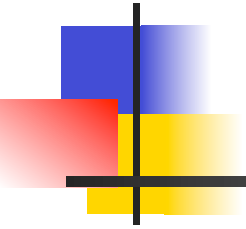


Building blocks and primary structure

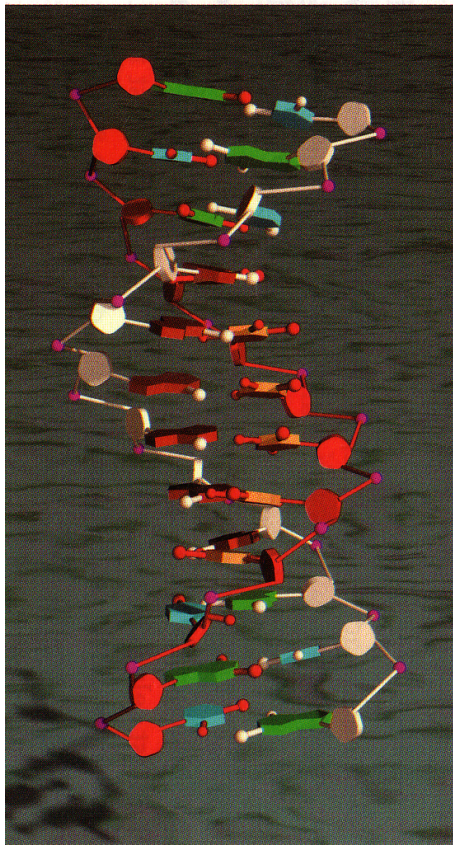


Why is the structure so important?

1. Interpretation of the mechanism of function of individual proteins
2. Approaches to the “protein folding problem”
3. Patterns of molecular evolution
4. Predictions of the structures of closely related proteins - homology modelling
5. Protein engineering:
 1. Modifications to probe mechanisms of function
 2. Attempts to enhance thermostability
 3. Clinical applications
 4. Catalytic antibodies
6. Drug design

Sequence = specific folding

A Sequence of Bases in DNA...



Triplets of bases read from one strand

Is Translated to a Sequence of Amino Acids in a Protein...

Three Bases

UUU	F	UCU	S	UAU	Y	UGU	C
UUC	F	UCC	S	UAC	Y	UGC	C
UUA	L	UCA	S	UAA	Stop	UGA	Stop
UUG	L	UCG	S	UAG	Stop	UGG	W
CUU	L	CCU	P	CAU	H	CGU	R
CUC	L	CCC	P	CAC	H	CGC	R
CUA	L	CCA	P	CAA	Q	CGA	R
CUG	L	CCG	P	CAG	Q	CGG	R
AUU	I	ACU	T	AAU	N	AGU	S
AUC	I	ACC	T	AAC	N	AGC	S
AUA	I	ACA	T	AAA	K	AGA	R
AUG	M	ACG	T	AAG	K	AGG	R
GUU	V	GCU	A	GAU	D	GGU	G
GUC	V	GCC	A	GAC	D	GGC	G
GUA	V	GCA	A	GAA	E	GGA	G
GUG	V	GCG	A	GAG	E	GGG	G

One Amino Acid

Genetic Code
'Translation Table'

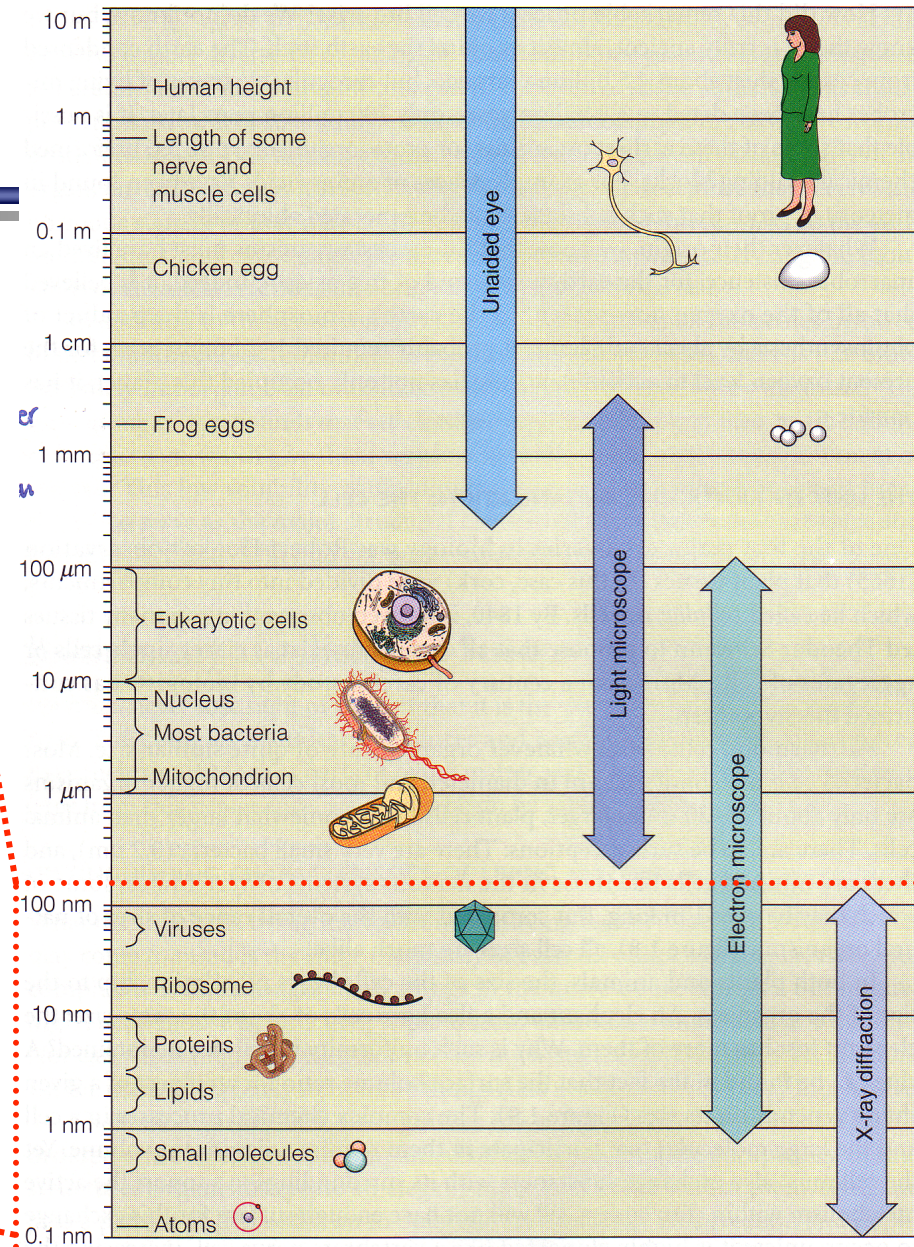
Which Folds Spontaneously to a Precise Three-Dimensional Structure

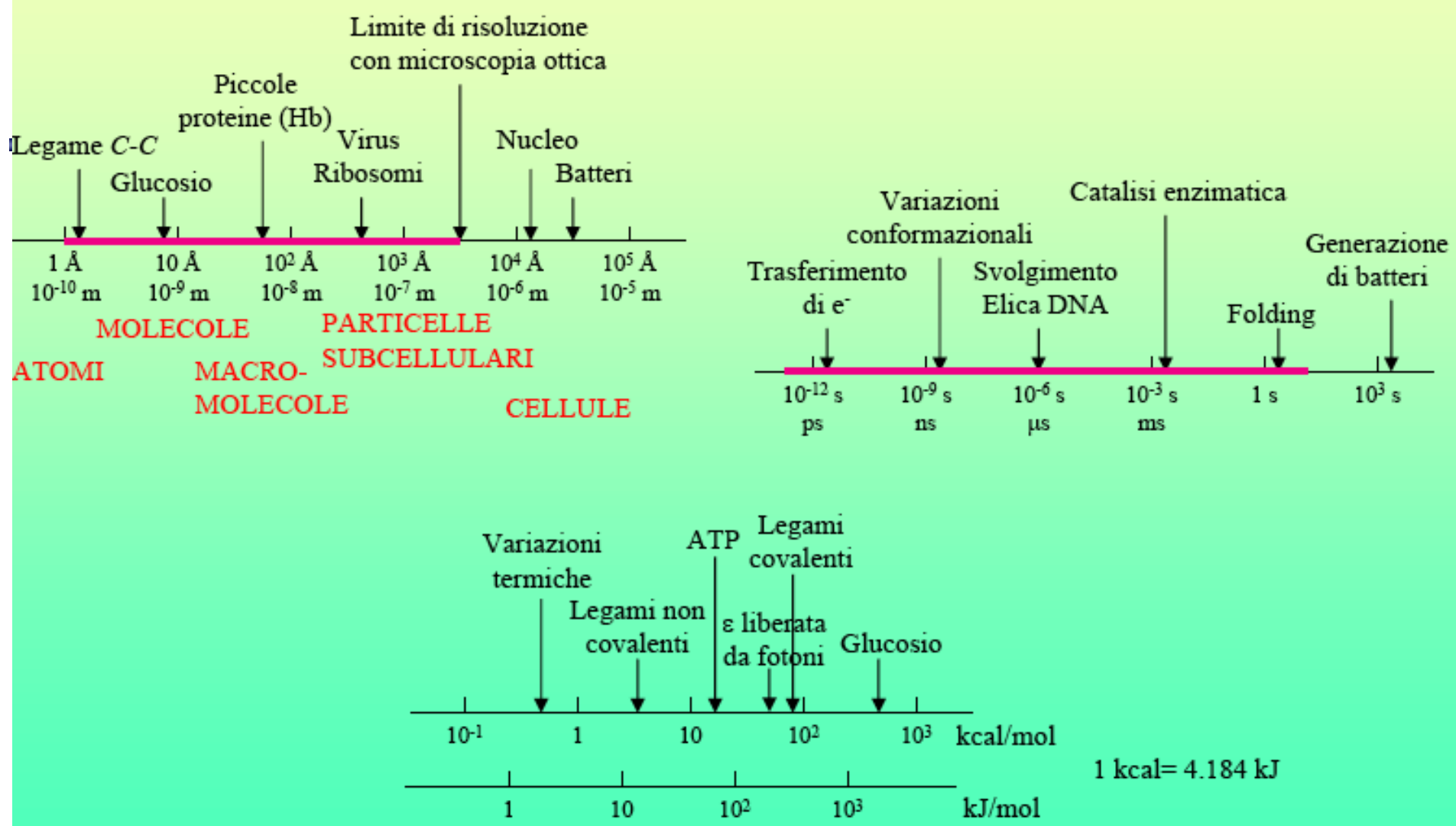


Range and sizes in biochemistry

Techniques for study in biochemistry / structural biology:

- Cryo-electron microscopy (Cryo-EM)
- Atomic Force Microscopy (AFM) and Scanning Tunneling Microscopy (STM)
- Crystallography and X-ray diffraction
- Nuclear Magnetic Resonance (NMR)
- Circular dichroism
- Fluorescence
- Raman scattering, Electron spin resonance, Mossbauer spectroscopy, Infra-red spectroscopy





Methods of structure determination

- X-ray crystallography and X-ray diffraction:
 - Crystallisation
 - Diffraction (information on intensity of scattered x-rays)
 - Phase solving:
 - Isomorphous replacement (heavy atoms)
 - Molecular replacement
 - Model building and refinement
 - Resolution (Å) = accuracy
 - R-factor (%) = how close the model reproduces the experimental electron density

Resolution of structures

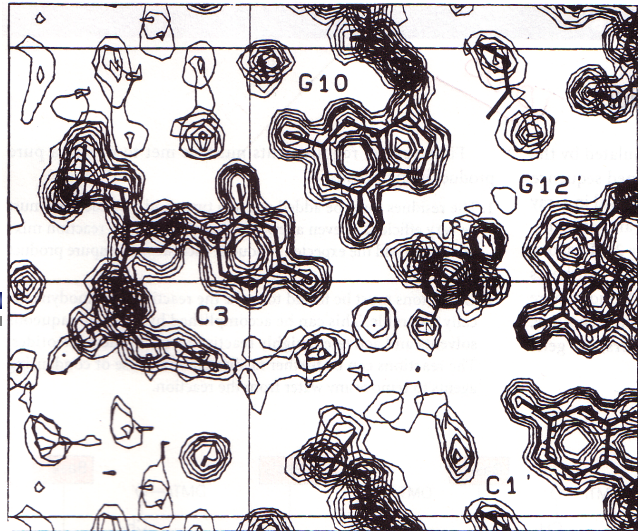


FIGURE 4A.6
Part of an electron density map derived from the DNA crystal diffraction pattern in Figure 4A.5.
Reprinted by permission from F. L. Suddath, *Nature* (1974) 248:20-24, fig. 3, p. 22. © 1974 Macmillan Magazines Ltd.

Confidence in structural features of proteins determined by X-ray crystallography

(These are *rough* estimates, and depend strongly on the quality of the data.)

