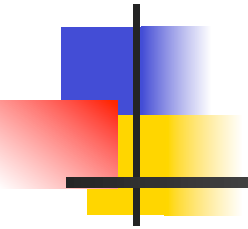


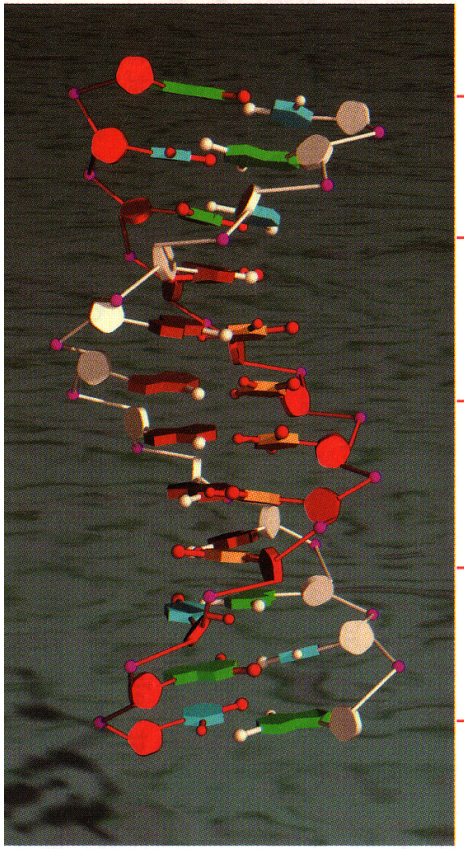
Building blocks and primary structure

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
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...

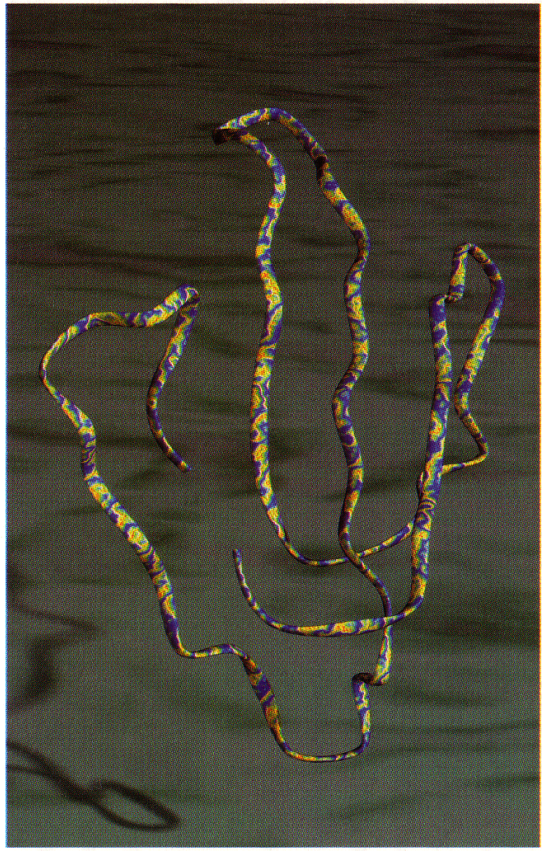
Three Bases

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

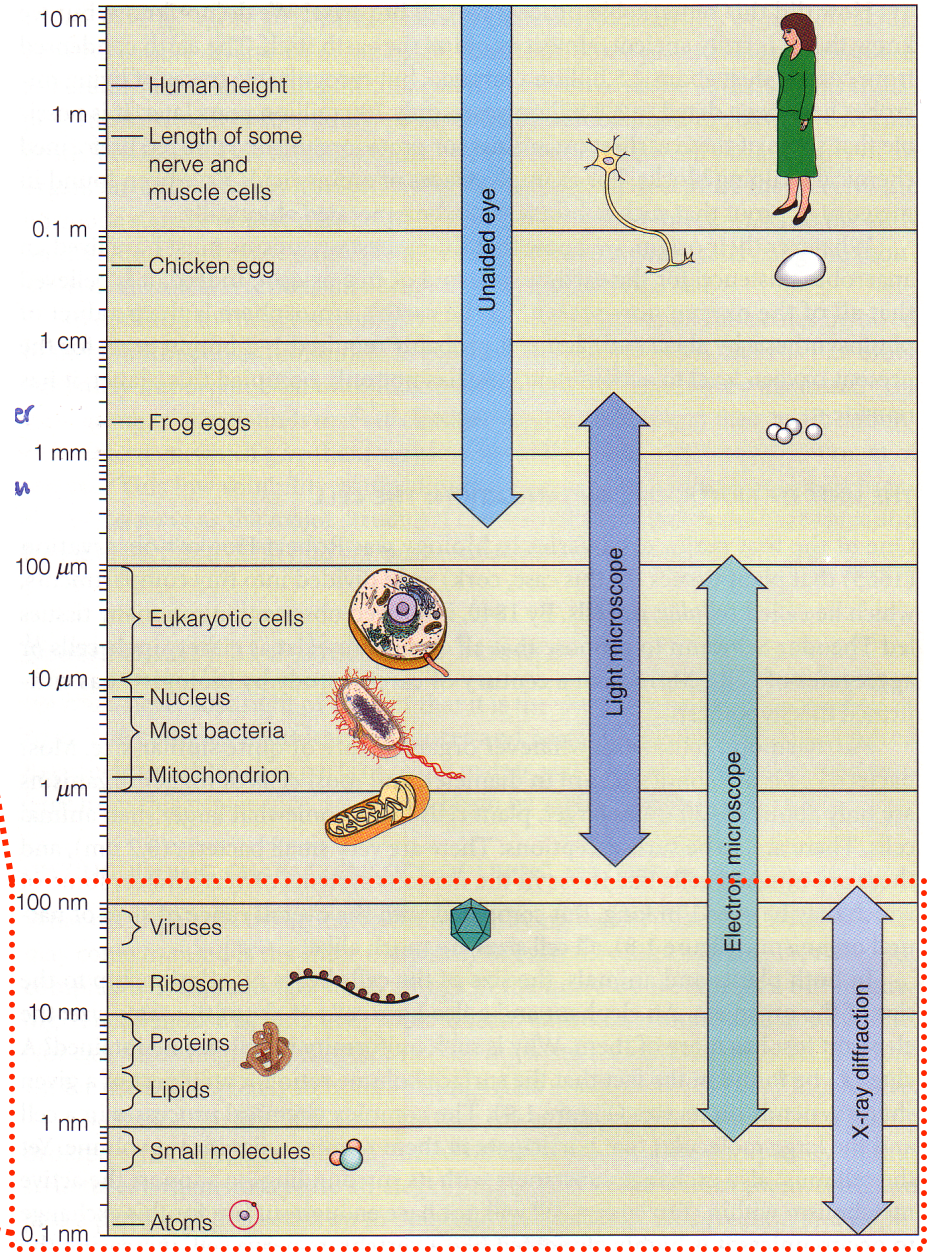
Genetic Code 'Translation Table'

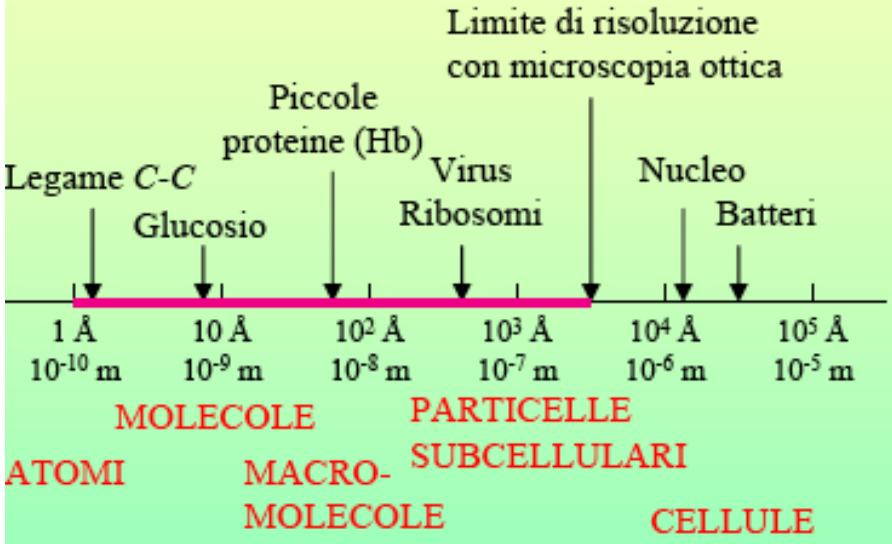
Which Folds Spontaneously to a Precise Three-Dimensional Structure

