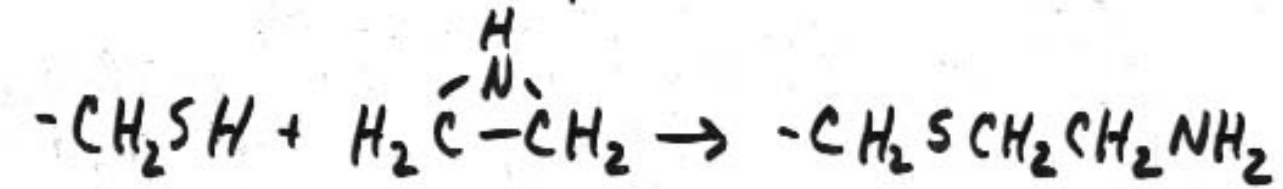
3. BNPS - Skatole - after Trp

4. o-Iodosobenzoic Acid -after Trp

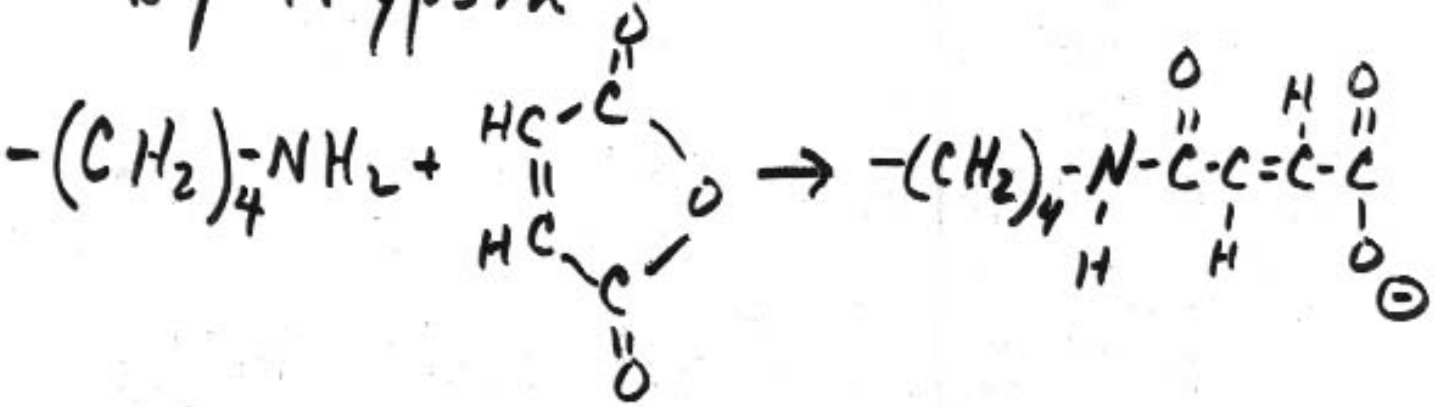
Reactions to create or block trypsin cleavage sites



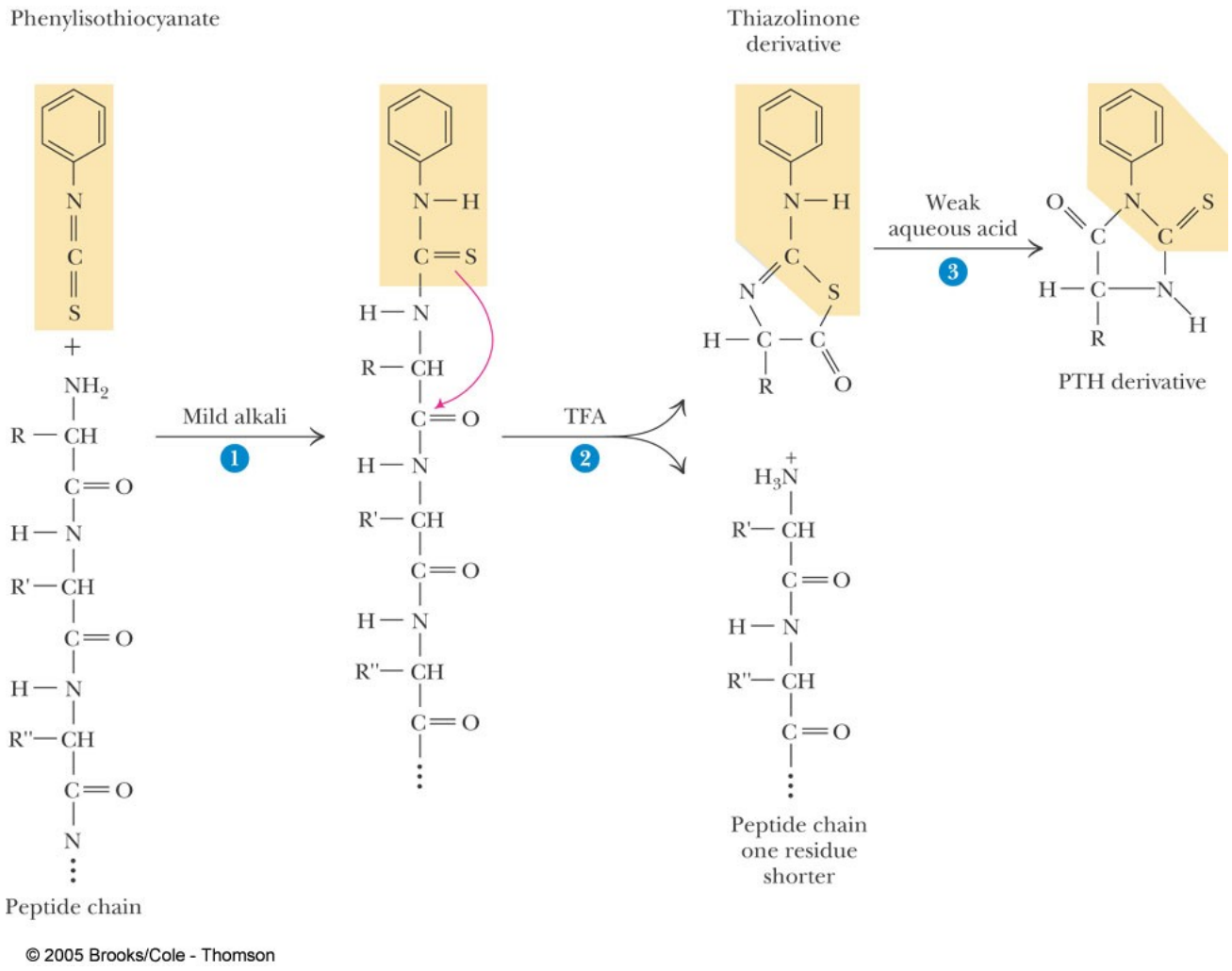
Conversion of Cys to Lys homolog by aminoethylation