Structural feature	Resolution				
	5 Å	3 Å	2.5 Å	2.0 Å	1.5 Å
Chain tracing	—	Fair	Good	Good	Good
Secondary structure	Helices	fair	Fair	Good	Good
Sidechain conformations	—	—	Fair	Good	Good
Orientation of peptide planes	—	—	Fair	Good	Good
Protein hydrogen atoms visible	—	—	—	—	Good

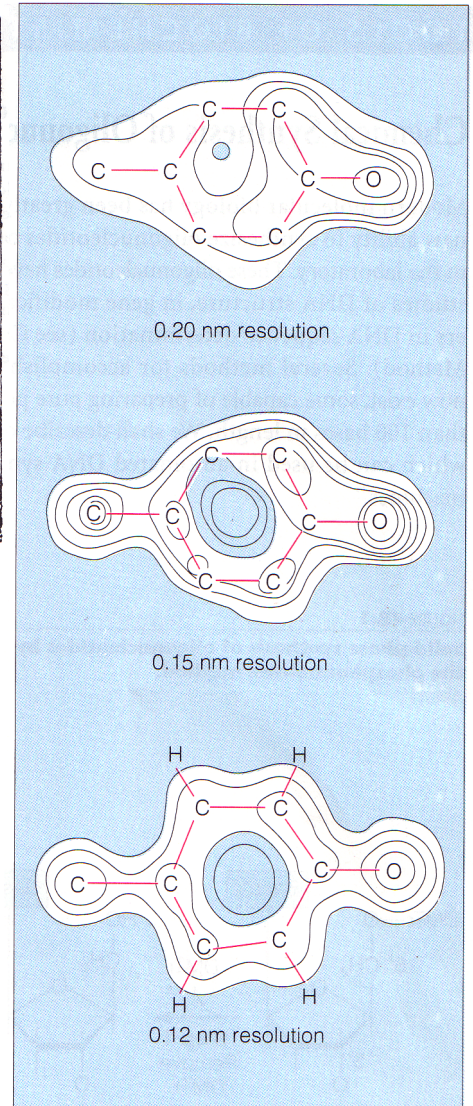


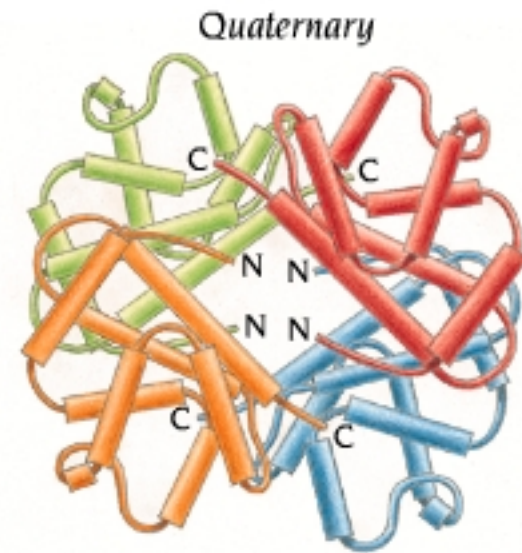
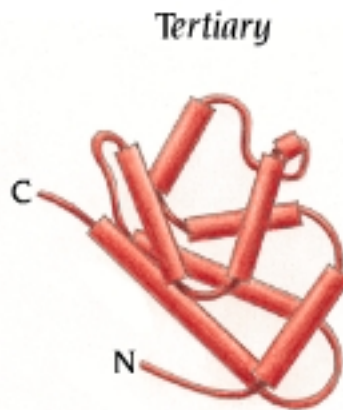
FIGURE 4A.7
Effect of increased resolution on molecular detail observed by x-ray diffraction.

From K. D. Watenpaugh, L. K. Sieker, L. H. Jensen, *J. Mol. Biol.* (1980) 138:615-633. © 1980 Academic Press.

cont.

- Electron diffraction:
 - Developed for membrane proteins that form regular 2D crystals (or arrays) - Henderson, bacterio-rhodopsin
- Neutron diffraction:
 - Can reveal H-atoms and can un-ambiguously identify water molecules
- Multi-dimensional NMR spectroscopy
 - Does not require crystallisation, done in solution, but size-limits
- Cryo-EM
 - Low temperature electron microscopy - Suitable for large assemblies (100-500 Å in diameter) - Van Heel, ribosome. No atoms are seen, but low resolution aggregates (3-4 Å).

Proteins are polymers of amino acids



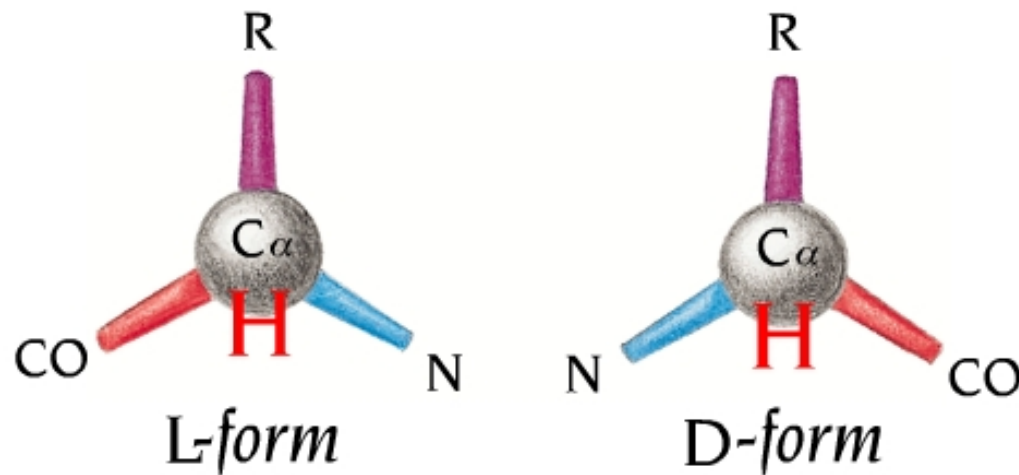
Regions of sequence with regular structure

AA sequence

Packing of elements of 2nd-ary structure into one or more compact units called domains

Polypeptide chains associated in functional assemblies

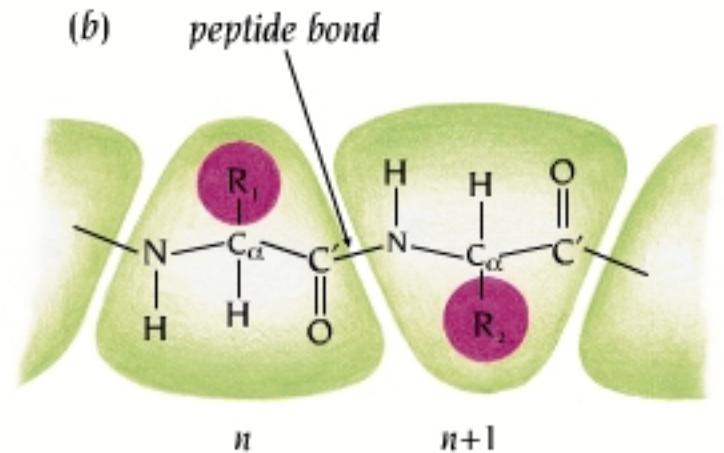
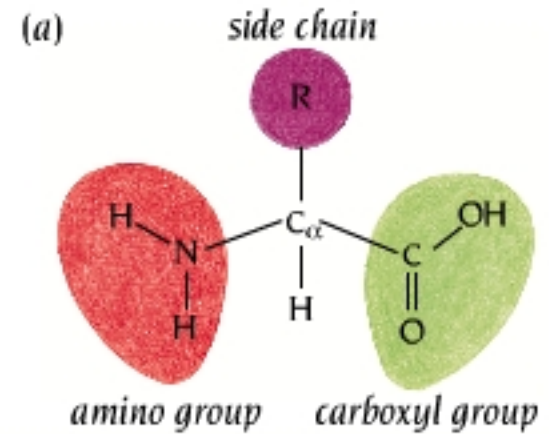
- AA are chiral



R
(CORN)

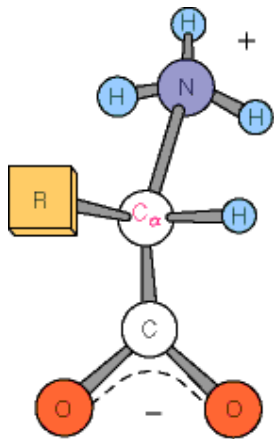
S

©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

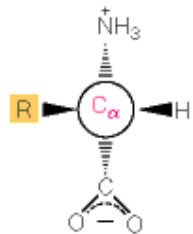


©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

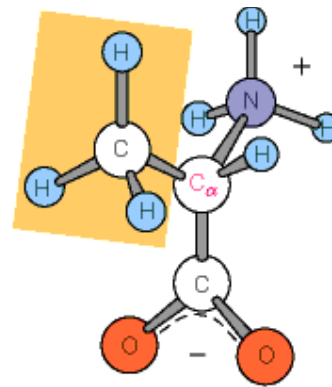
Stereochemistry of α -amino acids



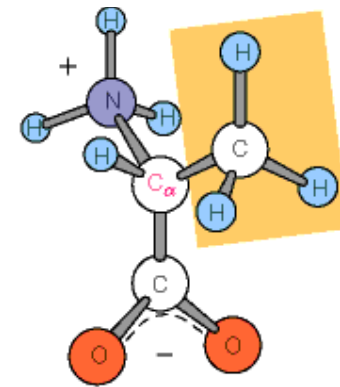
(a) α -Amino acid



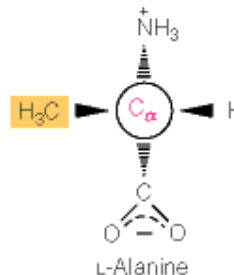
(b) Compact representation



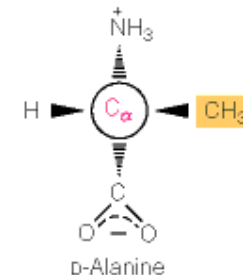
(a) L-Alanine



D-Alanine

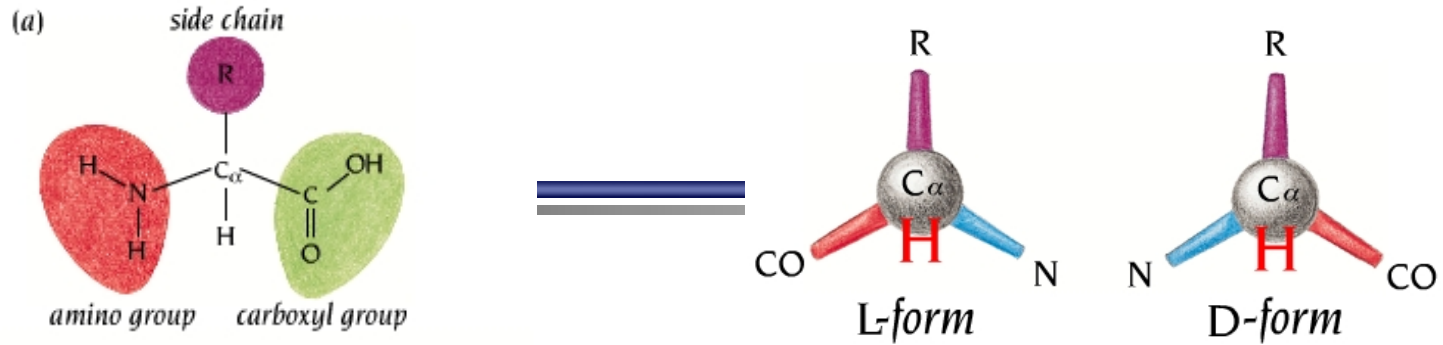


(b) L-Alanine



D-Alanine

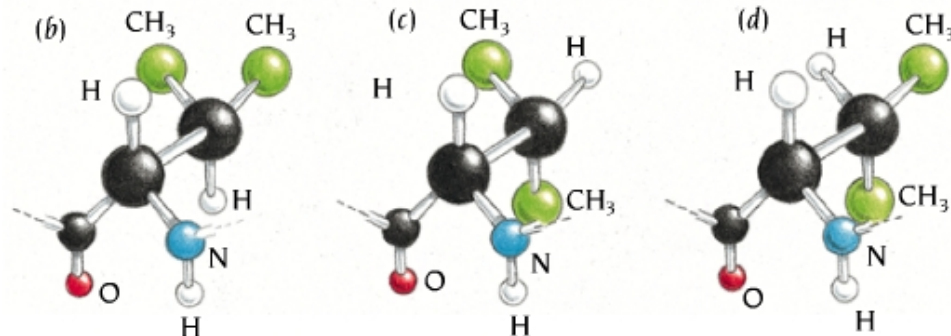
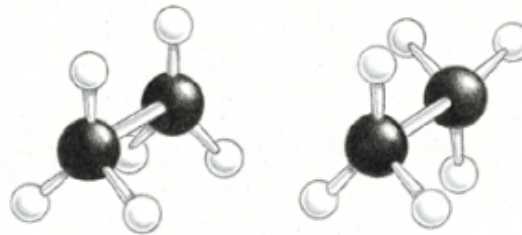
Aminoacids: classification and properties.