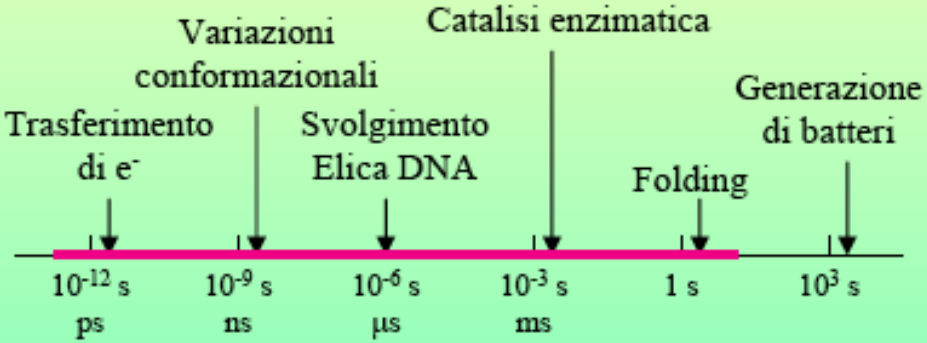


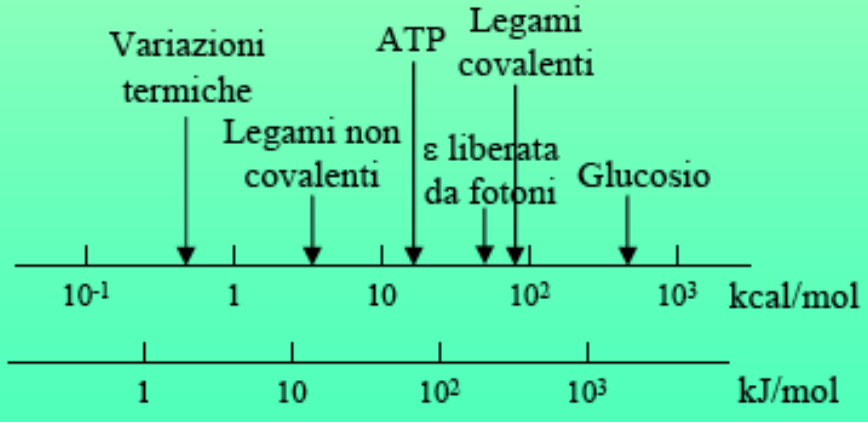
Range and sizes in biochemistry

- 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



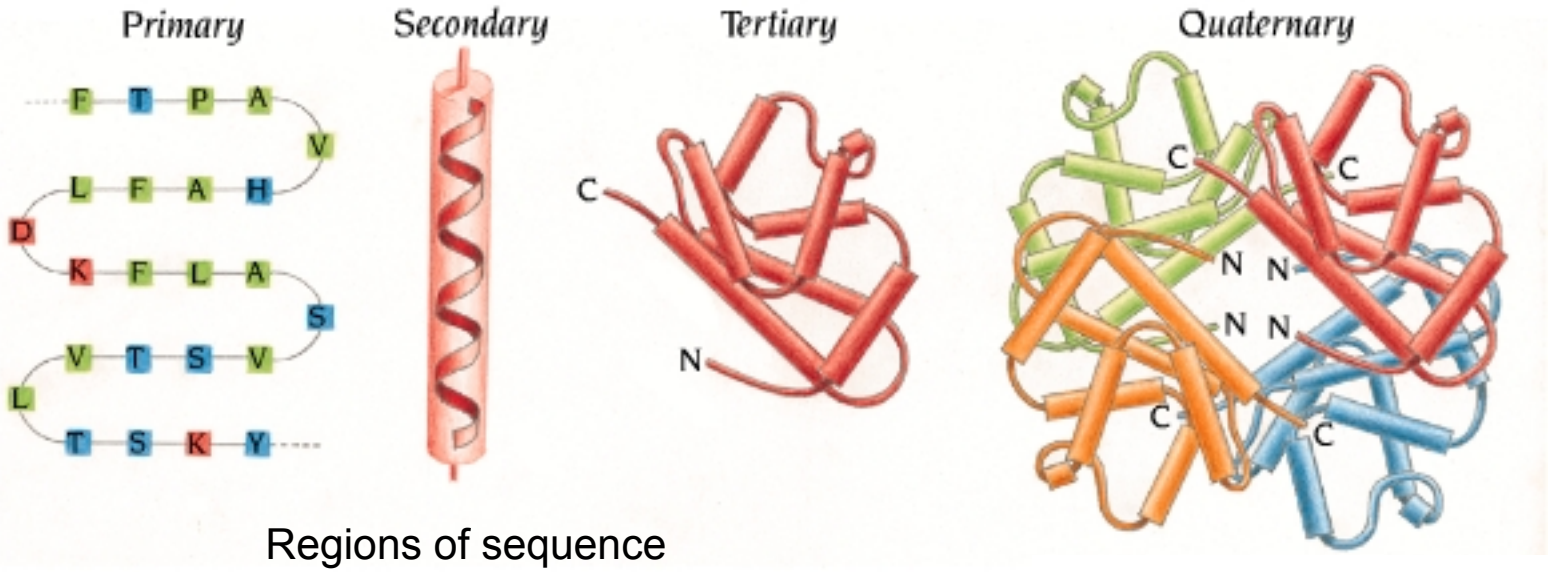






1 kcal= 4.184 kJ

Proteins are polymers of aminoacids



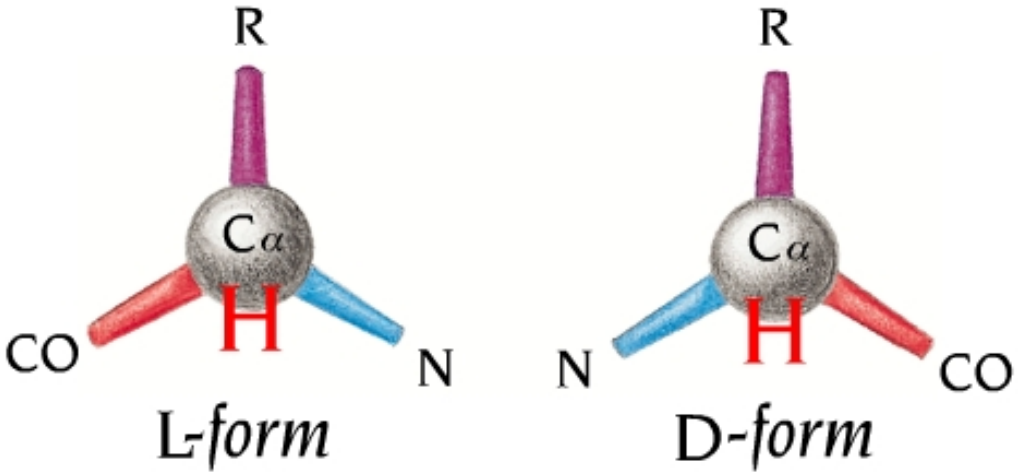
AA sequence

Regions of sequence with regular structure

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

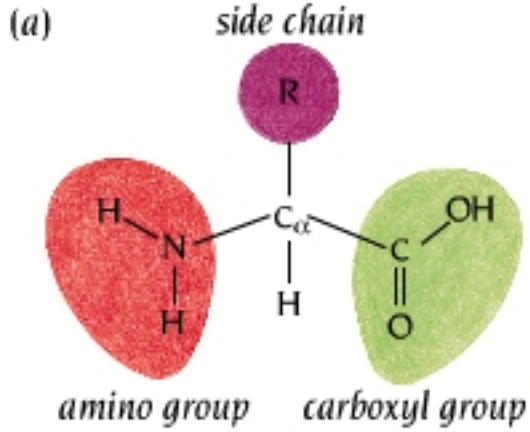
Polypeptide chains associated in functional assemblies

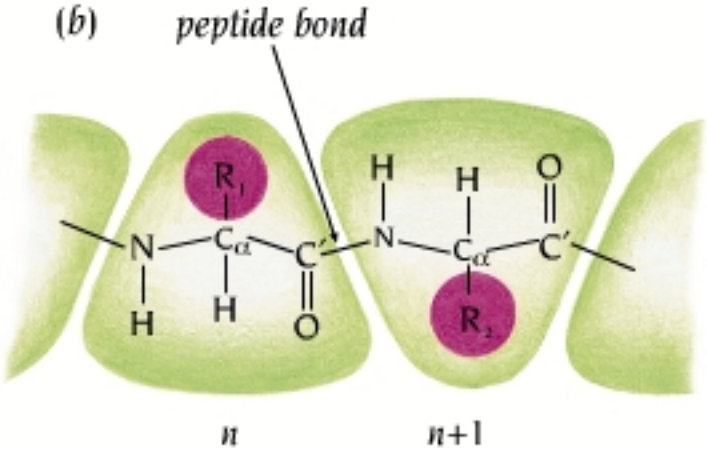
- AA are chiral



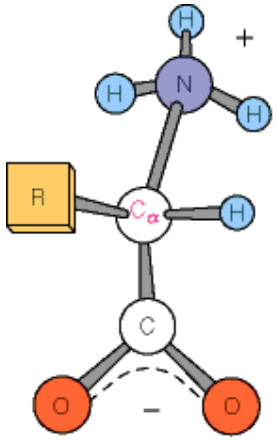
R
(CORN)

S

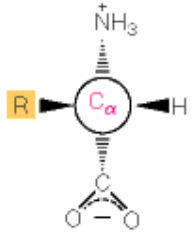




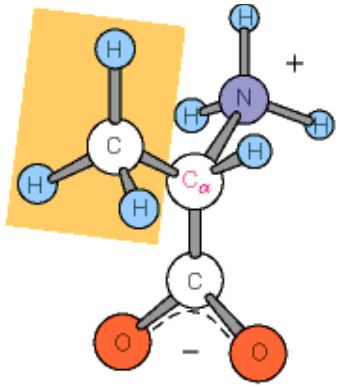
Stereochemistry of α-amino acids



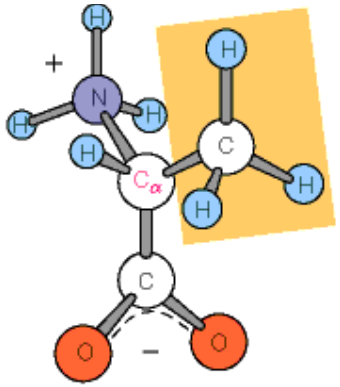
(a) α-Amino acid



(b) Compact representation

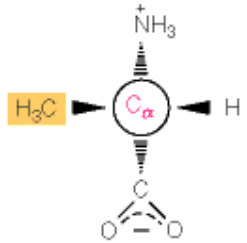


(a) L-Alanine

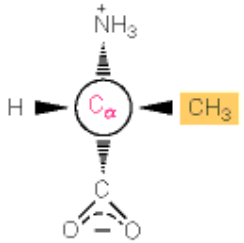


D-Alanine

✓

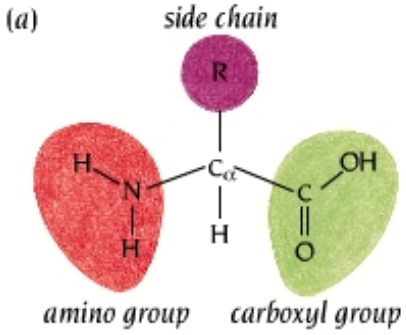


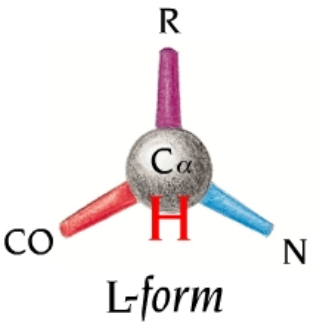
(b) L-Alanine

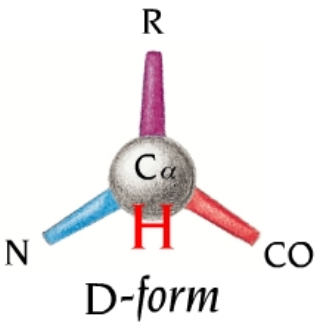


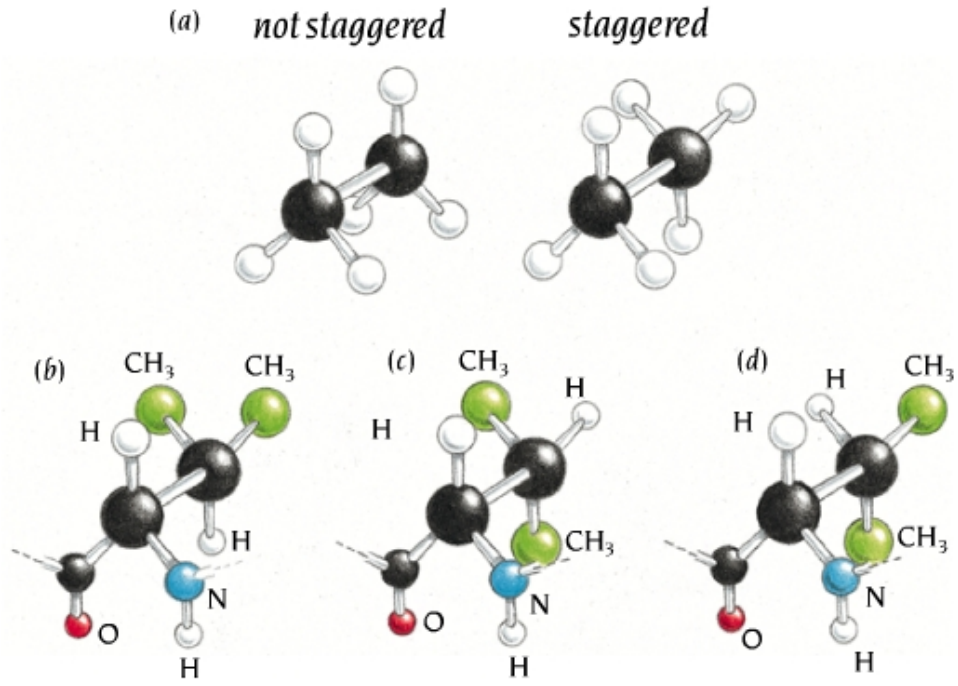
D-Alanine

Aminoacids: classification and properties.



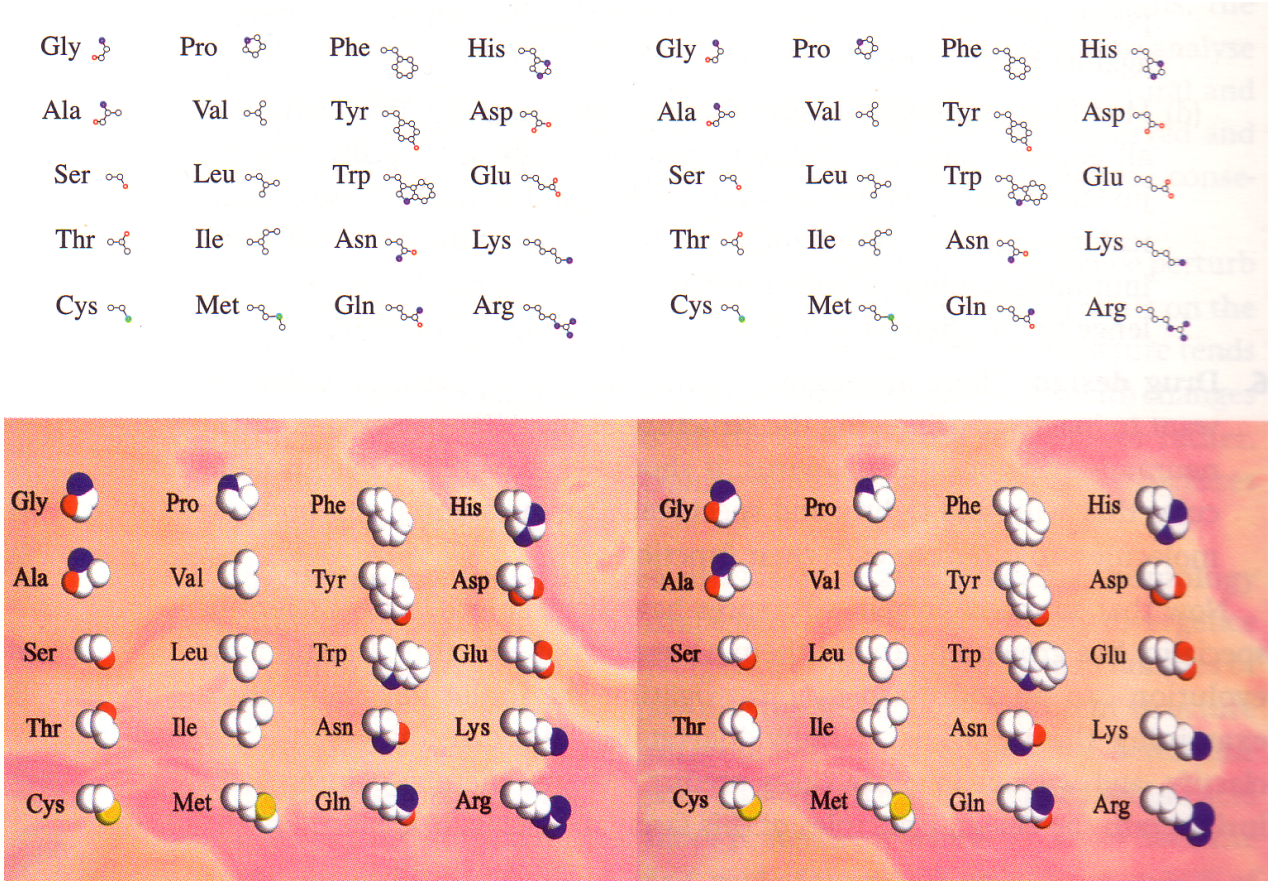




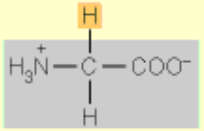


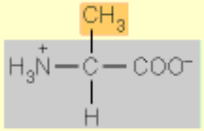
Different side-chains = different properties