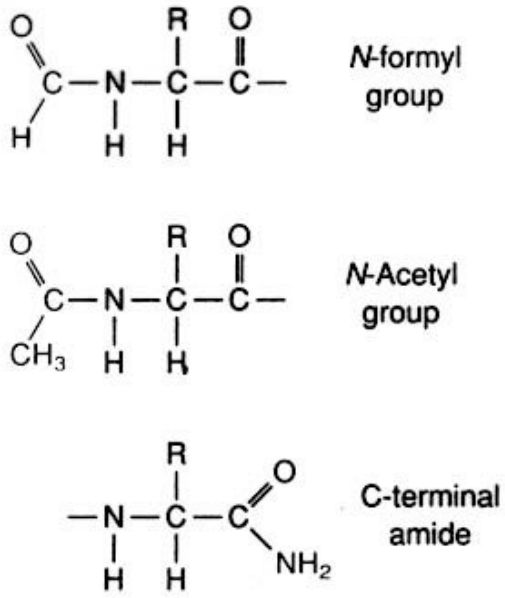


Maleylation of Lys to block cleavage by Trypsin



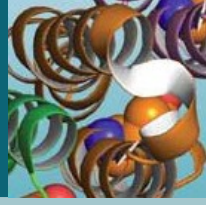
N-Terminal analysis using Edman's reagent, phenylisothiocyanate. Phenylisothiocyanate combines with the N-terminus of a peptide under mildly alkaline conditions to form a phenylthiocarbamoyl substitution. Upon treatment with TFA (trifluoroacetic acid), this cyclizes to release the N-terminal amino acid residue as a thiazolinone derivative, but the other peptide bonds are not hydrolyzed. Organic extraction and treatment with aqueous acid yield the N-terminal amino acid as a phenylthiohydantoin (PTH) derivative.





Groups that may block N- or C-termini in proteins.

How is the amino acid sequence of a protein determined?



Protein

↓ Break & Block Disulfide Bonds

↓ cleave into peptides with proteases and sequence-specific reagents

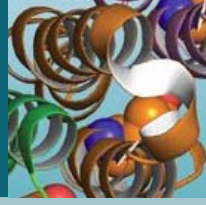
↓ Isolate peptides

↓ Sequence peptides by Edman degradation

↓ Assemble full Sequence

Alternative: Sequence peptides by tandem Mass Spectrometry

Amino acid sequences can be determined by mass spectrometry



- Mass spectrometry separates particles on the basis of their mass-to-charge ratios

The Nobel Prize in Chemistry 2002

"for the development of methods for identification and structure analyses of biological macromolecules"

"for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules"	"for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution"

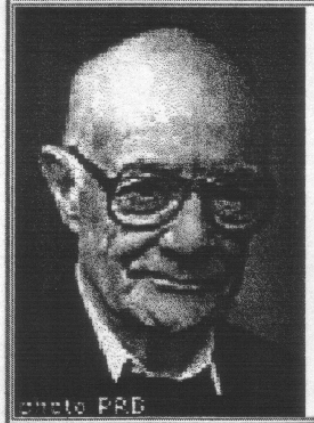
		
John B. Fenn	**Koichi Tanaka**	**Kurt Wüthrich**
½ 1/4 of the prize	½ 1/4 of the prize	½ 1/2 of the prize
USA	Japan	Switzerland
Virginia Commonwealth University Richmond, VA, USA	Shimadzu Corp. Kyoto, Japan	Eidgenössische Technische Hochschule (Swiss Federal Institute of Technology) Zürich, Switzerland; The Scripps Research Institute La Jolla, CA, USA
b. 1917	b. 1959	b. 1938

TABLE 5.3 The Two Most Common Methods of Mass Spectrometry for Protein Analysis

Electrospray Ionization (ESI-MS)

A solution of macromolecules is sprayed in the form of fine droplets from a glass capillary under the influence of a strong electrical field. The droplets pick up positive charges as they exit the capillary; evaporation of the solvent leaves multiply charged molecules. The typical 20-kD protein molecule will pick up 10 to 30 positive charges. The MS spectrum of this protein reveals all of the differently charged species as a series of sharp peaks whose consecutive m/z values differ by the charge and mass of a single proton (see Figure 5.14). Note that decreasing m/z values signify increasing number of charges per molecule, z. Tandem mass spectrometers downstream from the ESI source (ESI-MS/MS) can analyze complex protein mixtures (such as tryptic digests of proteins or chromatographically separated proteins emerging from a liquid chromatography column), selecting a single m/z species for collision-induced dissociation and acquisition of amino acid sequence information.

Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF MS)

The protein sample is mixed with a chemical matrix that includes a light-absorbing substance excitable by a laser. A laser pulse is used to excite the chemical matrix, creating a microplasma that transfers the energy to protein molecules in the sample, ionizing them and ejecting them into the gas phase. Among the products are protein molecules that have picked up a single proton. These positively charged species can be selected by the MS for mass analysis. MALDI-TOF MS is very sensitive and very accurate; as little as attomole (10^{-18} moles) quantities of a particular molecule can be detected at accuracies better than 0.001 atomic mass units (0.001 daltons). MALDI-TOF MS is best suited for very accurate mass measurements.

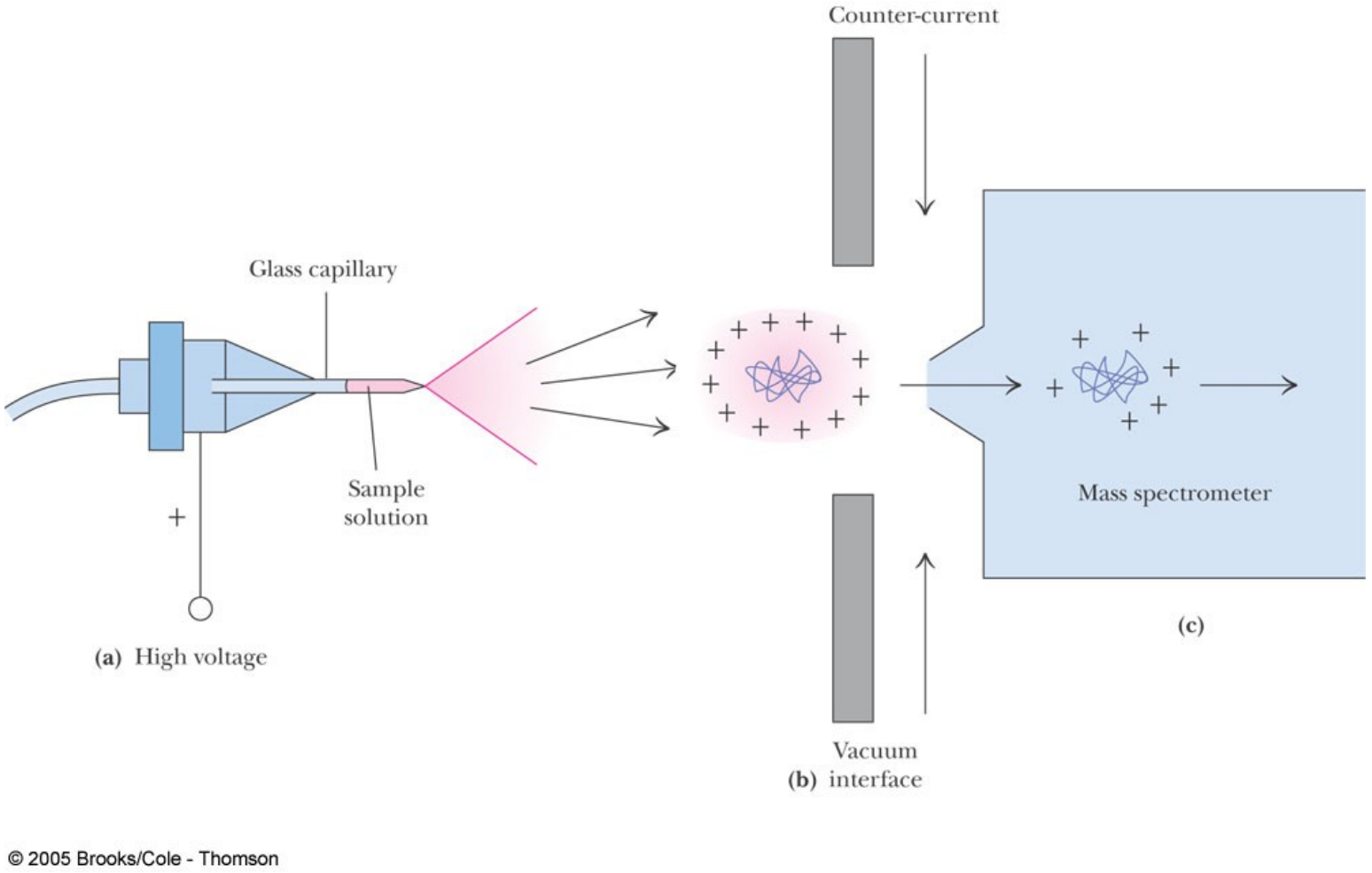
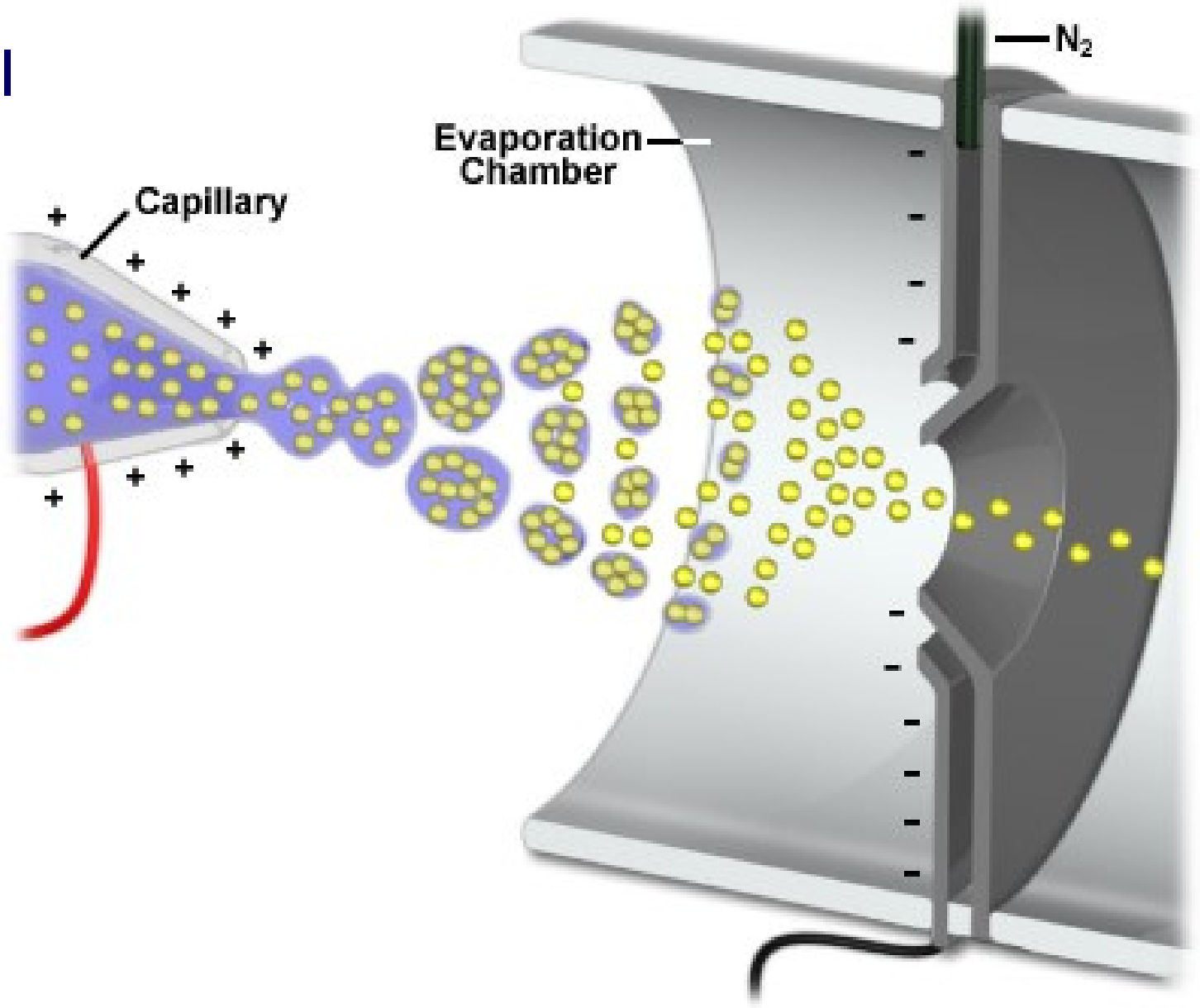