©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

(a) not staggered

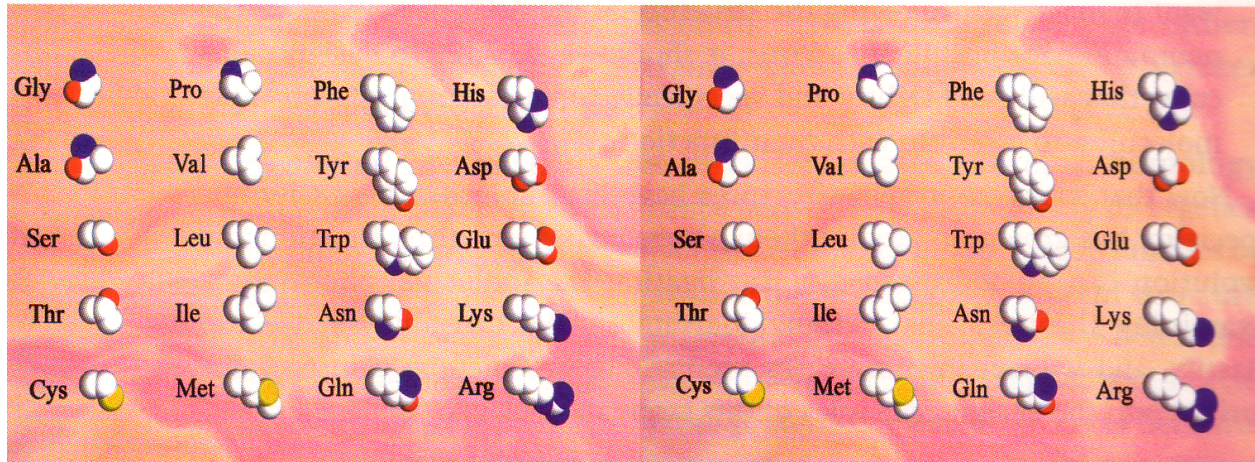
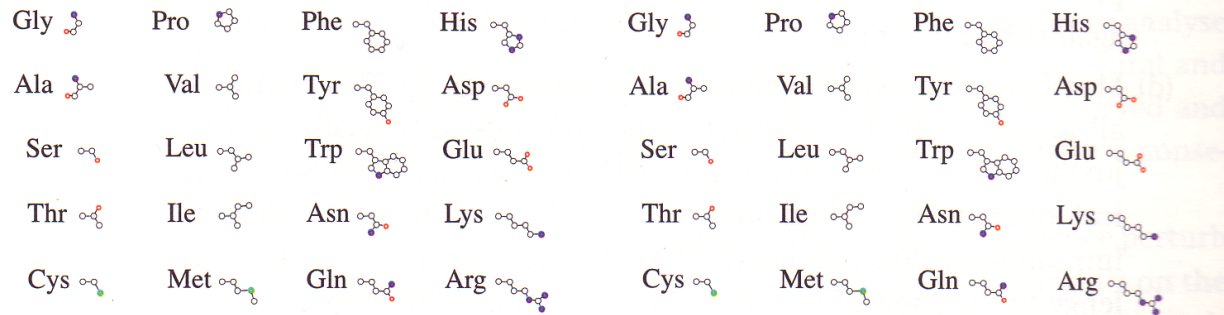
staggered



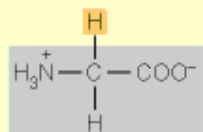
©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

Different side-chains = different properties

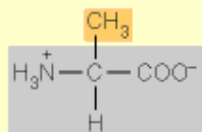
- 1 letter and 3 letters codes



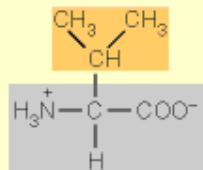
ALIPHATIC AMINO ACIDS



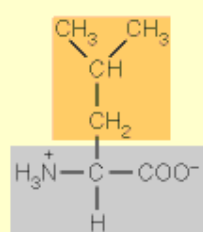
Glycine (Gly) G



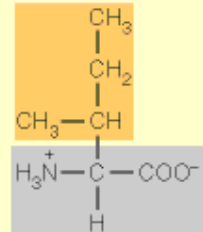
Alanine (Ala) A



Valine (Val) V

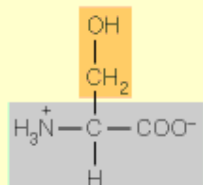


Leucine (Leu) L

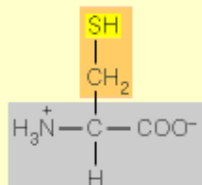


Isoleucine (Ile) I

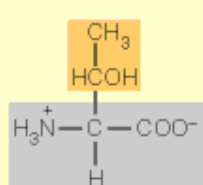
AMINO ACIDS WITH HYDROXYL- OR SULFUR-CONTAINING SIDE CHAINS



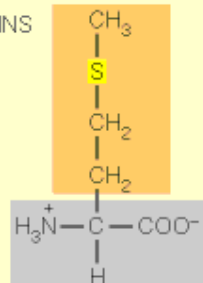
Serine (Ser) S



Cysteine (Cys) C

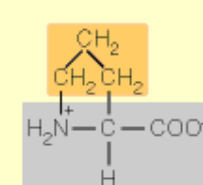


Threonine (Thr) T



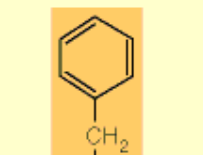
Methionine (Met) M

CYCLIC AMINO ACID

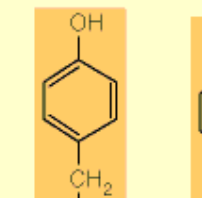


Proline (Pro) P

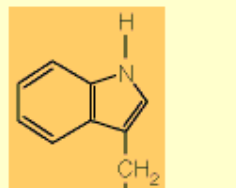
AROMATIC AMINO ACIDS



Phenylalanine (Phe) F

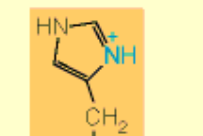


Tyrosine (Tyr) Y

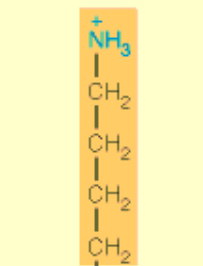


Tryptophan (Trp) W

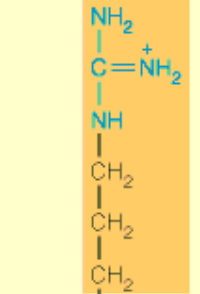
BASIC AMINO ACIDS



Histidine (His) H

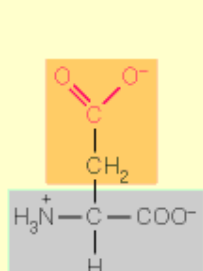


Lysine (Lys) K

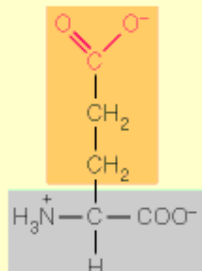


Arginine (Arg) R

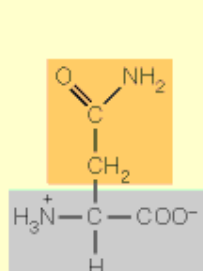
ACIDIC AMINO ACIDS AND THEIR AMIDES



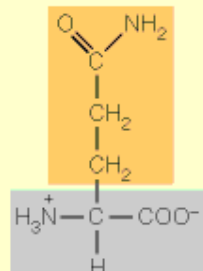
Aspartic acid (Asp) D



Glutamic acid (Glu) E



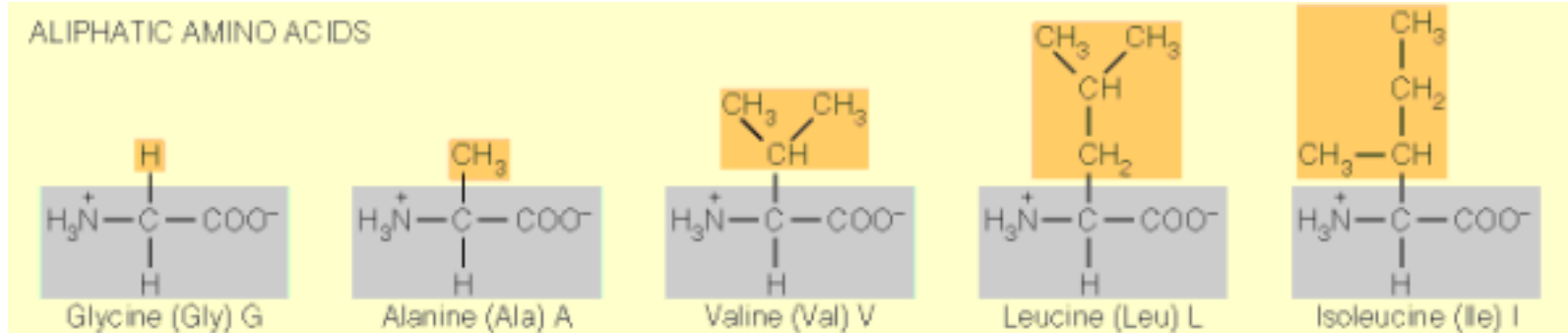
Asparagine (Asn) N



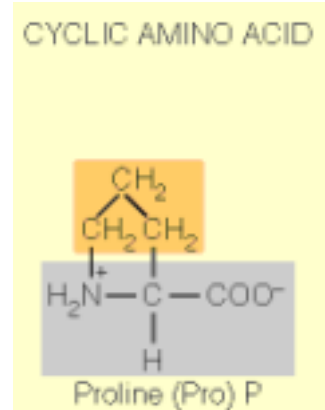
Glutamine (Gln) Q

Properties of aa side-chains

- Aliphatic aa:

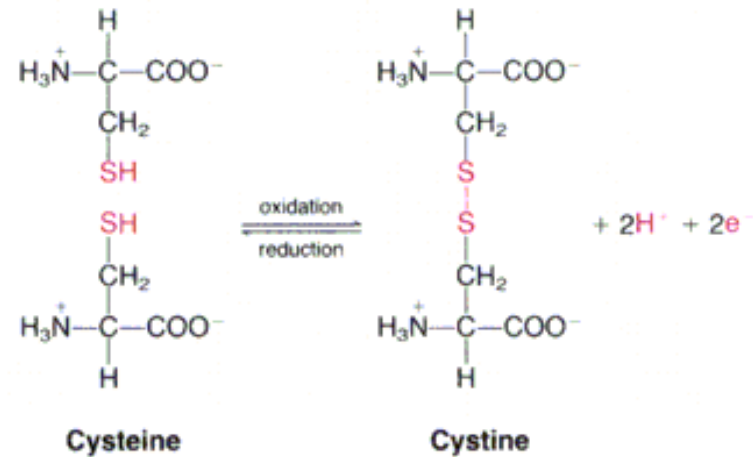
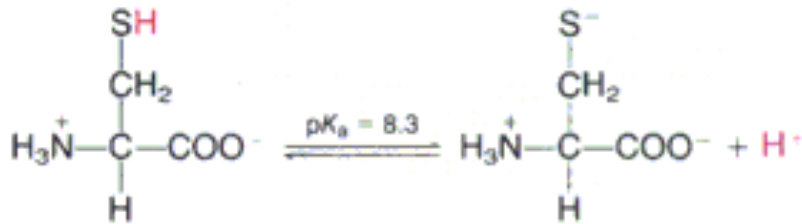
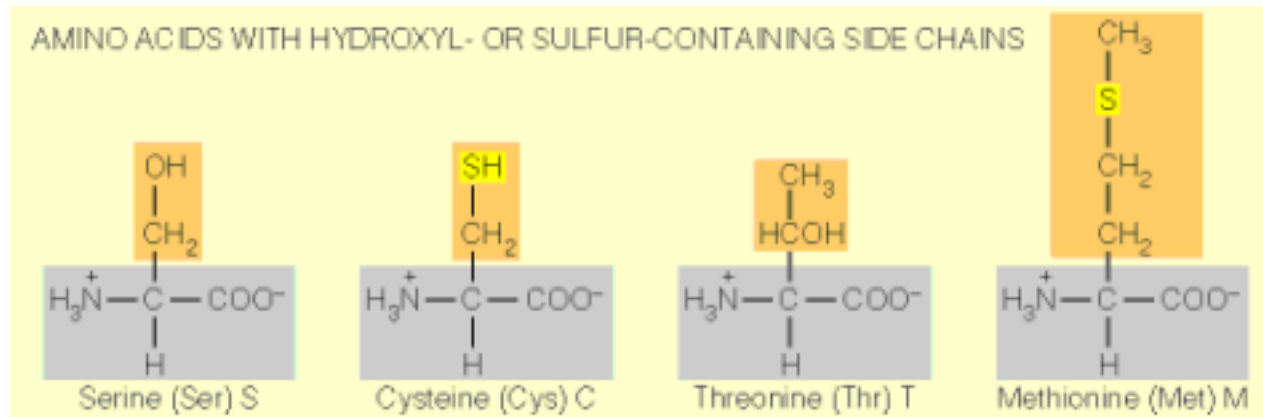


- Flexible and hydrophobic, the most Ile;
- Found in the hydrophobic core of proteins
- Pro: exception, rigid ring as side chain, often a structure breaker