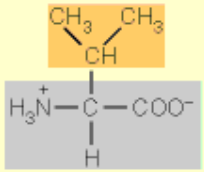
- 1 letter and 3 letters codes

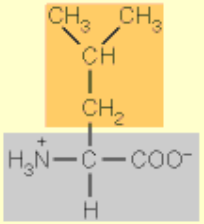


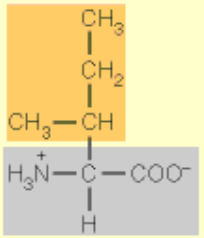
ALIPHATIC AMINO ACIDS


Glycine (Gly) G

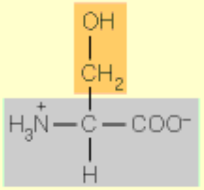

Alanine (Ala) A

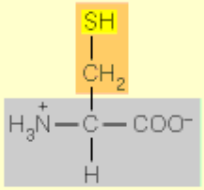

Valine (Val) V

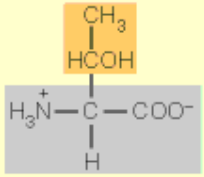

Leucine (Leu) L

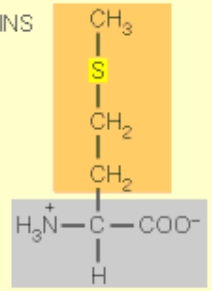

Isoleucine (Ile) I

AMINO ACIDS WITH HYDROXYL- OR SULFUR-CONTAINING SIDE CHAINS

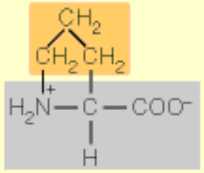

Serine (Ser) S


Cysteine (Cys) C

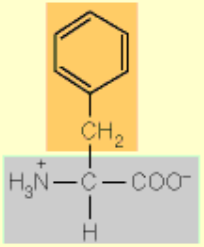

Threonine (Thr) T

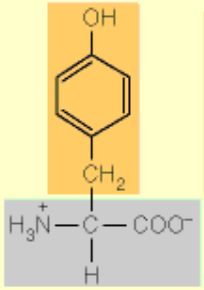

Methionine (Met) M

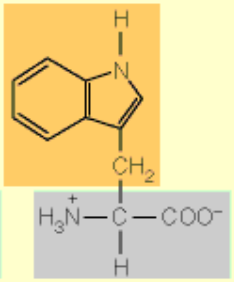
CYCLIC AMINO ACID


Proline (Pro) P

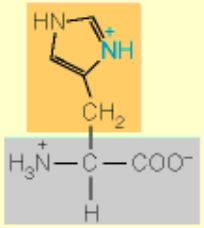
AROMATIC AMINO ACIDS

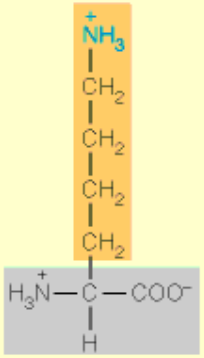

Phenylalanine (Phe) F

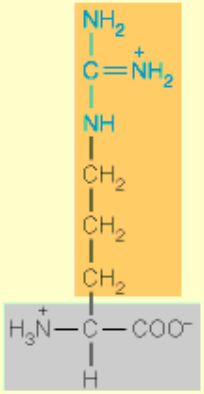

Tyrosine (Tyr) Y


Tryptophan (Trp) W

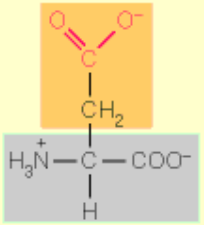
BASIC AMINO ACIDS

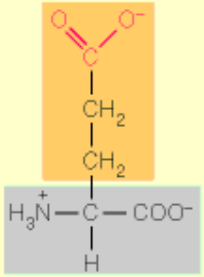

Histidine (His) H

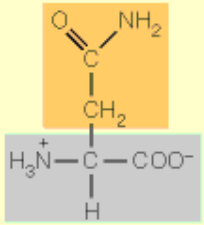

Lysine (Lys) K

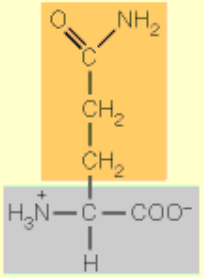

Arginine (Arg) R

ACIDIC AMINO ACIDS AND THEIR AMIDES


Aspartic acid (Asp) D

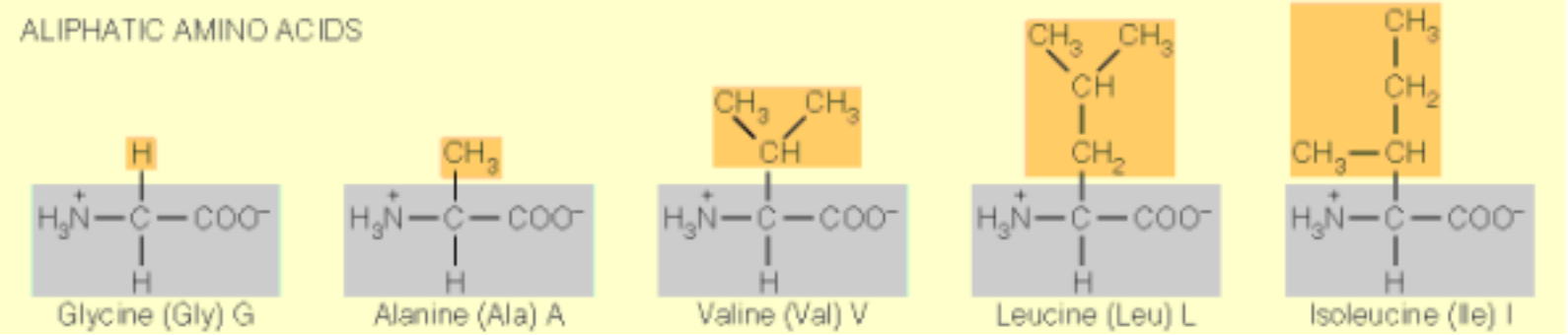

Glutamic acid (Glu) E


Asparagine (Asn) N

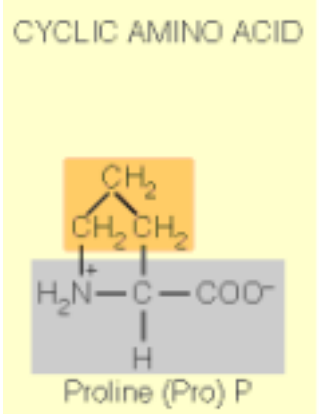

Glutamine (Gln) Q

Properties of aa side-chains

- Aliphatic aa:

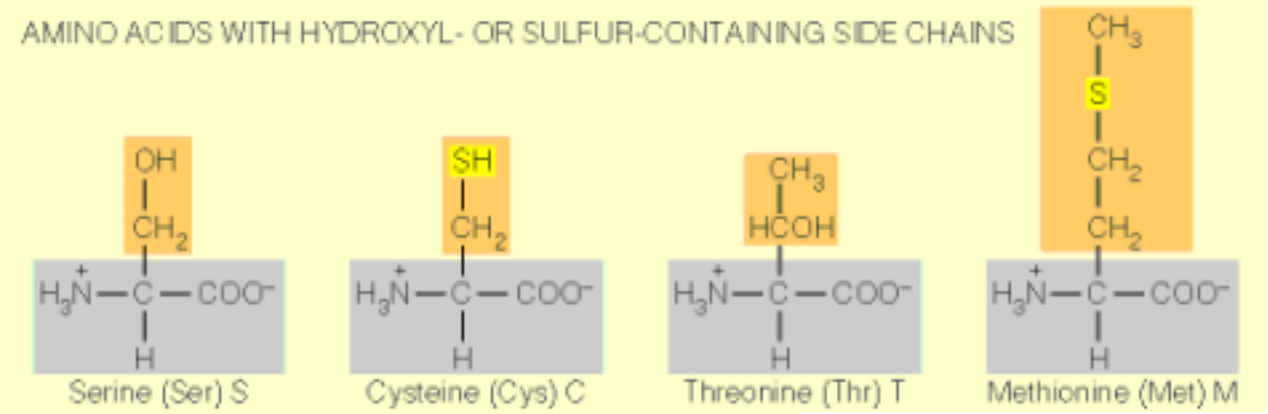


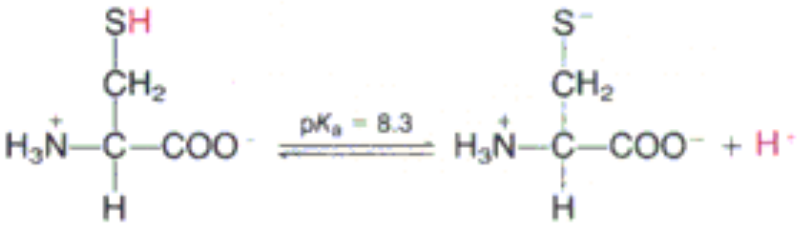
  - Flexible and hydrophobic, the most Ile;
  - Found in the hyodrophobic core of proteins
  - Pro: exception, rigid ring as side chain, often a structure breaker