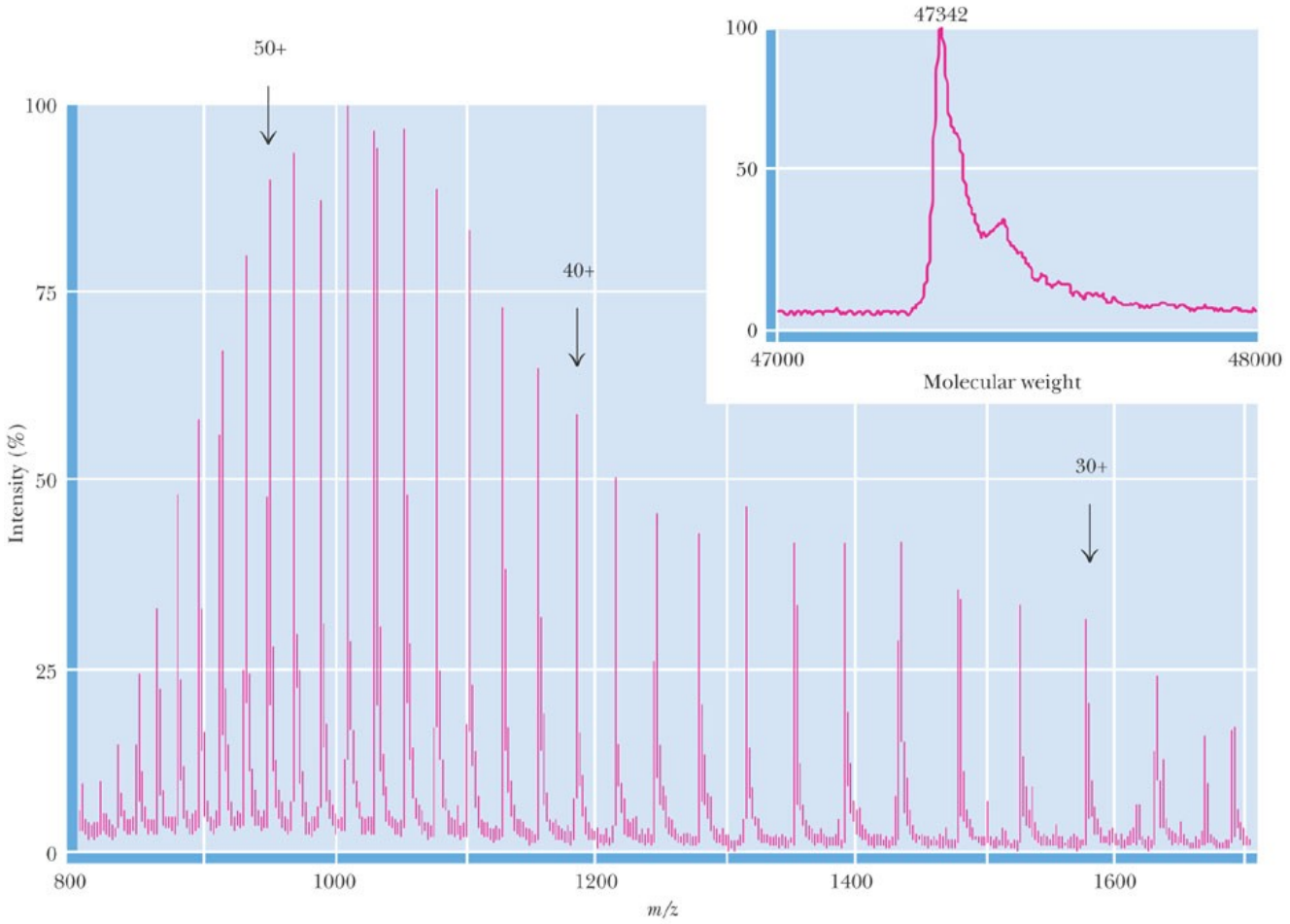
Figure 5.13 The three principal steps in electrospray mass spectrometry (ES-MS). **(a)** Small, highly charged droplets are formed by electrostatic dispersion of a protein solution through a glass capillary subjected to a high electric field; **(b)** protein ions are de-sorbed from the droplets into the gas phase (assisted by evaporation of the droplets in a stream of hot N_2 gas; and **(c)** analysis of the protein ions in a mass spectrometer. (*Adapted from Figure 1 in Mann, M., and Wilm, M., 1995* . Trends in Biochemical Sciences **20:***219–224*.)