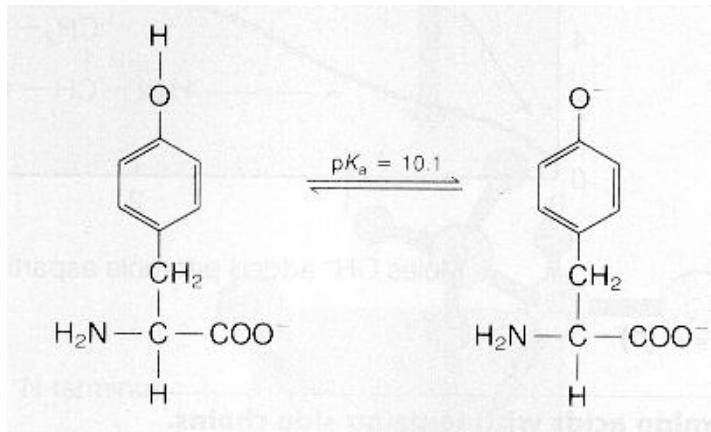
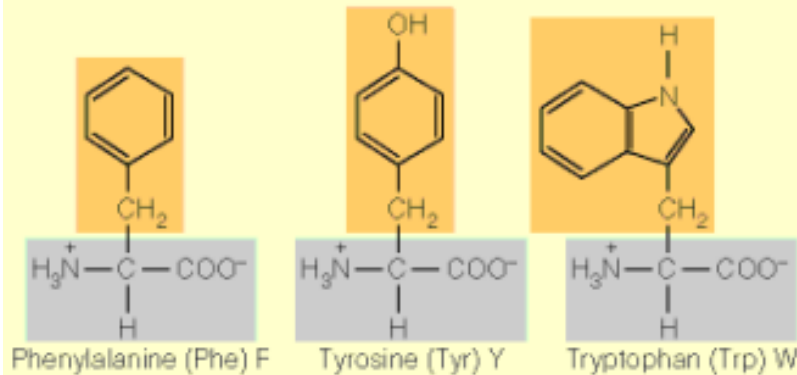
Hydroxyl-, sulfur- side chains

- Weakly polar side chains

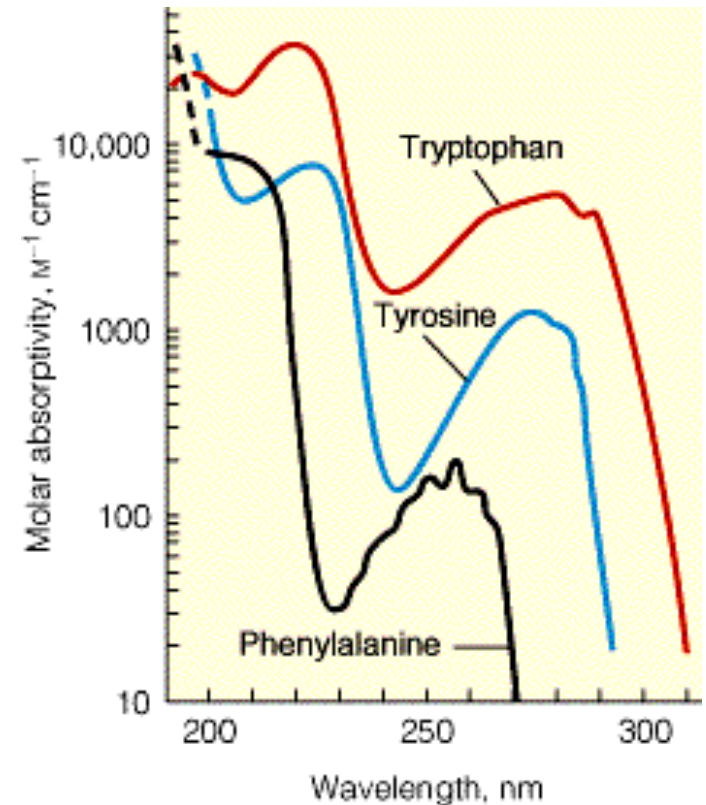


Aromatic side-chains

AROMATIC AMINO ACIDS



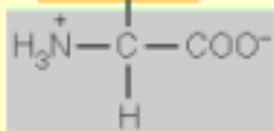
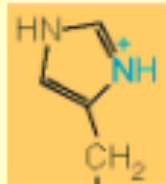
Absorbance



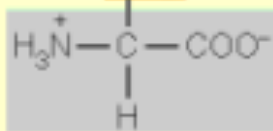
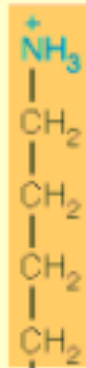
Detection of proteins

Basic side-chains

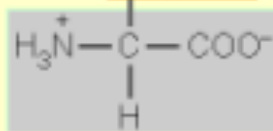
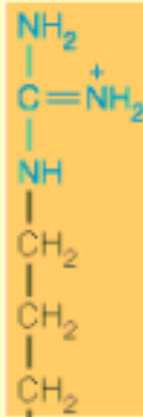
BASIC AMINO ACIDS



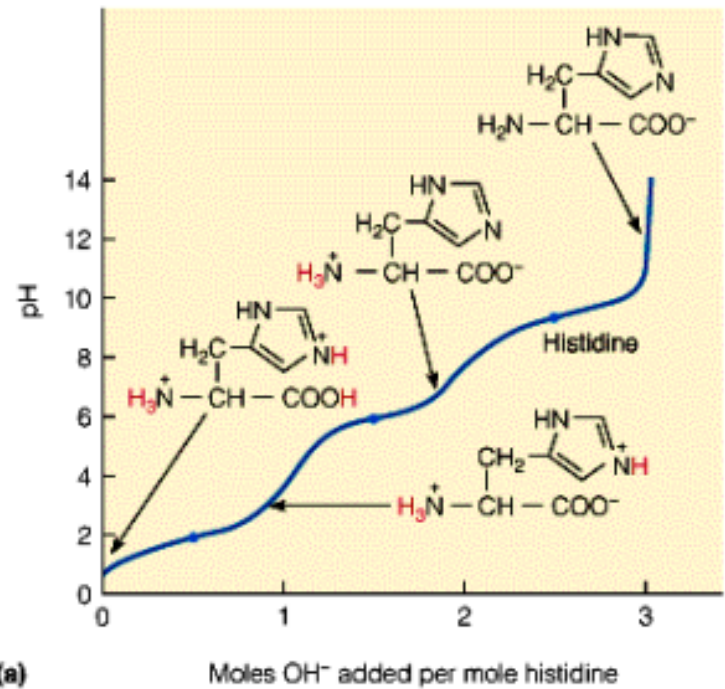
Histidine (His) H



Lysine (Lys) K



Arginine (Arg) R



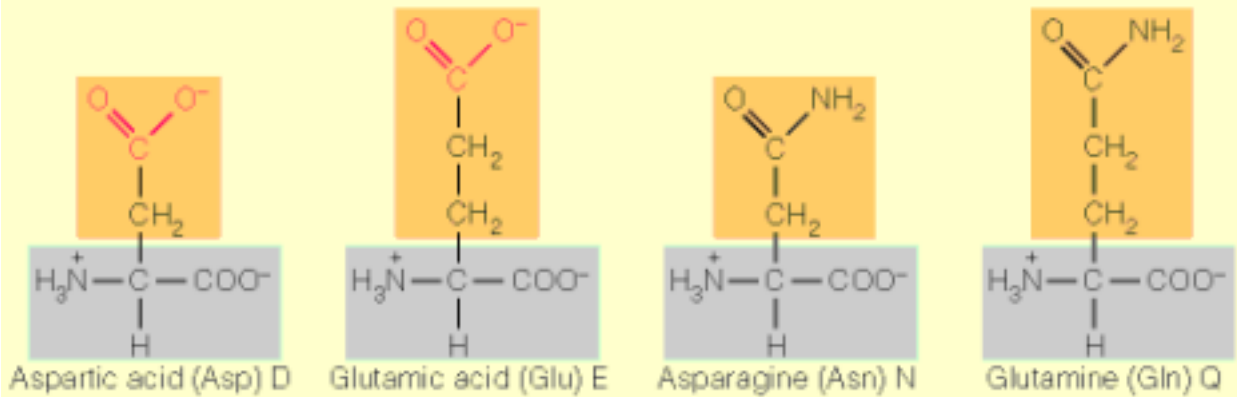
Typical pKa Ranges

Group Type	Typical pK _a Range ^a
α-Carboxyl	3.5–4.0
Side chain carboxyls of aspartic and glutamic acids	4.0–4.8
Imidazole (histidine)	6.5–7.4
Cysteine (—SH)	8.5–9.0
Phenolic (tyrosine)	9.5–10.5
α-Amino	8.0–9.0
Side chain amino (lysine)	9.8–10.4
Guanidinyll (arginine)	~12

^aValues outside these ranges are observed. For example, side chain carboxyls have been reported with pK_a values as high as 7.3.

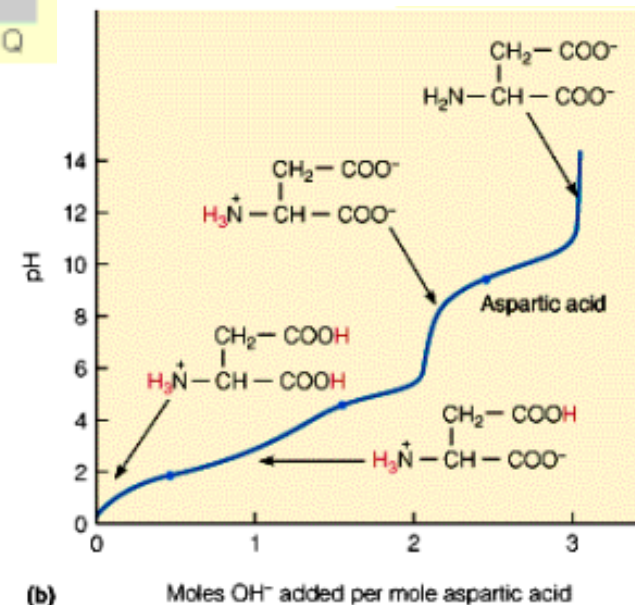
Acidic side-chains and amides

ACIDIC AMINO ACIDS AND THEIR AMIDES



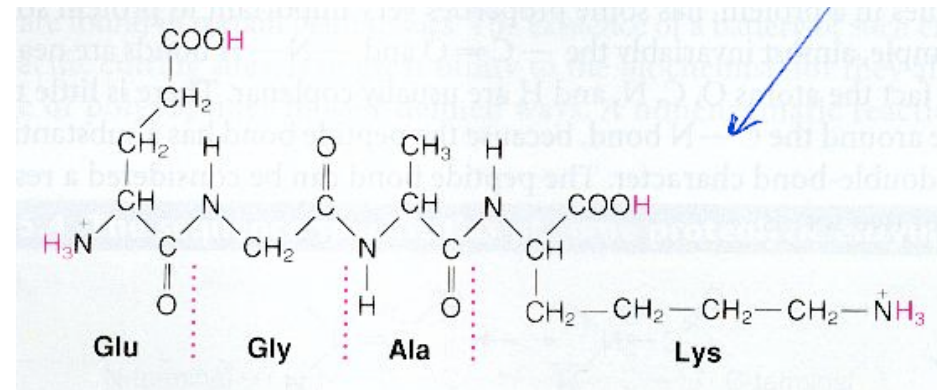
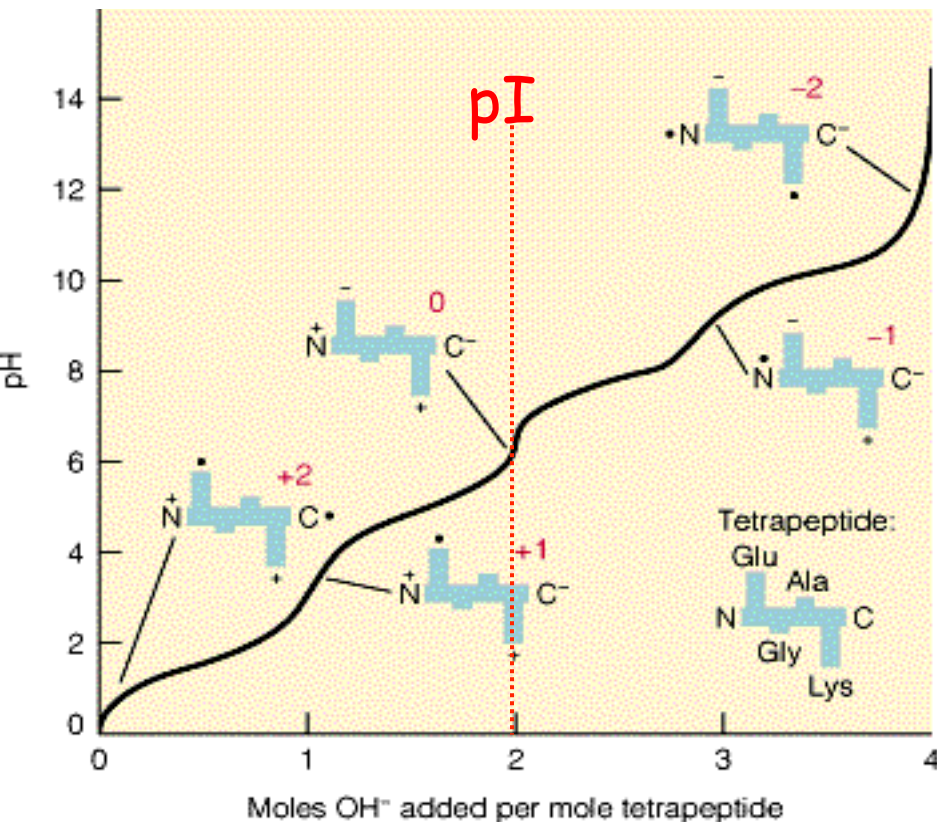
Group Type	Typical pK _a Range ^a
α-Carboxyl	3.5–4.0
Side chain carboxyls of aspartic and glutamic acids	4.0–4.8
Imidazole (histidine)	6.5–7.4
Cysteine (—SH)	8.5–9.0
Phenolic (tyrosine)	9.5–10.5
α-Amino	8.0–9.0
Side chain amino (lysine)	9.8–10.4
Guanidinyll (arginine)	~12