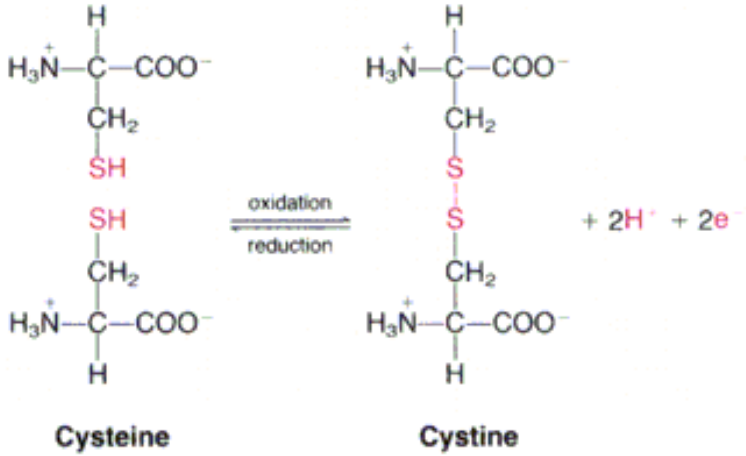


Hydroxyl-, sulfur- side chains

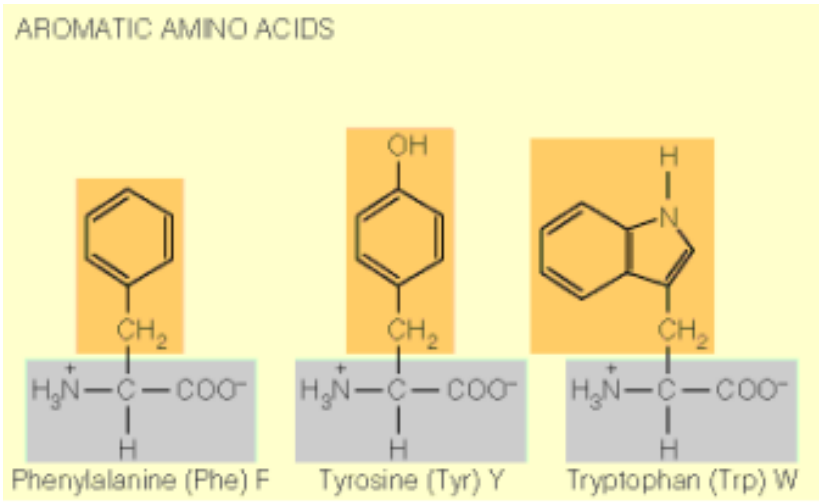
- Weakly polar side chains

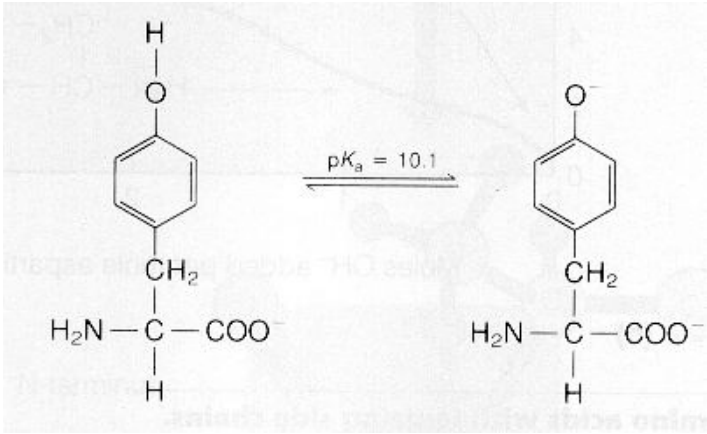




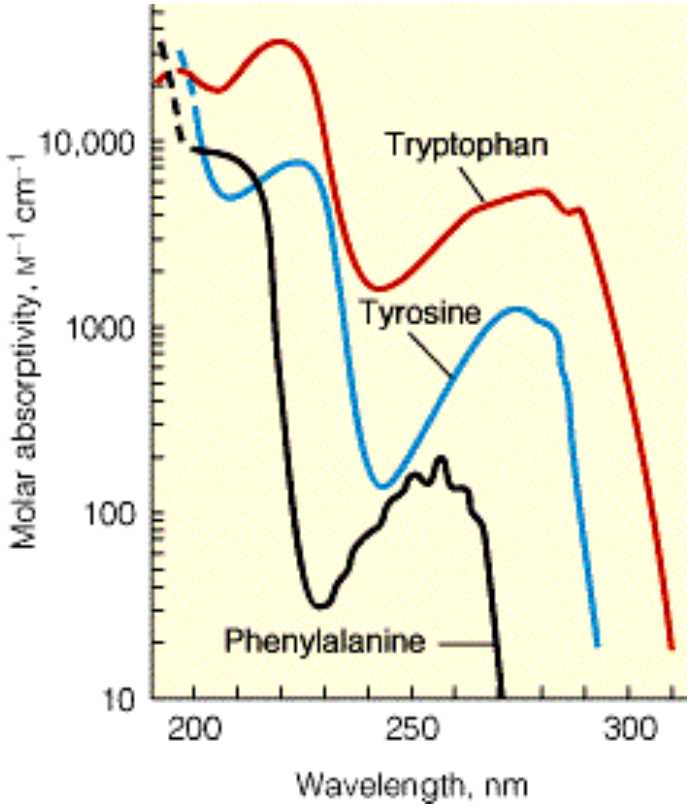


Aromatic side-chains



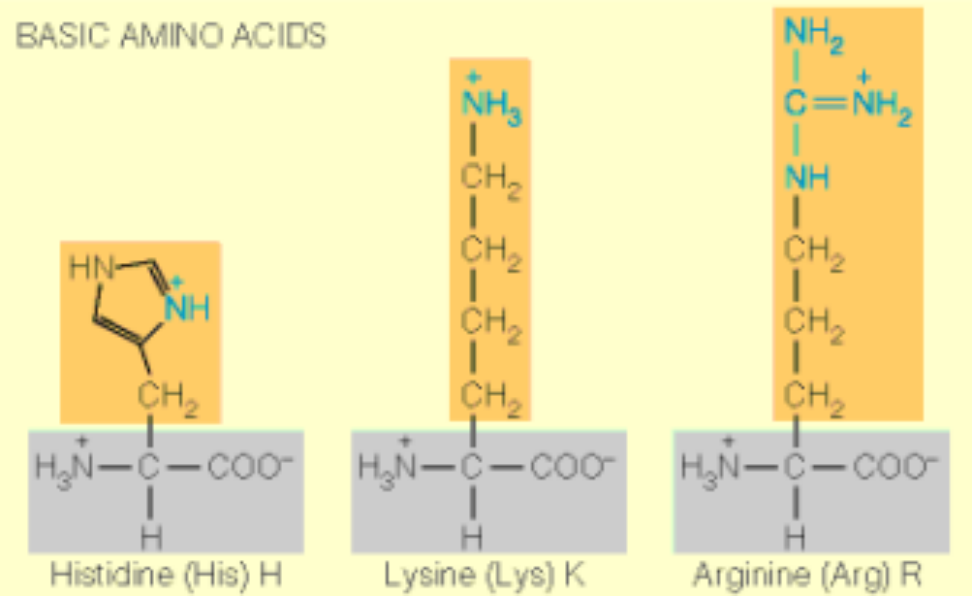


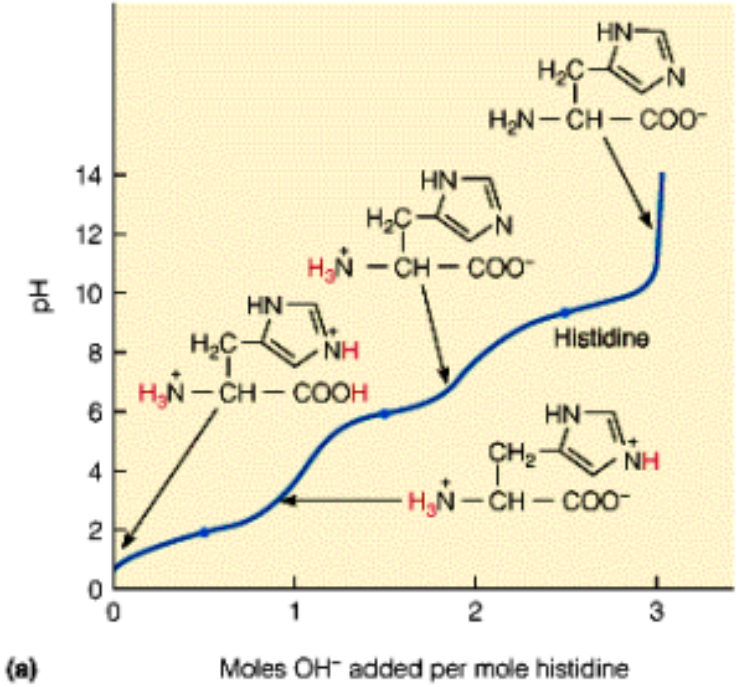
Absorbance



Detection of proteins

Basic side-chains

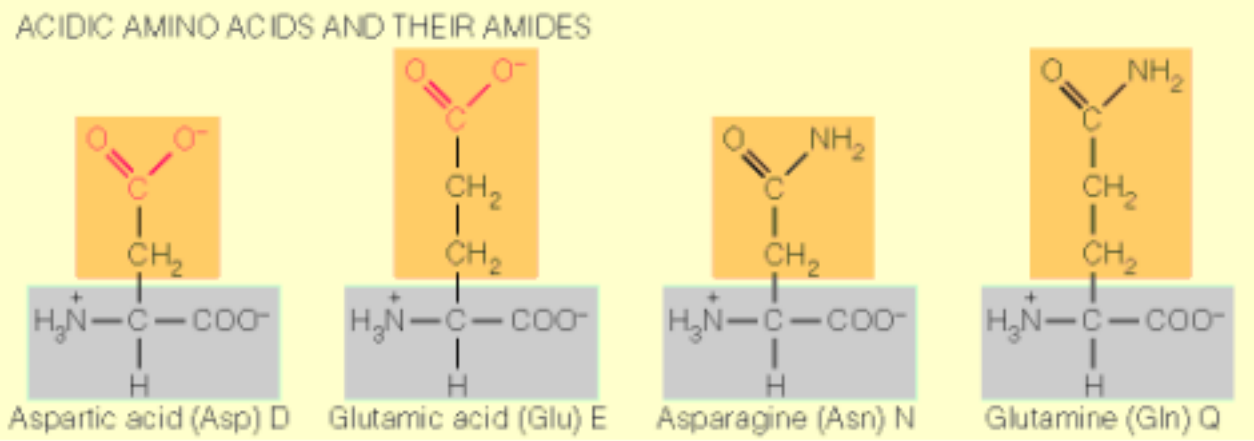




Group Type	Typical pK_a Range[a]
α-Carboxyl	3.5–4.0
Side chain carboxyls of aspartic and glutamic acids	4.0–4.8
Imidazole (histidine)	6.5–7.4
Cysteine (—SH)	8.5–9.0
Phenolic (tyrosine)	9.5–10.5
α-Amino	8.0–9.0
Side chain amino (lysine)	9.8–10.4
Guanidinyl (arginine)	~12

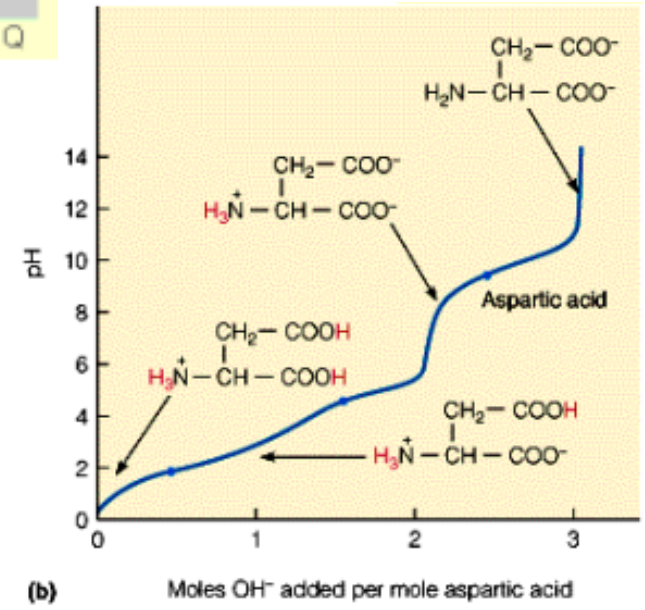
[a]Values outside these ranges are observed. For example, side chain carboxyls have been reported with pK_a values as high as 7.3.

Acidic side-chains and amides



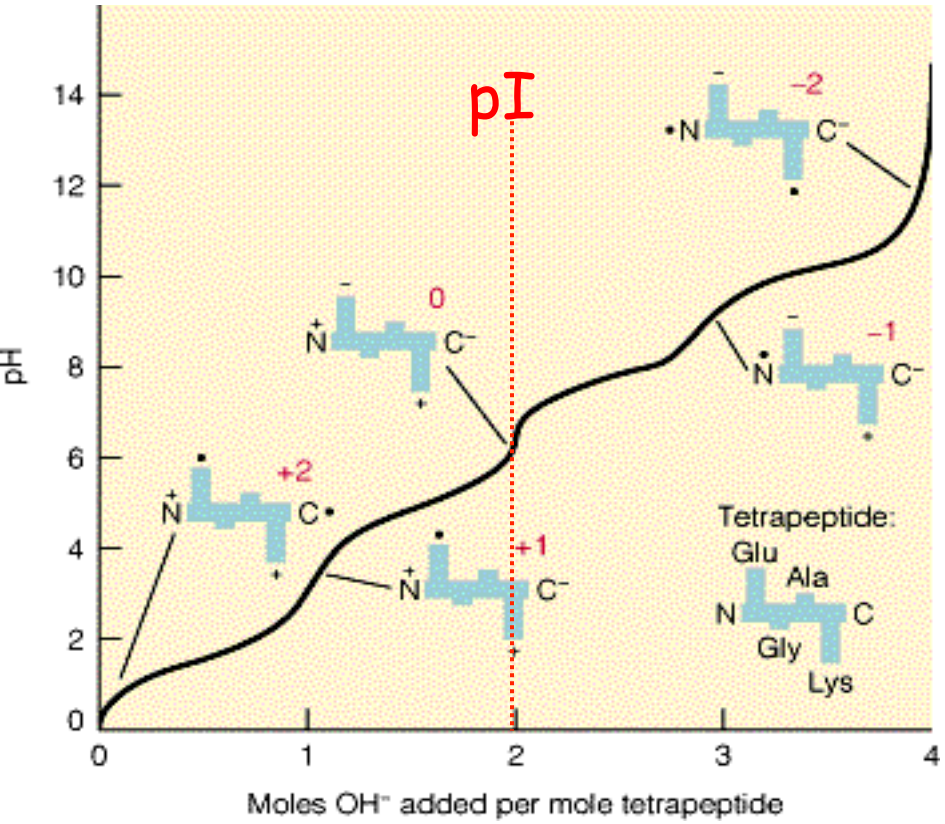
Group Type	Typical pK_a Range[a]
α-Carboxyl	3.5–4.0
Side chain carboxyls of aspartic and glutamic acids	4.0–4.8
Imidazole (histidine)	6.5–7.4
Cysteine (—SH)	8.5–9.0
Phenolic (tyrosine)	9.5–10.5
α-Amino	8.0–9.0
Side chain amino (lysine)	9.8–10.4
Guanidinyl (arginine)	~12

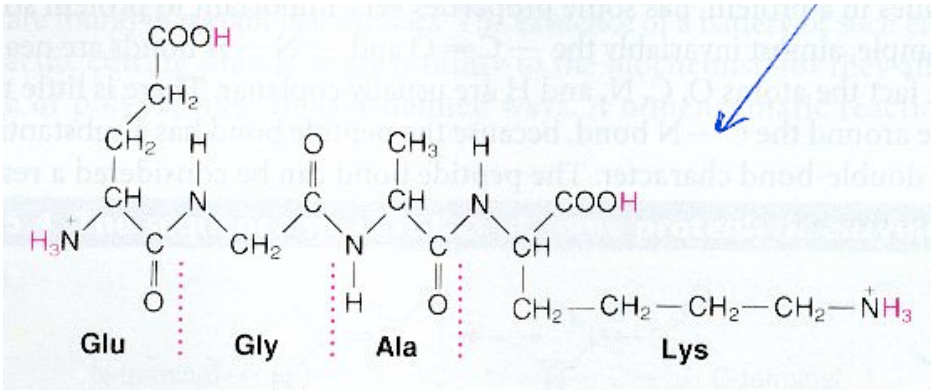
[a]Values outside these ranges are observed. For example, side chain carboxyls have been reported with pK_a values as high as 7.3.



Proteins are polyampholites

- pI: isoelectric point



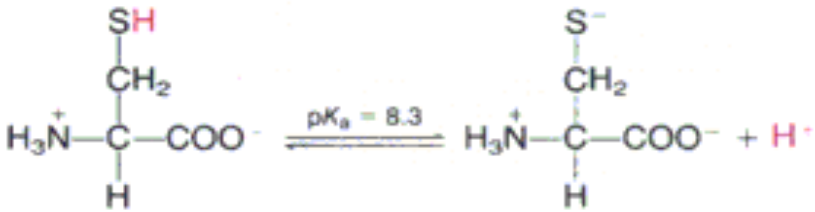


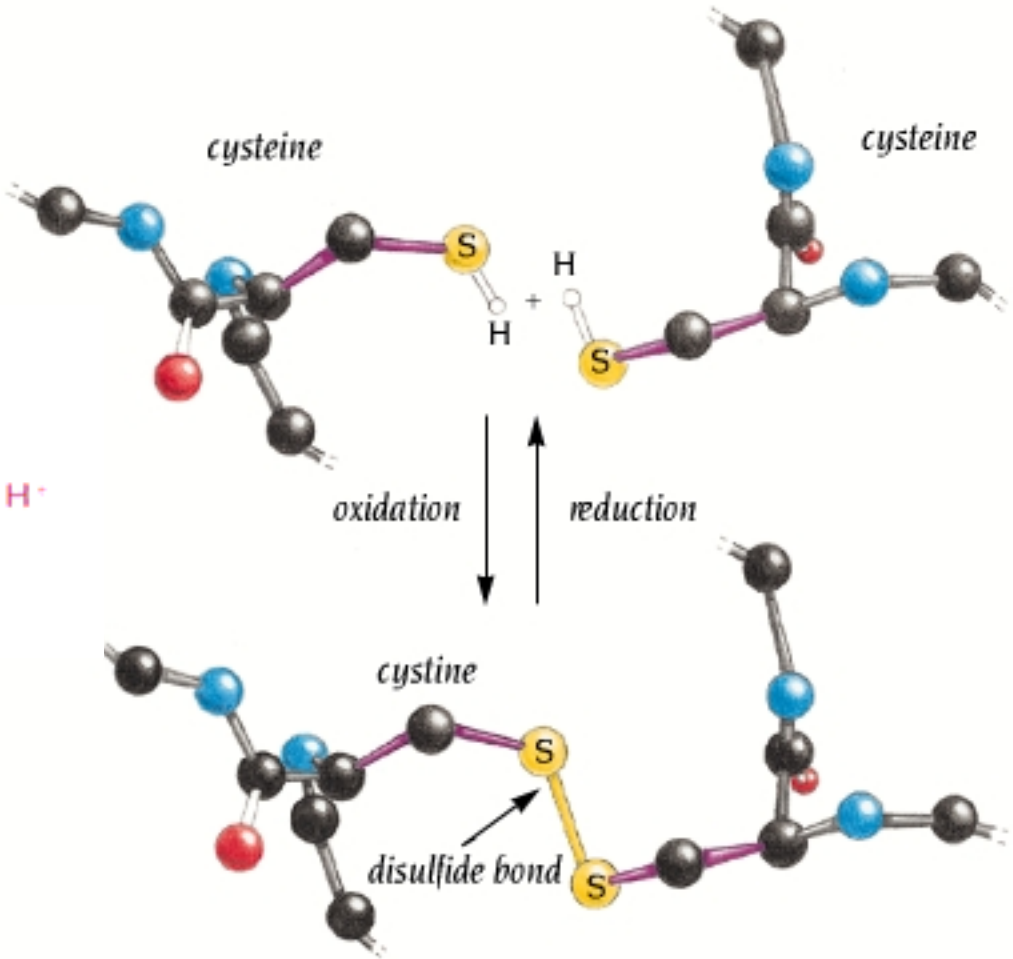
Name	Abbreviations	pK_a of α-COOH Group	pK_a of α-NH_3^+ Group	pK_a of Ionizing Side Chain[a]	Residue[b] Mass (daltons)	Occurrence[c] in Proteins (mol %)
Alanine	A, Ala	2.3	9.7	—	71.08	9.0
Arginine ✶	R, Arg	2.2	9.0	12.5 ✶	156.20	4.7
Asparagine	N, Asn	2.0	8.8	—	114.11	4.4
Aspartic acid ✶	D, Asp	2.1	9.8	3.9 ✶	115.09	5.5
Cysteine ✶	C, Cys	1.8	10.8	8.3	103.14	2.8 ✶
Glutamine	Q, Gln	2.2	9.1	—	128.14	3.9
Glutamic acid	E, Glu	2.2	9.7	4.2	129.12	6.2
Glycine	G, Gly	2.3	9.6	—	57.06	7.5
Histidine ✶	H, His	1.8	9.2	6.0	137.15	2.1 ✶
Isoleucine	I, Ile	2.4	9.7	—	113.17	4.6
Leucine	L, Leu	2.4	9.6	—	113.17	7.5
Lysine	K, Lys	2.2	9.0	10.0	128.18	7.0
Methionine	M, Met	2.3	9.2	—	131.21	1.7
Phenylalanine	F, Phe	1.8	9.1	—	147.18	3.5
Proline	P, Pro	2.0	10.6	—	97.12	4.6
Serine	S, Ser	2.2	9.2	—	87.08	7.1
Threonine	T, Thr	2.6	10.4	—	101.11	6.0
Tryptophan ✶	W, Trp	2.4	9.4	—	186.21	1.1 ✶
Tyrosine	Y, Tyr	2.2	9.1	10.1	163.18	3.5
Valine	V, Val	2.3	9.6	—	99.14	6.9