ESI



http://www.magnet.fsu.edu/education/tutorials/tools/ionization_esi.html

Figure 5.14 Electrospray mass spectrum of the protein, aerolysin K. The attachment of many protons per protein molecule (from less than 30 to more than 50 here) leads to a series of *m*/*z* peaks for this single protein. The inset shows a computer analysis of the data from this series of peaks that generates a single peak at the correct molecular mass of the protein. (*Adapted from Figure 2 in Mann, M., and Wilm, M., 1995*. Trends in Biochemical Sciences **20:***219–224*.)



Amino Acid Sequence Can Be Determined by Mass Spectro-metry

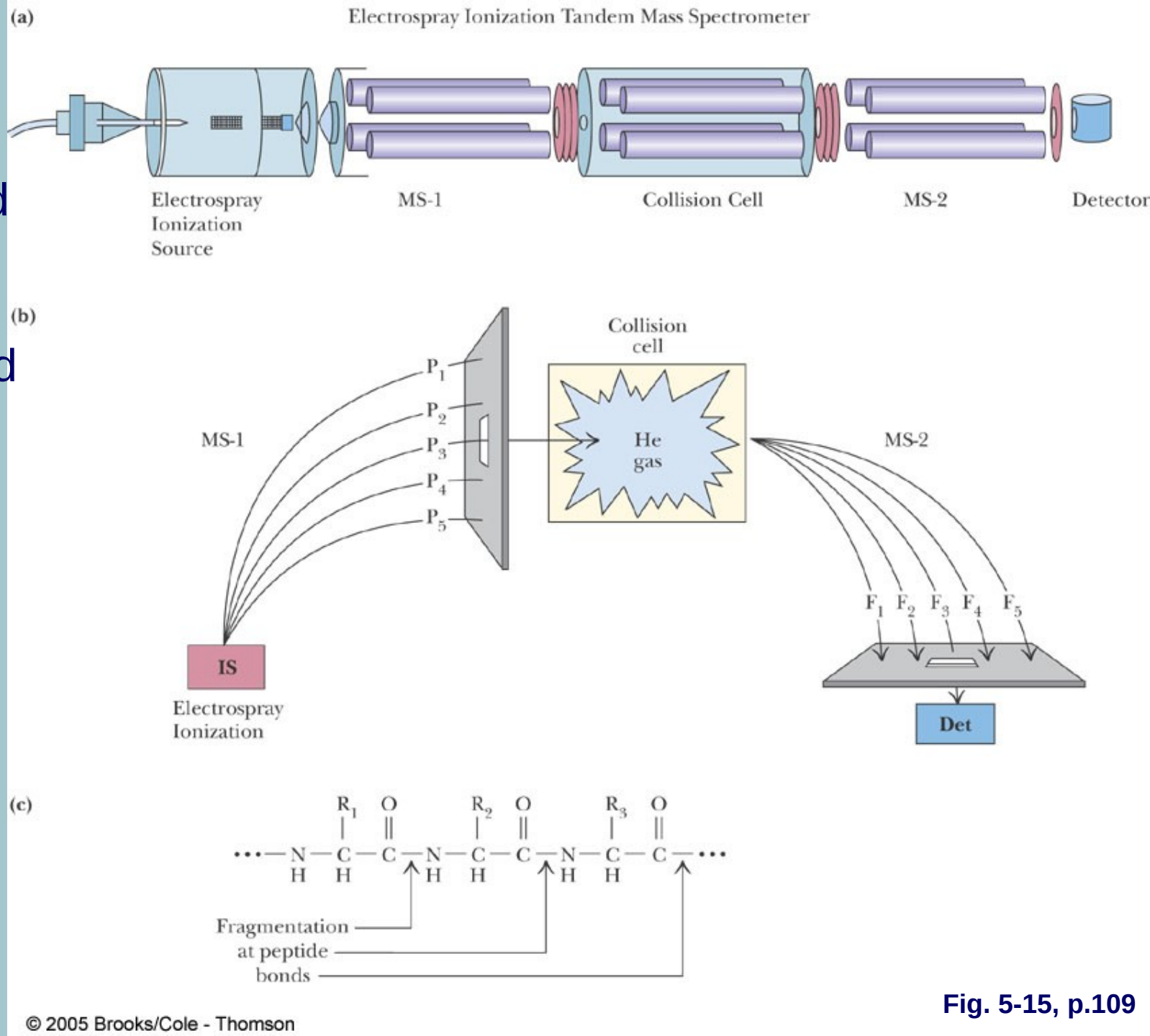
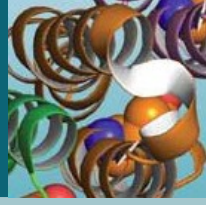


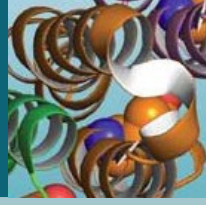
Fig. 5-15, p.109

5.5 What is the Nature of Amino Acid Sequences?



- Sequence and composition reflect the function of the protein, e.g., membrane proteins have more hydrophobic residues; whereas, fibrous proteins may have atypical sequences.
- Homologous proteins from different organisms have similar shapes and sequences which can be aligned.
- Mutations represent changes at one or more sites in the amino acid sequence of a protein.