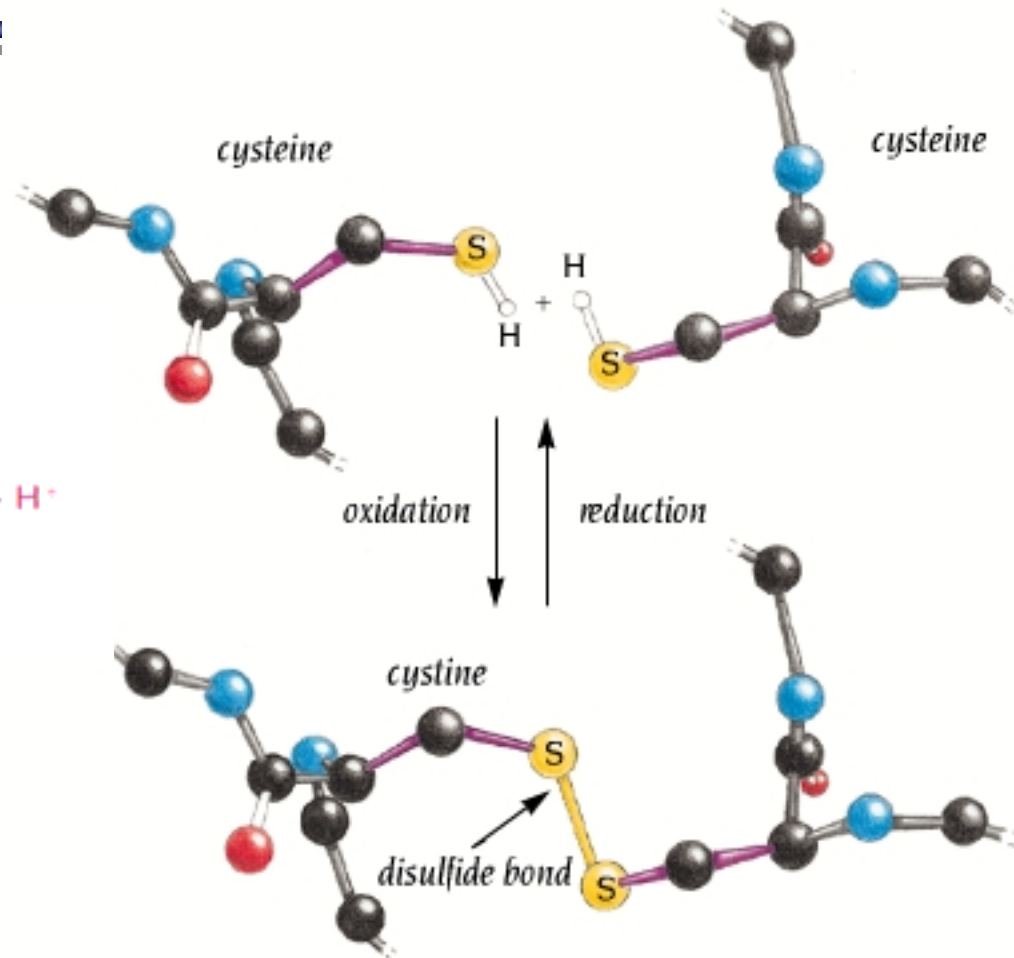
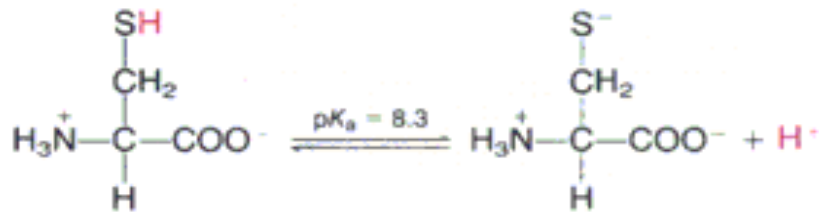
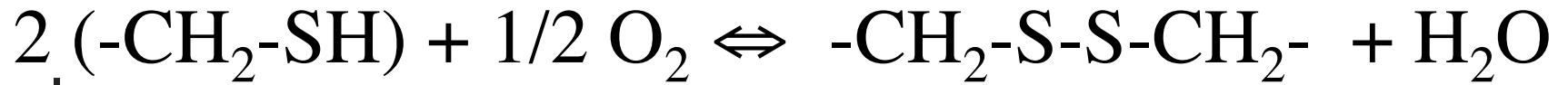
^aValues outside these ranges are observed. For example, side chain carboxyls have been reported with pK_a values as high as 7.3.



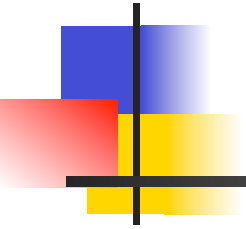
Proteins are polyampholites

- pI: isoelectric point





Analysis of proteins' primary structure:



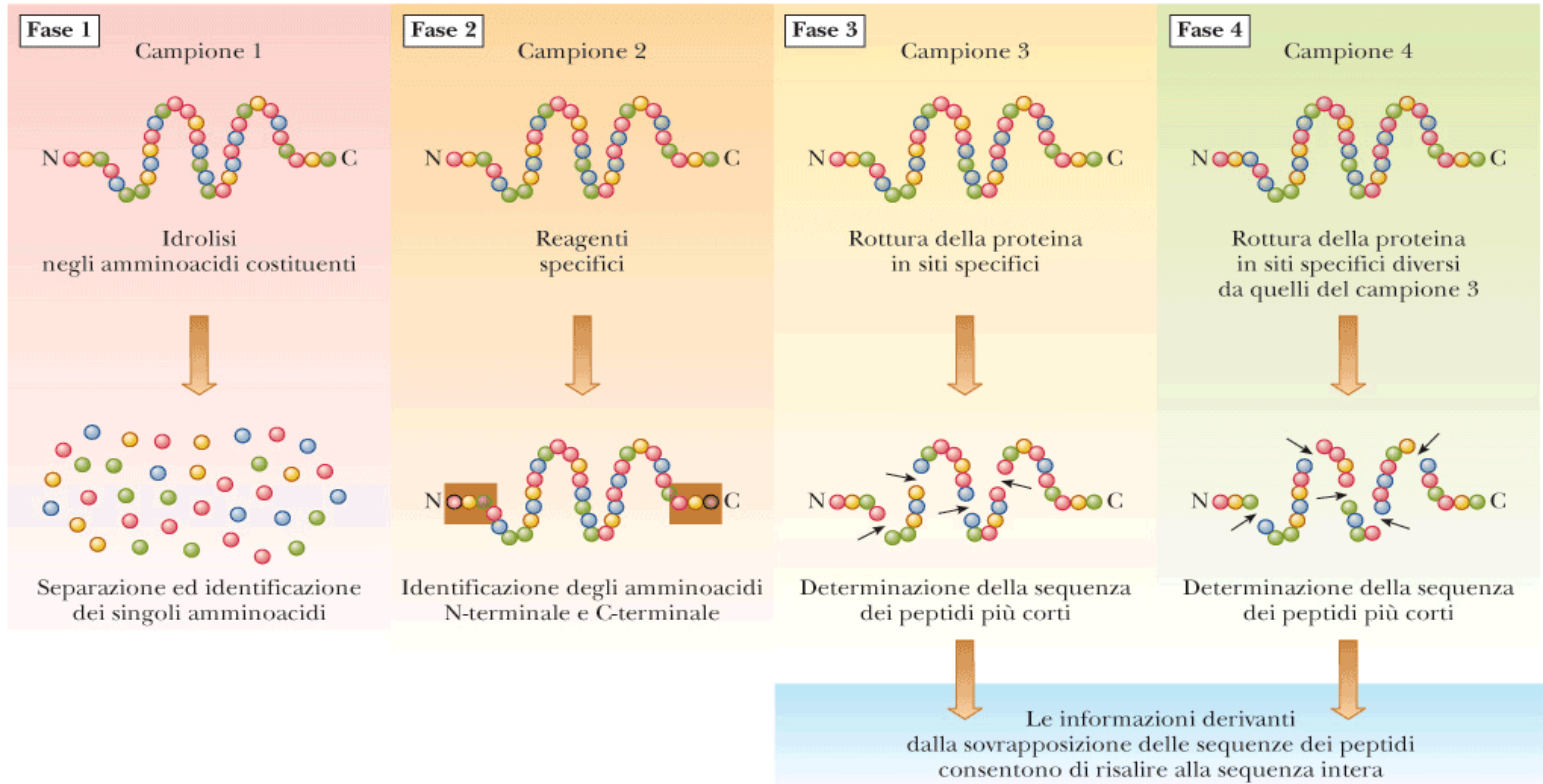
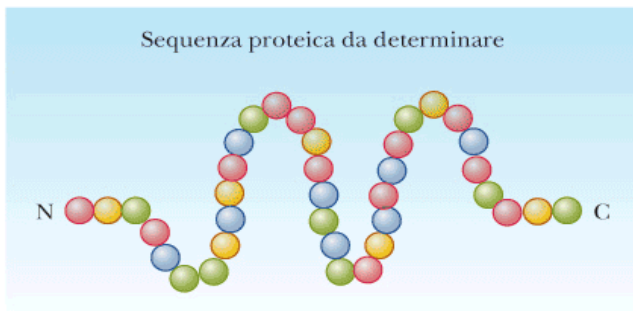
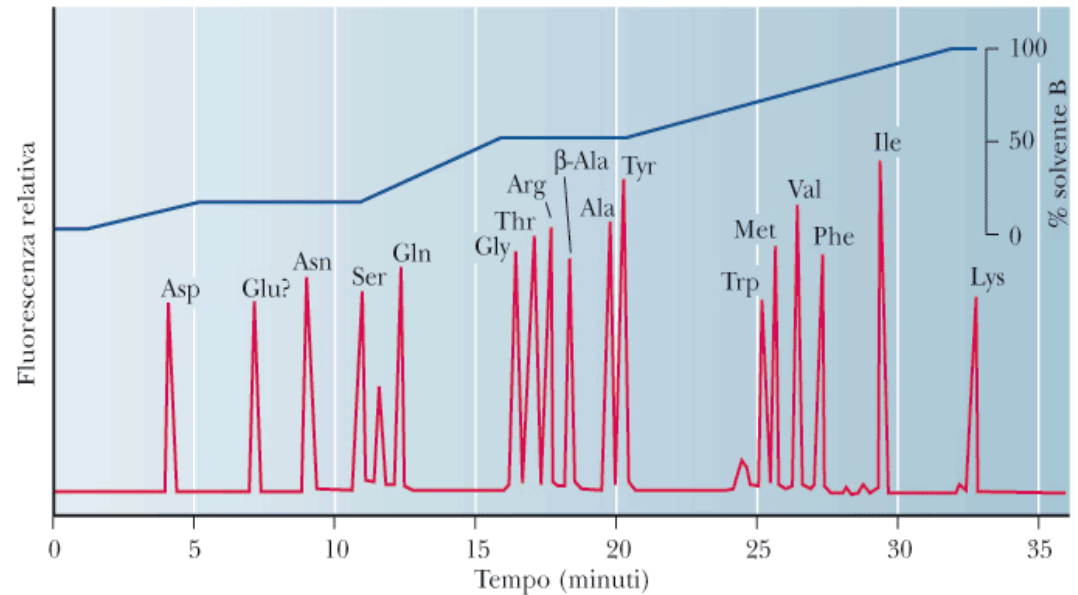


FIGURA 5.14 La strategia per determinare la struttura primaria di una data proteina. La sequenza di aminoacidi può essere determinata con quattro analisi differenti, eseguite su quattro campioni separati della stessa proteina.

Chemical degradation

1. HCl:
Complete degradation,
6M HCl at
100-110 °C
for 12-36 h



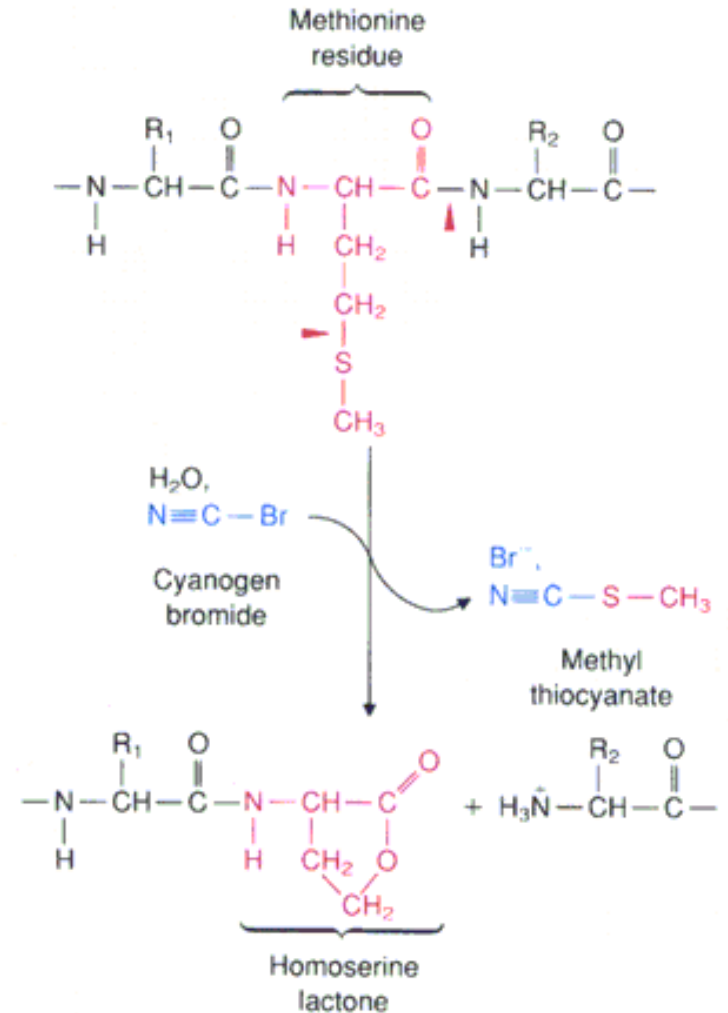
■ **FIGURA 5.15** Cromatogramma HPLC relativo alla separazione di una miscela di aminoacidi.