[a] Approximate values found for side chains on the *free* amino acids.

[b] To obtain the mass of the amino acid itself, add the mass of a mole of water, 18.02 g. The values given are for neutral side chains; slightly different values will apply at pH values where protons have been gained or lost from the side chains.

$$2\ (\text{-}CH_2\text{-}SH) + 1/2\ O_2 \Leftrightarrow\ \text{-}CH_2\text{-}S\text{-}S\text{-}CH_2\text{-}\ + H_2O$$





Analysis of proteins’ primary structure:

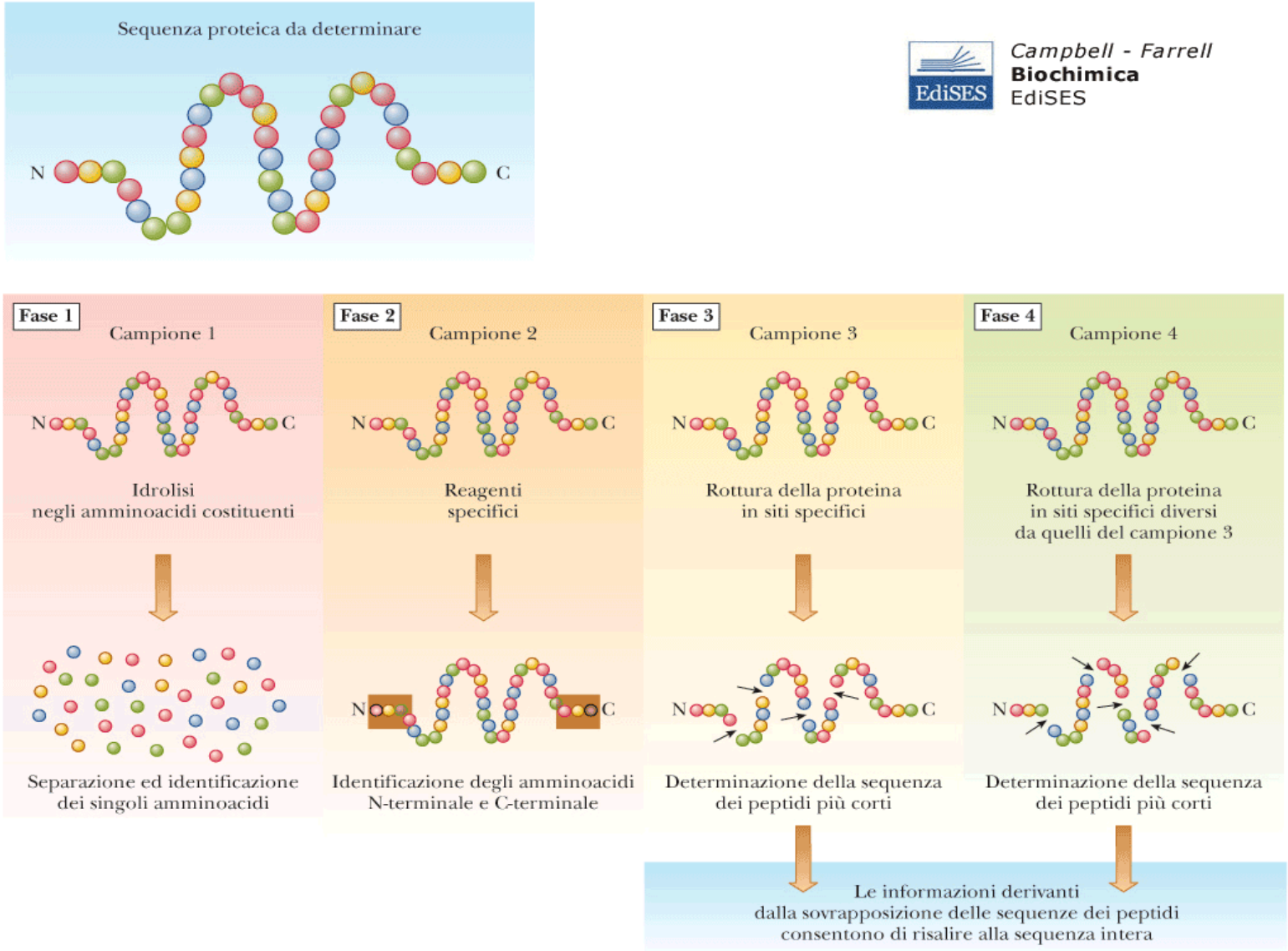


FIGURA 5.14 **La strategia per determinare la struttura primaria di una data proteina.** La sequenza di amminoacidi può essere determinata con quattro analisi differenti, eseguite su quattro campioni separati della stessa proteina.

Chemical degradation

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

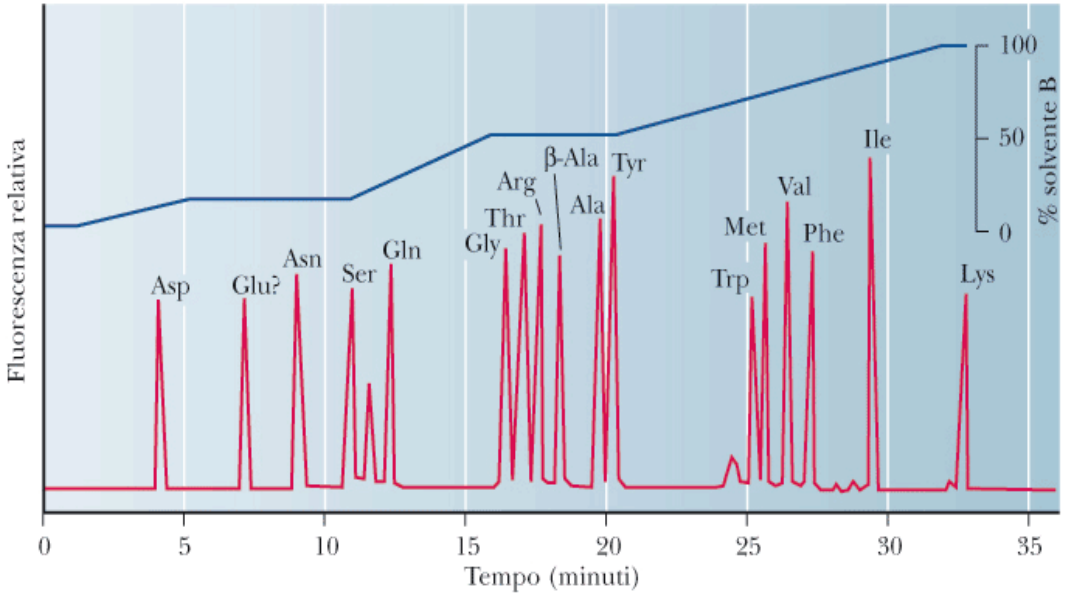
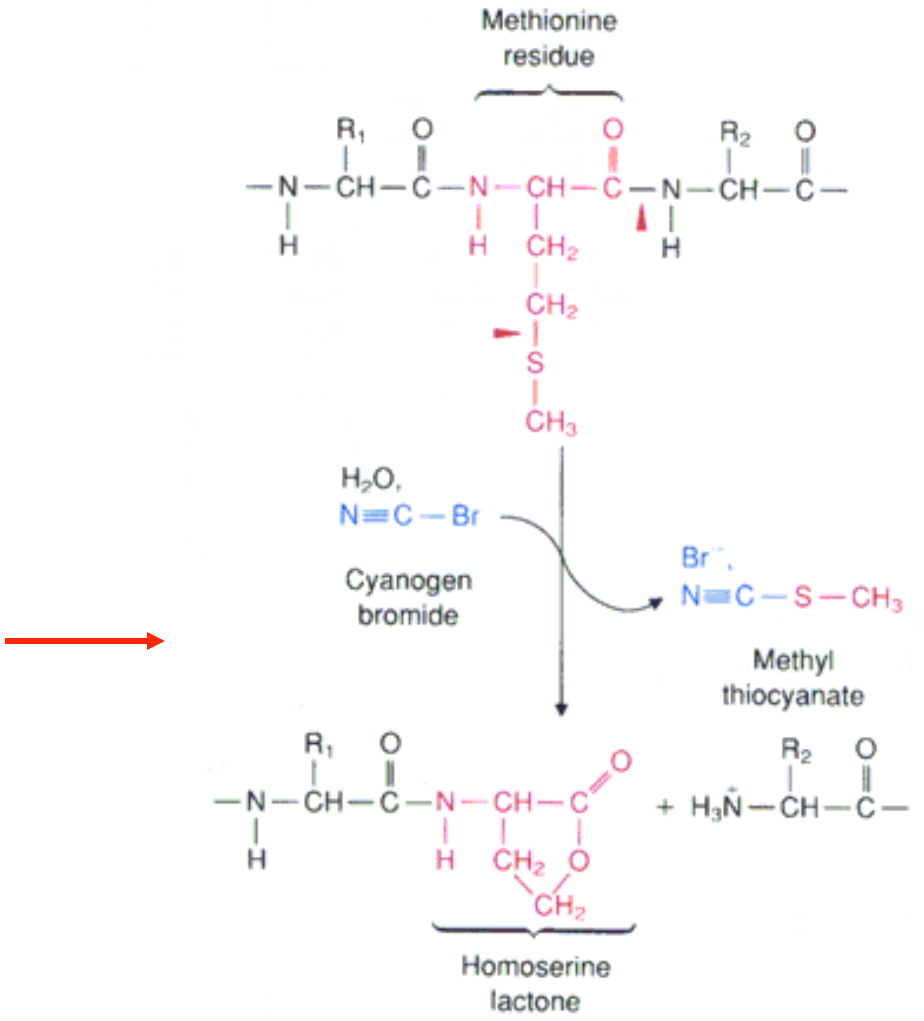


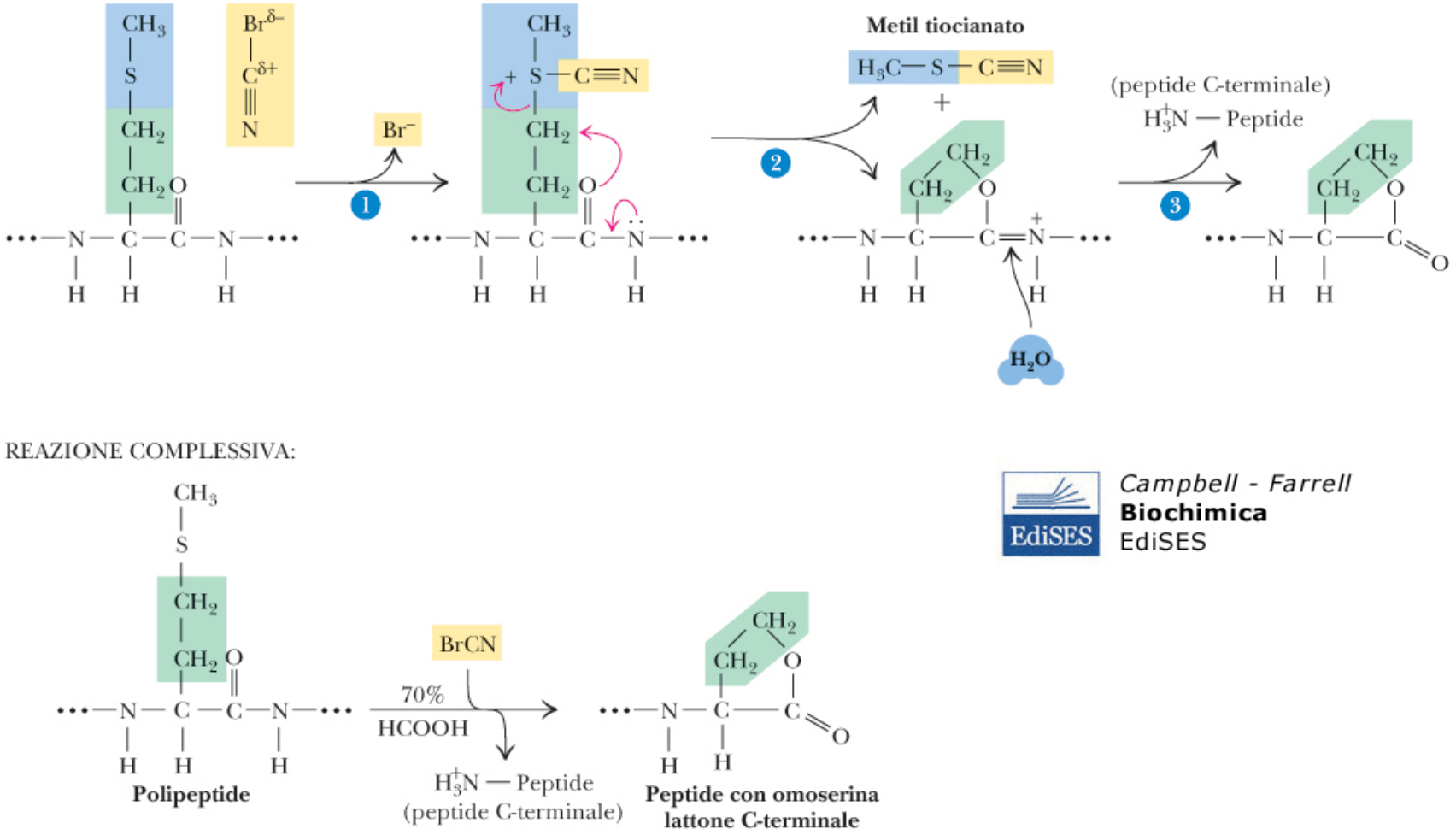
FIGURA 5.15 Cromatogramma HPLC relativo alla separazione di una miscela di amminoacidi.

