What are the proteins?

Dr. Abel MORENO, Institute of Chemistry, UNAM. MEXICO.

E-mail address: carcamo@unam.mx or abel.moreno@mac.com

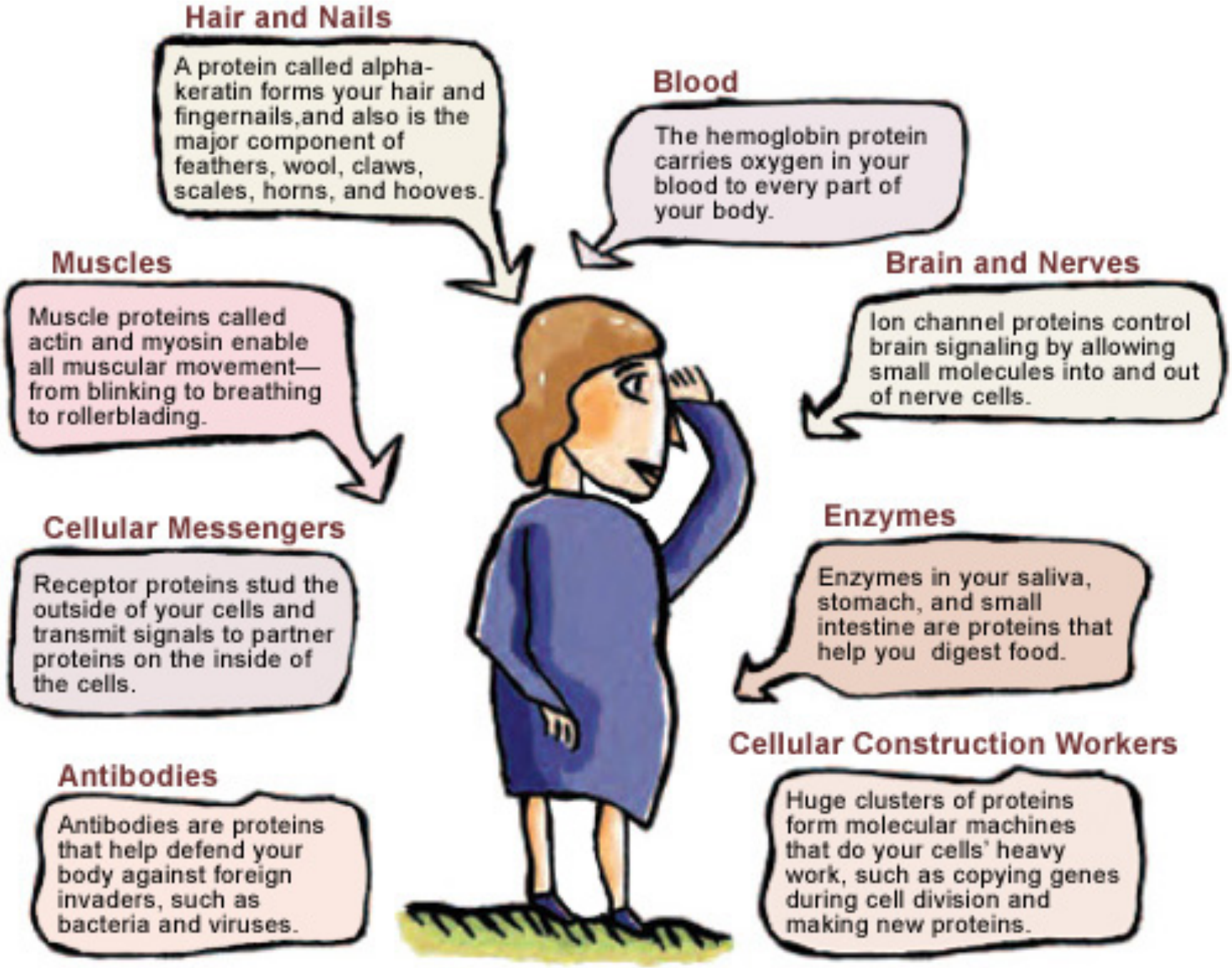
Outline of the talk and the course

- Amino acids, proteins, definition, function and structure.
- Chemical structure and functionality compared to the DNA, ARN, Membrane Proteins.
- Applications to biotechnology, medicine, materials science
- Protein Crystallography, Protein Crytallogenesis.

VIDEO about Protein biosynthesis and the cell function.

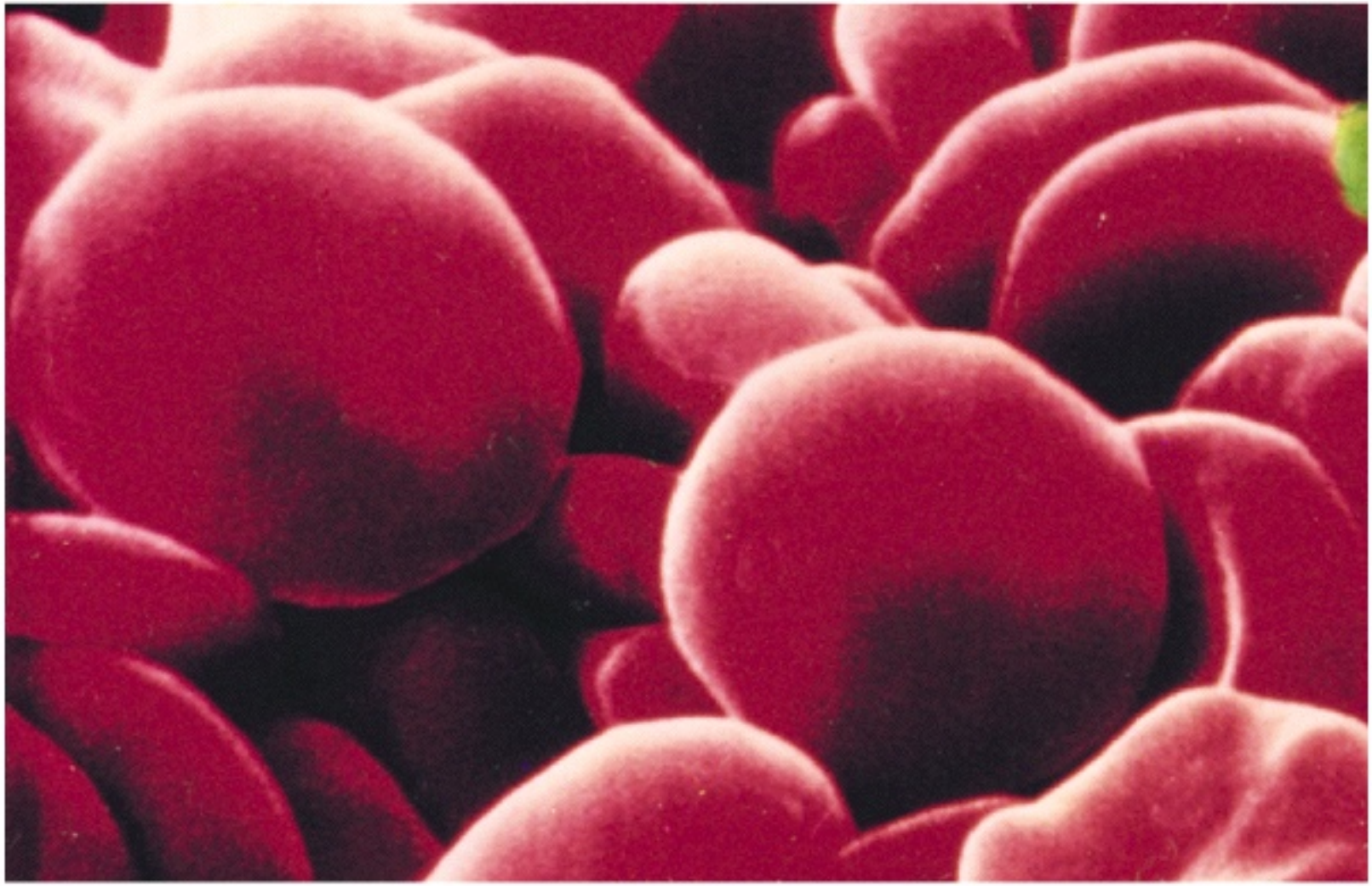
“Quisiera que la palabra que le propongo, proteína, se entendiera como derivada de PROTEIOS, porque describe la sustancia primitiva o principal de la nutrición animal que las plantas elaboran para los herbívoros y que estos proporcionan a los carnívoros”

J. J. Berzelius, carta a G. J. Mulder, 1838.


Hair and Nails
A protein called alpha-keratin forms your hair and fingernails,and also is the major component of feathers, wool, claws, scales, horns, and hooves.
Blood
The hemoglobin protein carries oxygen in your blood to every part of your body.
Muscles
Muscle proteins called actin and myosin enable all muscular movement—from blinking to breathing to rollerblading.
Brain and Nerves
Ion channel proteins control brain signaling by allowing small molecules into and out of nerve cells.
Cellular Messengers
Receptor proteins stud the outside of your cells and transmit signals to partner proteins on the inside of the cells.
Enzymes
Enzymes in your saliva, stomach, and small intestine are proteins that help you digest food.
Antibodies
Antibodies are proteins that help defend your body against foreign invaders, such as bacteria and viruses.
Cellular Construction Workers
Huge clusters of proteins form molecular machines that do your cells' heavy work, such as copying genes during cell division and making new proteins.

(a)

(c)



(b)

Protein

20 L-amino acids

Essential	Nonessential
Isoleucine***	Alanine
Leucine***	Arginine* #
Lysine***	Aspartate
Methionine***	Cysteine* #
Phenylalanine ***	Glutamate
Threonine***	Glutamine*
Tryptophan***	Glycine*
Valine***	Proline*
Histidine***	Serine
Asparagine*	Tyrosine* #
Selenocysteine**	Pyrrolysine**

(*) Essential only in certain cases.[1][2]
(**) Truly unclassified. Added to sustain the 22 Numbers of Essential Amino Acids.
(***) Amino acids regarded as essential for humans
(#) Amino acids required by infants and growing children.

Some important points:

- Essential amino acids are "essential" not because they are more important to life than the others, but because the body does not synthesize them, making it essential to include them in one's diet in order to obtain them.

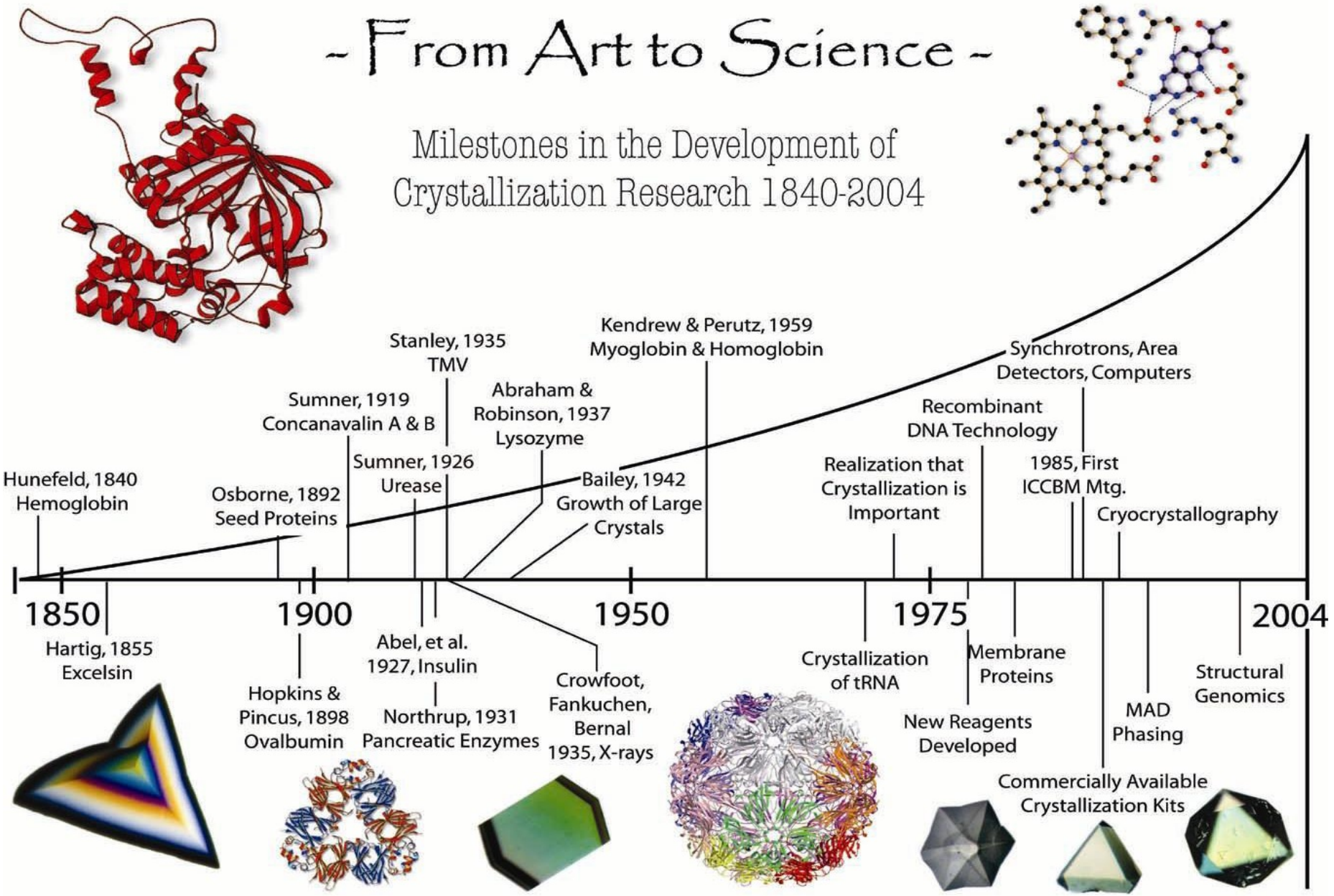
- The amino acids arginine, cysteine, glycine, glutamine, histidine, proline, serine, and tyrosine are considered conditionally essential, meaning they are not normally required in the diet, but must be supplied exogenously to specific populations that do not synthesize it in adequate amounts.

- A very interesting point is that lysine is not easily synthesized by humans in a huge amounts, bacteria, fungus, and some proteins from animals as well as cereals are rich in this essential amino acid.

- Amino acids containing sulfur can be converted each other.

Protein Crystallogenesis

This is part of the Crystallography devoted to understand the physicochemical aspects related to the nucleation, and crystal growth of biological macromolecules. The methodological approach uses tools from Chemistry, Physics and Biology.



Alex McPherson's Lecture at RAMC meeting (Le Bischenberg France, 2011).

Protein crystallization what for?

At the beginning (latter half of the 19th century):

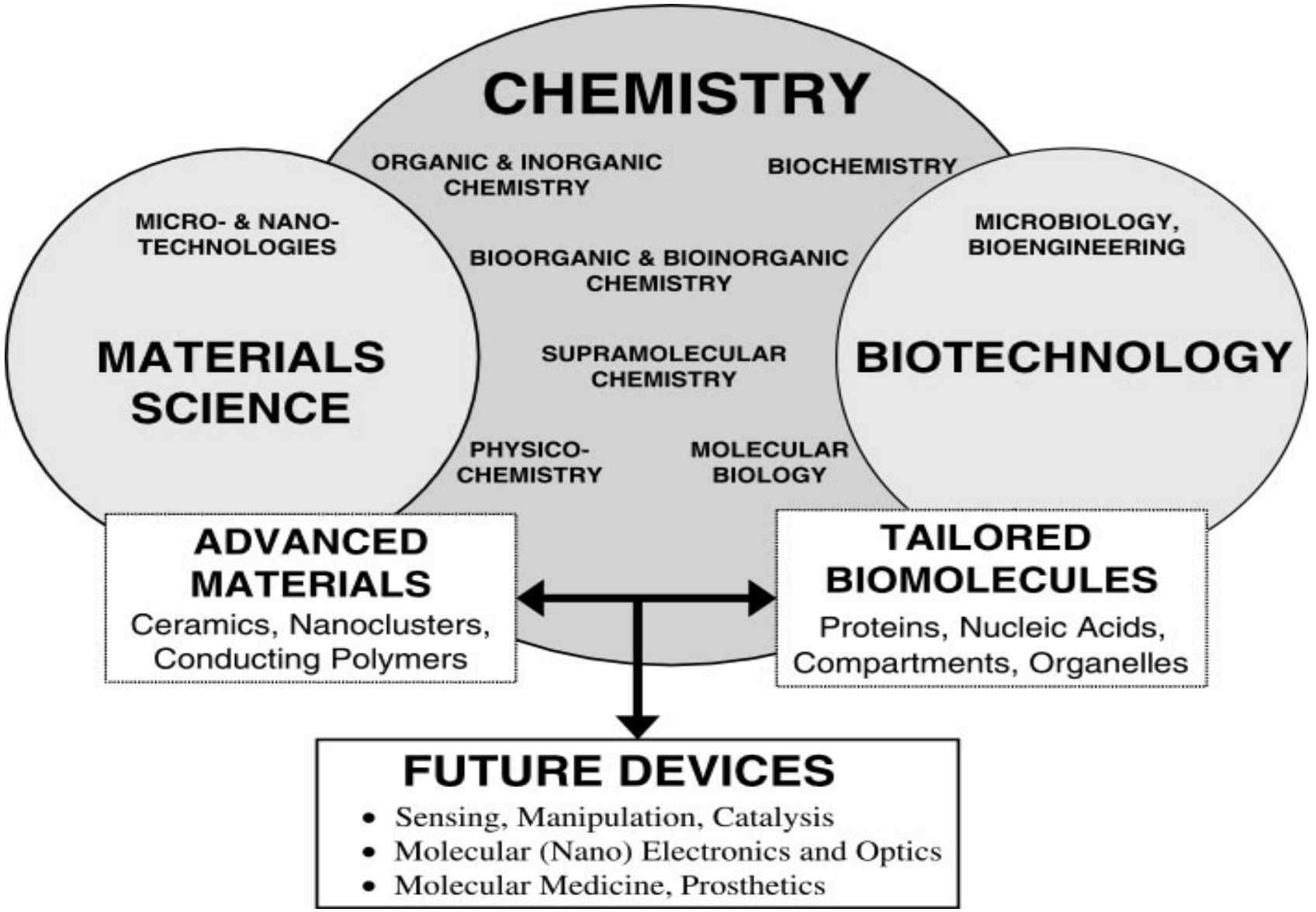
1. It provided a means for the purification of specific proteins.
2. It served as a demonstration that a protein had been purified.
3. It was an interesting laboratory curiosity.

Between 1900 and 1940

1. Emphasis was on enzymes to prove properties and nature of catalytic macromolecules.
2. Protein crystallization for X-ray diffraction.

In the 1980's and up to now

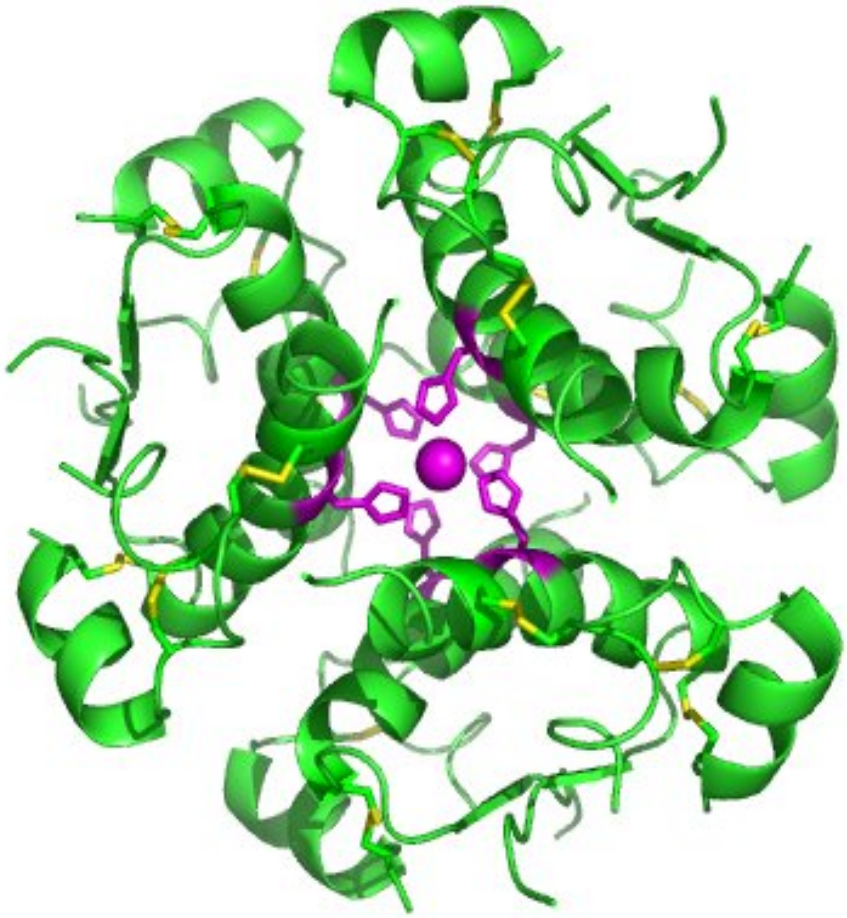
1. Due to the development of the recombinant DNA technology permitted researchers, for the first time, to prepare ample amounts of otherwise rare and elusive proteins.
2. Structural biologists would like to describe all living systems, and the materials they produce, in molecular and even atomic terms.



Reference: Nanoparticles, Proteins, and Nucleic Acids: Biotechnology Meets Materials Science Christof M. Niemeyer* Angew. Chem. Int. Ed. 2001, 40, 4128-4158.

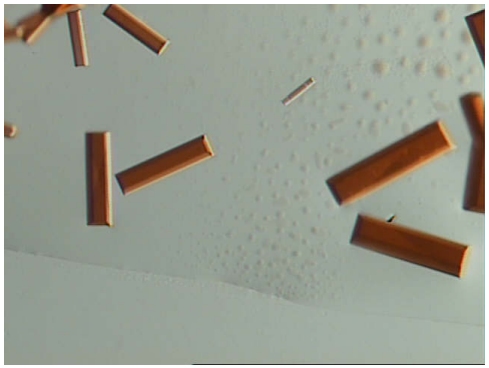
Insulin a recombinant protein available in the market thanks to Biotechnology

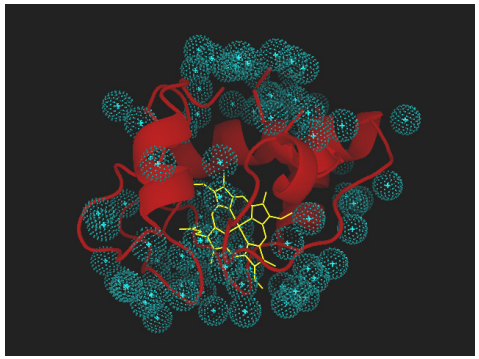
In 1969, after decades of work, Dorothy Crowfoot Hodgkin determined the spatial conformation of the molecule, the so-called tertiary structure, by means of X-ray diffraction studies. She had been awarded a Nobel Prize in Chemistry in 1964 for the development of crystallography.