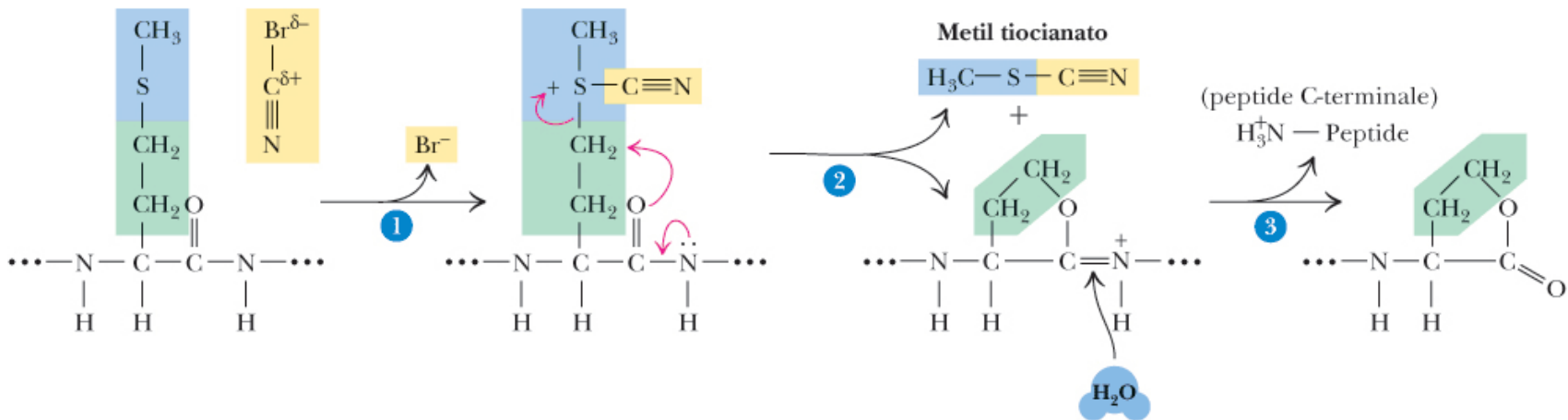


Campbell - Farrell
Biochimica
EdiSES

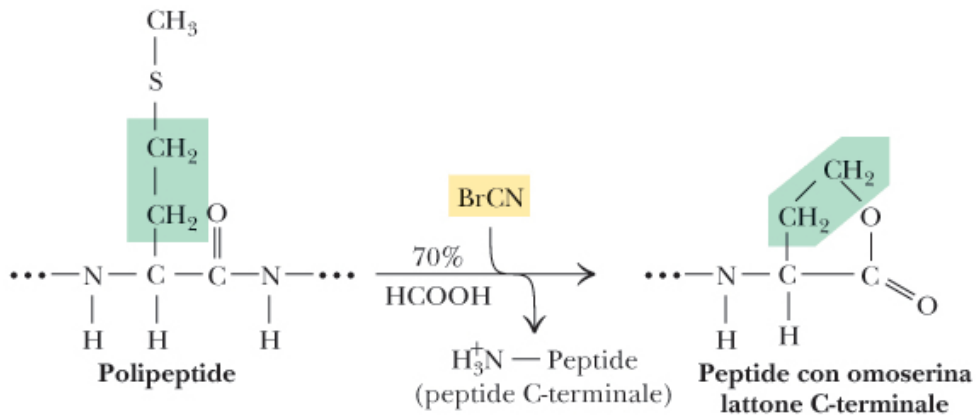
Chemical degradation

2. Cyanogen bromide



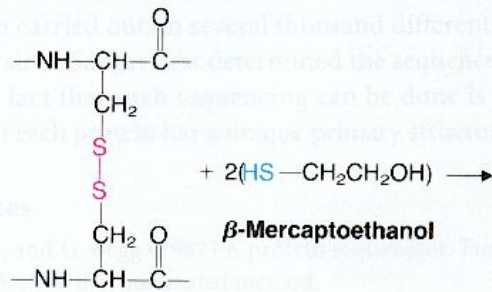


REAZIONE COMPLESSIVA:



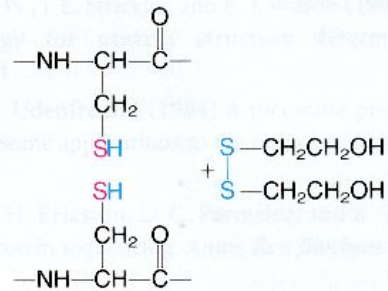
Campbell - Farrell
Biochimica
 EdiSES

Protein hydrolysis by cyanogen bromide takes place where methionine residues are

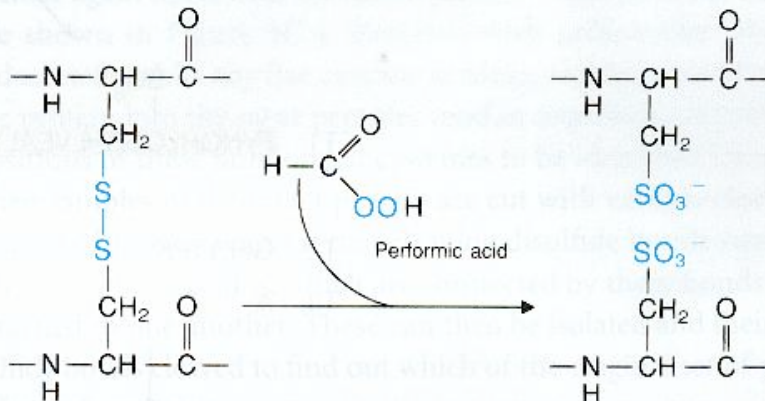


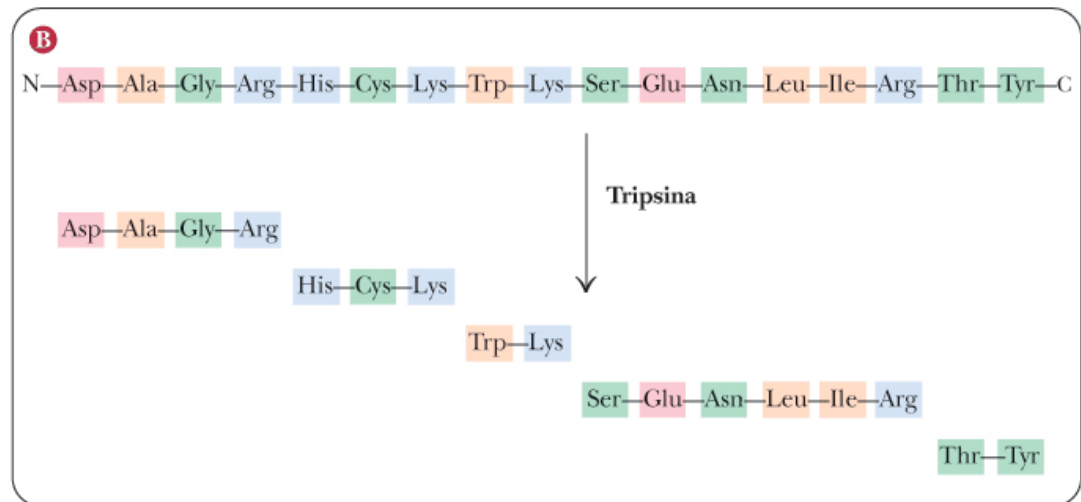
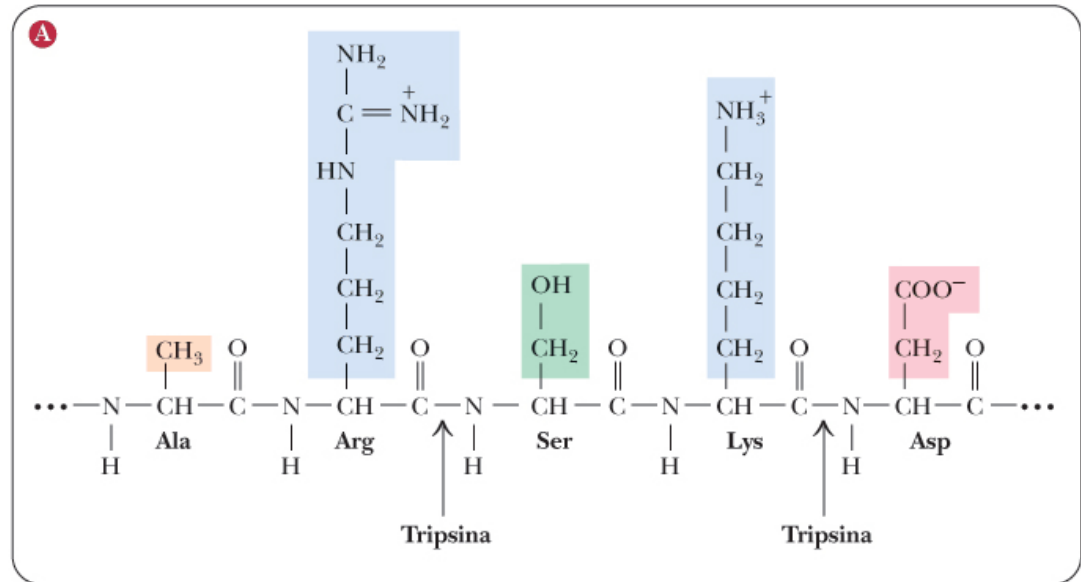
β-Mercaptoethanol

3. β-mercaptoethanol

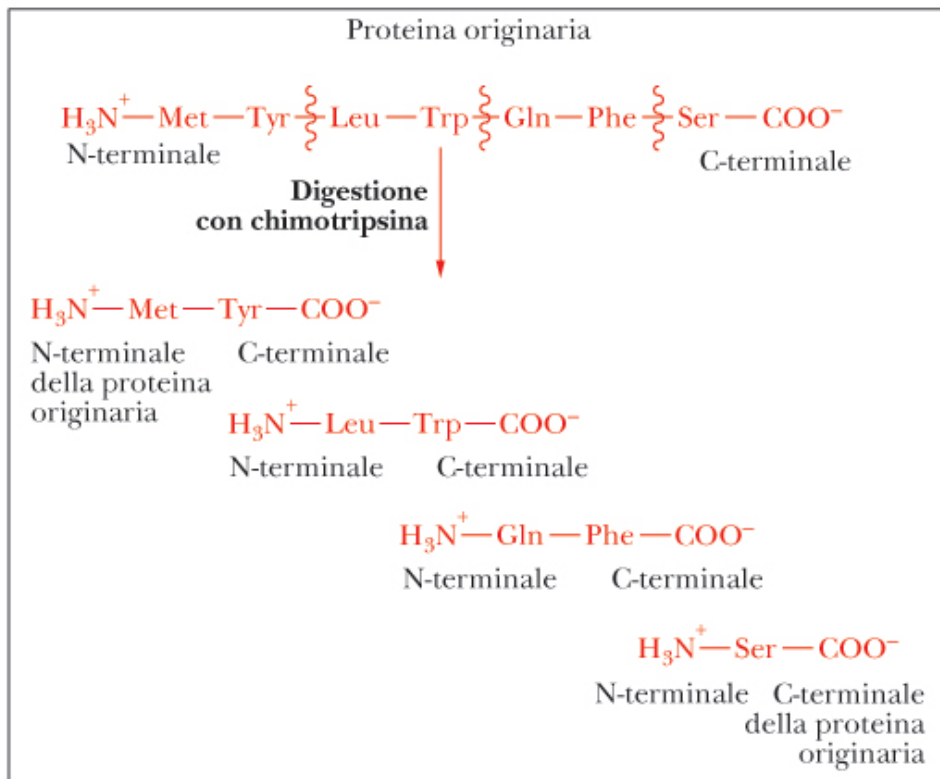


4. Performic acid





Peptide digestion with trypsin. A) Trypsin is a proteolytic enzyme, or protease, that cuts specifically only peptide bonds where arginine or lysine provide the carbonyl group. B) The reaction products are a mixture of peptide fragments with Arg and Lys as C-term amino acids and a single peptide deriving from the C-terminal of the polypeptide chain.