Campbell - Farrell
Biochimica
EdiSES

Chemical degradation

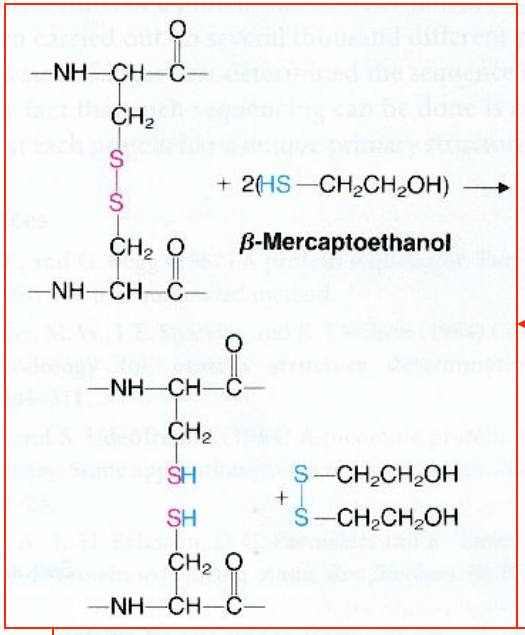
2. Cyanogen bromide





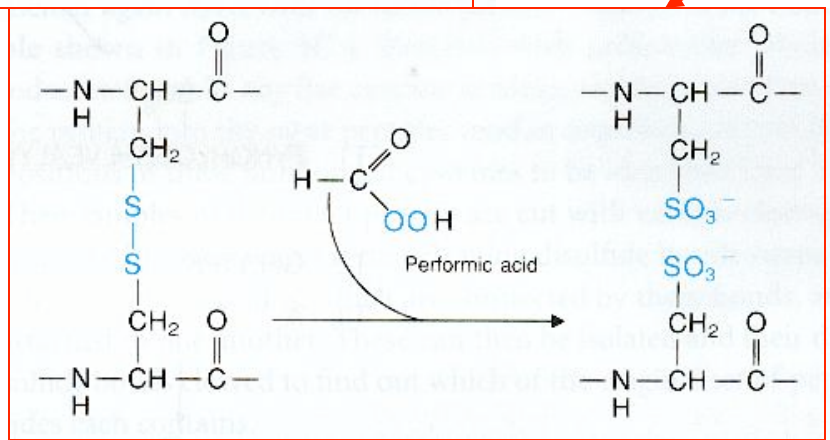
Campbell - Farrell
Biochimica
EdiSES

Protein hydrolysis by cyanogen bromide takes place where methionine residues are

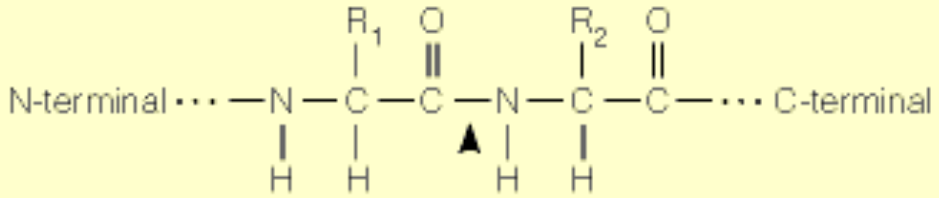


3. ß-mercaptoethanol

4. Performic acid



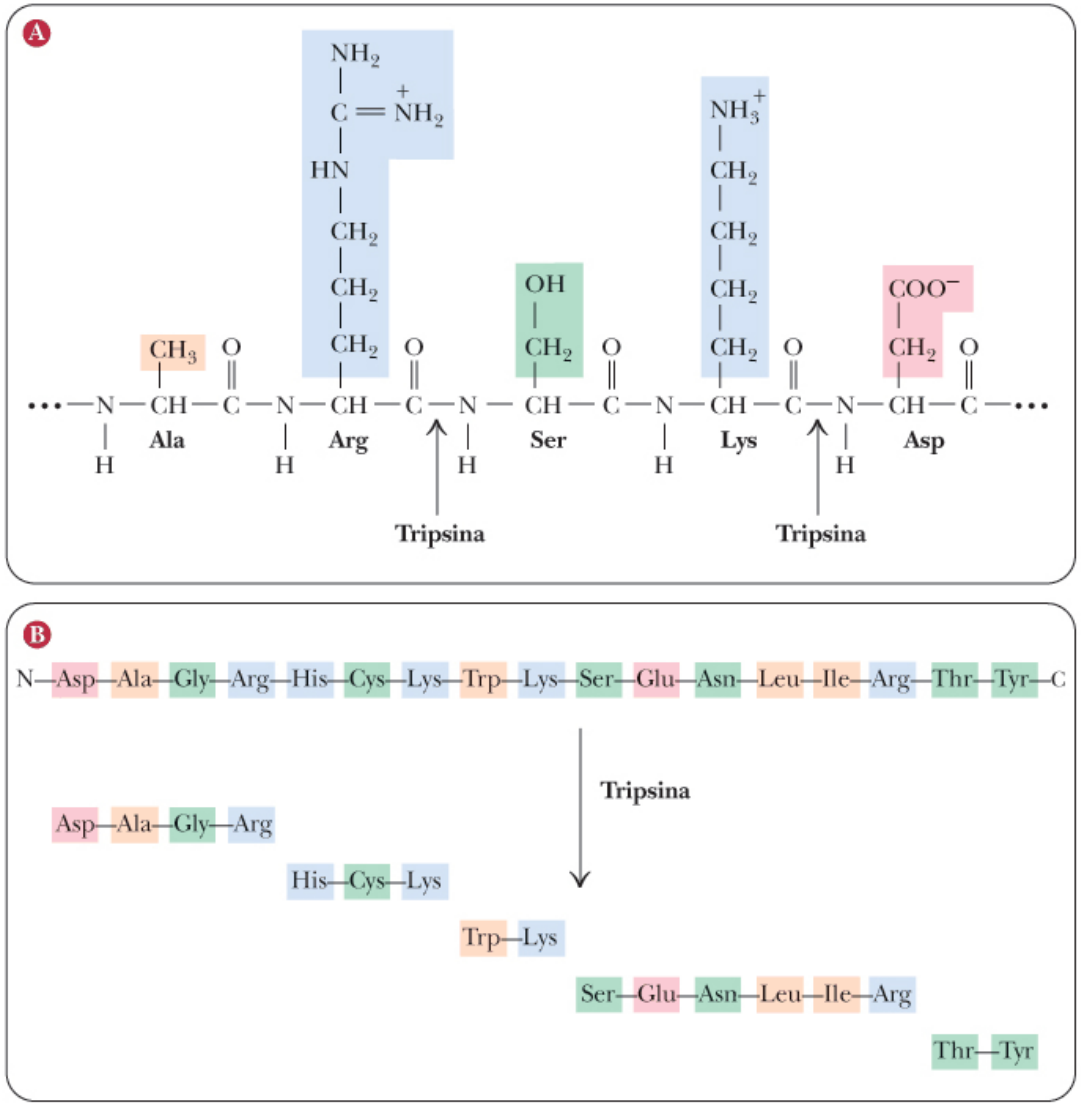
Enzymatic degradation: Proteases:



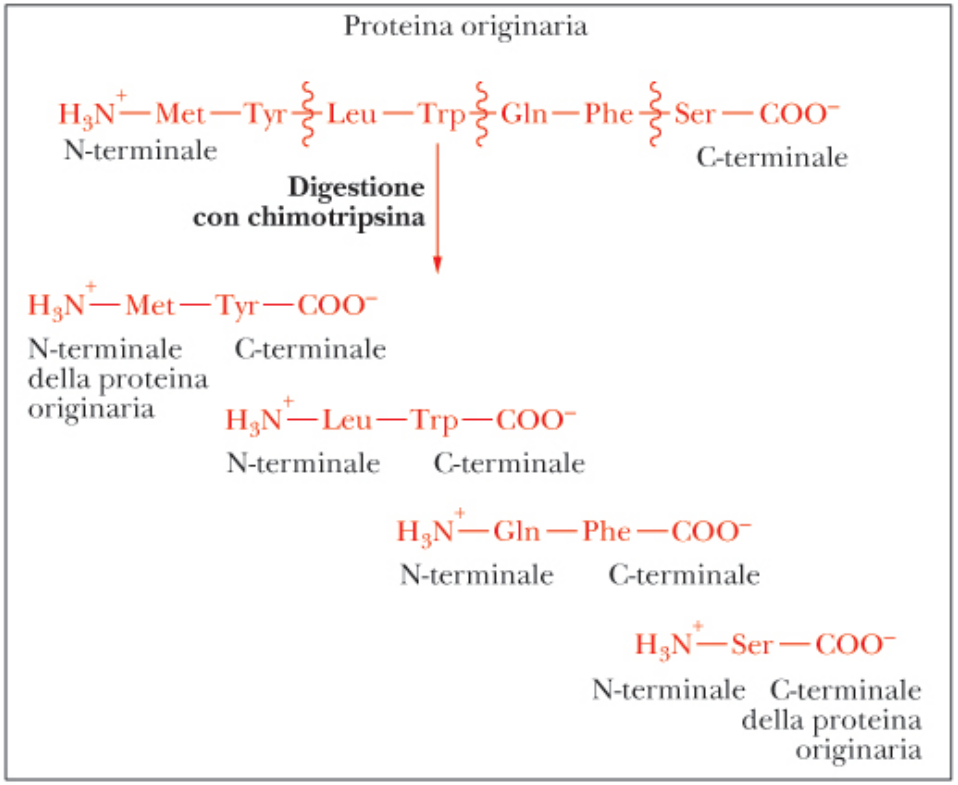
Enzyme	Preferred Site[a]	Source
Trypsin	R_1 = Lys, Arg	From digestive systems of animals, many other sources
Chymotrypsin	R_1 = Tyr, Trp, Phe, Leu	Same as trypsin
Thrombin	R_1 = Arg	From blood; involved in coagulation
V-8 protease	R_1 = Asp, Glu	From *Staphylococcus aureus*
Prolyl endopeptidase	R_1 = Pro	Lamb kidney, other tissues
Subtilisin	Very little specificity	From various bacilli
Carboxypeptidase A	R_2 = C-terminal amino acid	From digestive systems of animals
Thermolysin	R_2 = Leu, Val, Ile, Met	From *Bacillus thermoproteolyticus*

[a]The residues indicated are those next to which cleavage is most likely. Note that 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 R_2 = Pro.