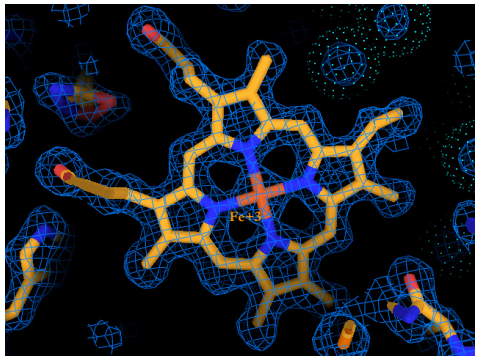


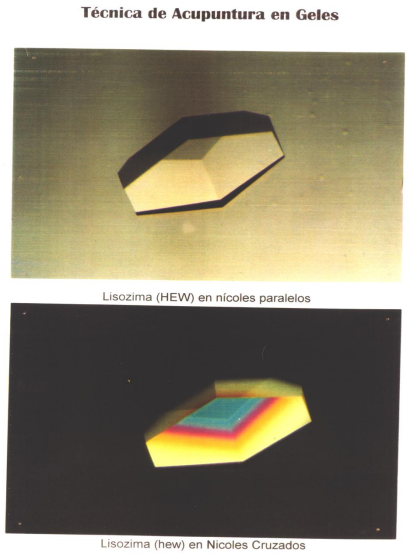
e.g Probiomed S.A. Mexican Company.

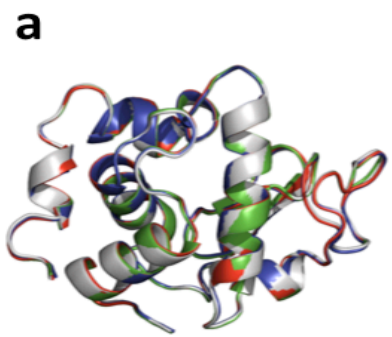
PROTEINS AND ENZYMES AT A GLANCE FROM X-RAY CRYSTALLOGRAPHY

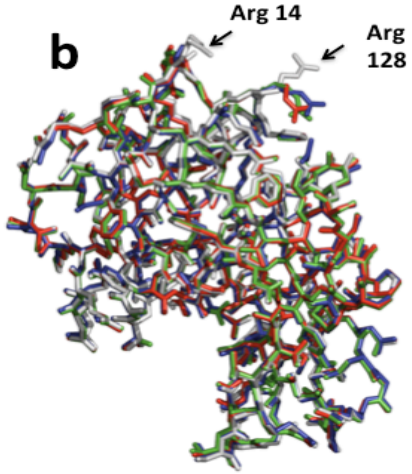


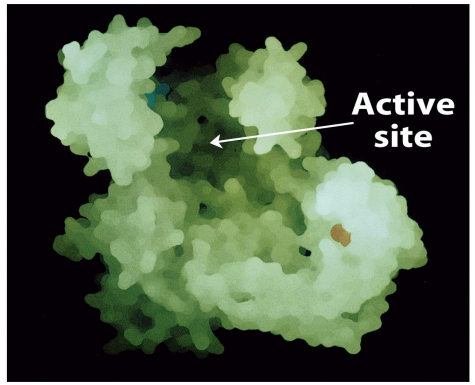




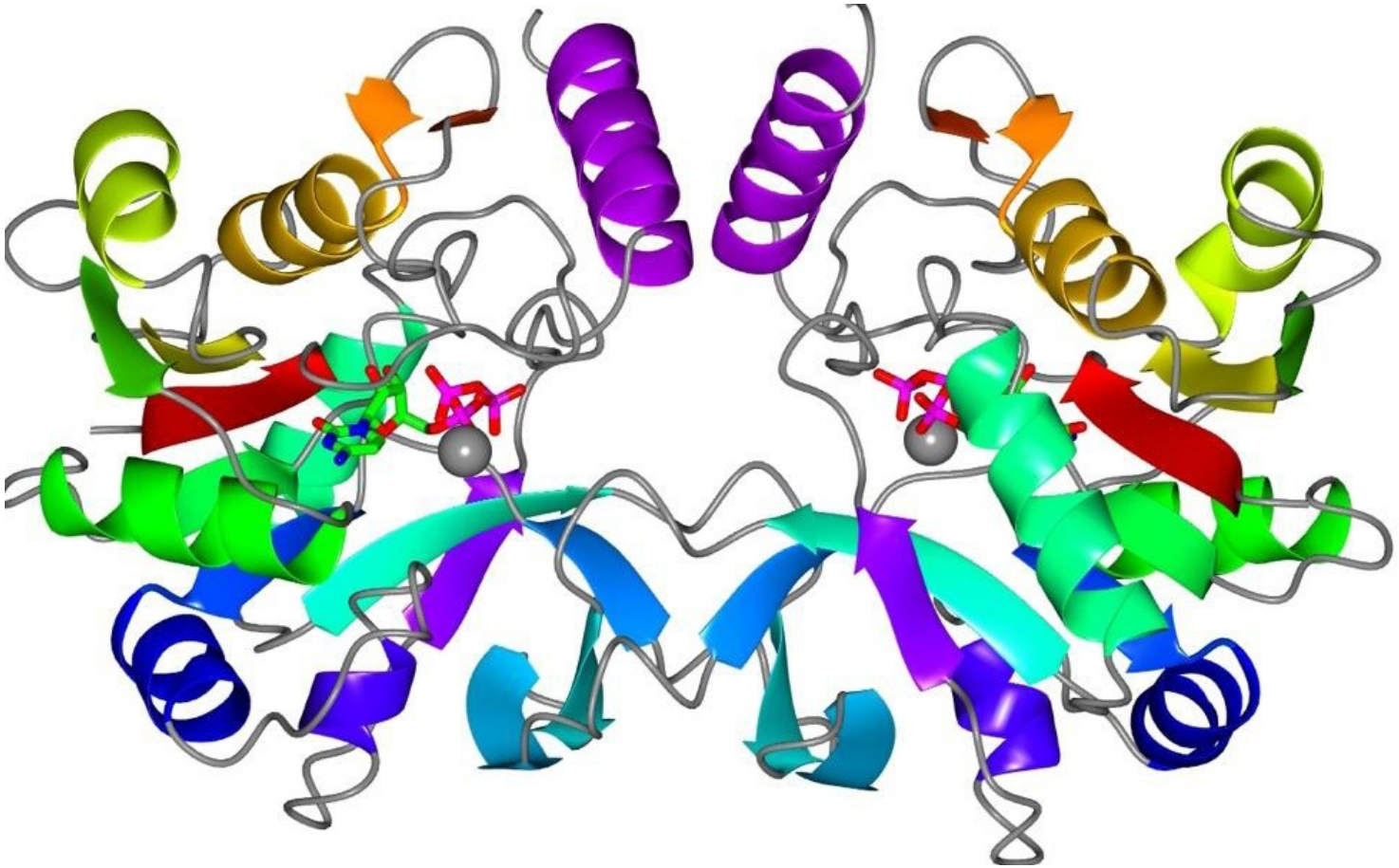




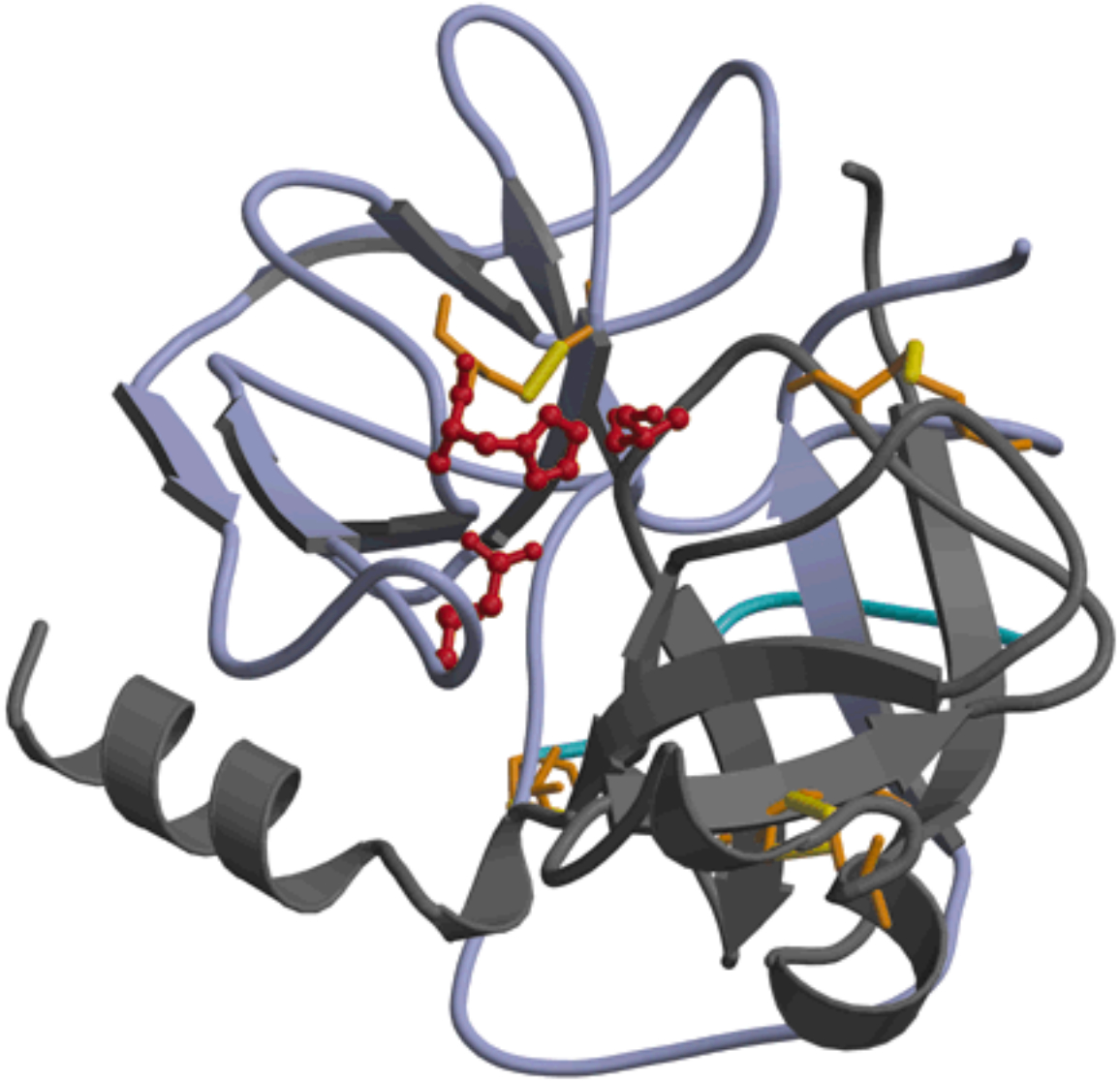


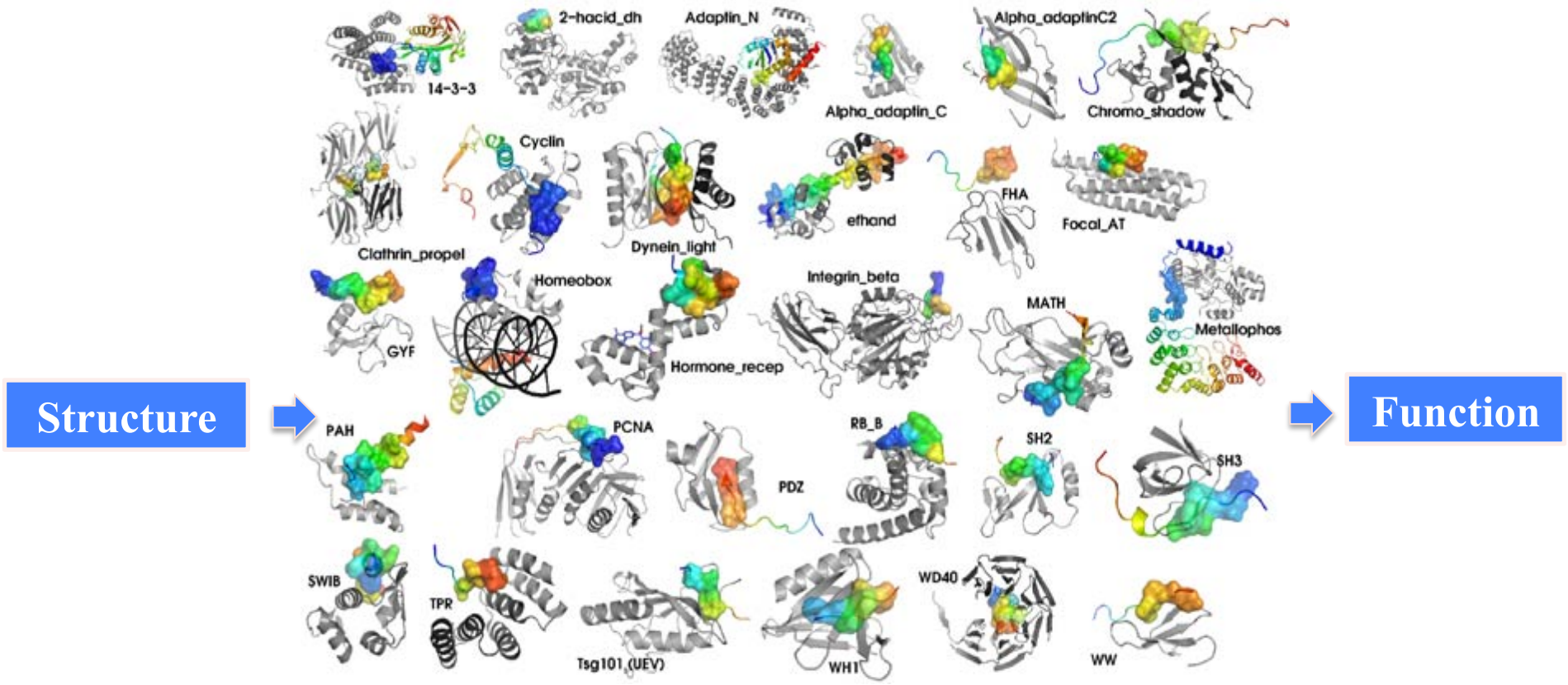


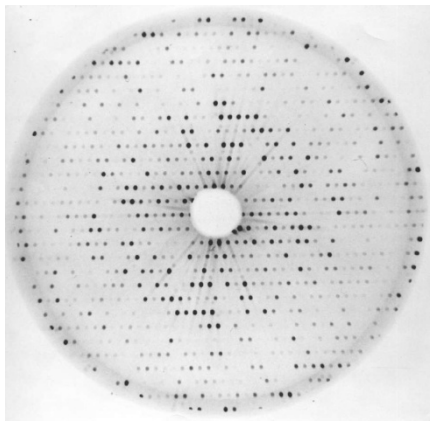
M. smegmatis IspD + CTP
M. smegmatis IspD + CMP
M. tuberculosis IspD + CTP