Computer Programs Can Align Sequences and Discover Homology Between Proteins

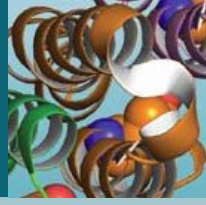


S. acidocaldarius	FPIAKGGTAAIPGPFGSGKTVTLQSLAKWSAAK---VVIYVGCGERGNEMTD
E. coli	CPFAKGGKVGLFGGAGVGKTVNMMELIRNIAIEHSGYSVFAGVGERTREGND

Alignment of the amino acid sequences of two protein homologs using gaps. Shown are parts of the amino acid sequences of the catalytic subunits from the major ATP-synthesizing enzyme (ATP synthase) in a representative archaea and a bacterium. These protein segments encompass the nucleotide-binding site of these enzymes.

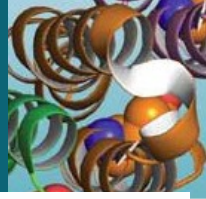
Identical residues in the two sequences are shown in red. Introduction of a three-residue-long gap in the archaeal sequence optimizes the alignment of the two sequences.

Blocks **S**ubstitution **M**atrix (BLOSUM)



- Methods for alignment and comparison of protein sequences depend upon some quantitative measure of how similar two sequences are.
- One way to measure similarity is to use a matrix that assigns scores for all possible substitutions of one amino acid for another.
- BLOSUM62 is the substitution matrix most often used with the popular BLAST alignment program.
- BLOSUM62 assigns a probability score for each position in an alignment based on the frequency with which that substitution occurs in the consensus sequences of related proteins.

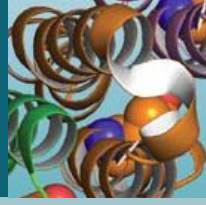
Blocks **S**ubstitution **M**atrix (BLOSUM)



	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	–1	5																		
N	–2	0	6																	
D	–2	–2	1	6																
C	0	–3	–3	–3	9															
Q	–1	1	0	0	–3	5														
E	–1	0	0	2	–4	2	5													
G	0	–2	0	–1	–3	–2	–2	6												
H	–2	0	1	–1	–3	0	0	–2	8											
I	–1	–3	–3	–3	–1	–3	–3	–4	–3	4										
L	–1	–2	–3	–4	–1	–2	–3	–4	–3	2	4									
K	–1	2	0	–1	–3	1	1	–2	–1	–3	–2	5								
M	–1	–1	–2	–3	–1	0	–2	–3	–2	1	2	–1	5							
F	–2	–3	–3	–3	–2	–3	–3	–3	–1	0	0	–3	0	6						
P	–1	–2	–2	–1	–3	–1	–1	–2	–2	–3	–3	–1	–2	–4	7					
S	1	–1	1	0	–1	0	0	0	–1	–2	–2	0	–1	–2	–1	4				
T	0	–1	0	–1	–1	–1	–1	–2	–2	–1	–1	–1	–1	–2	–1	1	5			
W	–3	–3	–4	–4	–2	–2	–3	–2	–2	–3	–2	–3	–1	1	–4	–3	–2	11		
Y	–2	–2	–2	–3	–2	–1	–2	–3	2	–1	–1	–2	–1	3	–3	–2	–2	2	7	
V	0	–3	–3	–3	–1	–2	–2	–3	–3	3	1	–2	1	–1	–2	–2	0	–3	–1	4

Figure 5.18 The BLOSUM62 substitution matrix provides scores for all possible exchanges of one amino acid with another.