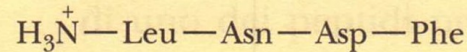


Campbell - Farrell
Biochimica
EdiSES

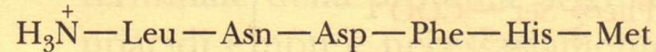
Chimotrypsin digestion of a protein. Chimotrypsin hydrolyses proteins where aromatic amino acids are

The overlap of the sequences of fragments allows to determine the protein sequence

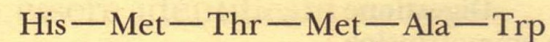
Chimotripsina



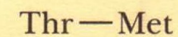
Bromuro di cianogeno



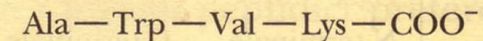
Chimotripsina



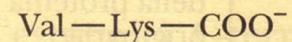
Bromuro di cianogeno



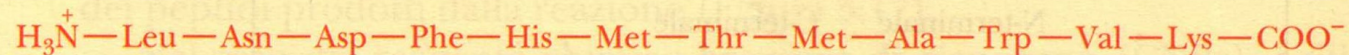
Bromuro di cianogeno



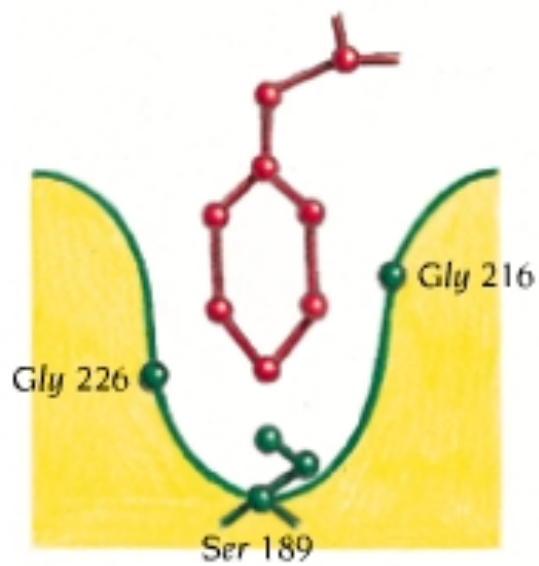
Chimotripsina



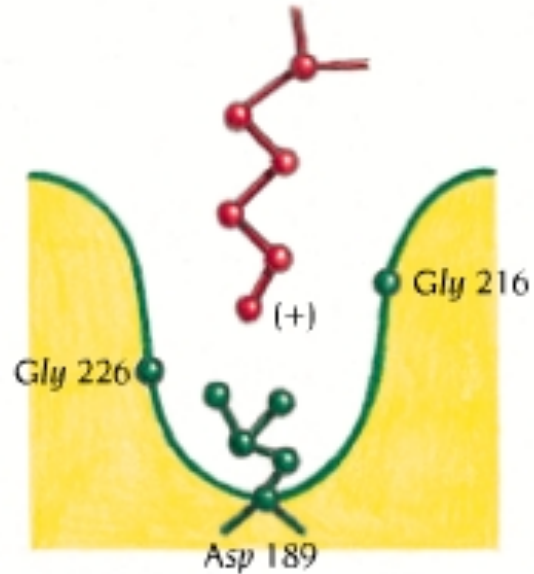
Sequenza complessiva



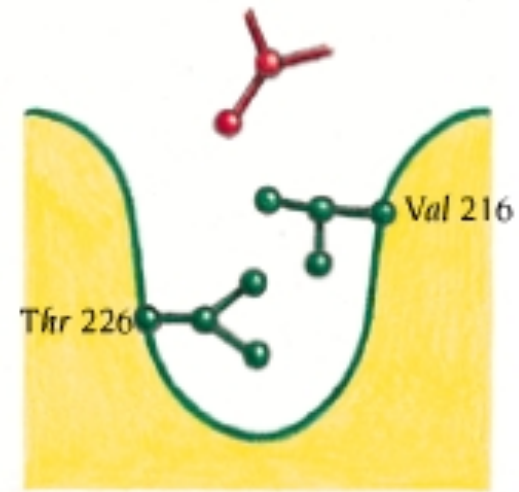
Protease active sites



Chymotrypsin



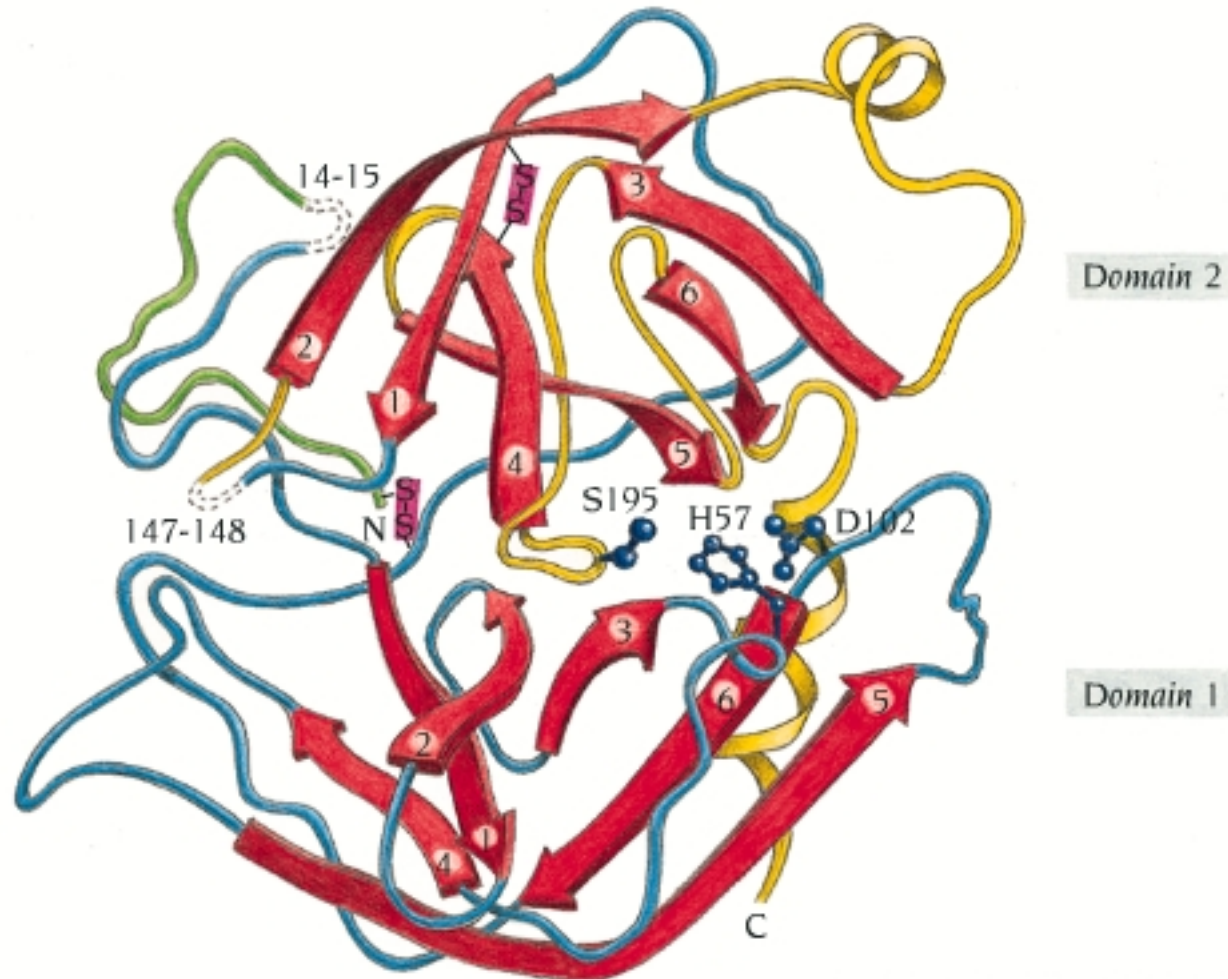
Trypsin



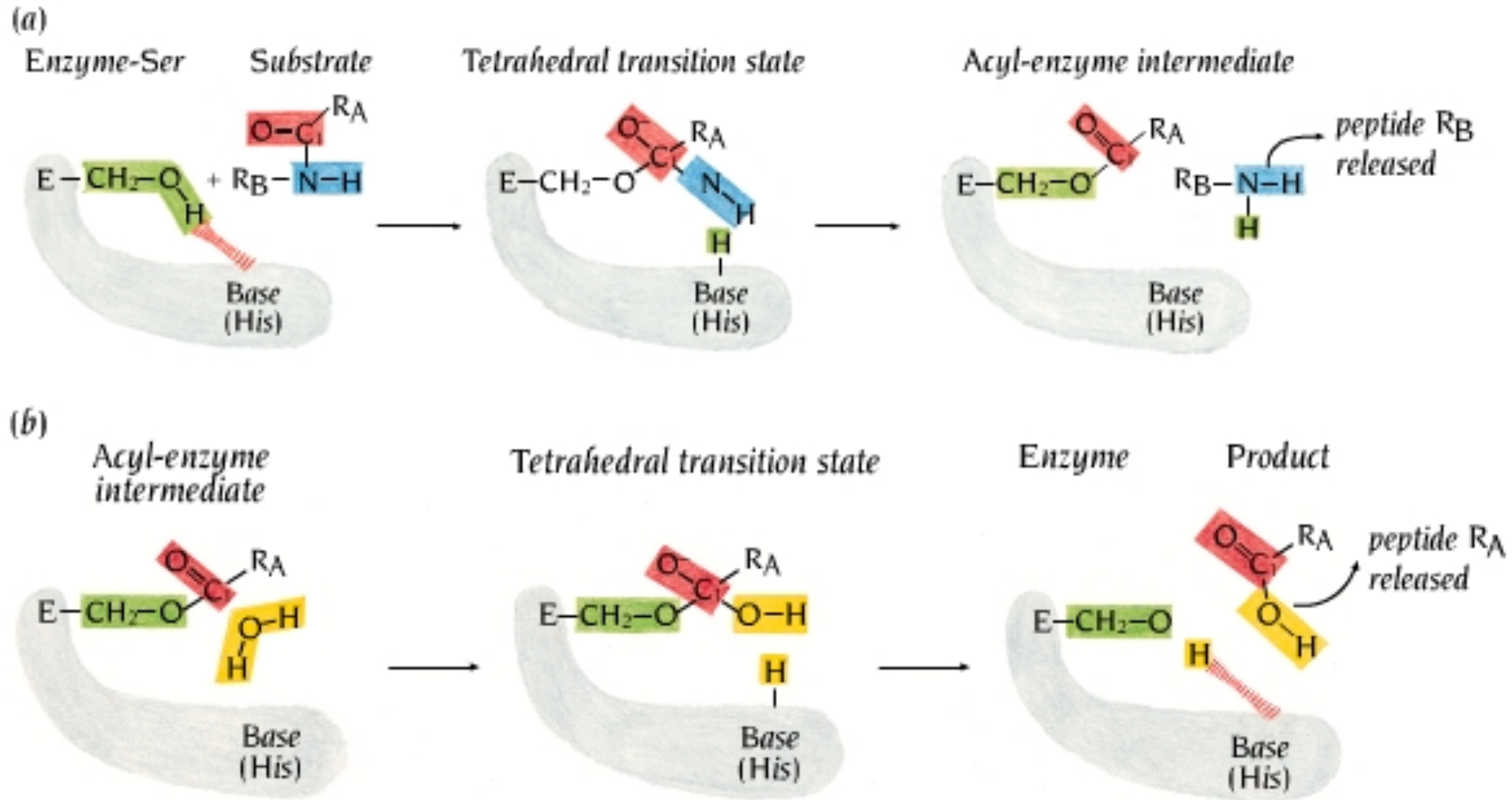
Elastase

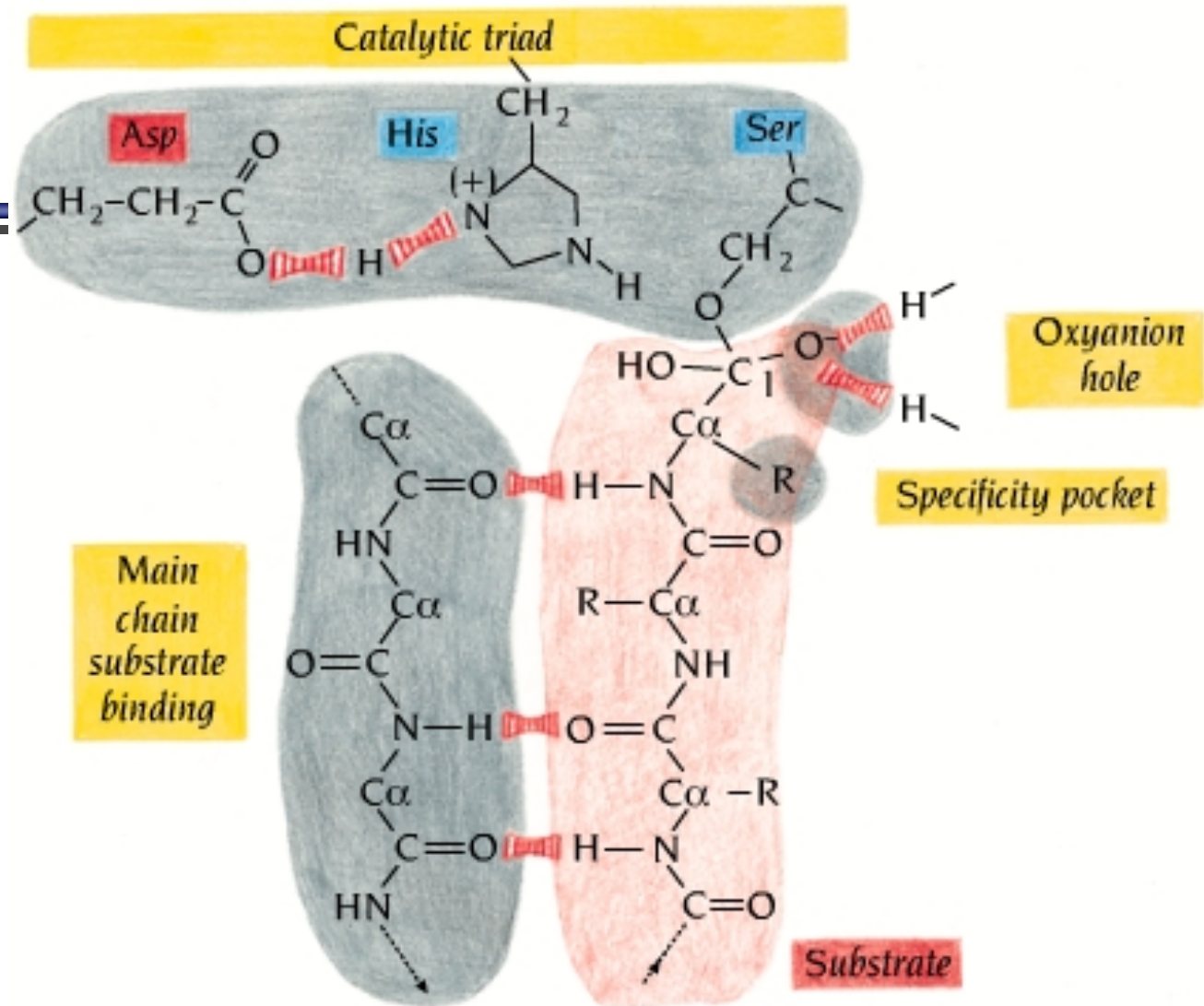
©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