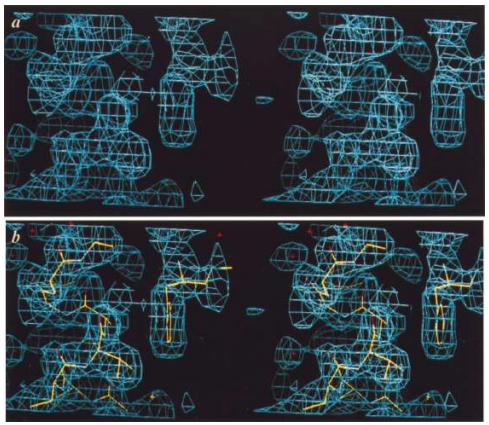


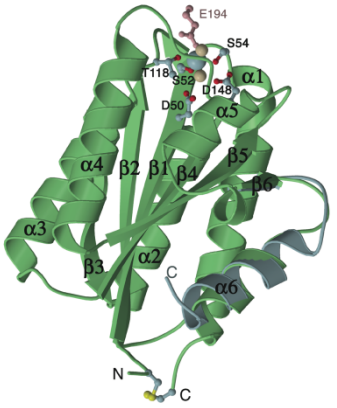
Björkelid, Bergfors et al Acta D 2011





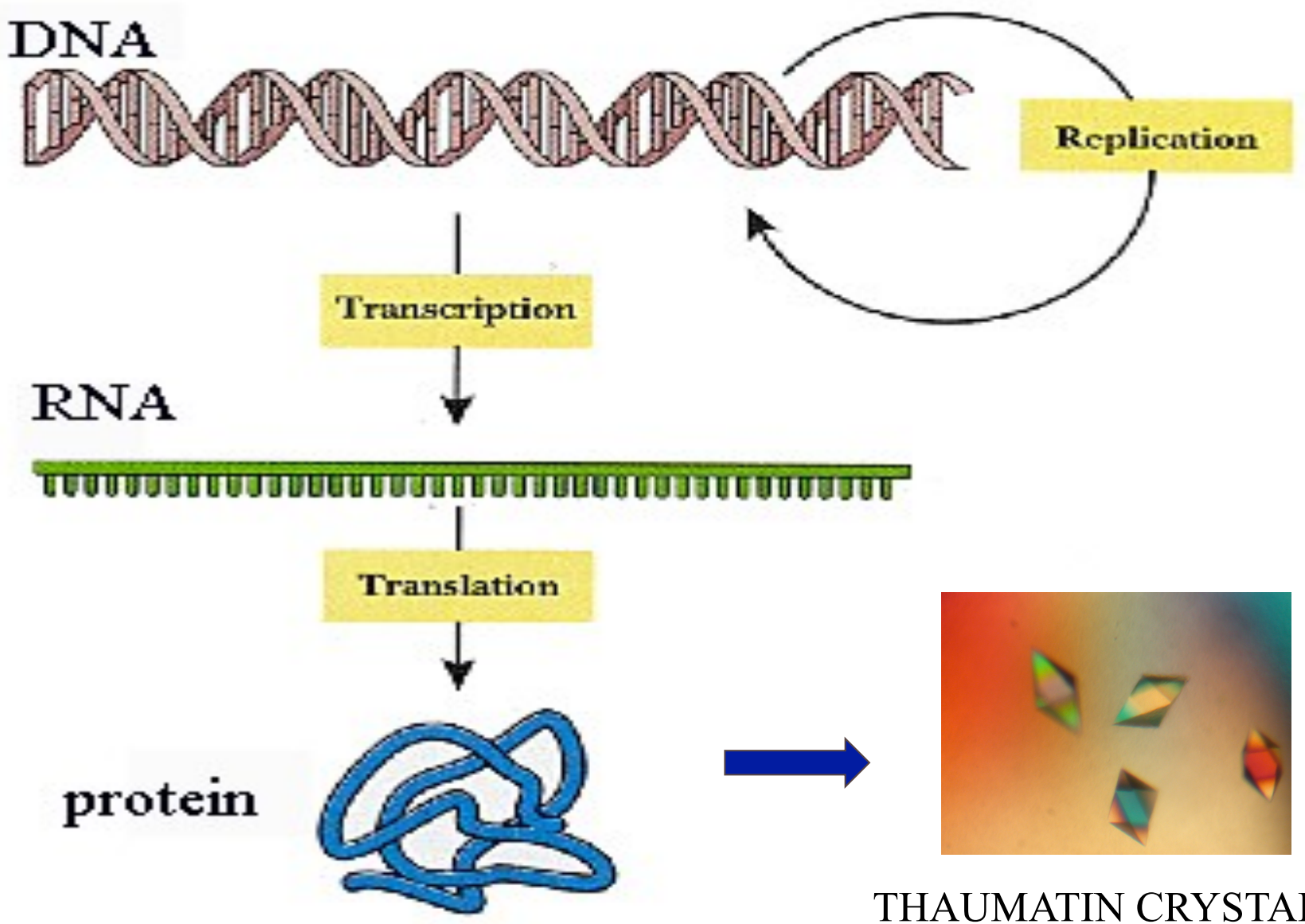




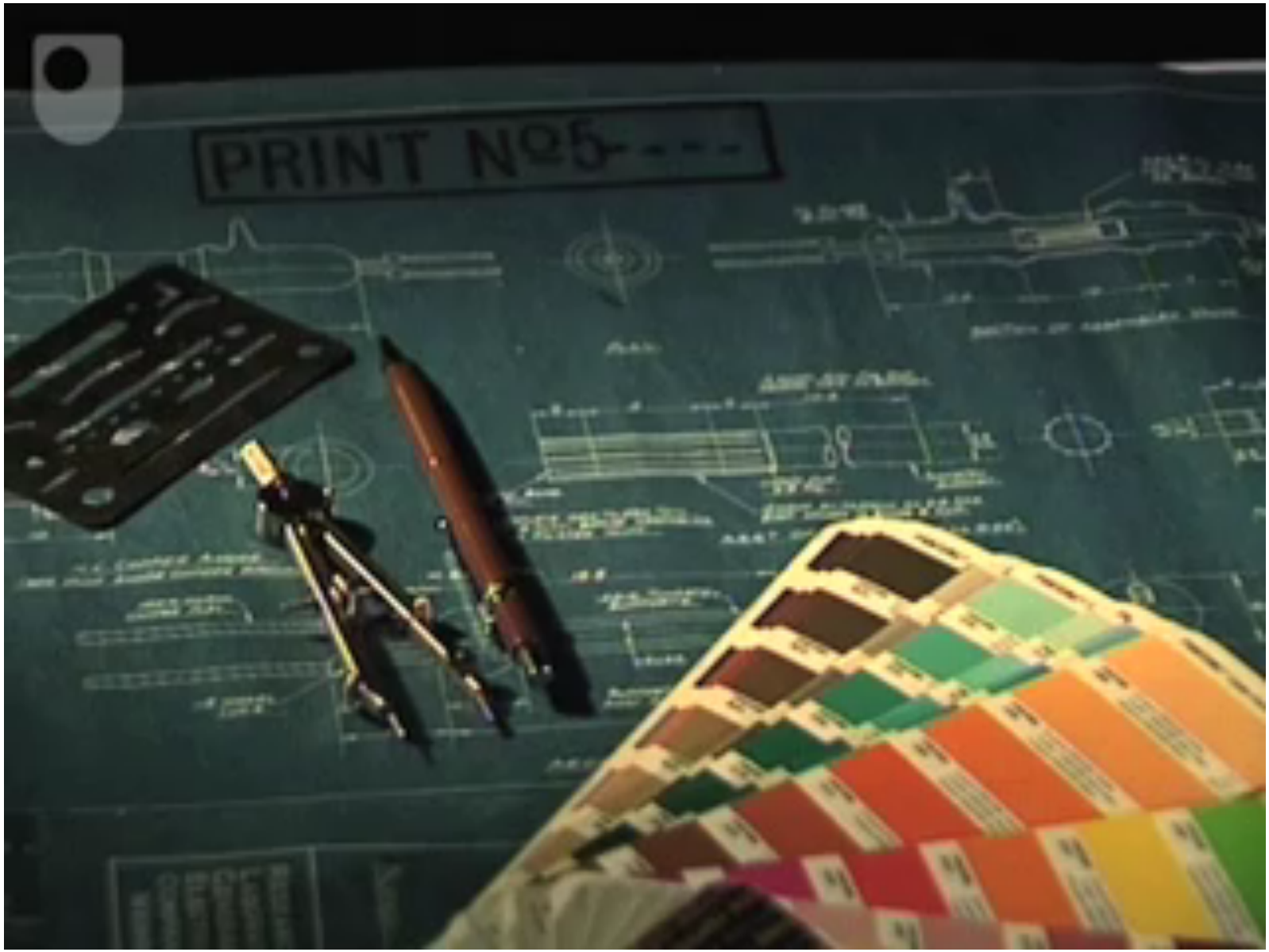


Crystal -> Diffraction pattern -> Electron density -> Model

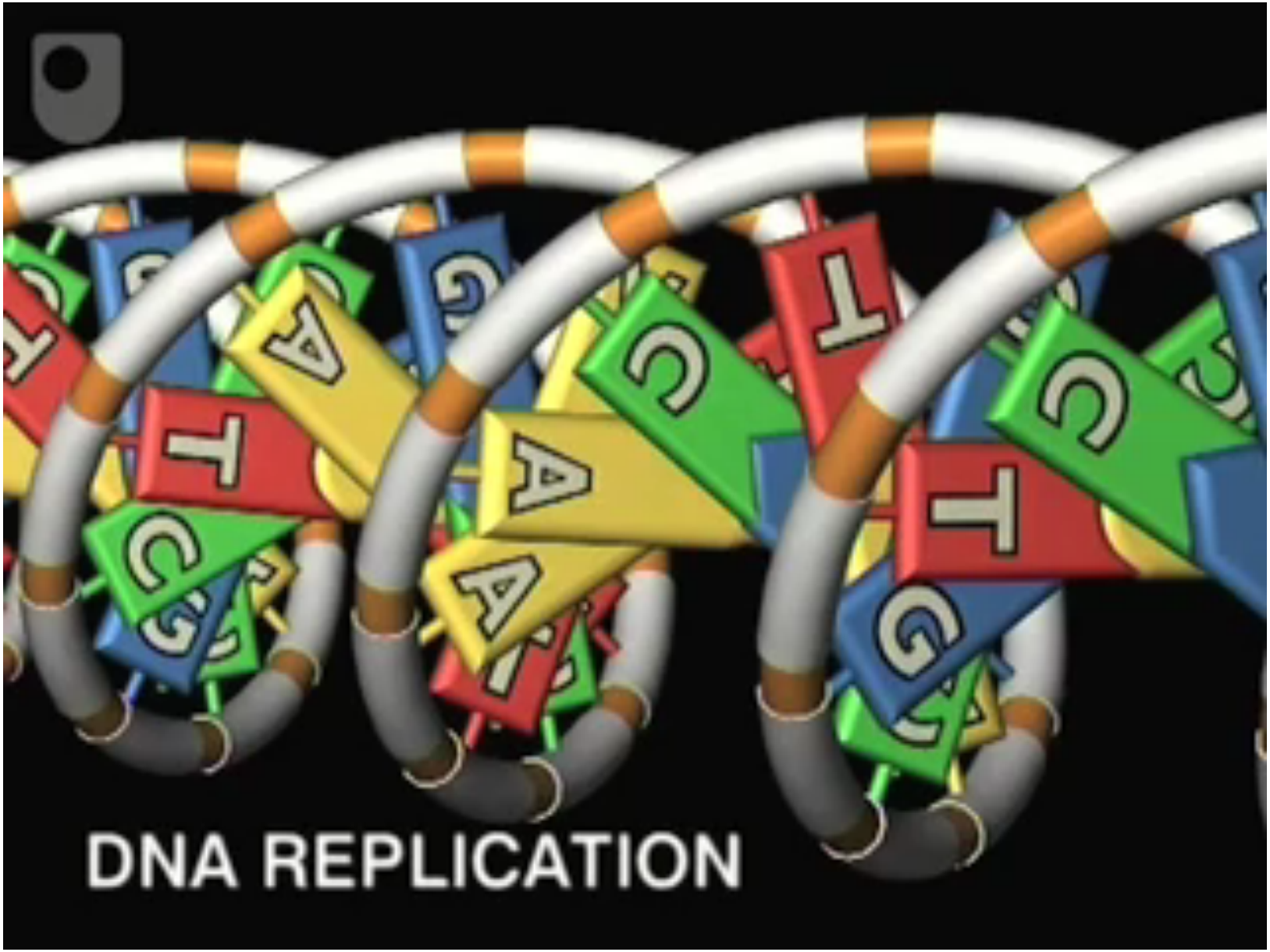
DNA/RNATHE ORIGIN OF LIFE AND THE BIOSYNTHESIS OF PROTEINS



THAUMATIN CRYSTALS

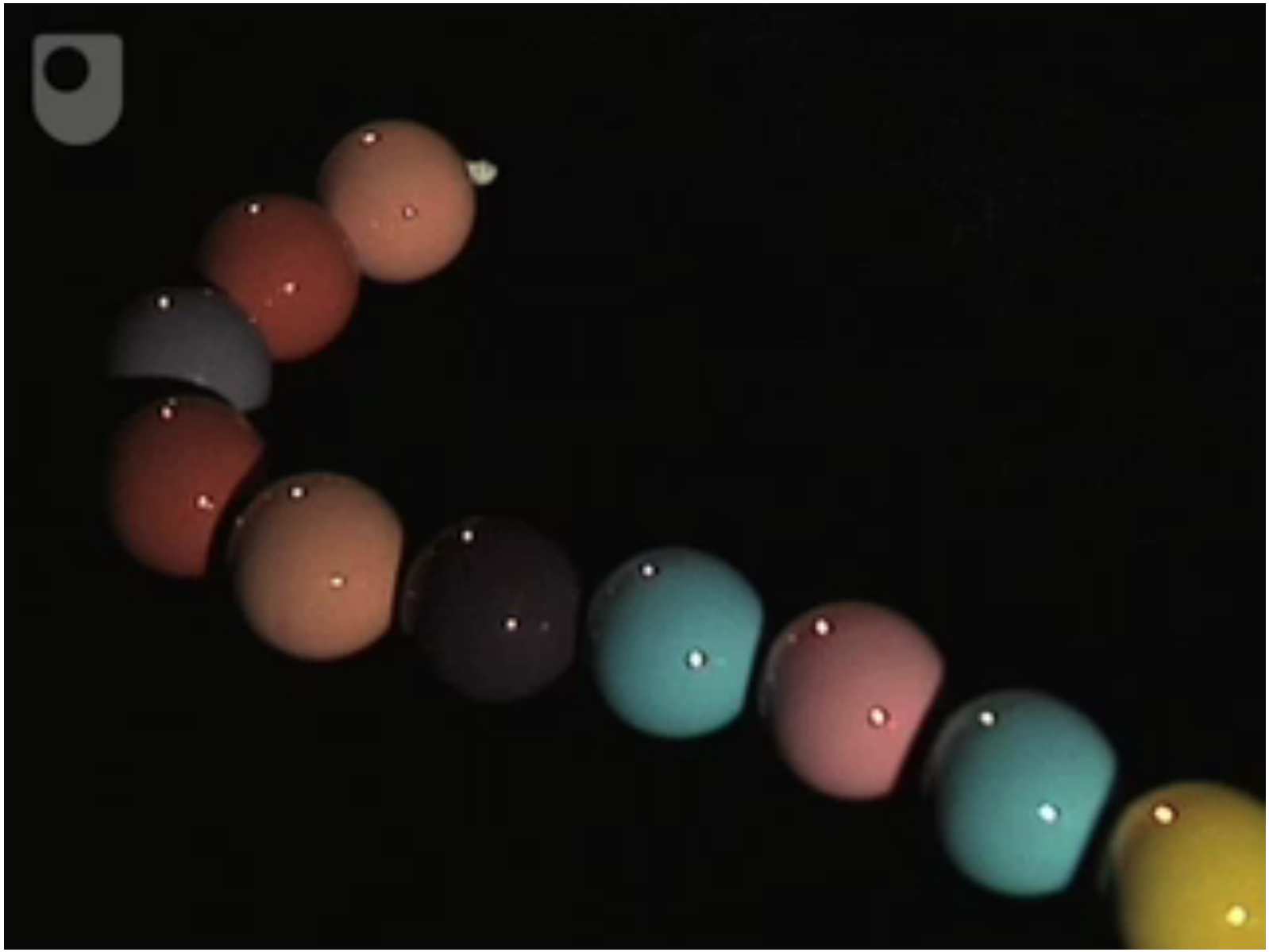

PRINT Nº5


GENE TO PROTEIN
AN OVERVIEW

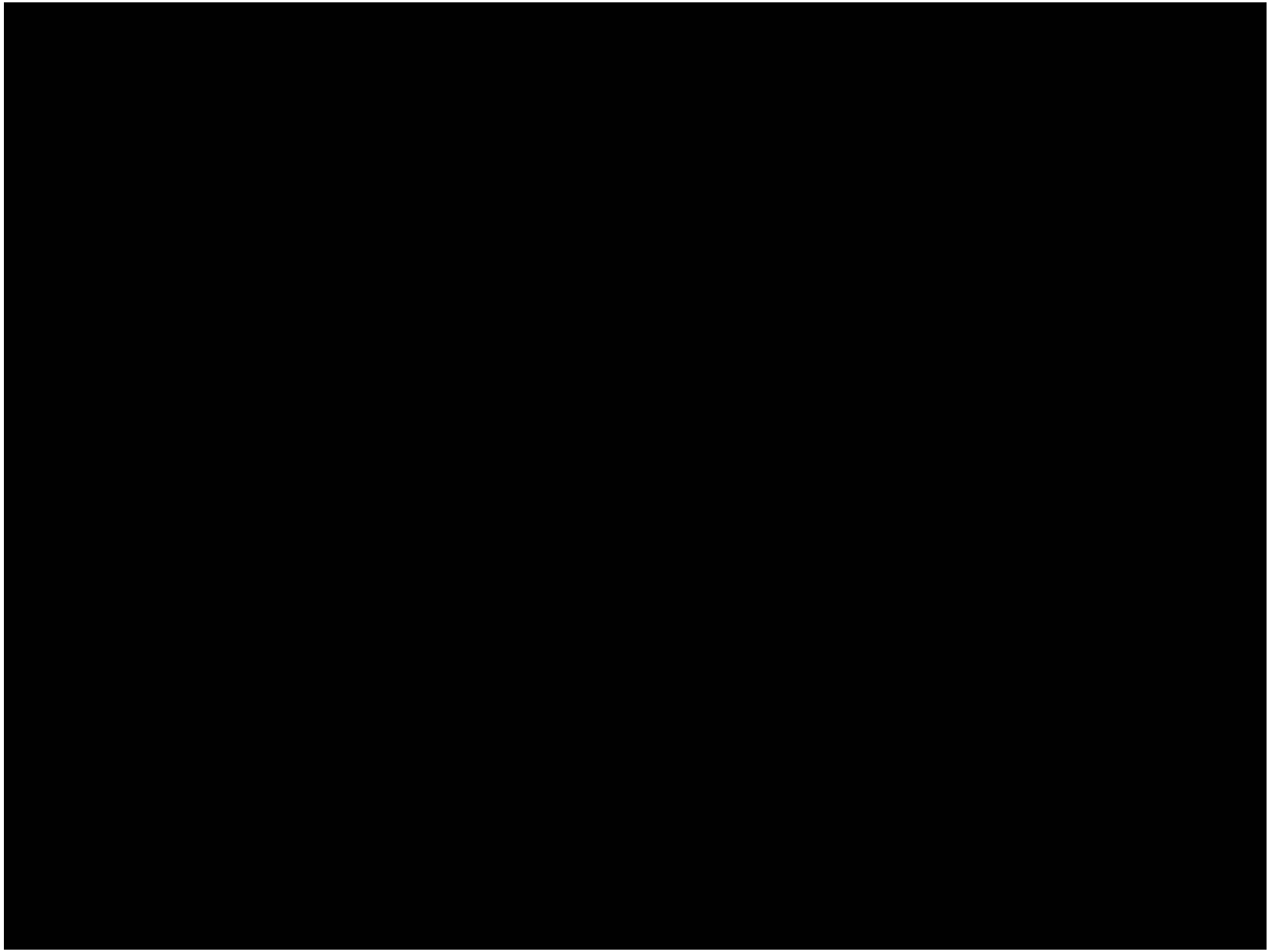

DNA REPLICATION


TRANSCRIPTION
WRITING THE MESSAGE

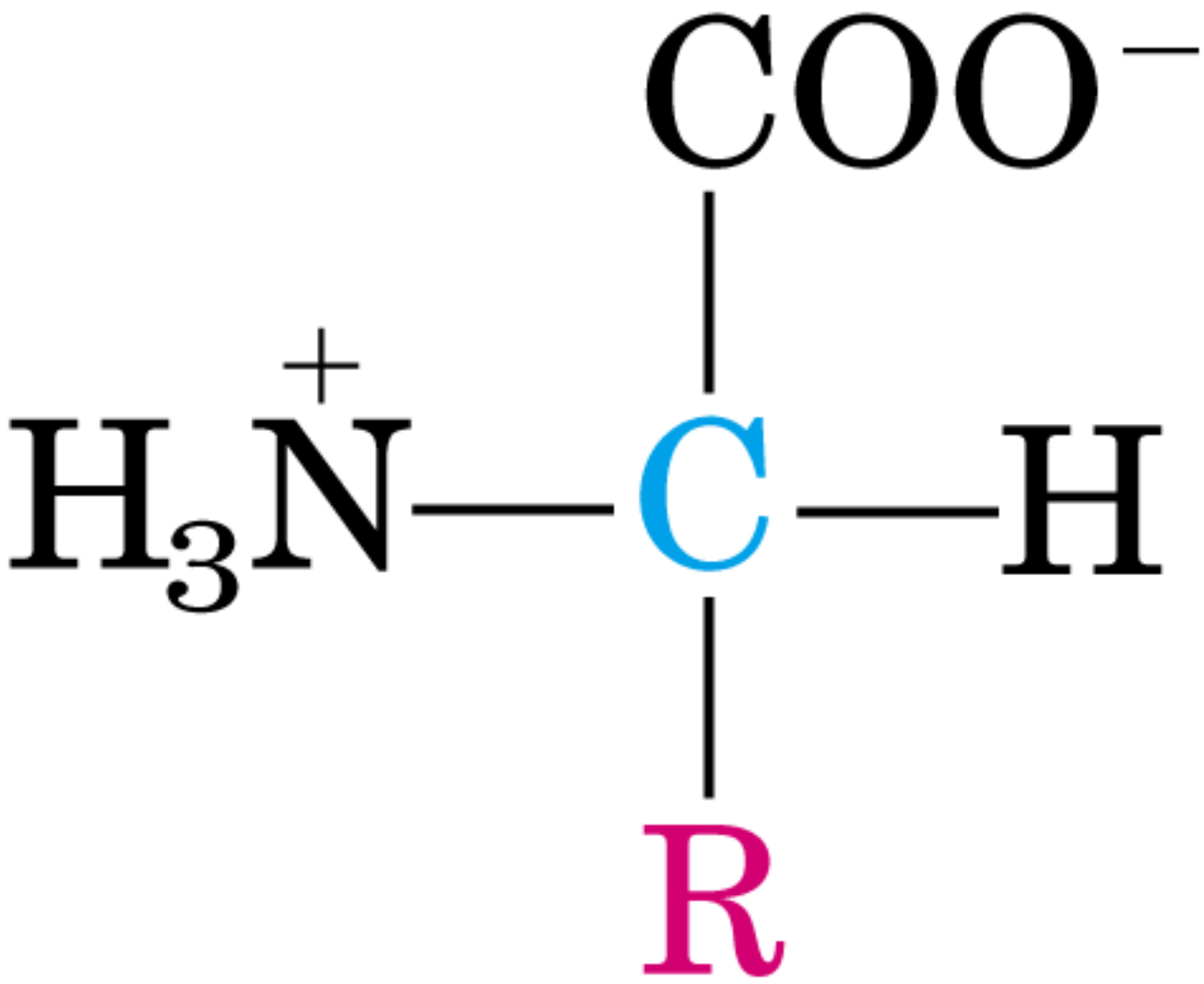

U
C
G
C
A
G
A
U
G
G
C
U
G
C
U
TRANSLATION
READING THE MESSAGE

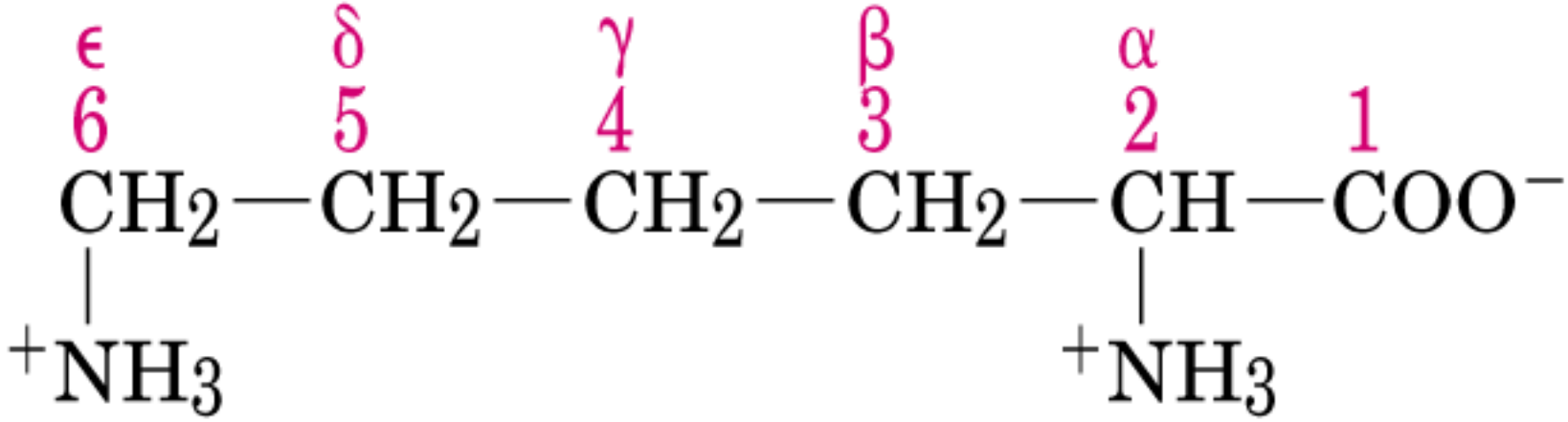


ACT I

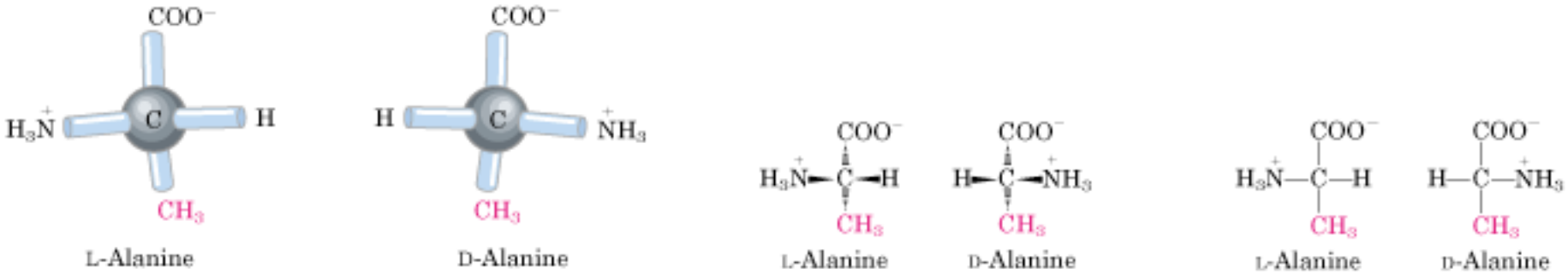


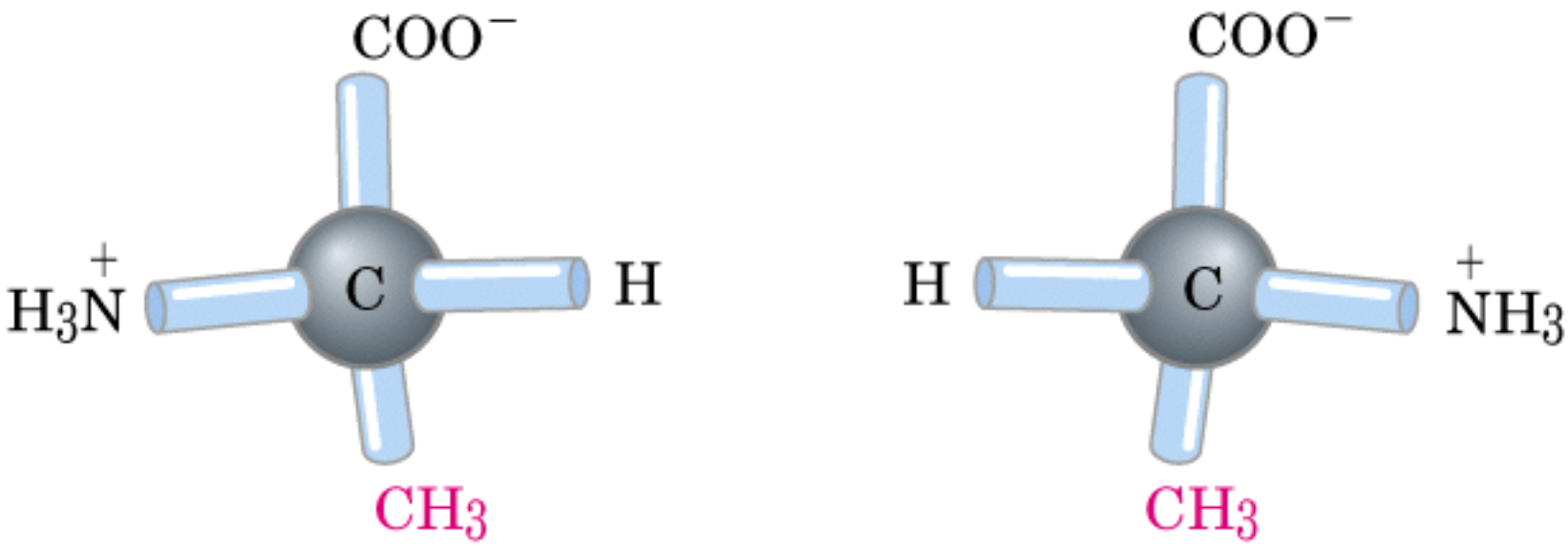
Amino acids, peptides and proteins


COO⁻
H3N+
C
H
R



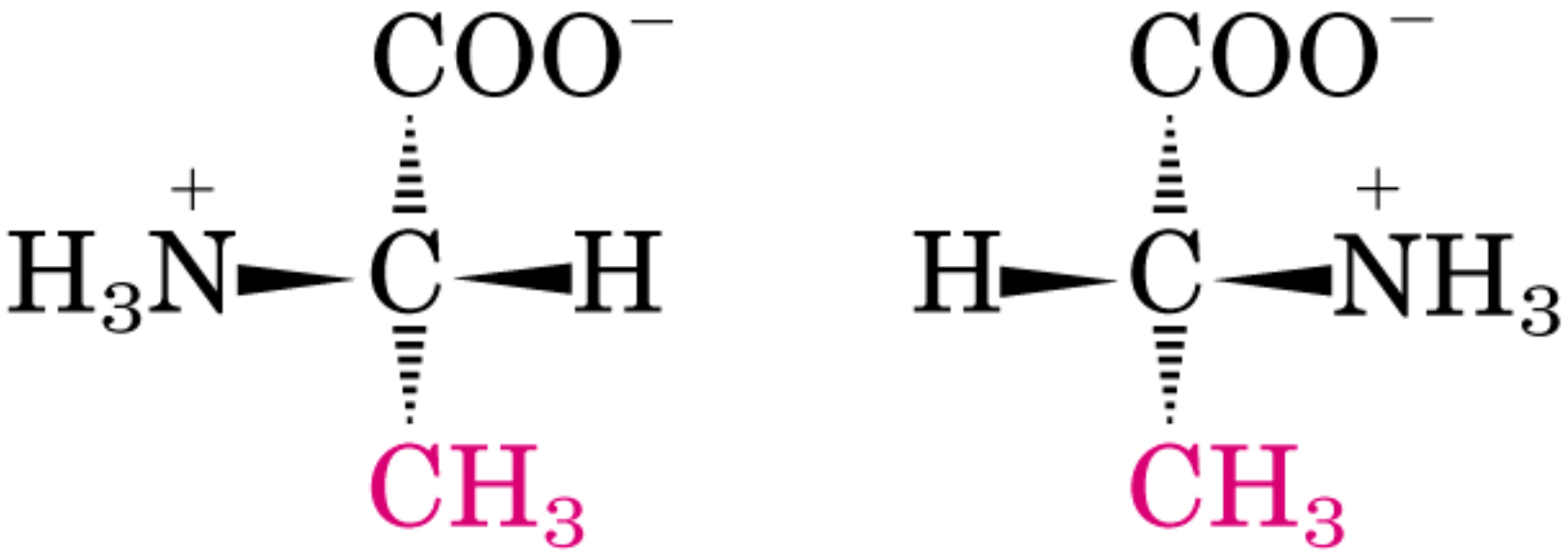
Lysine


COO⁻
H₃N⁺
C
H
CH₃
L-Alanine
COO⁻
H
C
NH₃⁺
CH₃
D-Alanine
(a)
COO⁻
H₃N⁺
C
H
CH₃
L-Alanine
COO⁻
H
C
NH₃⁺
CH₃
D-Alanine
(b)
COO⁻
H₃N⁺
C
H
CH₃
L-Alanine
COO⁻
H
C
NH₃⁺
CH₃
D-Alanine
(c)