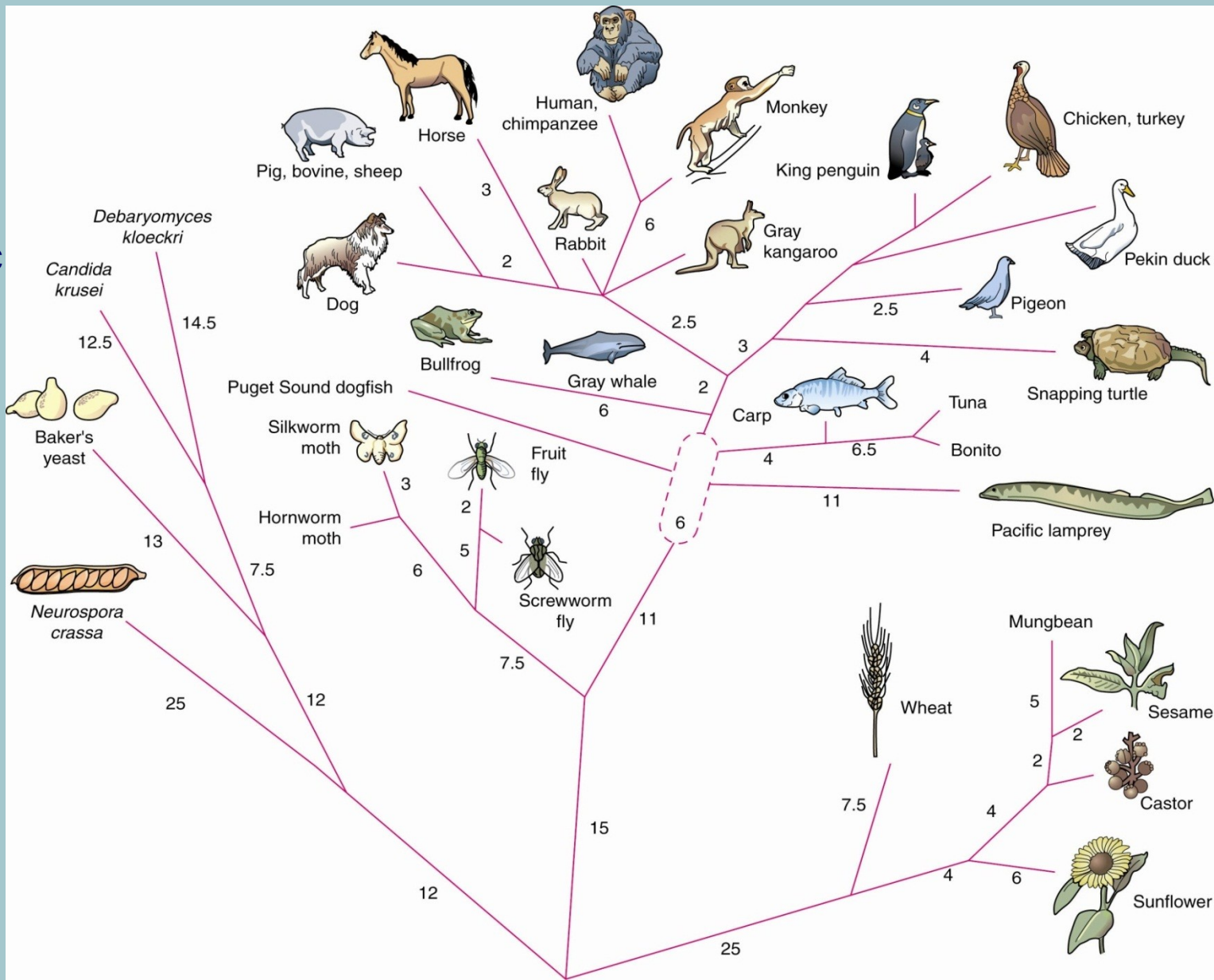
Phylogeny of Cytochrome c



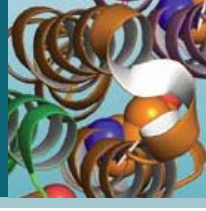
- The number of amino acid differences between two cytochrome c sequences is proportional to the phylogenetic difference between the species from which they are derived.
- This observation can be used to build phylogenetic trees of proteins.
- This is the basis for studies of molecular evolution.

Related Proteins Show a Common Evolutionary Origin

Fig. 5.20 This phylogenetic tree depicts the evolutionary relationships among eukaryotic organisms as determined by the similarity of their cytochrome c sequences. Numbers are the amino acid changes between hypothetical branch points and to the extant species at the tips.