Mechanism of protease activity



Mechanism of protease activity





© 1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

5. Edman degradation

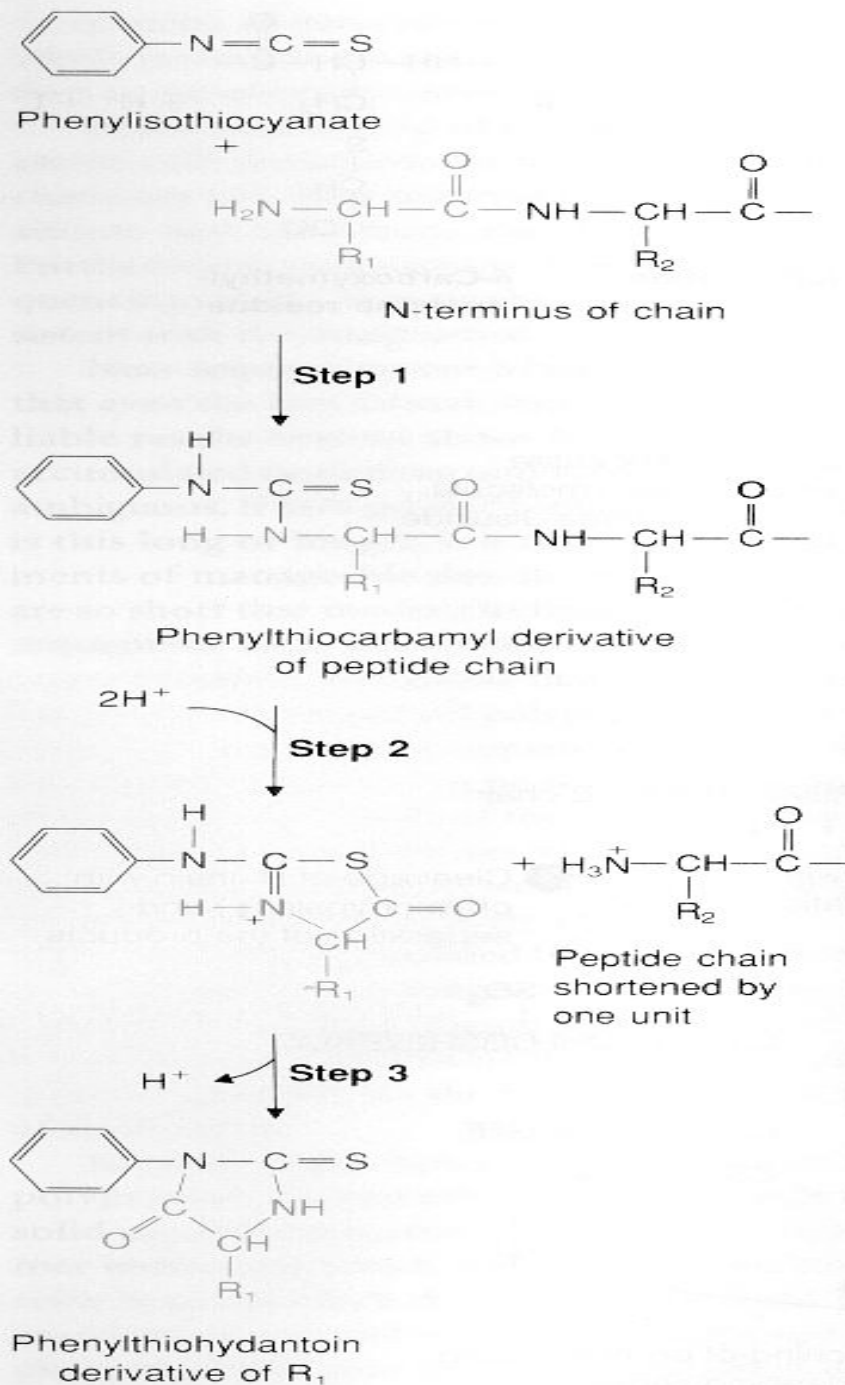
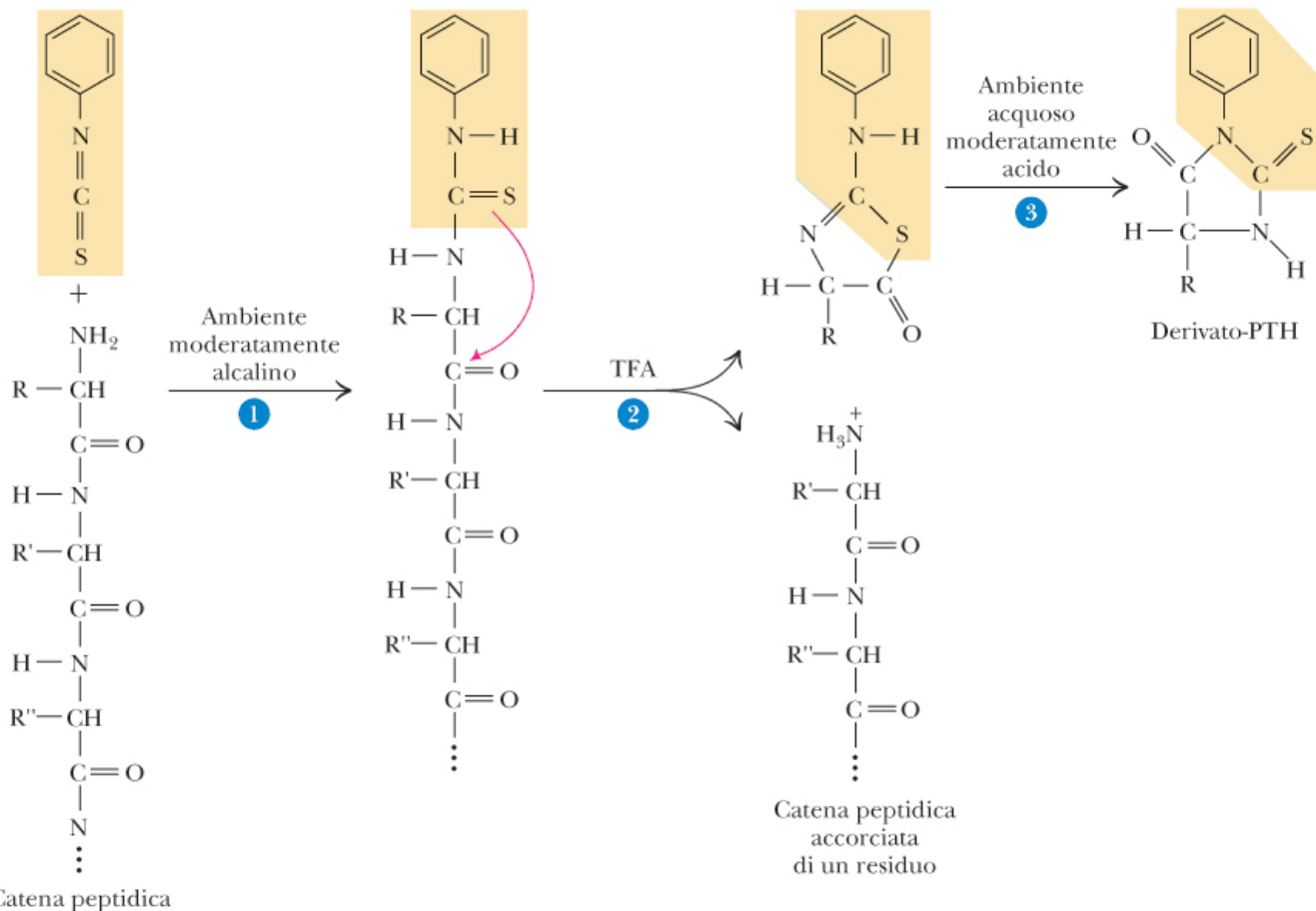


FIGURE 5C.1

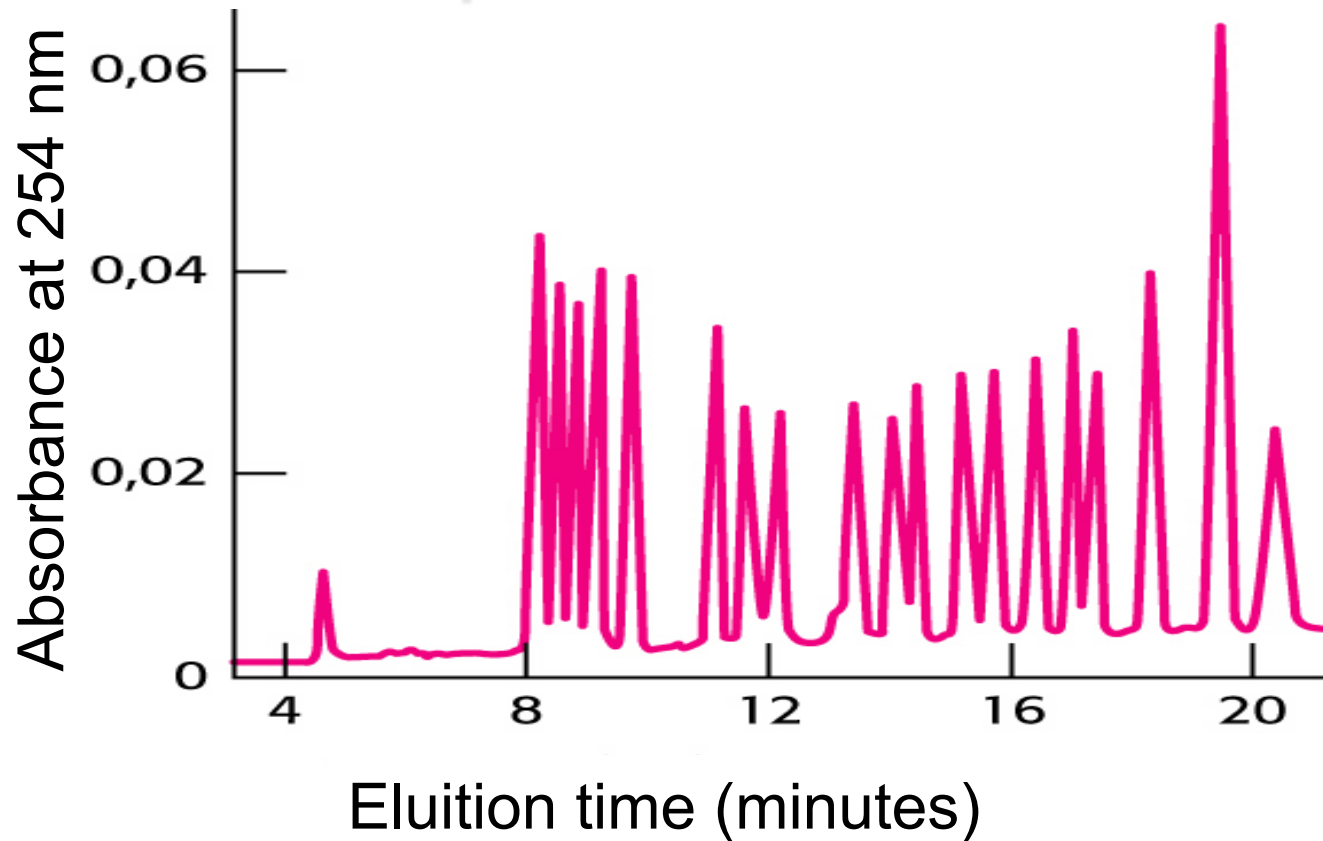


Edman degradation. 1) In moderate alkaline conditions, phenylthioisocyanate combines with the N-terminal of the peptide to form a phenyl thiocarbamoyl (PTC) derivative. 2) After treatment with trifluoroacetic acid (TFA), a cyclic compound is formed and the first N-term amino acid is released as a thiazoline derivative, whereas the other peptide bonds are not hydrolyzed. 3) After organic extraction and treatment with an aqueous and acidic solution, a phenylthiohydantoin derivative (PTH) of the N-term amino acid is formed. The process is repeated several times to determine the amino acid at each step until the sequence of the peptide is complete.



Identification of the N-term PTH-aa N-terminale through chromatography

Separation of the 20 PTH-aa



Peptide analysers

- The two major direct methods of protein sequencing are the Edman degradation reaction and mass spectrometry.
- Proteomics