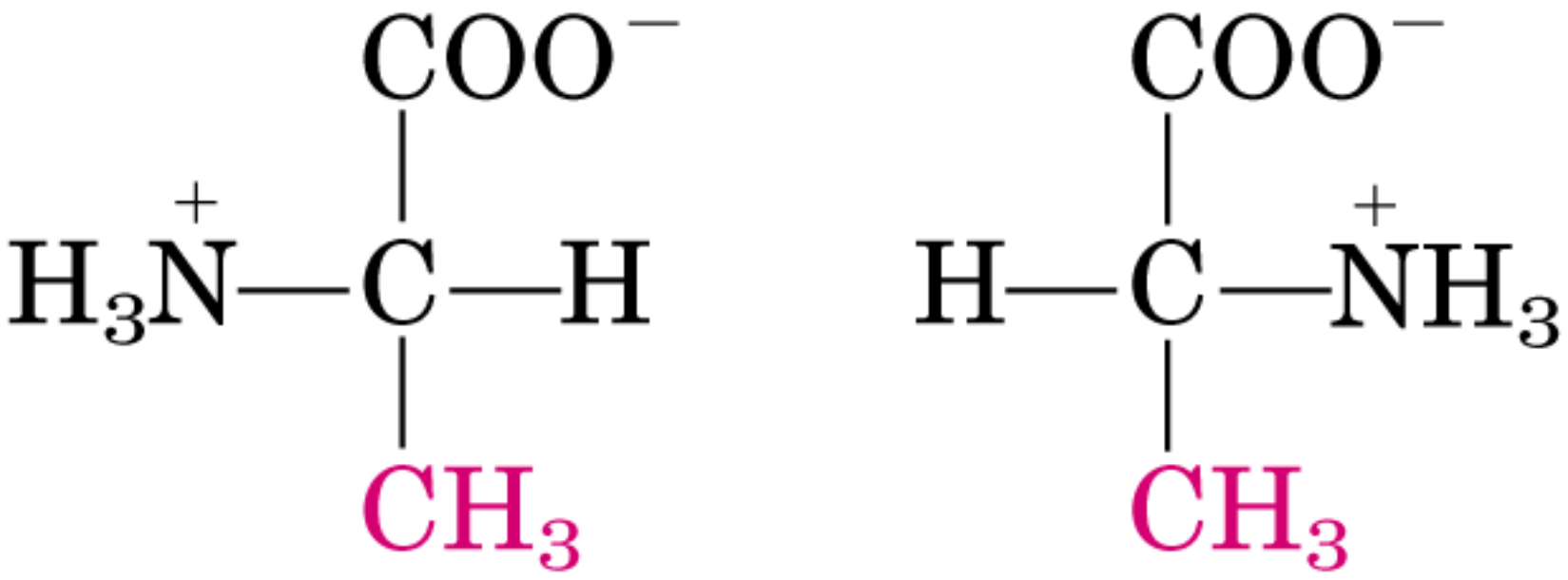
L-Alanine D-Alanine

(a)



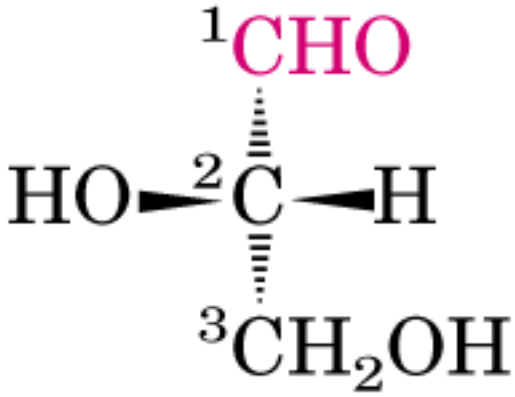
L-Alanine D-Alanine

(b)

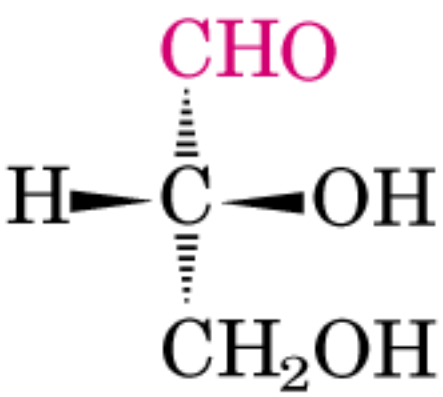


L-Alanine D-Alanine

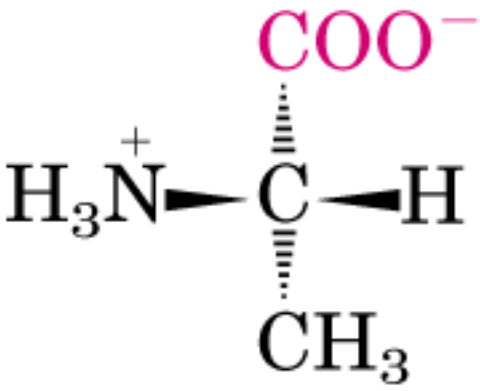
(c)



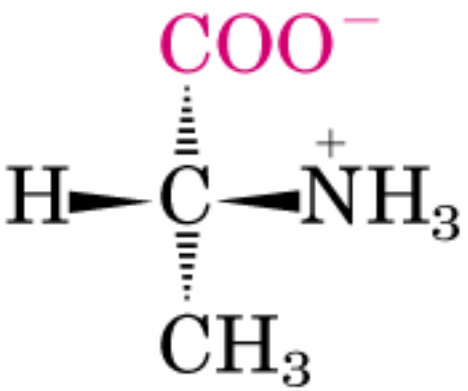
L-Glyceraldehyde



D-Glyceraldehyde



L-Alanine



D-Alanine

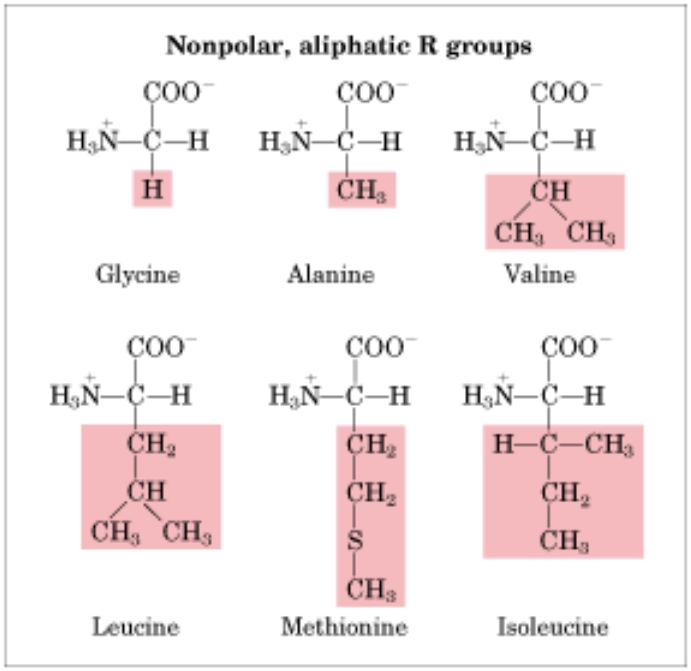
table 5–1

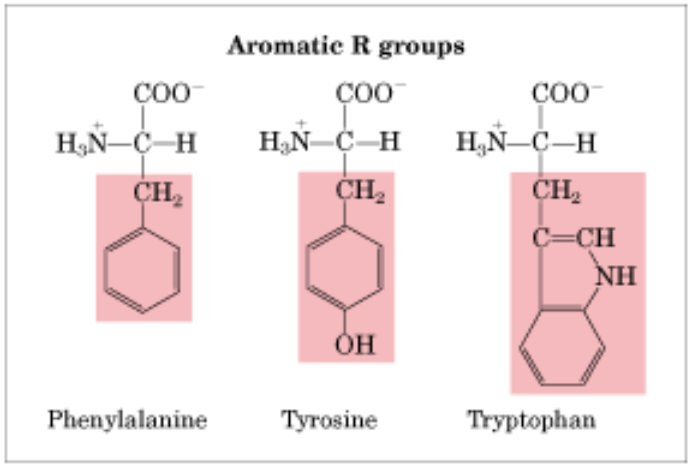
Properties and Conventions Associated with the Standard Amino Acids

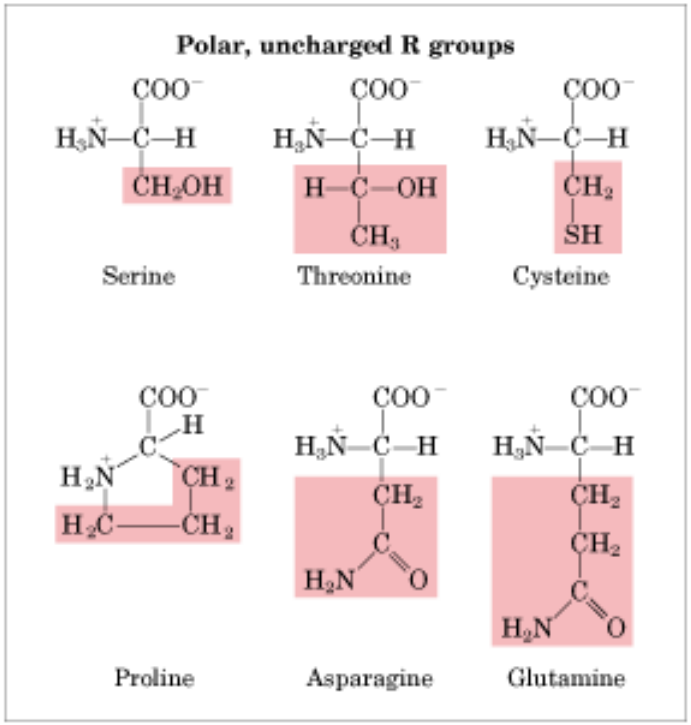
Amino acid	Abbreviated names		M_r	pK_a values: pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)†
Nonpolar, aliphatic R groups									
Glycine	Gly	G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala	A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val	V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu	L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile	I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met	M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups									
Phenylalanine	Phe	F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr	Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp	W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups									
Serine	Ser	S	105	2.21	9.15		5.68	−0.8	6.8
Proline	Pro	P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr	T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys	C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn	N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln	Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups									
Lysine	Lys	K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His	H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg	R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups									
Aspartate	Asp	D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu	E	147	2.19	9.67	4.25	3.22	−3.5	6.3

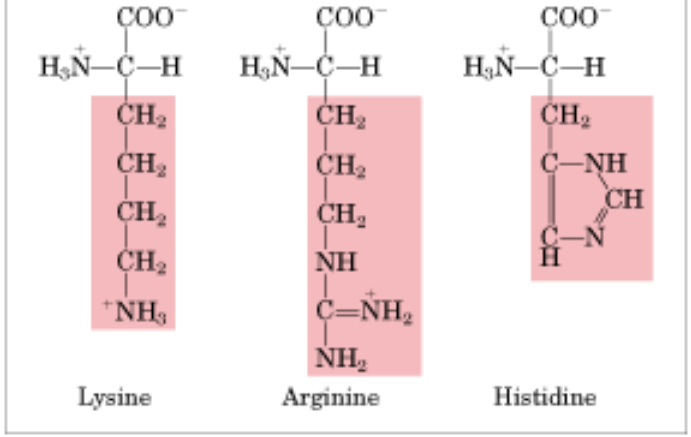
*A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (− values) or a hydrophobic environment (+ values). See Chapter 12. From Kyte, J. & Doolittle, R.F. (1982) *J. Mol. Biol.* **157**, 105 – 132.

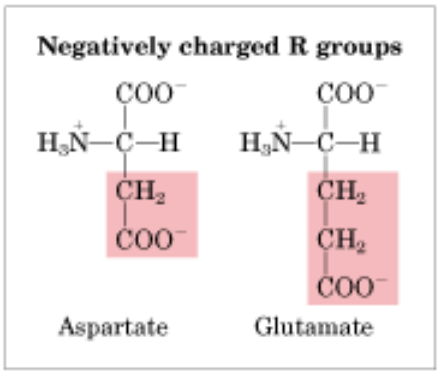
†Average occurrence in over 1150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed) Plenum Press, NY, pp. 599–623.


Nonpolar, aliphatic R groups
Glycine
Alanine
Valine
Leucine
Methionine
Isoleucine