Campbell - Farrell
Biochimica
EdiSES



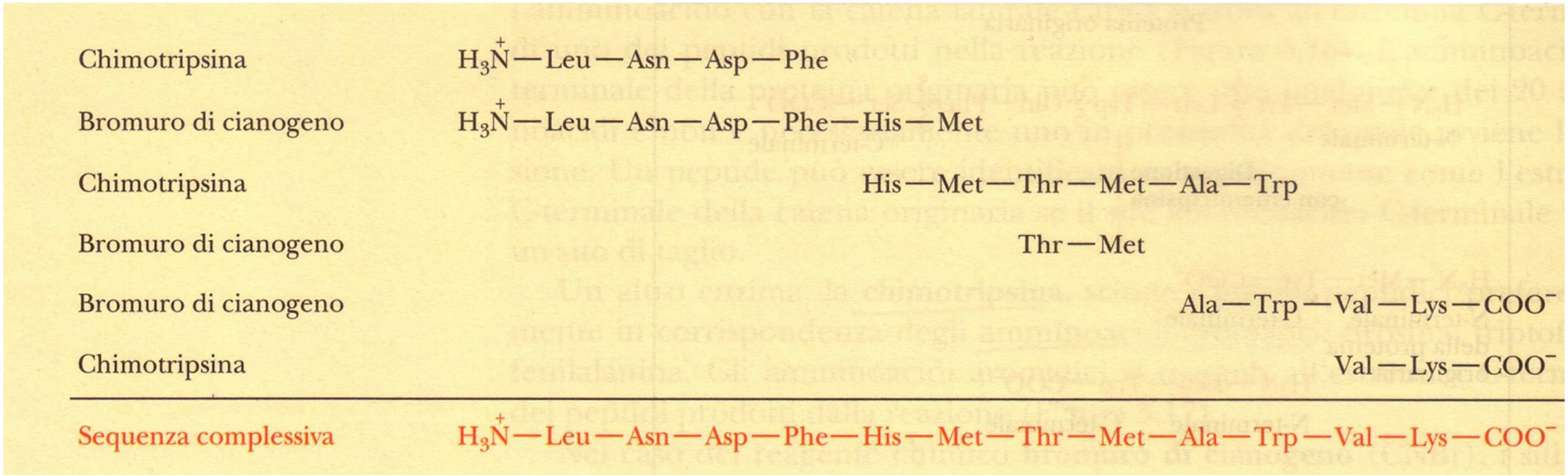
Peptide digestion with trypsin. A) Trypsin is a proteolytic enzyme, or protease, that cuts specifically only peptide bonds where arginine or lysine provide the carbonyl group. B) The reaction products are a mixture of peptide fragments with Arg and Lys as C-term amino acids and a single peptide deriving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



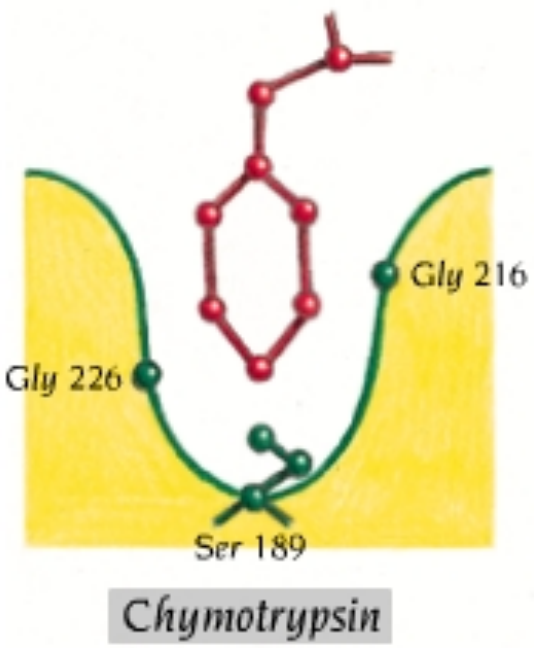
Campbell - Farrell
Biochimica
EdiSES

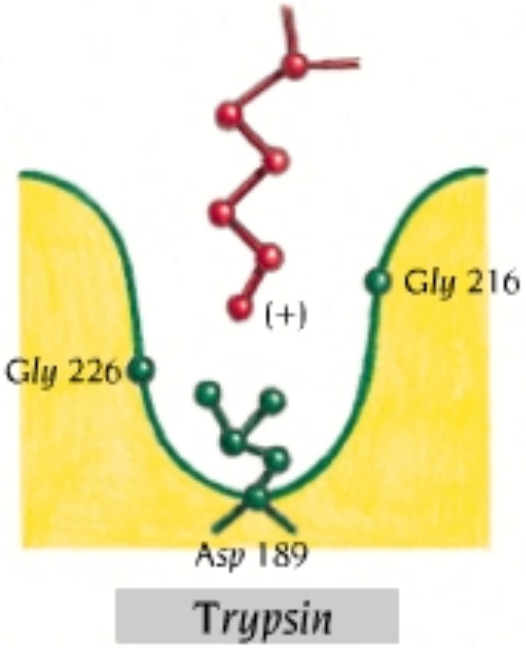
Chimotrypsin digestion of a protein. Chimoptrypsin hydrolyses proteins where aromatic amino acids are

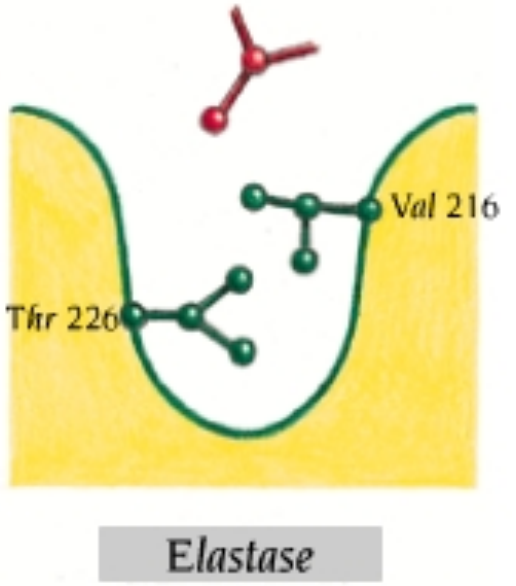
The overlap of the sequences of fragments allows to determine the protein sequence