Related Proteins Show a Common Evolutionary Origin



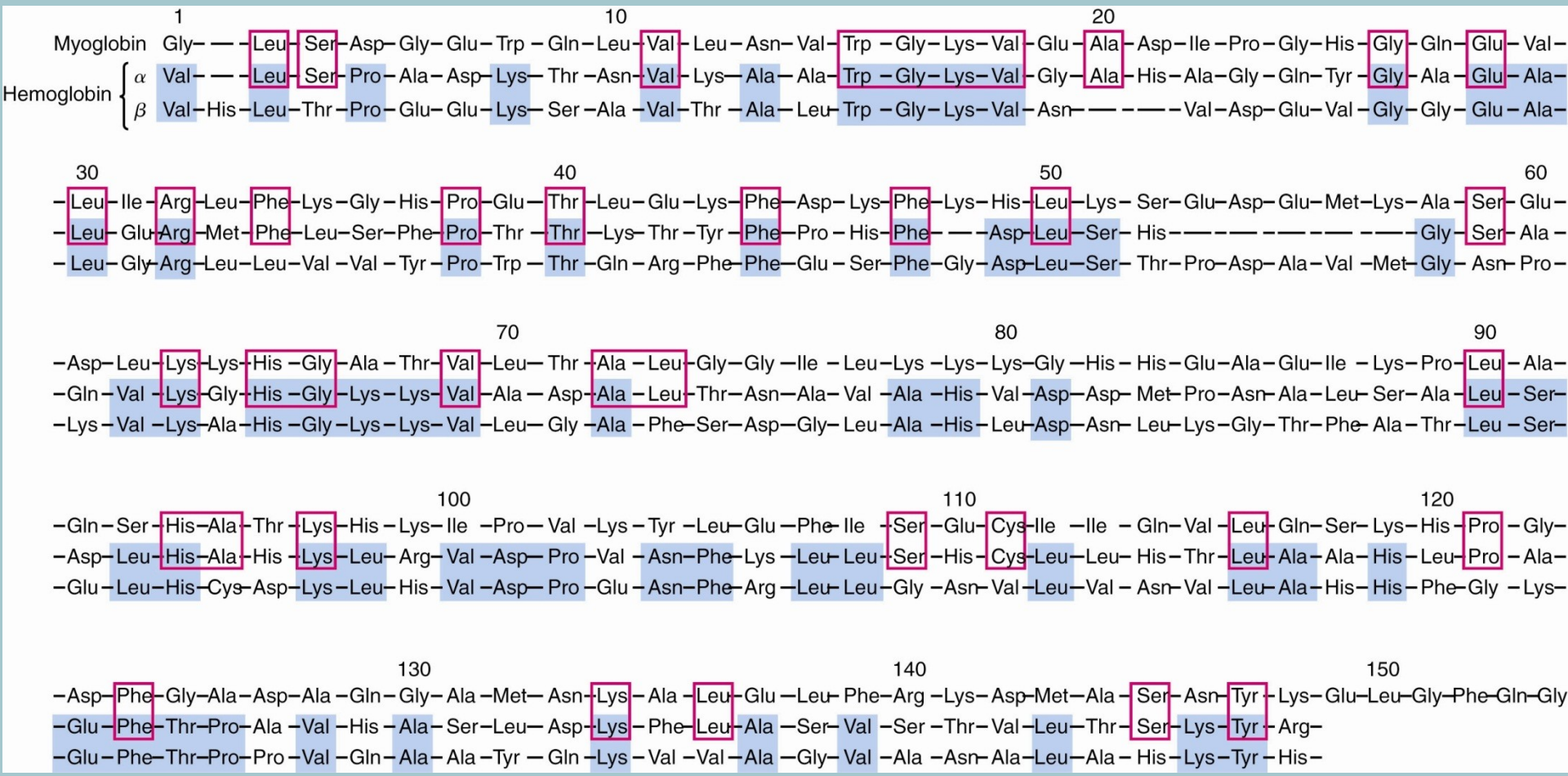


Figure 5.21 The amino acid sequences of the globin chains of human hemoglobin and myoglobin show a strong degree of homology.

Proteins with Different Function May Share a Common Answer

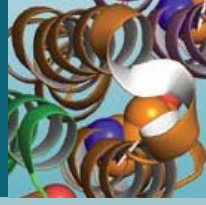
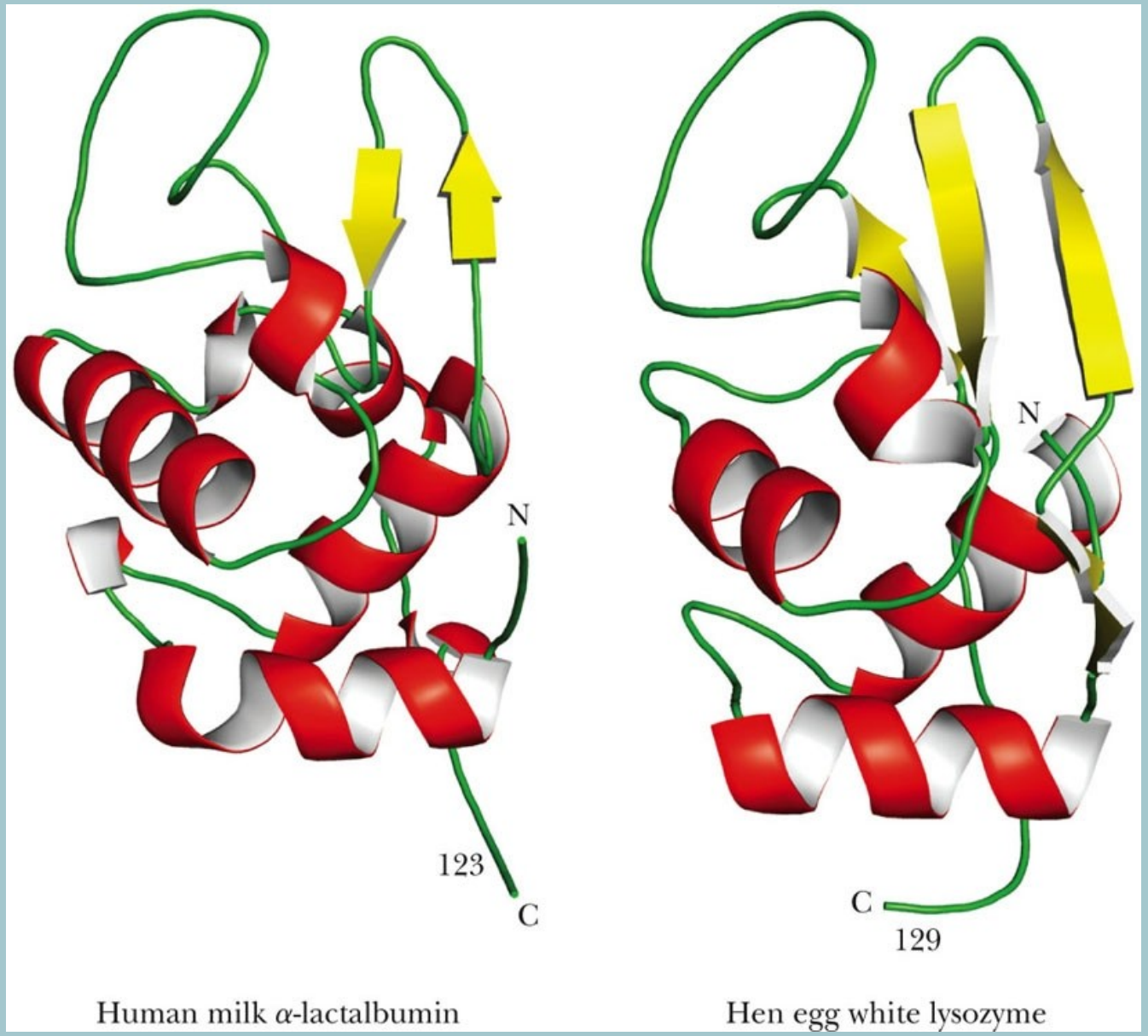
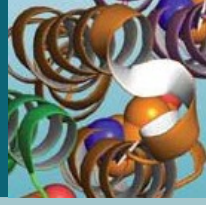


Figure 5.23 The tertiary structures of hen egg white lysozyme and human α-lactalbumin are very similar, and their sequences are identical at 48 positions.



Solid Phase Synthesis of Peptides



- R. Bruce Merrifield and his collaborators pioneered the solid-phase synthesis of polypeptides in the laboratory.
- Synthesis proceeds from C terminus to N terminus.
- Blocking groups must be added to functional groups of amino acids and later removed, to prevent unwanted side chain reactions.
- To begin, the carboxyl terminus of a residue is covalently anchored to an insoluble resin.
- After each addition of a residue, the resin particles are collected by filtration to remove reagents.
- Automation and computer control now permit synthesis of peptides of 30 residues or more, relatively cheaply.