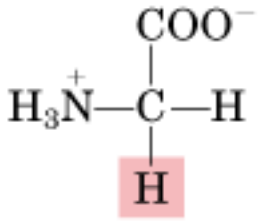

Aromatic R groups
Phenylalanine
Tyrosine
Tryptophan


Polar, uncharged R groups
Serine
Threonine
Cysteine
Proline
Asparagine
Glutamine

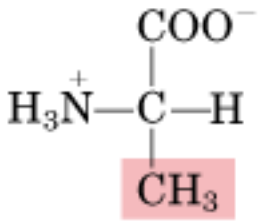

Positively charged R groups
Lysine
Arginine
Histidine


Negatively charged R groups
Aspartate
Glutamate

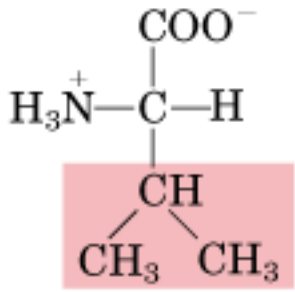
Nonpolar, aliphatic R groups



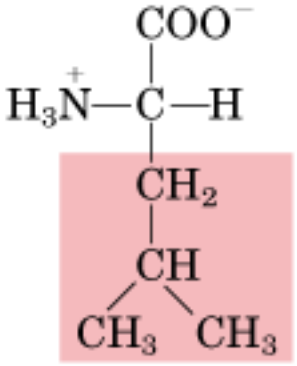
Glycine



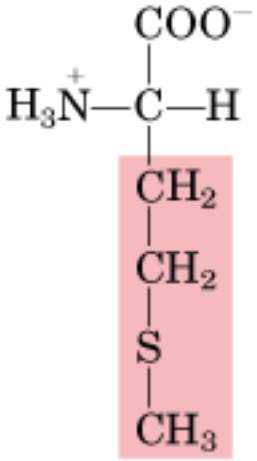
Alanine



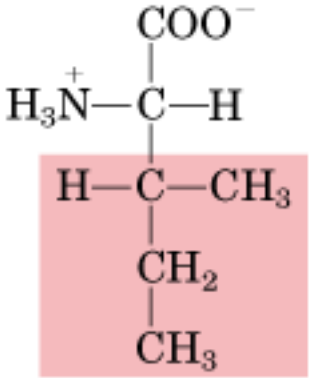
Valine



Leucine

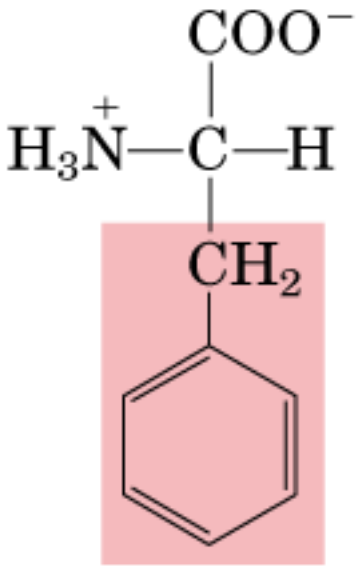


Methionine

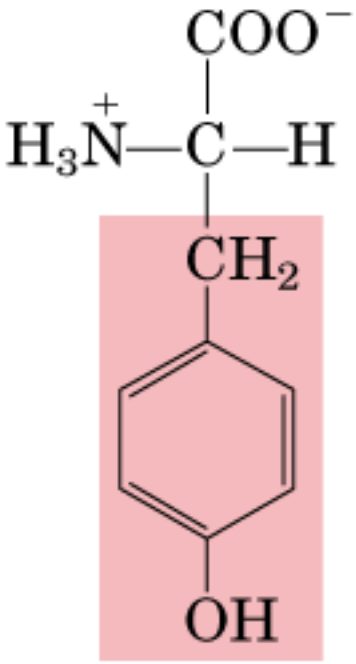


Isoleucine

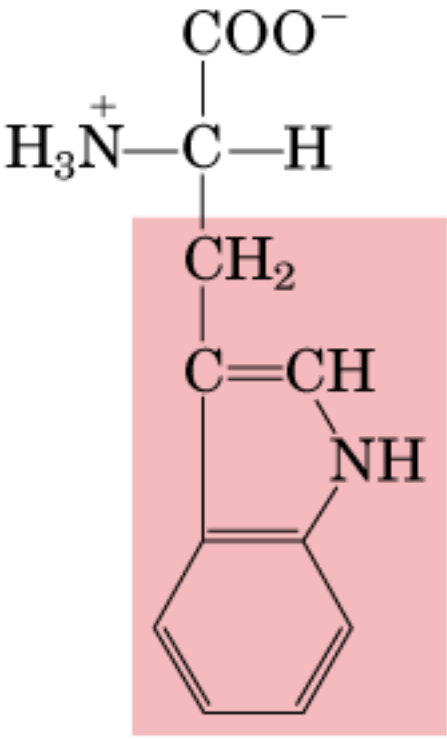
Aromatic R groups



Phenylalanine

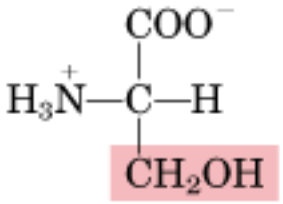


Tyrosine

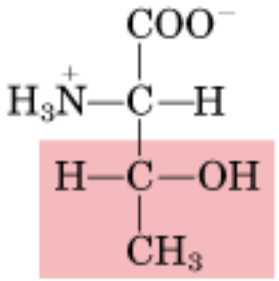


Tryptophan

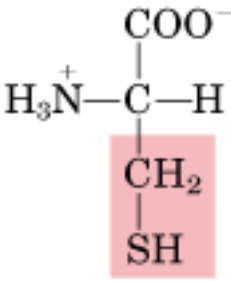
Polar, uncharged R groups



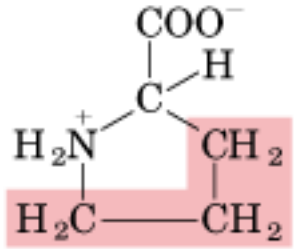
Serine



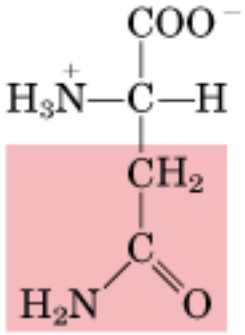
Threonine



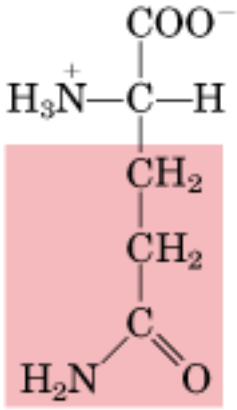
Cysteine



Proline

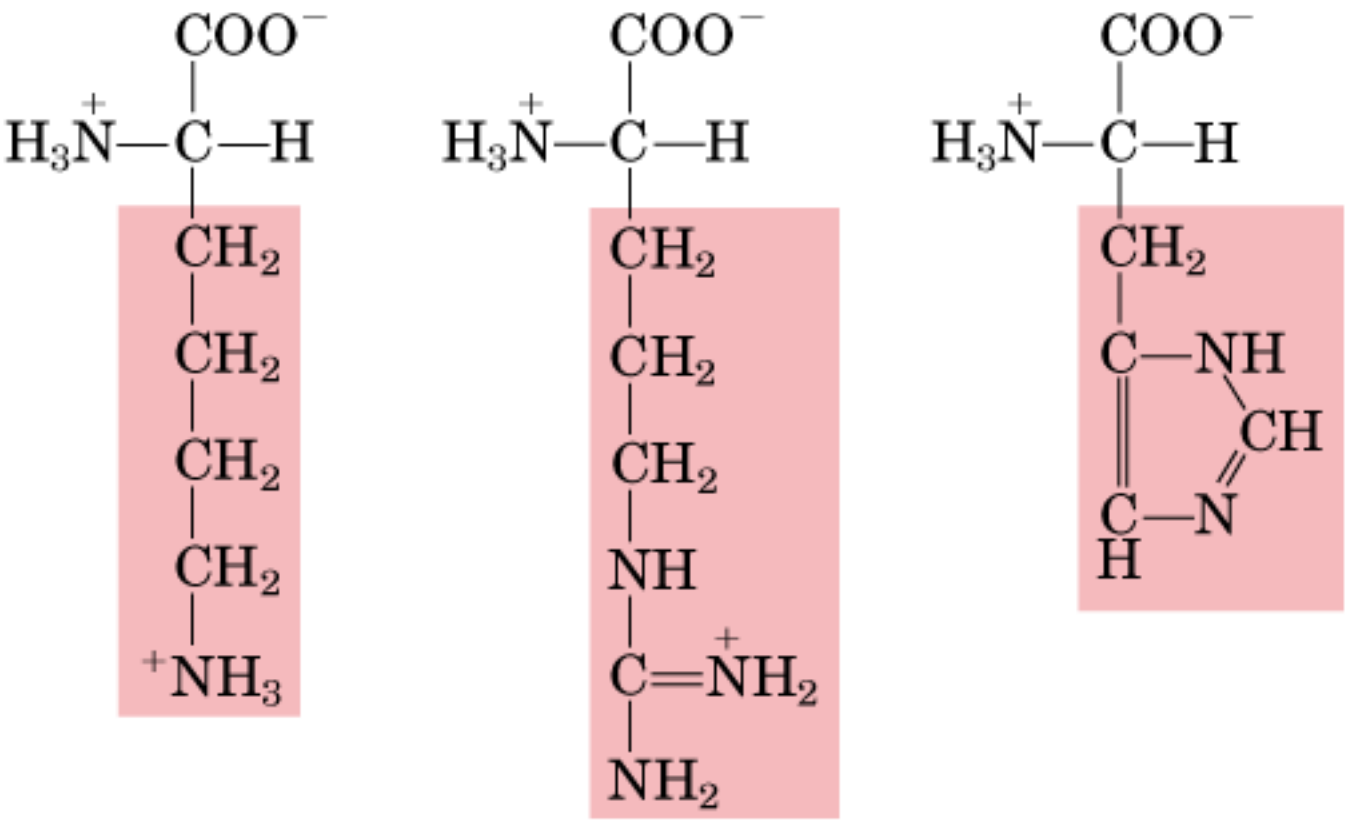


Asparagine

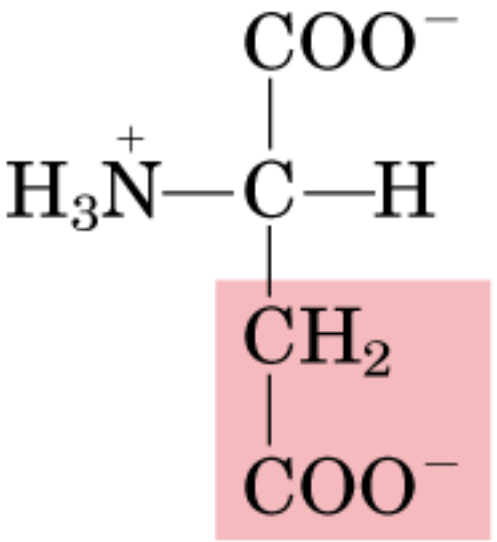


Glutamine

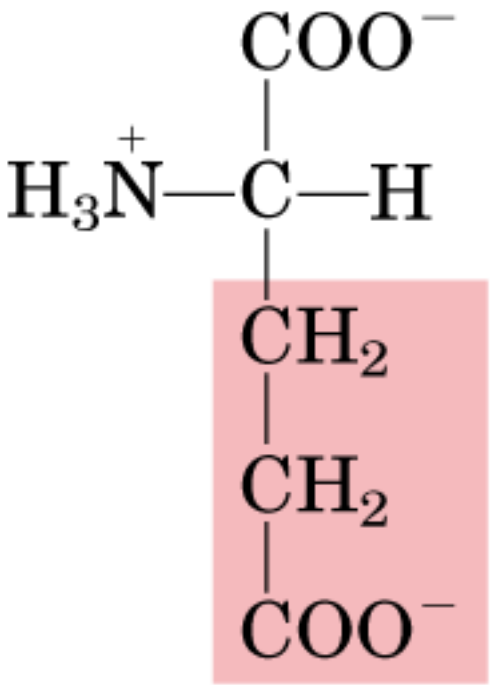
Positively charged R groups



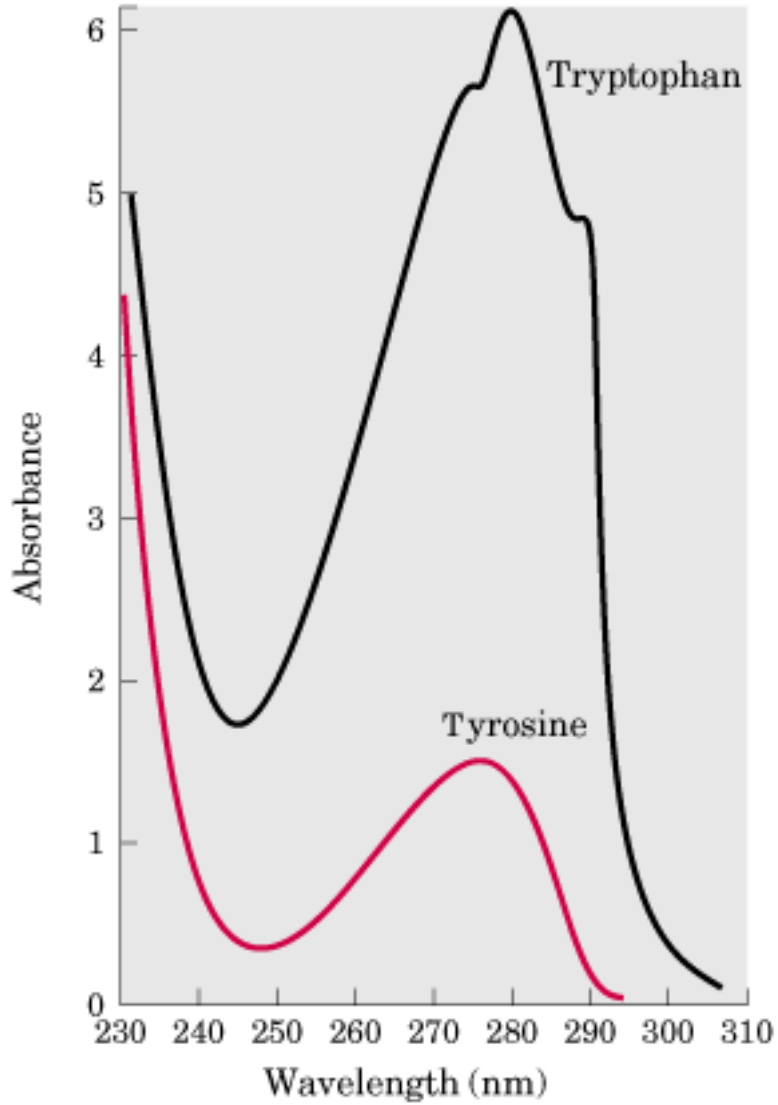
Negatively charged R groups

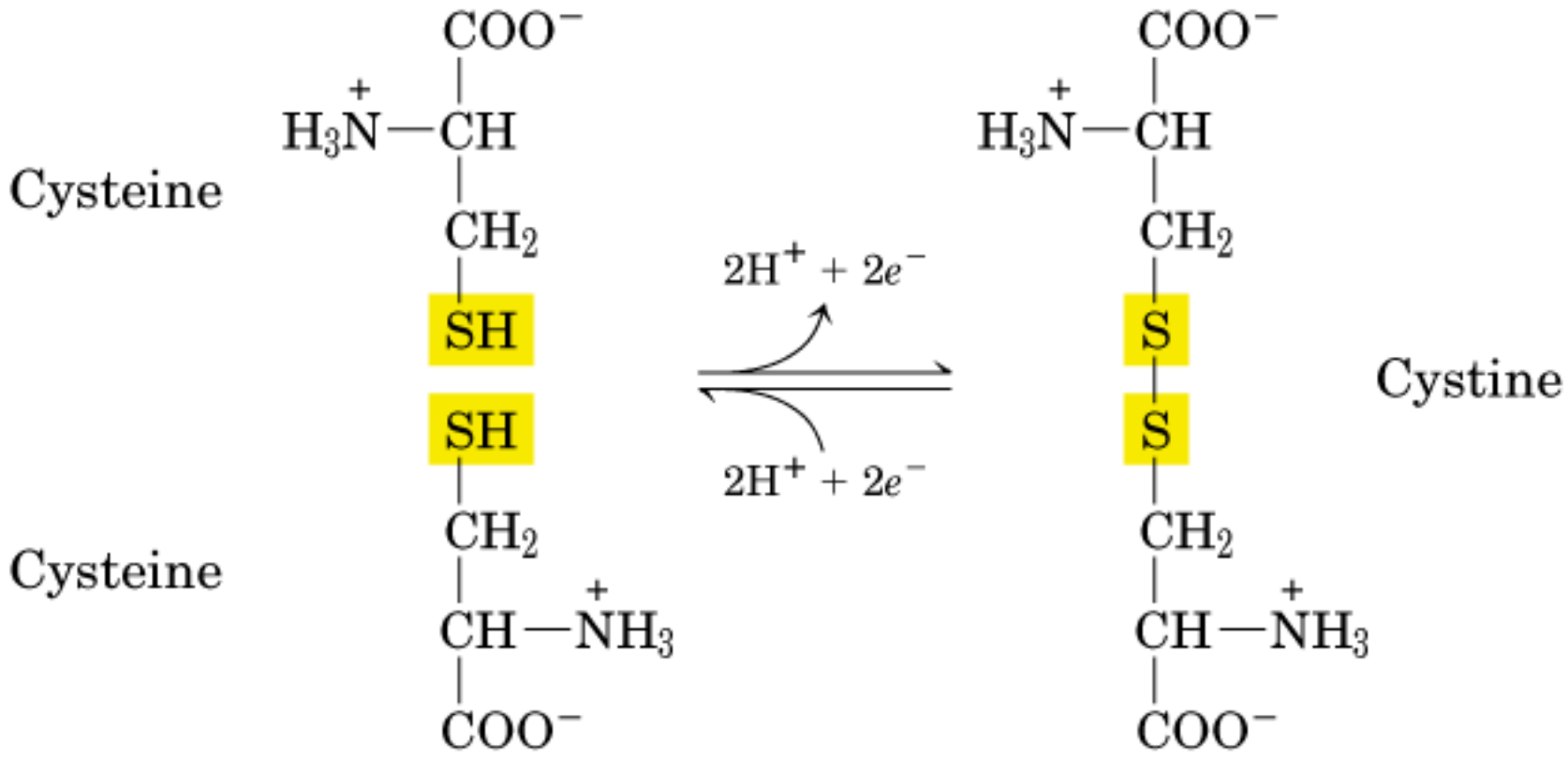


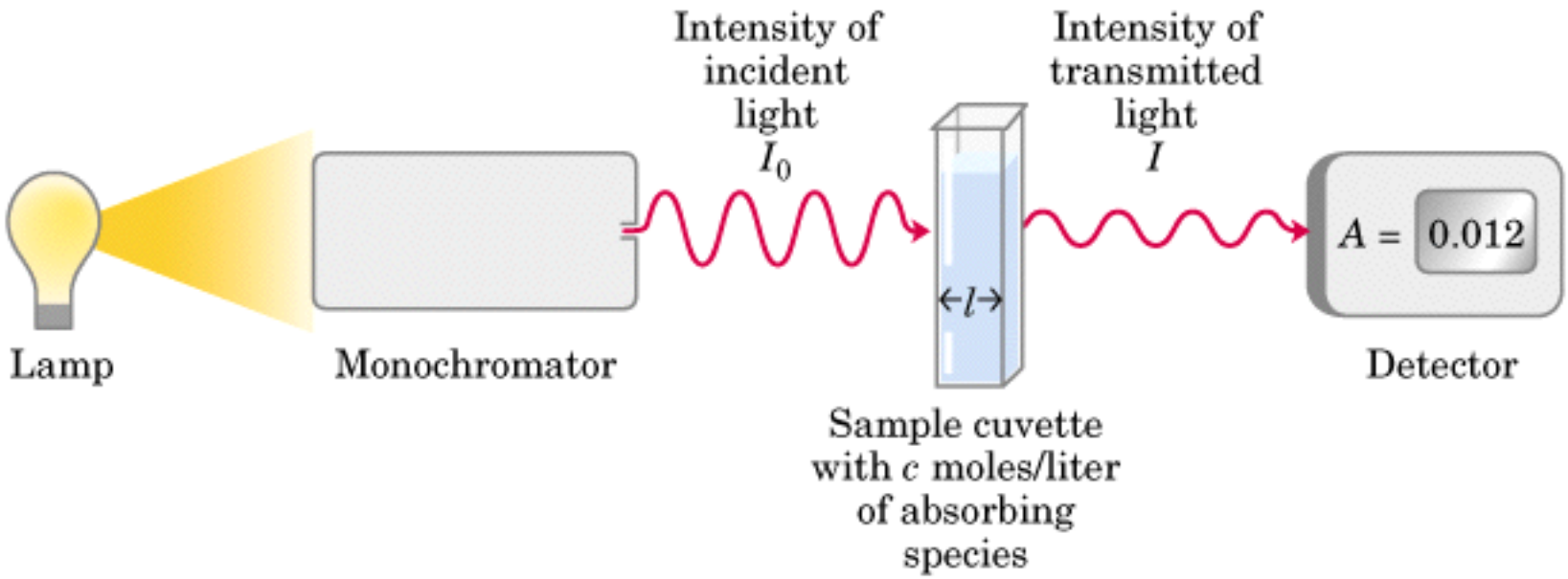
Aspartate

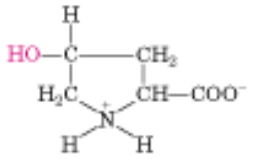


Glutamate

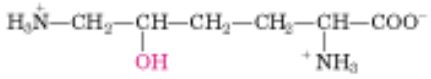

Tryptophan
Tyrosine
Absorbance
0
1
2
3
4
5
6
230
240
250
260
270
280
290
300
310
Wavelength (nm)


Cysteine
Cysteine
Cystine
COO⁻
H₃N⁺—CH
CH₂
SH
SH
CH₂
CH—NH₃⁺
COO⁻
2H⁺ + 2e⁻
2H⁺ + 2e⁻
COO⁻
H₃N⁺—CH
CH₂
S
S
CH₂
CH—NH₃⁺
COO⁻

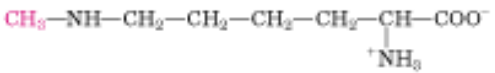

Intensity of incident light I_0
Intensity of transmitted light I
A = 0.012
$\leftarrow l \rightarrow$
Lamp
Monochromator
Sample cuvette with c moles/liter of absorbing species
Detector



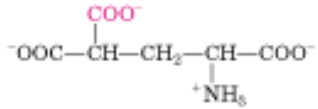
4-Hydroxyproline



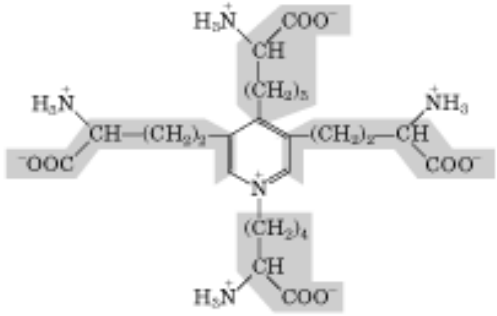
5-Hydroxylysine



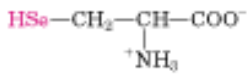
6-*N*-Methyllysine



γ-Carboxyglutamate



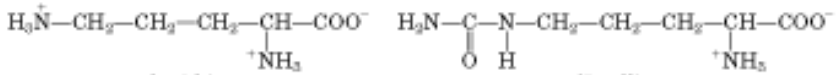
Desmosine



Selenocysteine

Nonstandard amino acids not found in proteins

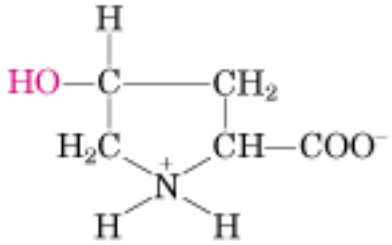
(a)



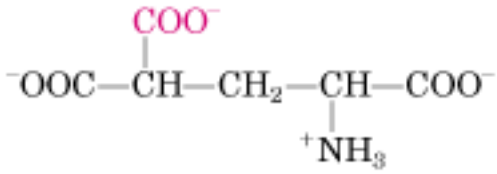
Ornithine Citrulline

Nonstandard amino acids not found in proteins

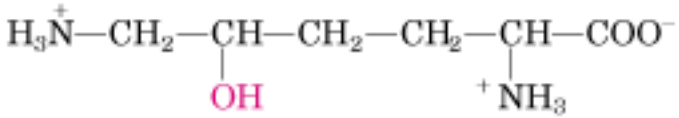
(b)



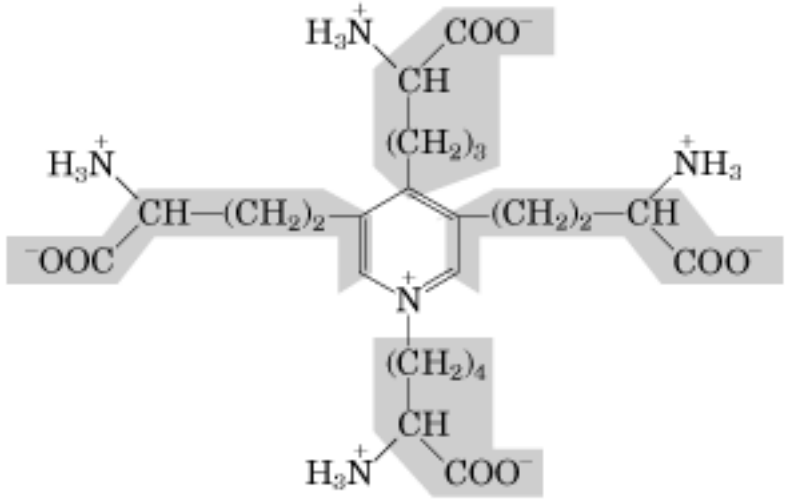
4-Hydroxyproline



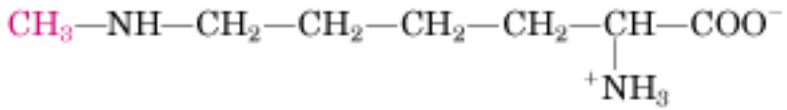
γ-Carboxyglutamate



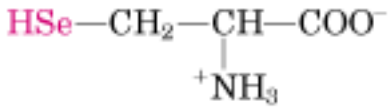
5-Hydroxylysine



Desmosine

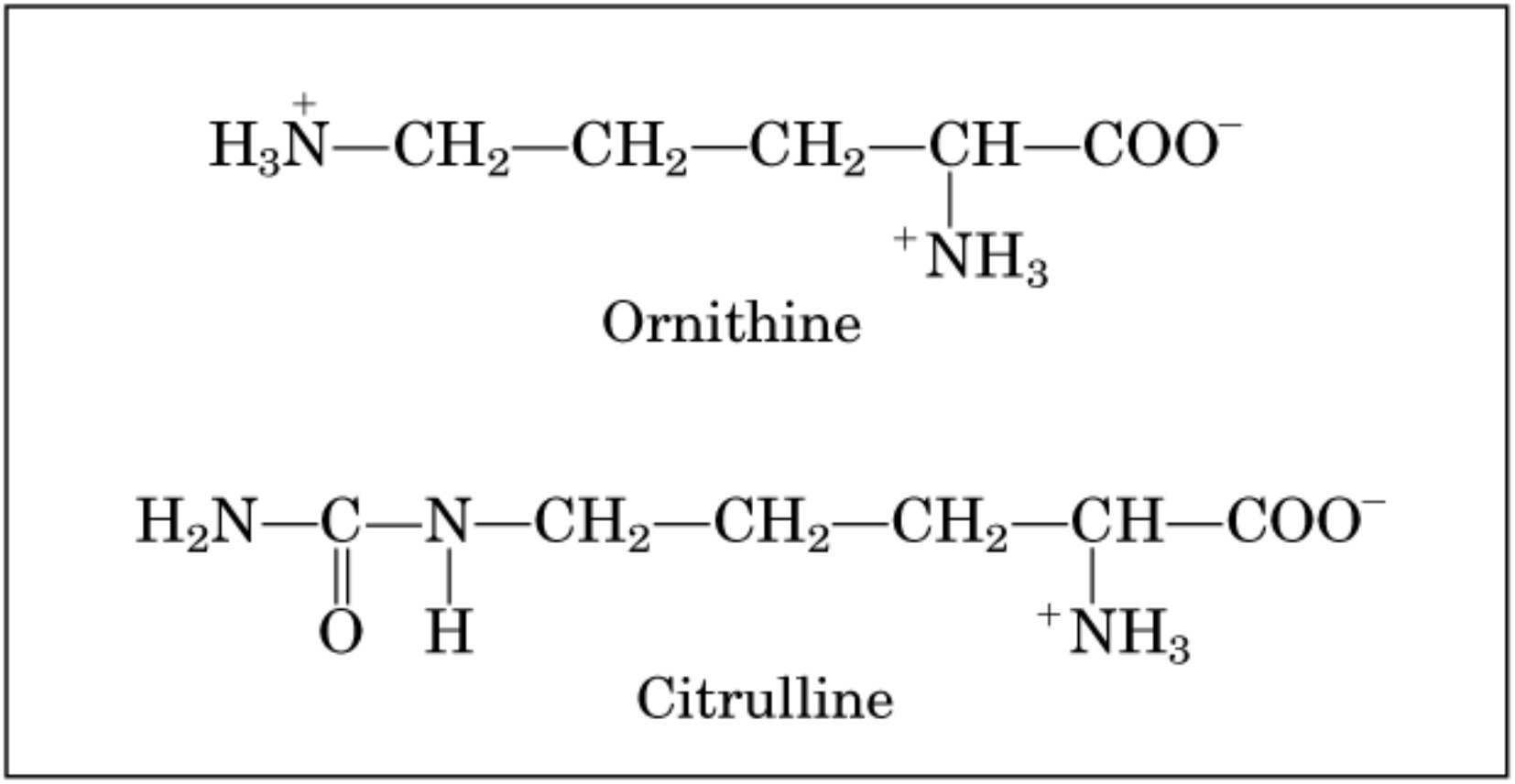


6-*N*-Methyllysine

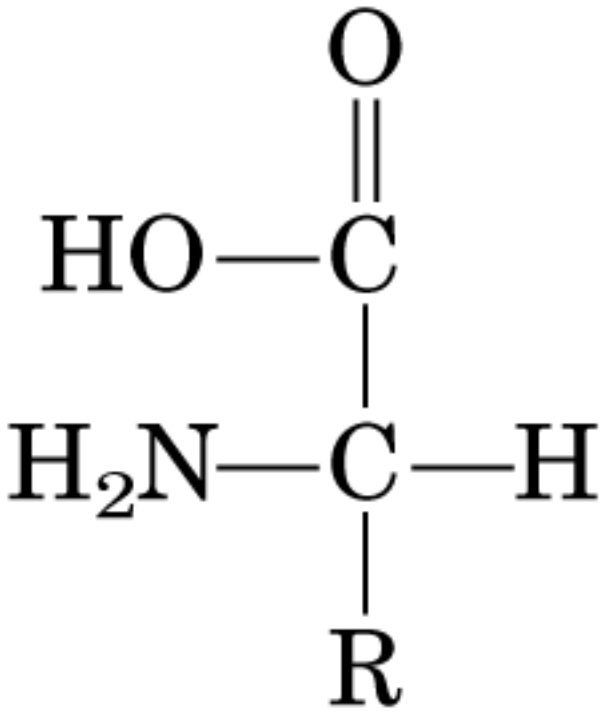


Selenocysteine

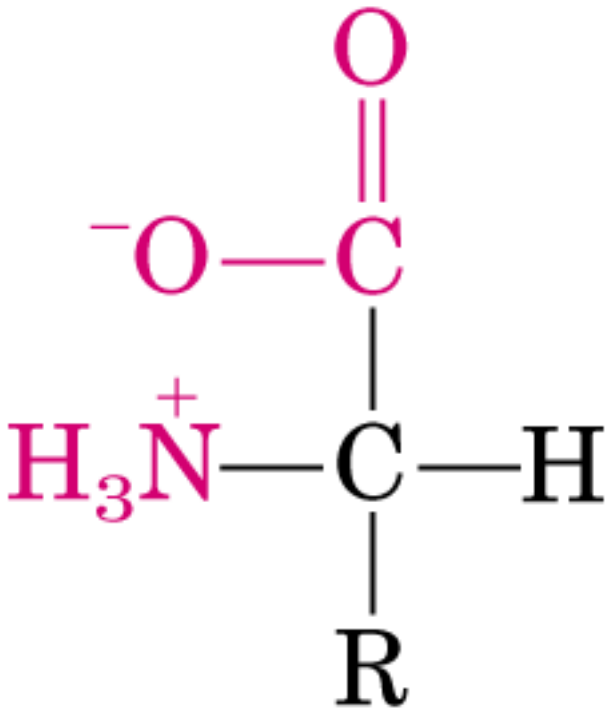
(a)