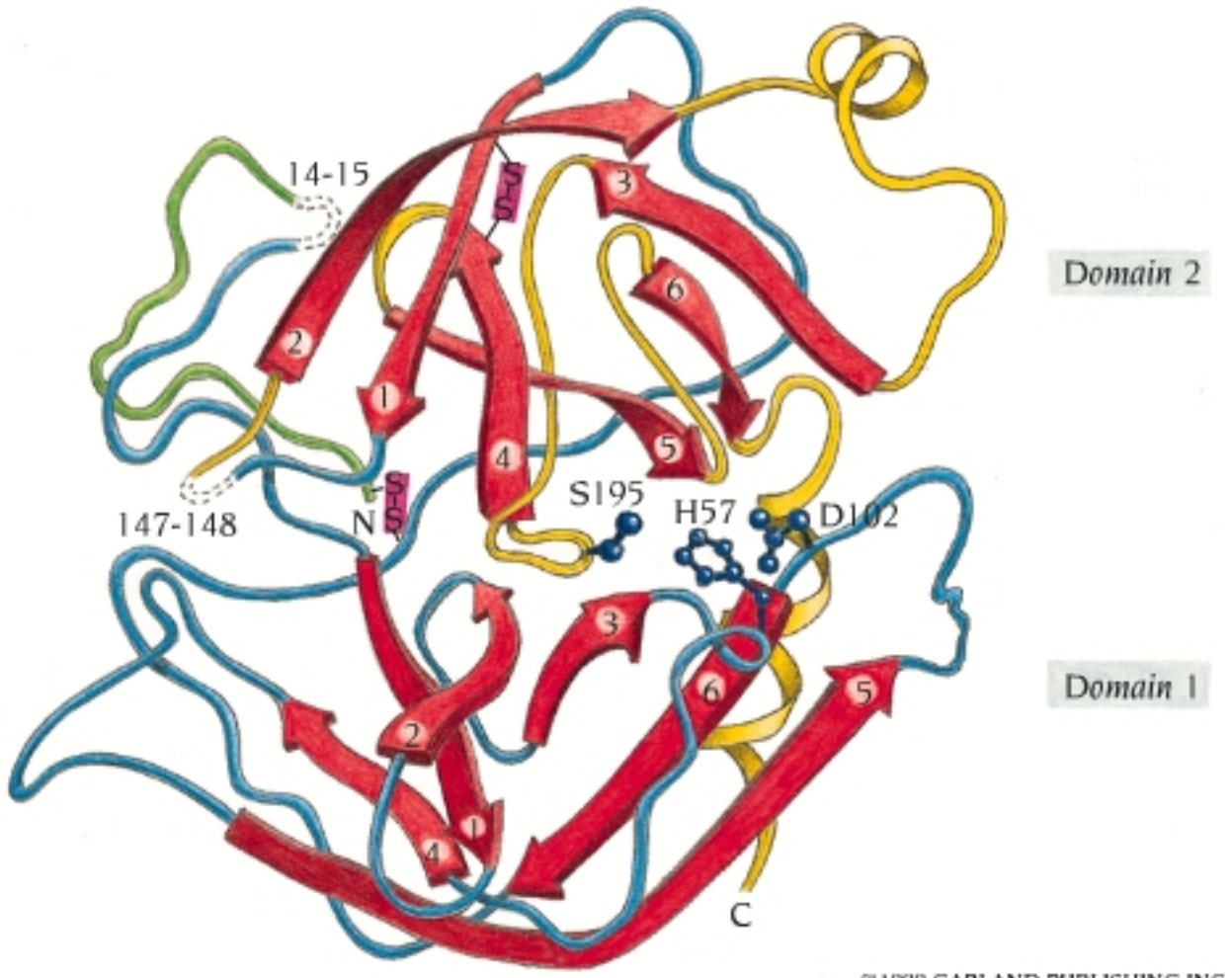
Protease active sites



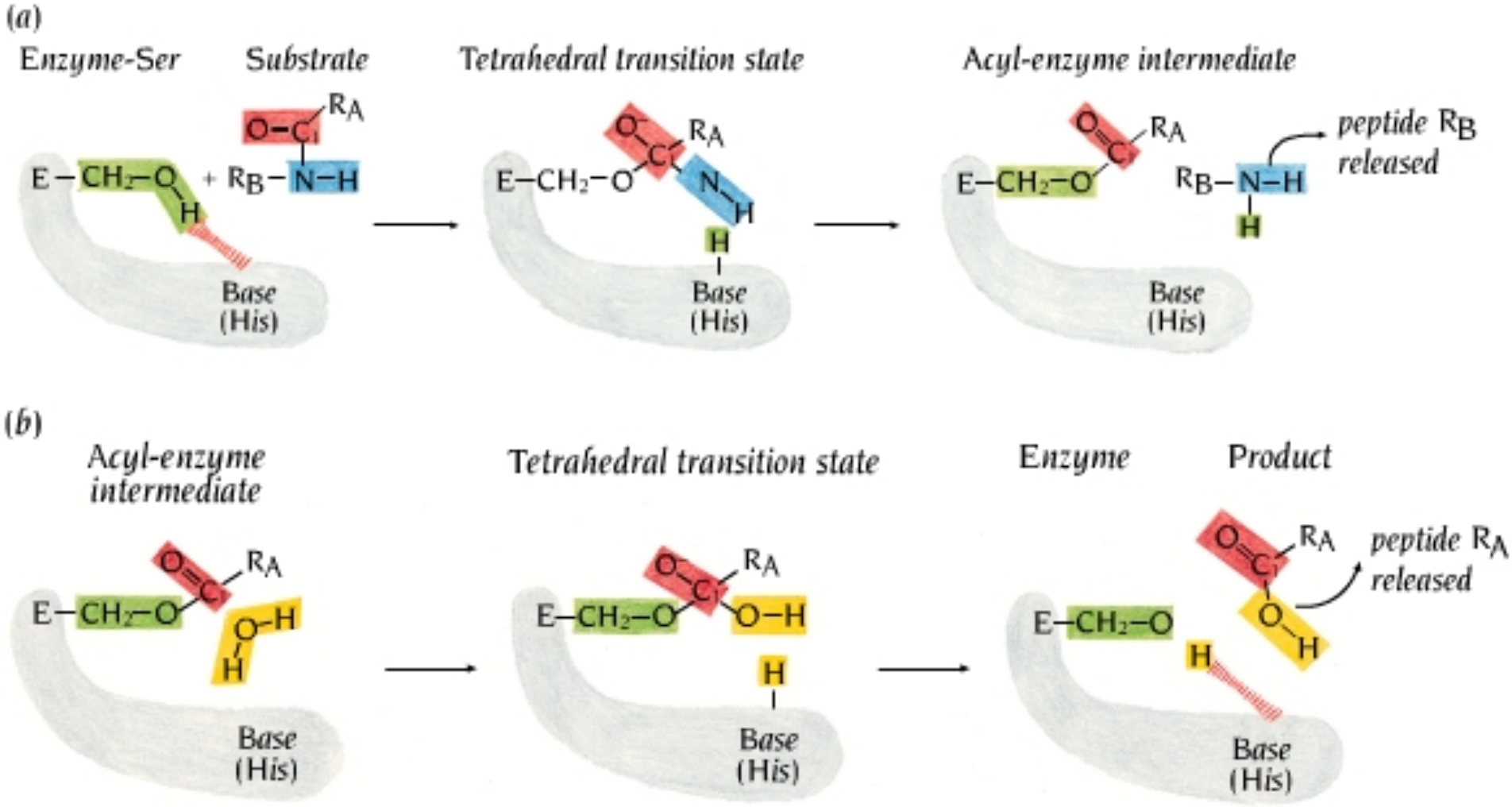


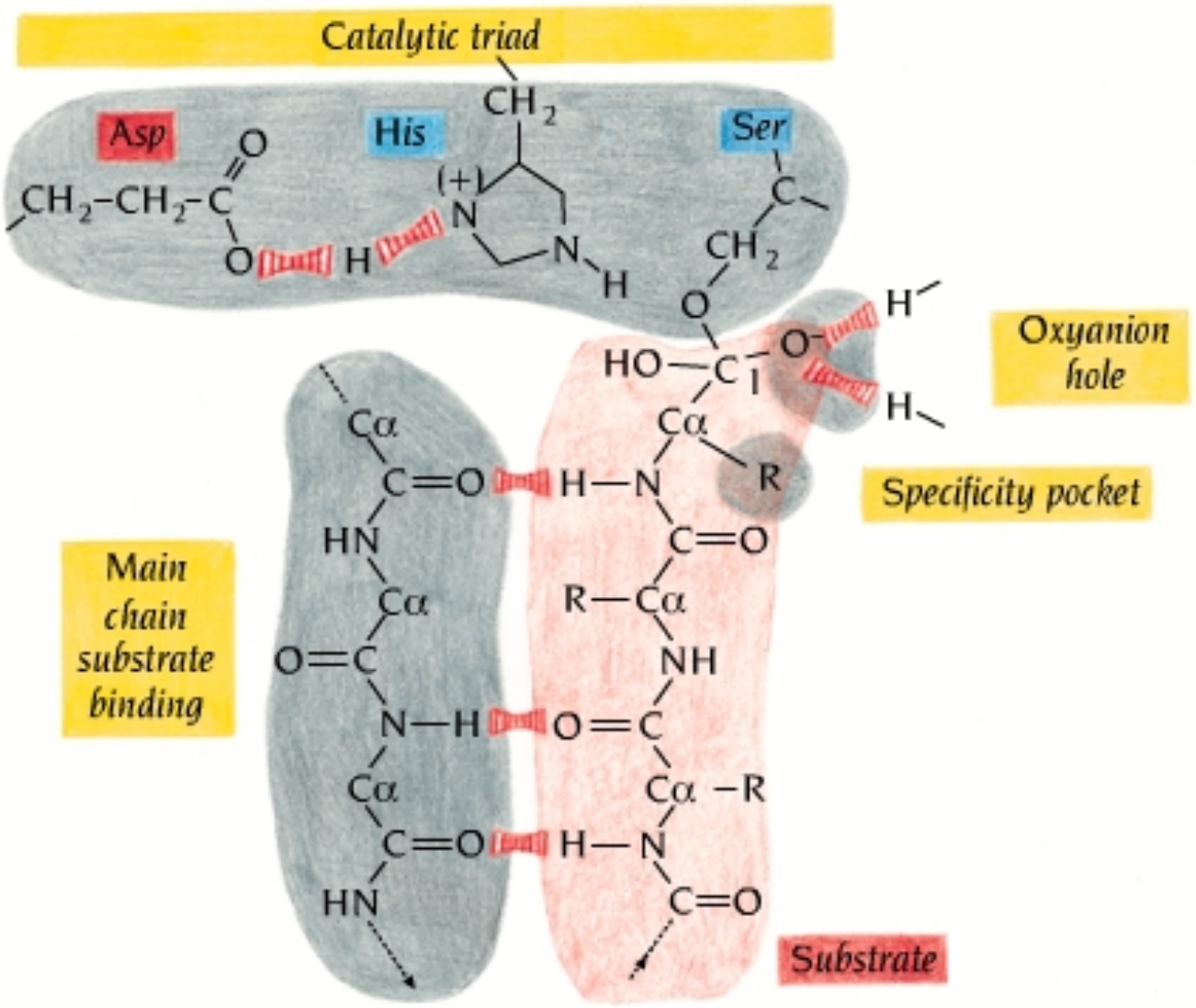


Mechanism of protease activity



Mechanism of protease activity





5. Edman degradation

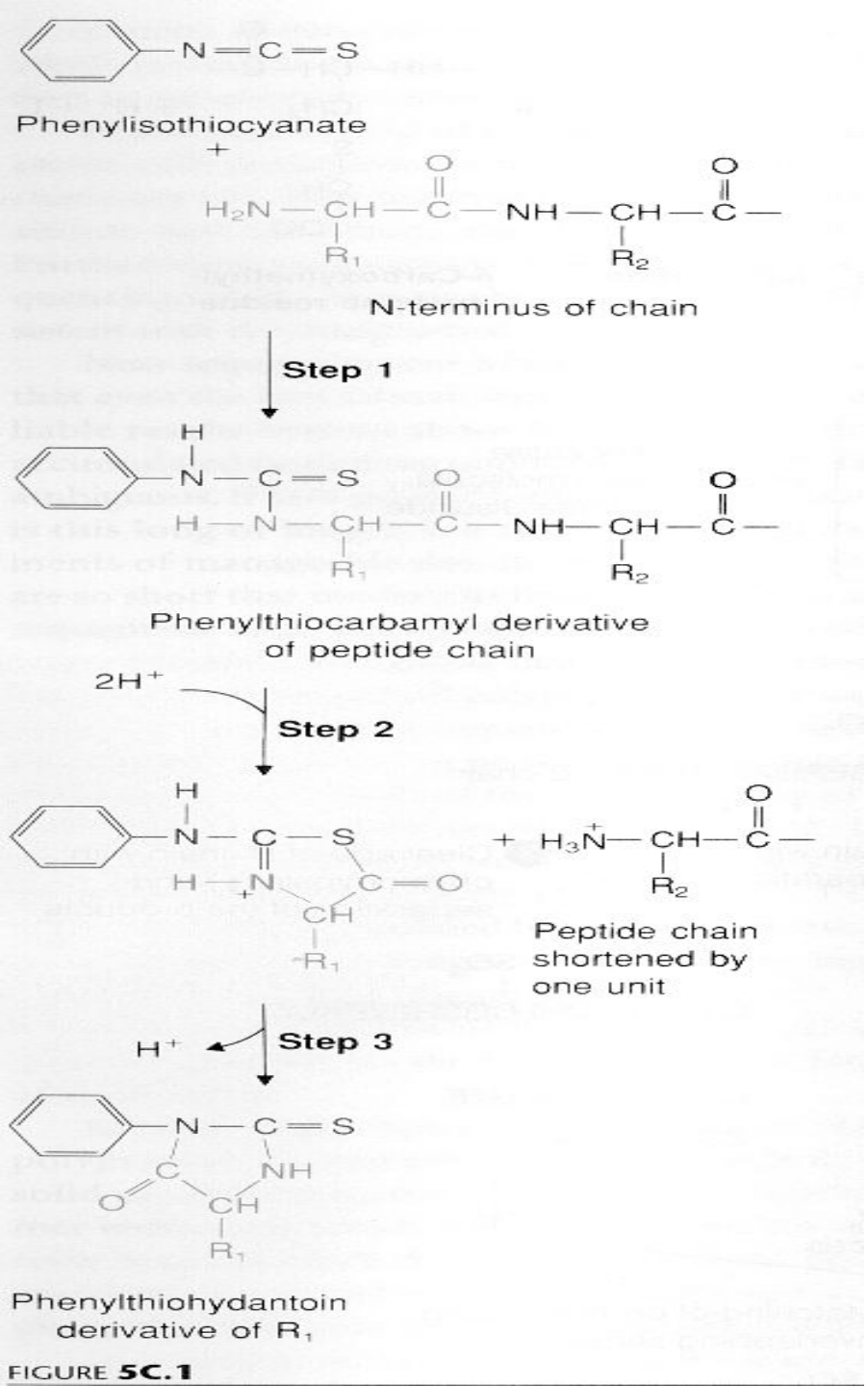
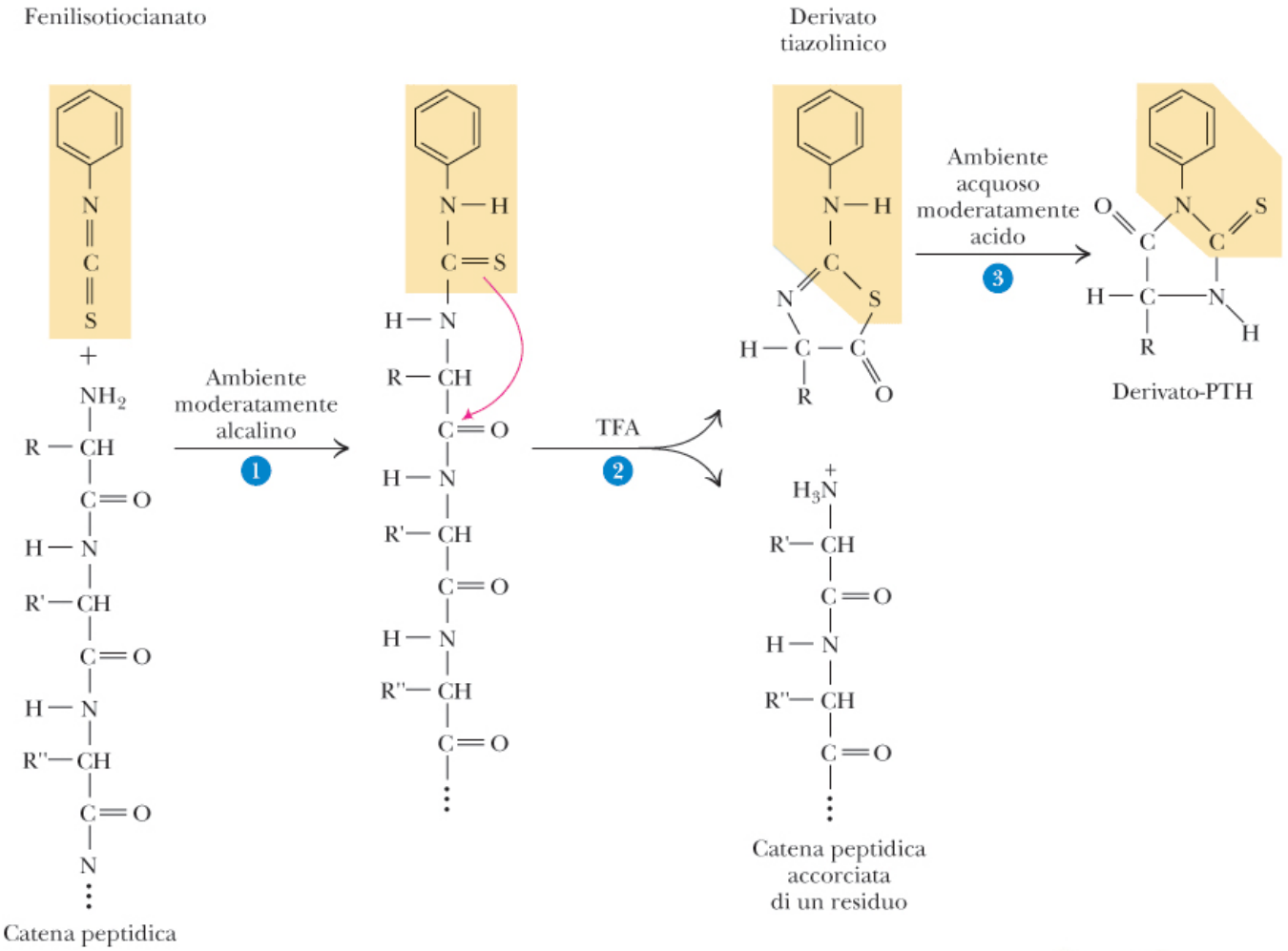


FIGURE 5C.1

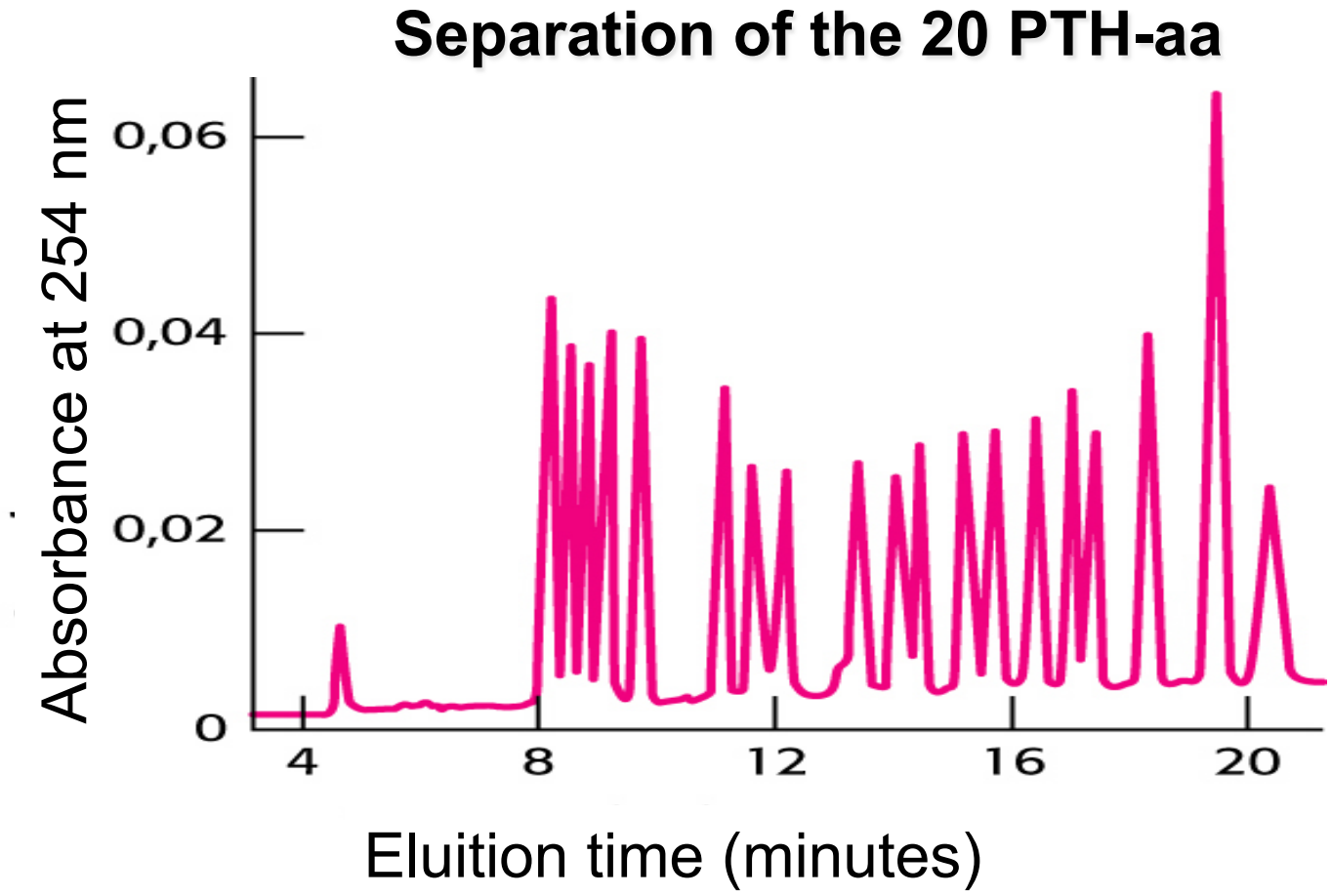


Edman degradation. 1) In moderate alkaline conditions, phenylthioisocianate combines with the N-terminal of the peptide to form a phenyl thiocarbamoyl (PTC) derivative. 2) After treatment with trifluoracetic acid (TFA), a cyclic compound is formed and the first N-term amino acid is released as a thiazoline derivative, whereas the other peptide binds are not hydrolized. 3) After organic extraction and treatment with 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.

Campbell - Farrell
Biochimica
EdiSES

http://www.youtube.com/watch?v=dom2oii_AHm

Identification of theN-termPTH-aa N-terminale through chromatography



Peptide analysers

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