H3N+—CH2—CH2—CH2—CH—COO−
+NH3
Ornithine
H2N—C—N—CH2—CH2—CH2—CH—COO−
O
H
+NH3
Citrulline

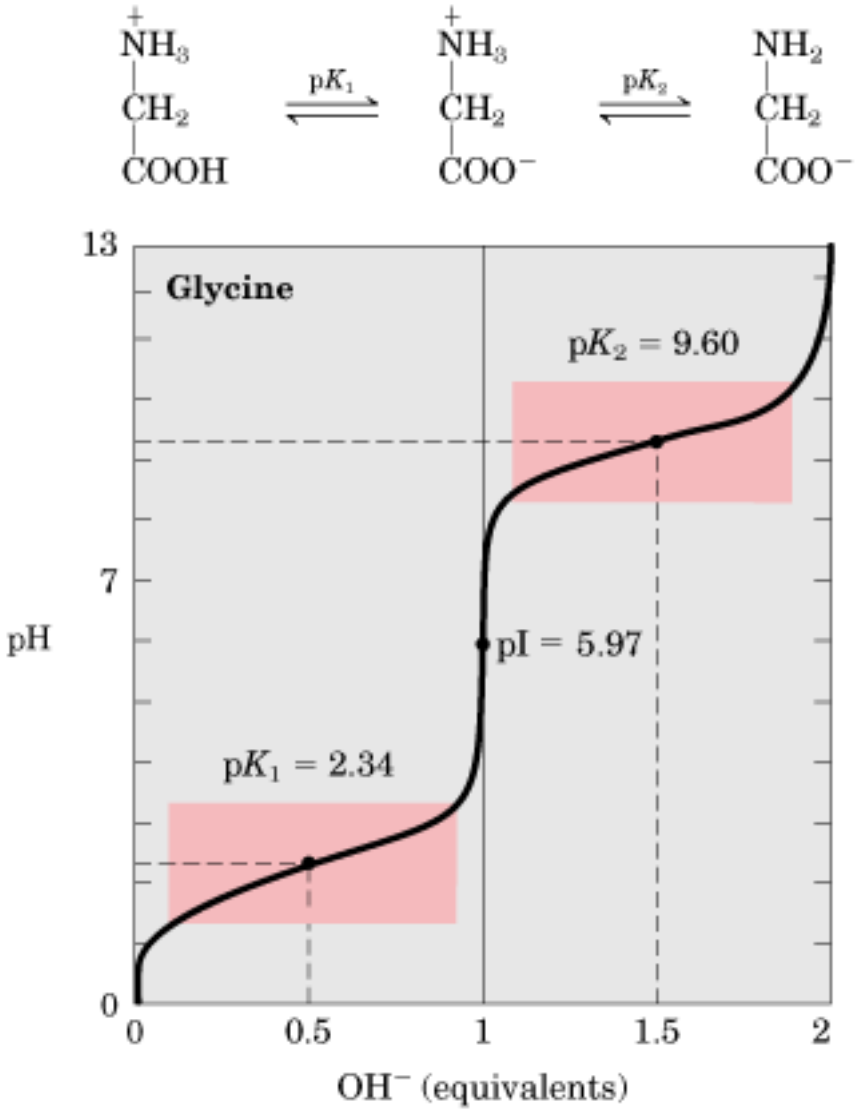
(b)

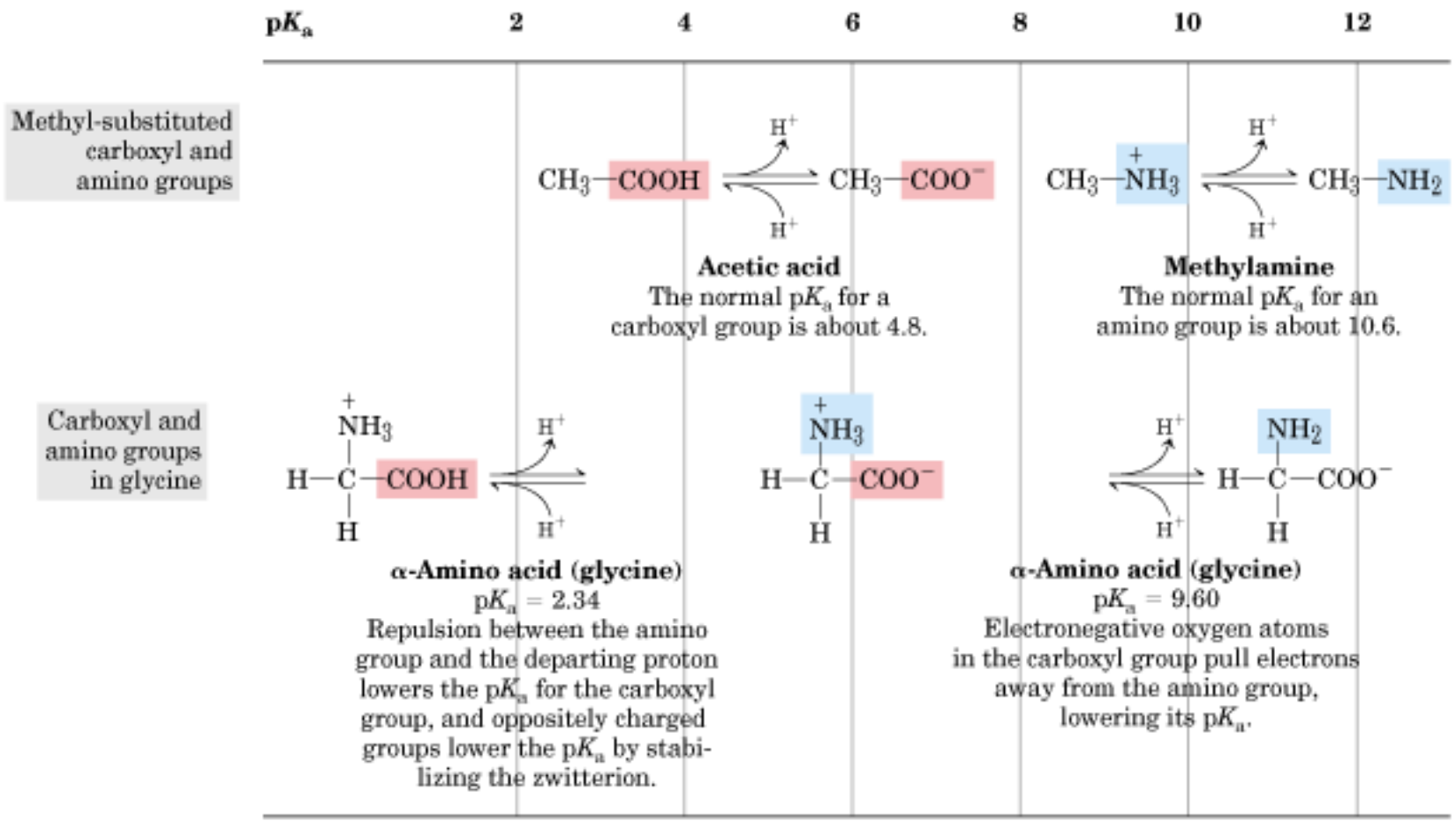


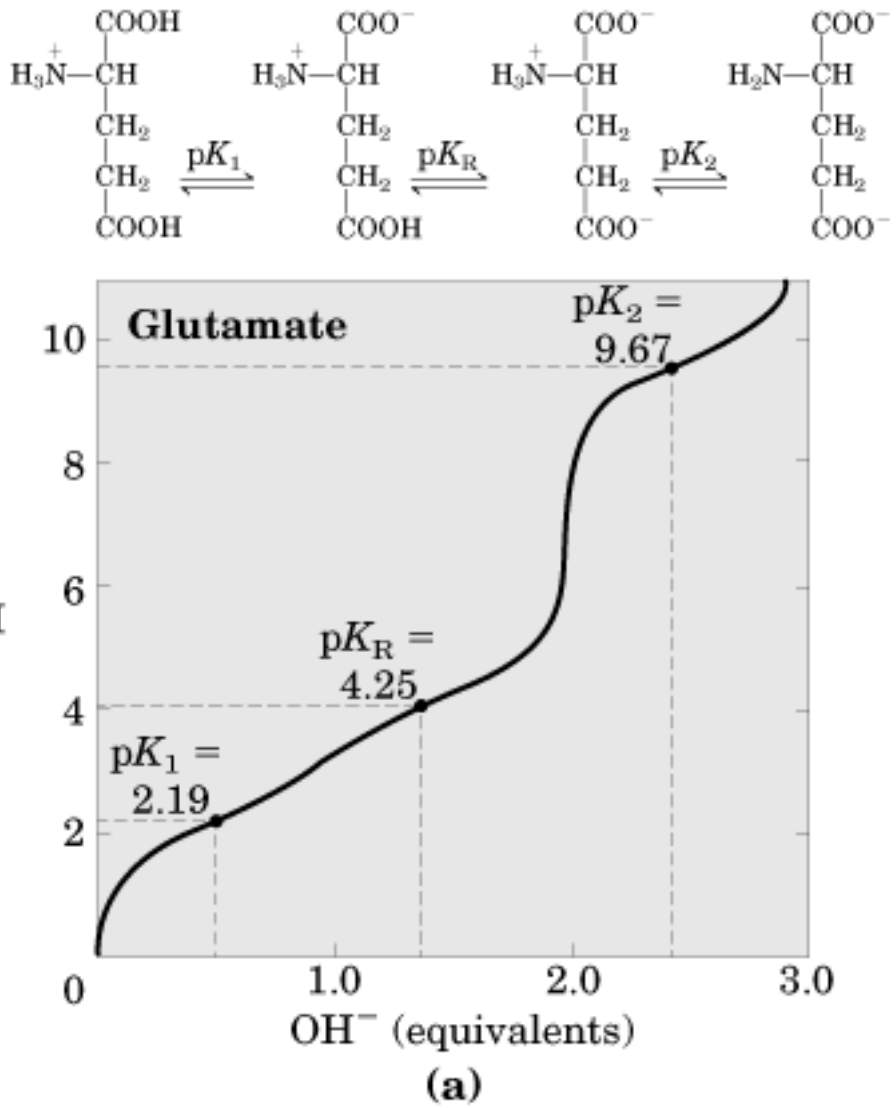
Nonionic form

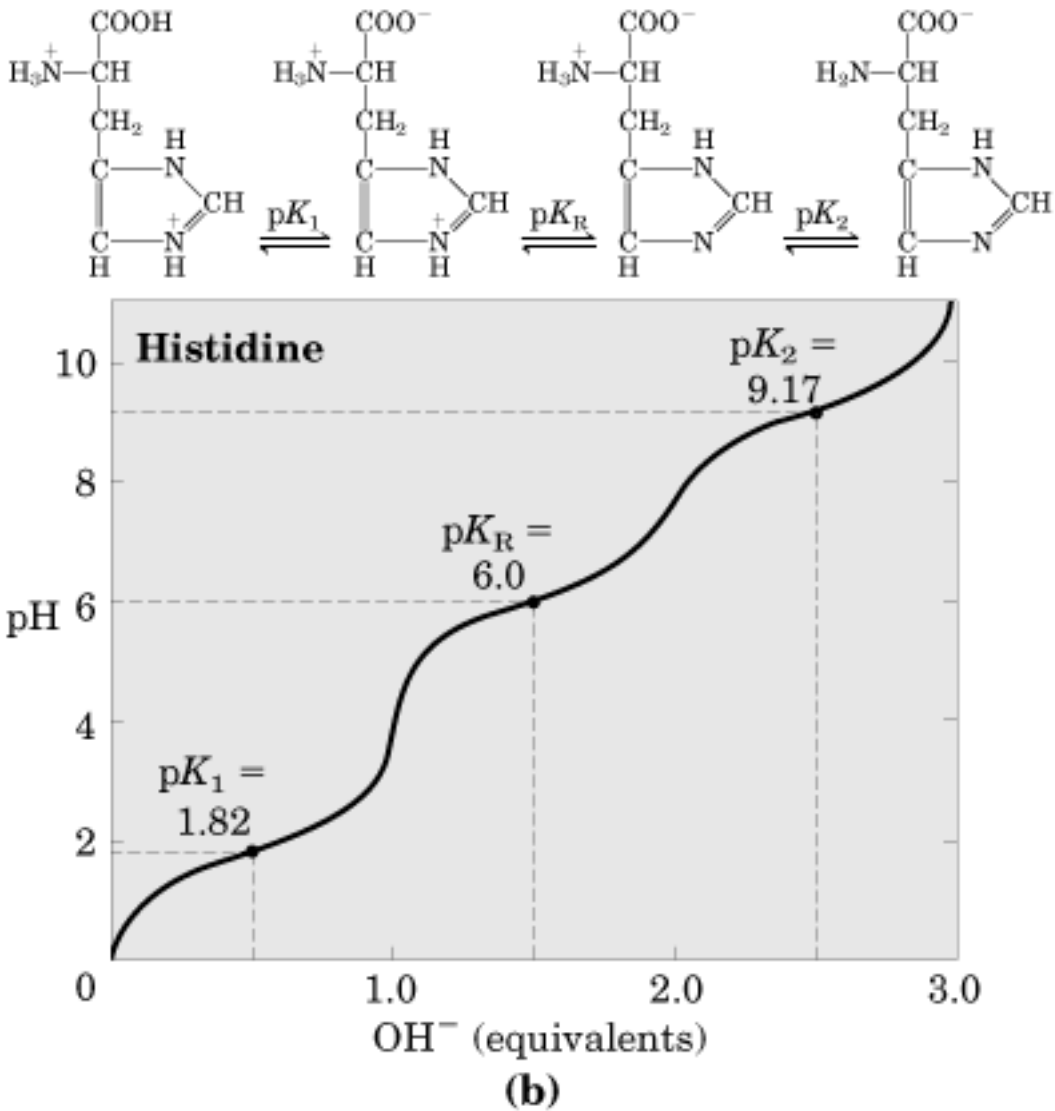


Zwitterionic form

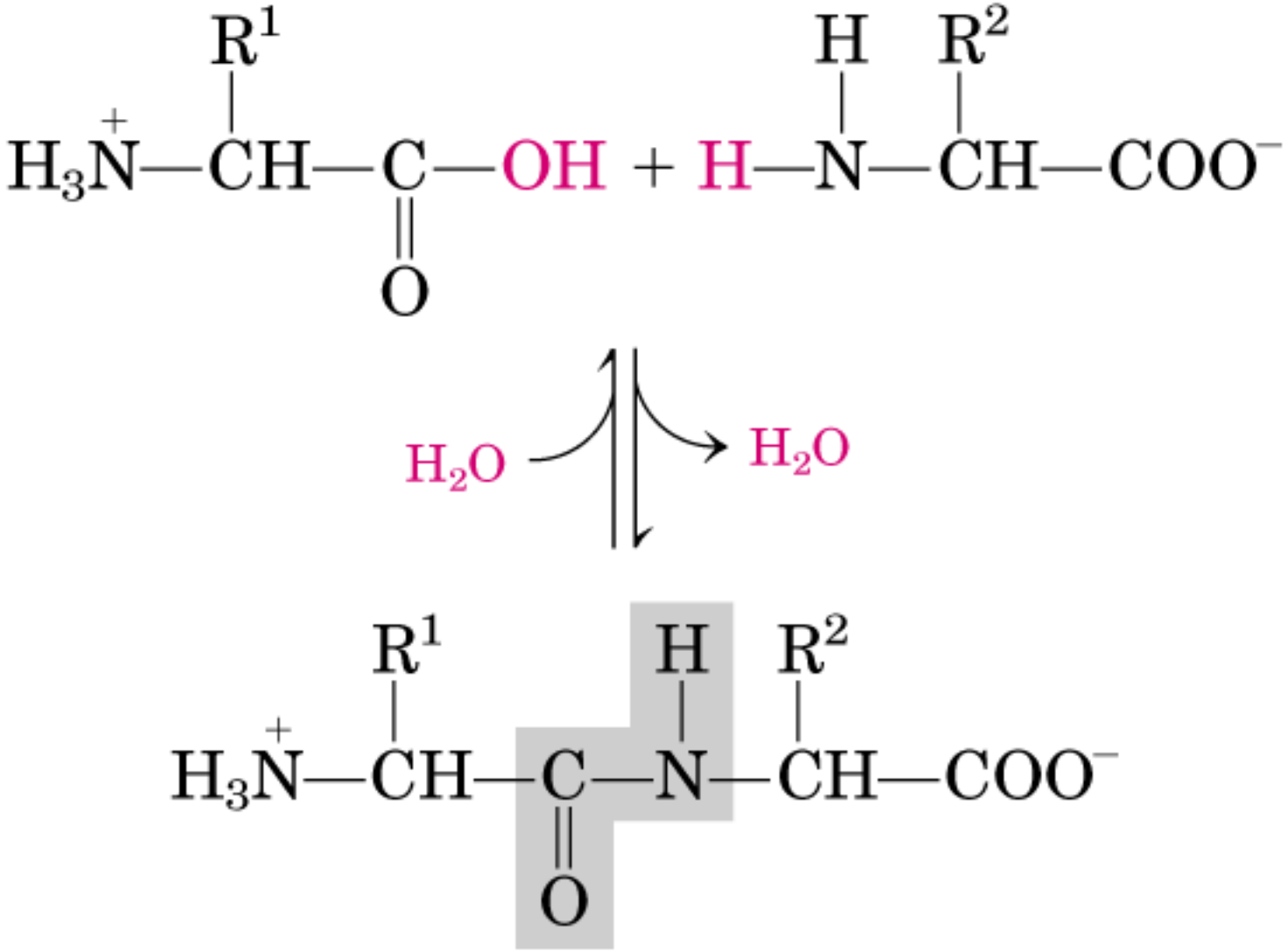

NH3+
CH2
COOH
pK1
NH3+
CH2
COO−
pK2
NH2
CH2
COO−
Glycine
pK2 = 9.60
pI = 5.97
pK1 = 2.34
pH
13
7
0
0
0.5
1
1.5
2
OH− (equivalents)

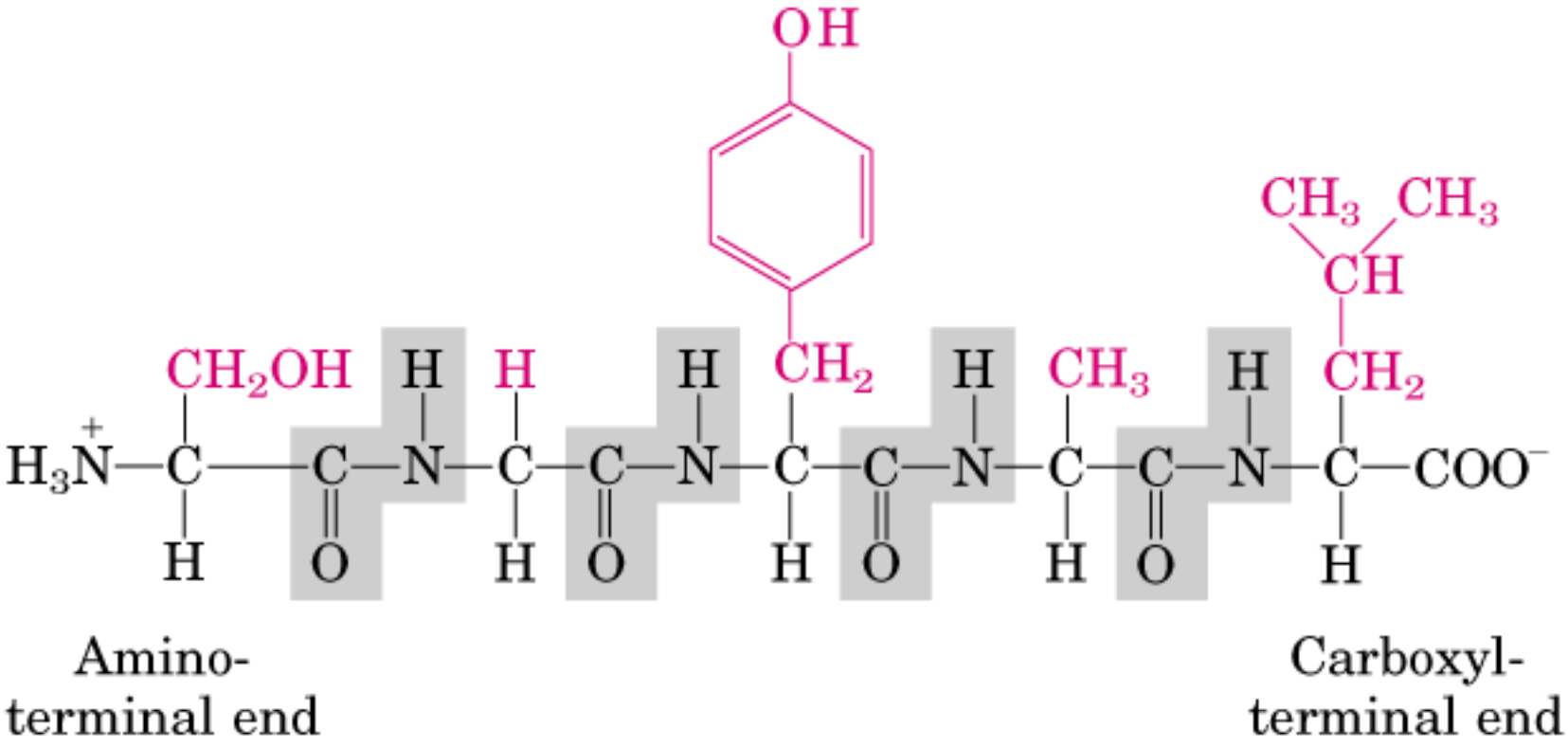

pKa
2
4
6
8
10
12
Methyl-substituted carboxyl and amino groups
CH3—COOH
CH3—COO⁻
H+
H+
Acetic acid
The normal pKa for a carboxyl group is about 4.8.
CH3—NH3+
CH3—NH2
Methylamine
The normal pKa for an amino group is about 10.6.
Carboxyl and amino groups in glycine
α-Amino acid (glycine)
pKa = 2.34
Repulsion between the amino group and the departing proton lowers the pKa for the carboxyl group, and oppositely charged groups lower the pKa by stabilizing the zwitterion.
α-Amino acid (glycine)
pKa = 9.60
Electronegative oxygen atoms in the carboxyl group pull electrons away from the amino group, lowering its pKa.

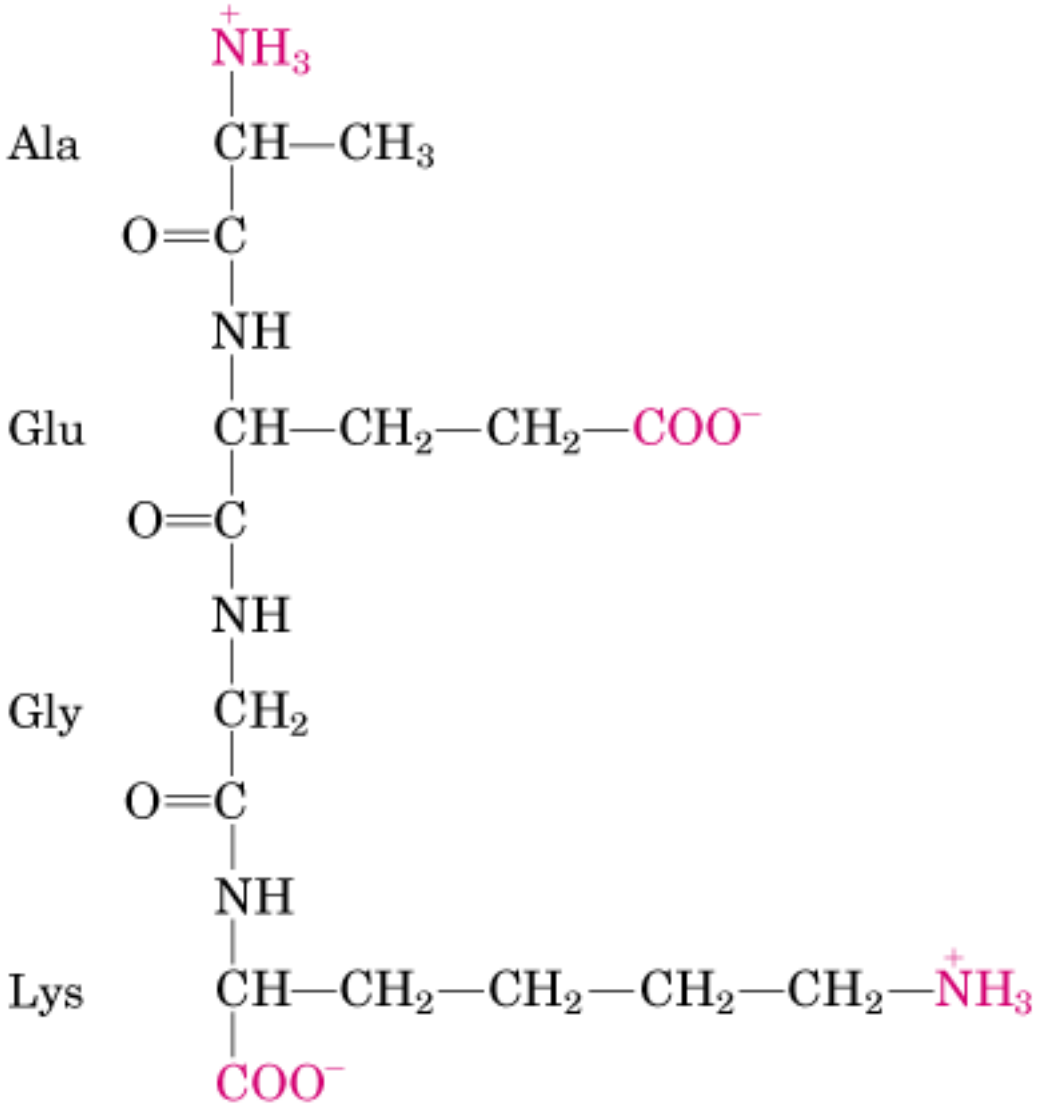

Glutamate
pK2 = 9.67
pKR = 4.25
pK1 = 2.19
pH
10
8
6
4
2
0
1.0
2.0
3.0
OH⁻ (equivalents)
(a)

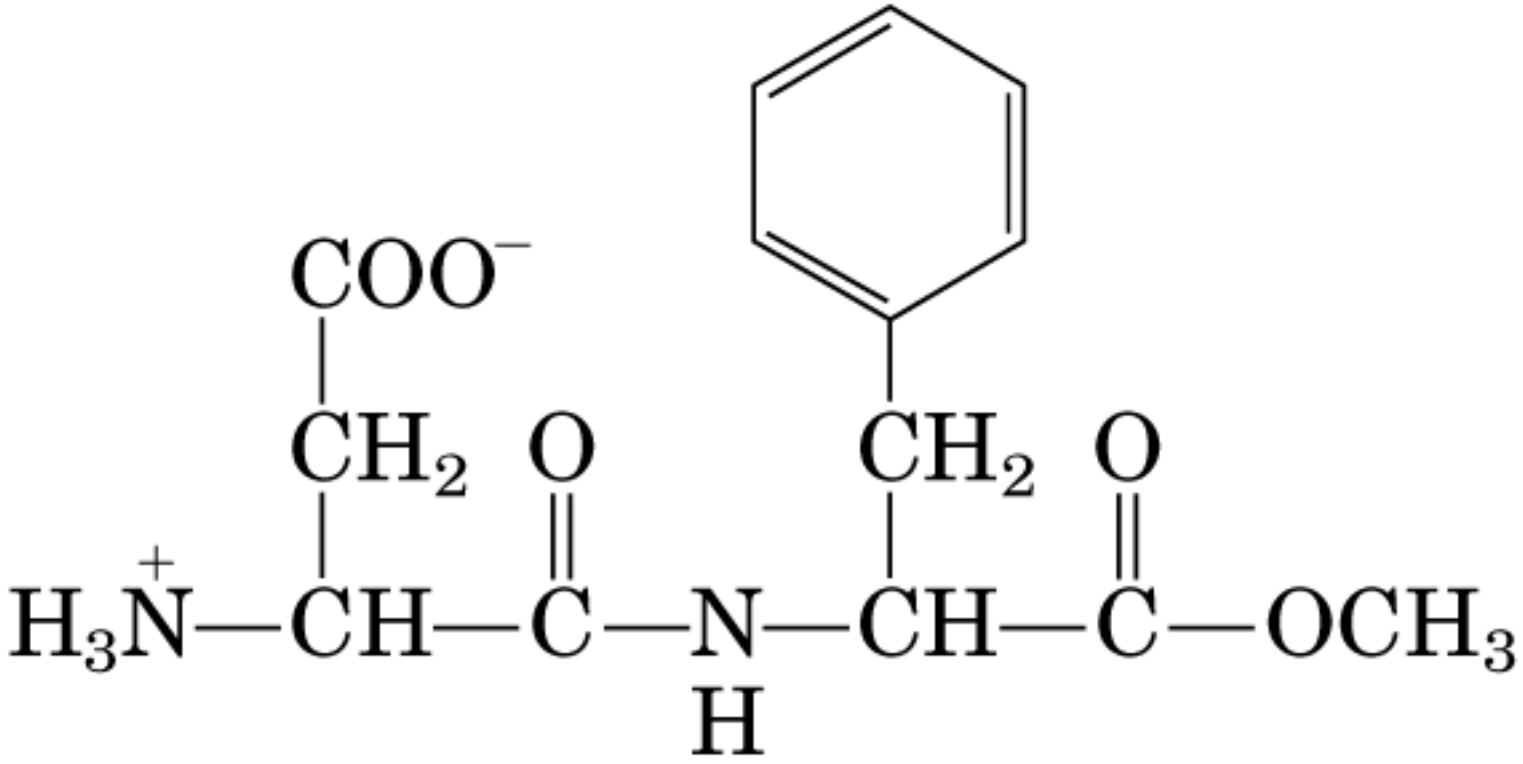

COOH
COO⁻
COO⁻
COO⁻
pK1
pKR
pK2
Histidine
pK2 = 9.17
pKR = 6.0
pK1 = 1.82
pH
10
8
6
4
2
0
1.0
2.0
3.0
OH⁻ (equivalents)

(b)


R1
H3N+—CH—C—OH + H—N—CH—COO−
H
R2
O
H2O
H2O
R1
H
R2
H3N+—CH—C—N—CH—COO−
O


OH
CH3 CH3
CH
CH2OH
H
H
H
CH2
H
CH3
H
CH2
H3N+
C
C
N
C
C
N
C
C
N
C
C
N
C
COO−
H
O
H
O
H
O
H
O
H
O
H
Amino-
terminal end
Carboxyl-
terminal end


NH3
Ala
CH—CH3
O=C
NH
Glu
CH—CH2—CH2—COO−
O=C
NH
Gly
CH2
O=C
NH
Lys
CH—CH2—CH2—CH2—CH2—NH3
COO−



L-Aspartyl-L-phenylalanine methyl ester
(aspartame)

table 5–2

Molecular Data on Some Proteins

	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome *c* (human)	13,000	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (egg white)	13,930	129	1
Myoglobin (equine heart)	16,890	153	1
Chymotrypsin (bovine pancreas)	21,600	241	3
Chymotrypsinogen (bovine)	22,000	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	68,500	609	1
Hexokinase (yeast)	102,000	972	2
RNA polymerase (*E. coli*)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (*E. coli*)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

table 5–3

Amino Acid Composition of Two Proteins*

Amino acid	Number of residues per molecule of protein	
	Bovine cytochrome *c*	Bovine chymotrypsinogen
Ala	6	22
Arg	2	4
Asn	5	15
Asp	3	8
Cys	2	10
Gln	3	10
Glu	9	5
Gly	14	23
His	3	2
Ile	6	10
Leu	6	19
Lys	18	14
Met	2	2
Phe	4	6
Pro	4	9
Ser	1	28
Thr	8	23
Trp	1	8
Tyr	4	4
Val	3	23
Total	104	245

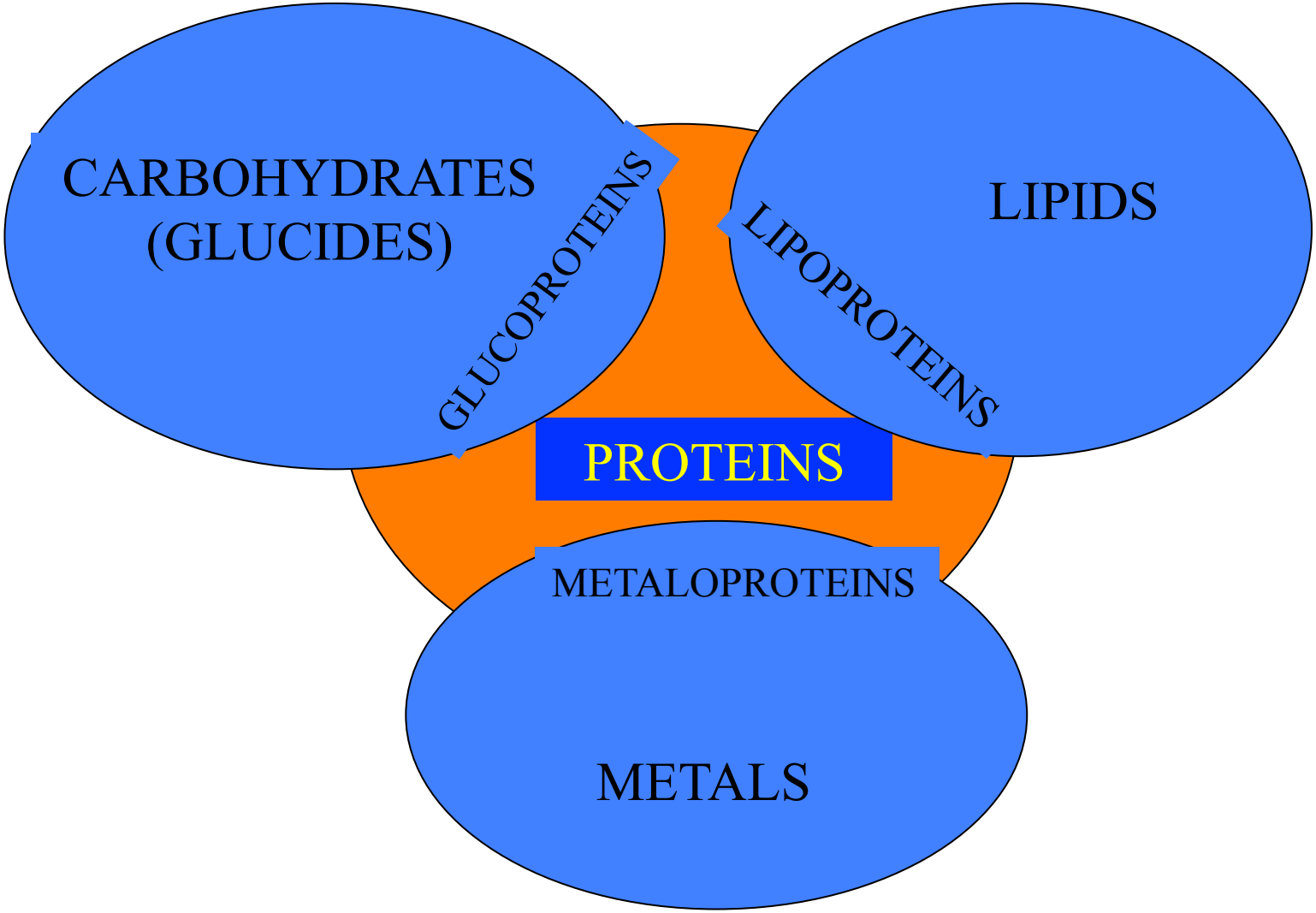
*Note that standard procedures for the acid hydrolysis of proteins convert Asn and Gln to Asp and Glu, respectively. In addition, Trp is destroyed. Special procedures must be employed to determine the amounts of these amino acids.

table 5–4

Conjugated Proteins

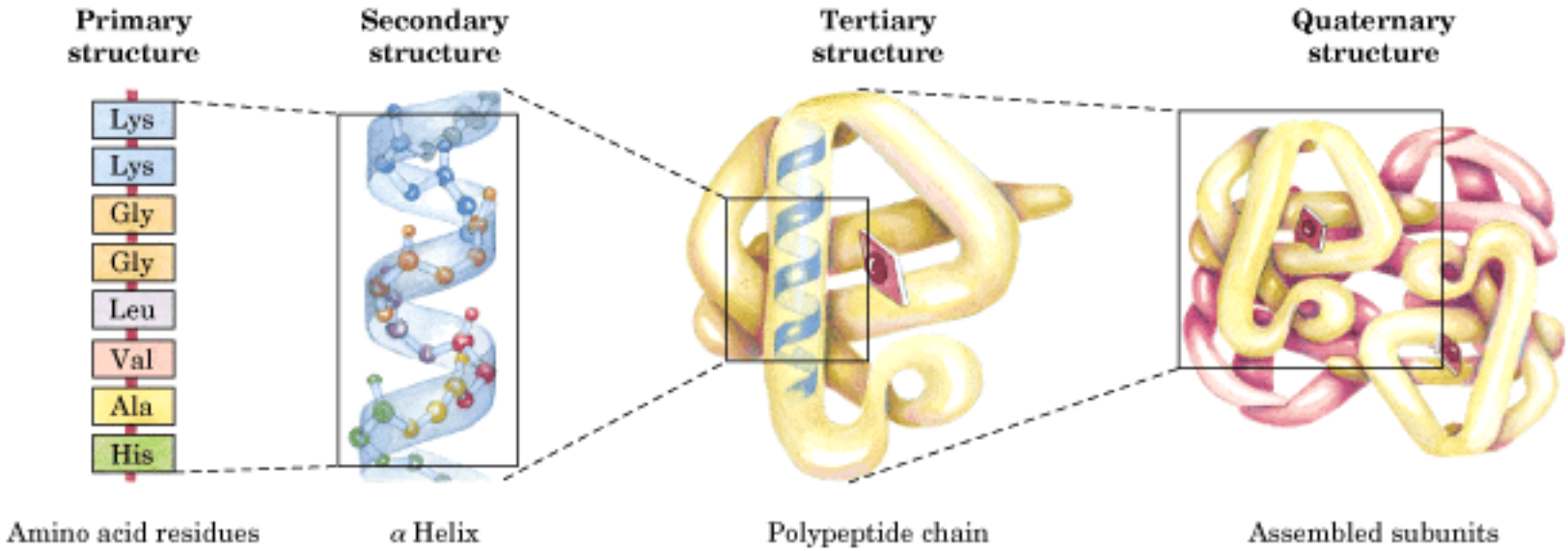
Class	Prosthetic group(s)	Example
Lipoproteins	Lipids	β_1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

CONJUGATED PROTEINS

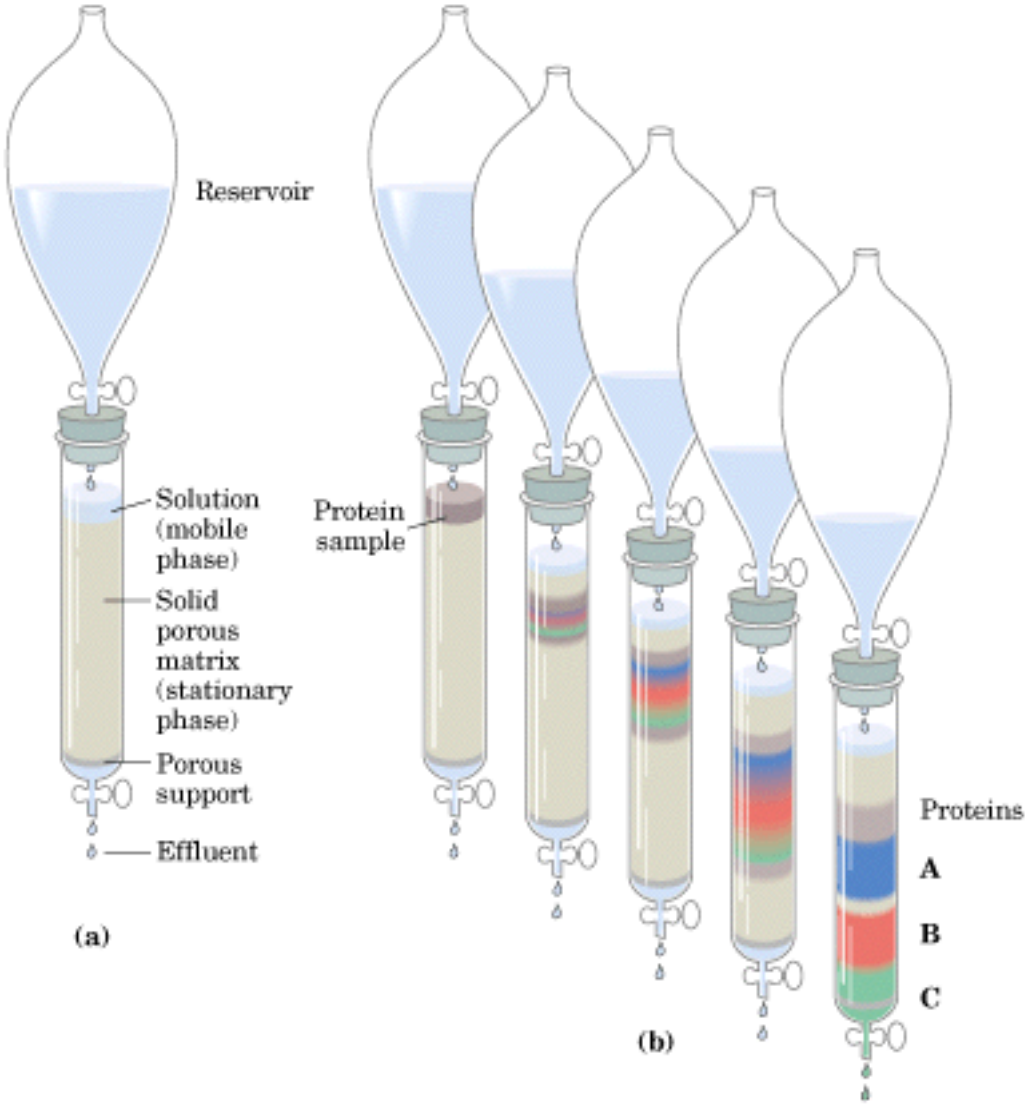


conjugated proteins:

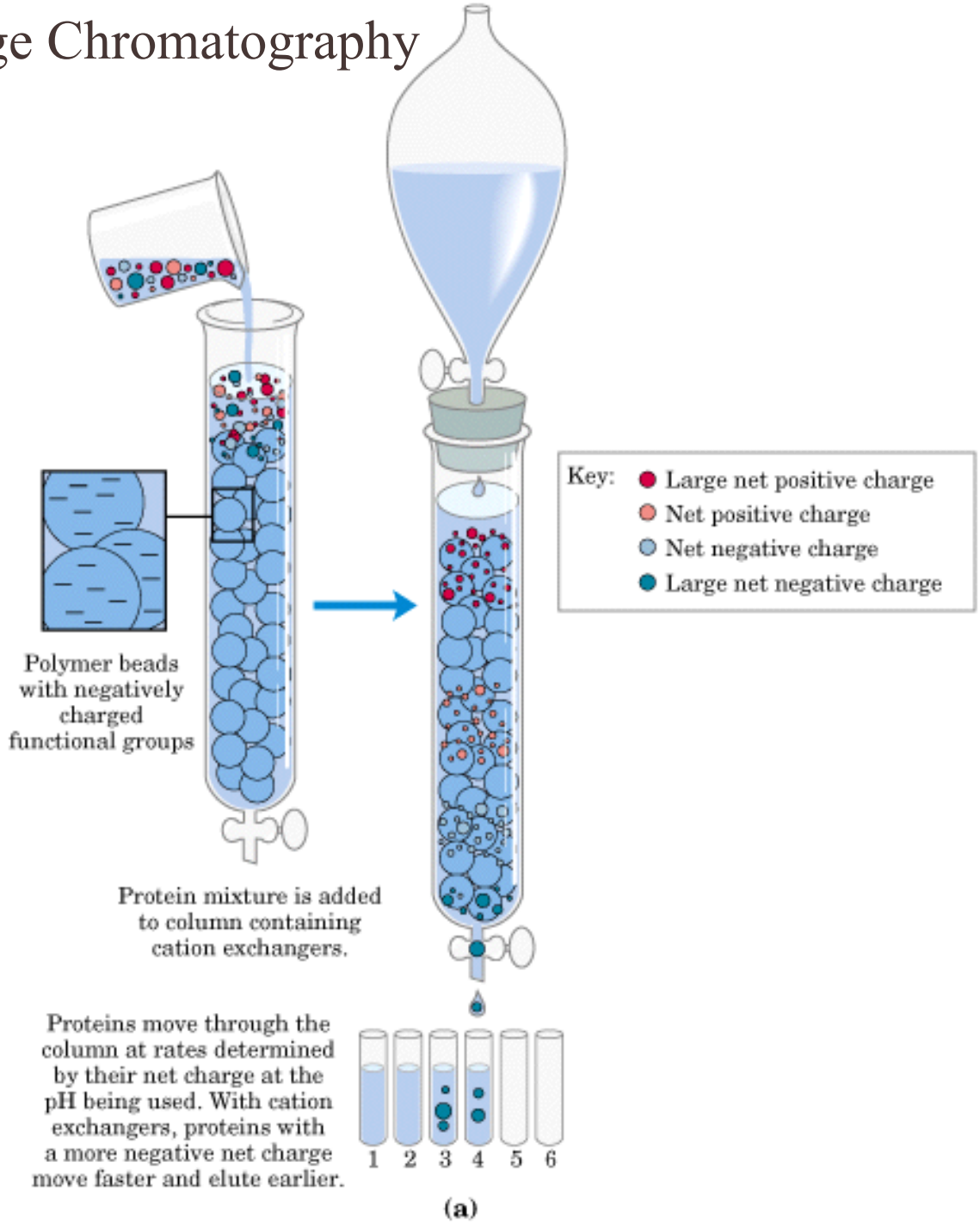
lipoproteins, glycoproteins, phosphoproteins, hemoproteins, flavoproteins, metalloproteins, phytochromes, cytochromes and opsins.


Primary structure
Secondary structure
Tertiary structure
Quaternary structure
Lys
Lys
Gly
Gly
Leu
Val
Ala
His
Amino acid residues
α Helix
Polypeptide chain
Assembled subunits

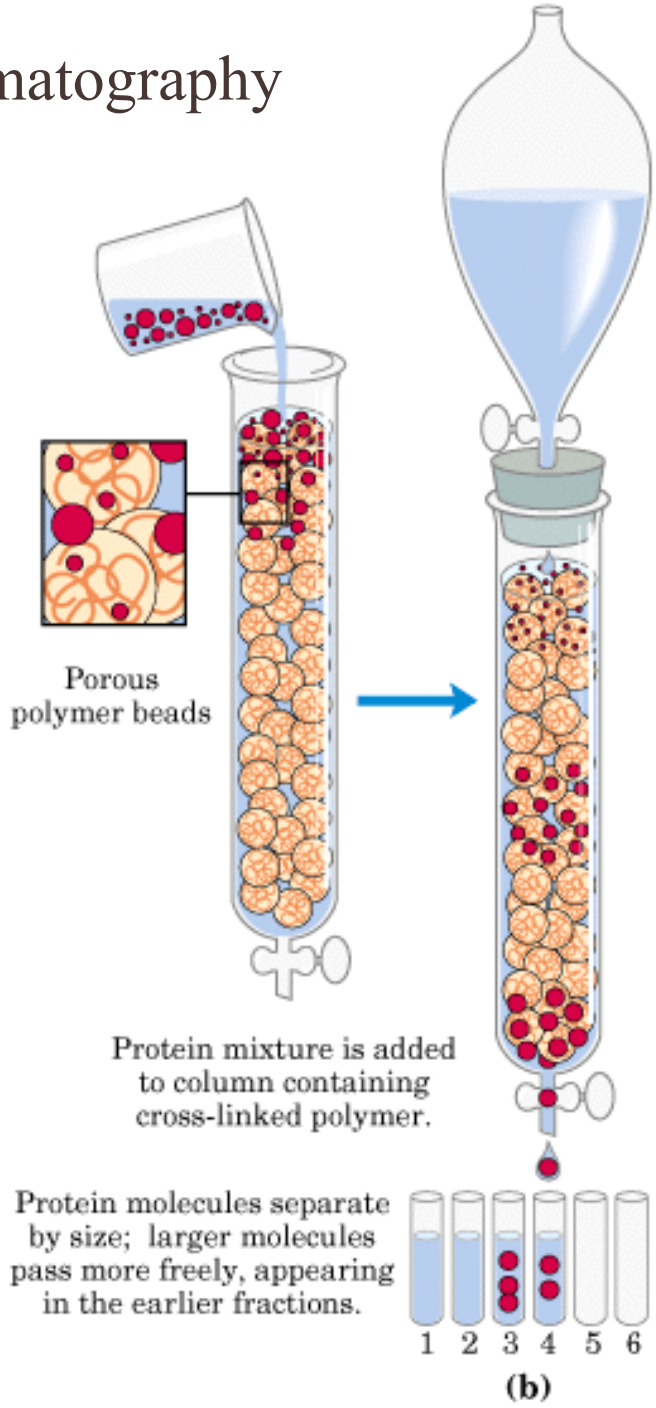
Column Chromatography



Ionic Exchange Chromatography



Size Exclusion Chromatography



Affinity Chromatography

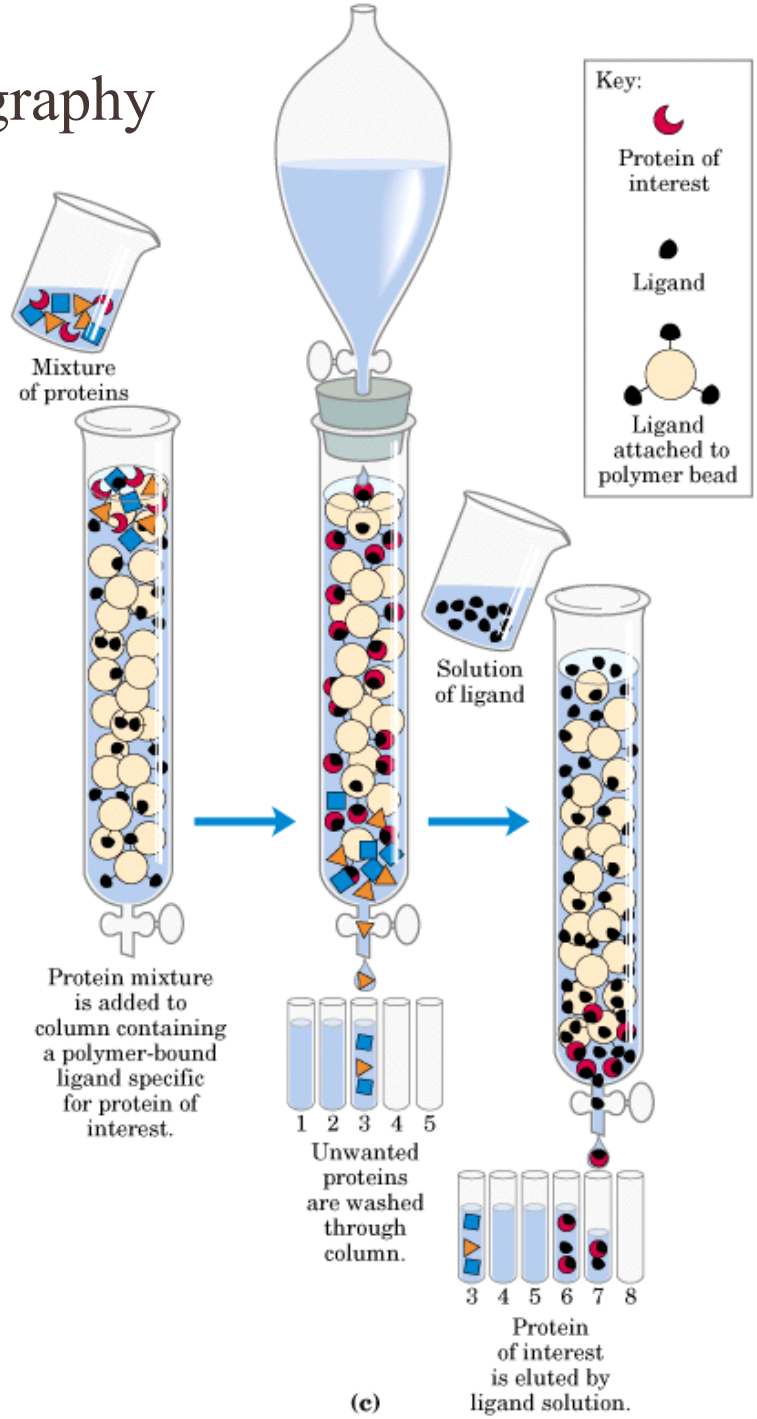


table 5–5

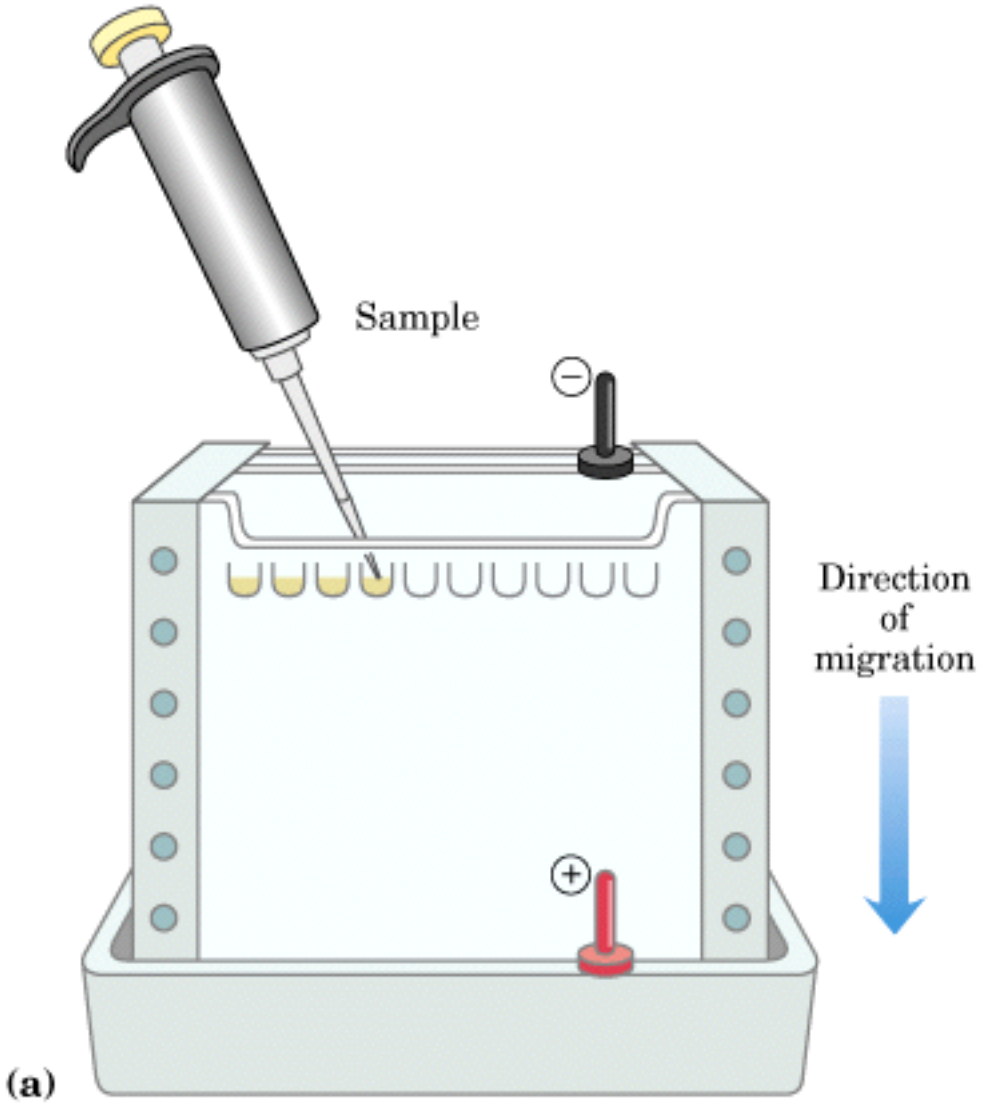
A Purification Table for a Hypothetical Enzyme*

Procedure or step	Fraction volume (ml)	Total protein (mg)	Activity (units)	Specific activity (units/mg)
1. Crude cellular extract	1,400	10,000	100,000	10
2. Precipitation with ammonium sulfate	280	3,000	96,000	32
3. Ion-exchange chromatography	90	400	80,000	200
4. Size-exclusion chromatography	80	100	60,000	600
5. Affinity chromatography	6	3	45,000	15,000

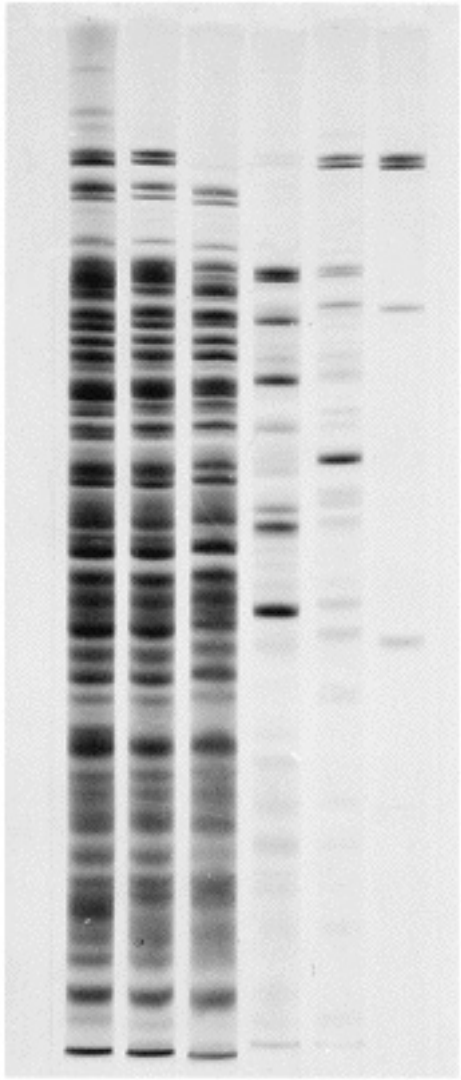
*All data represent the status of the sample *after* the designated procedure has been carried out. Activity and specific activity are defined on page 137.

HPLC: High Performance Liquid Chromatography
UFLC: Ultra Fast Liquid Chromatography

❖ Electrophoresis

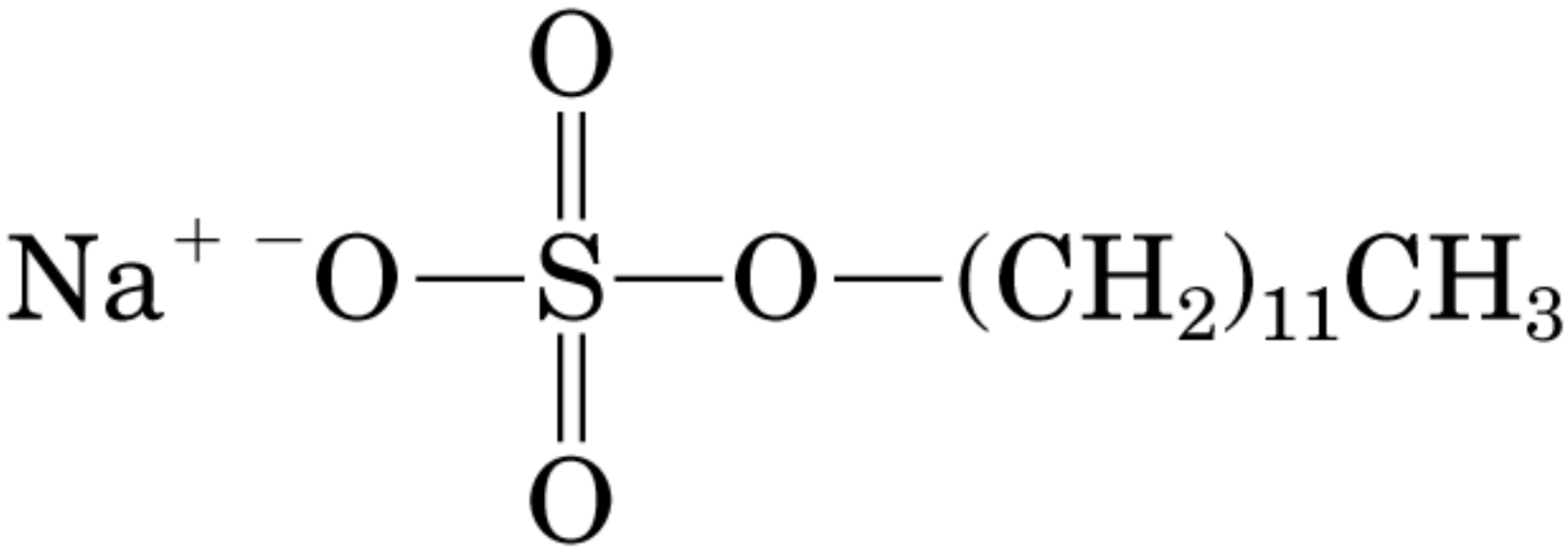


- Different Steps in Protein Purification and characterization.

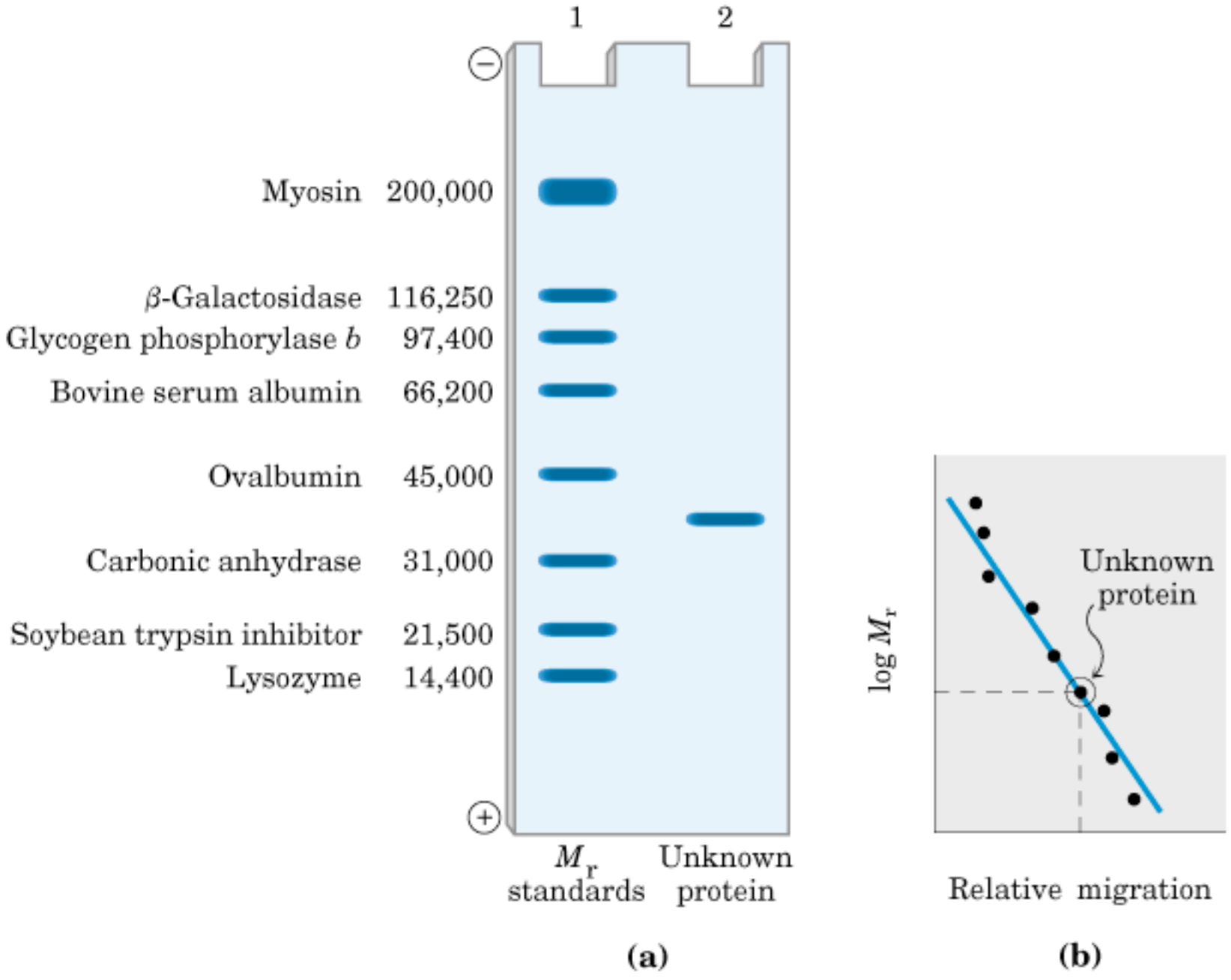


(b)

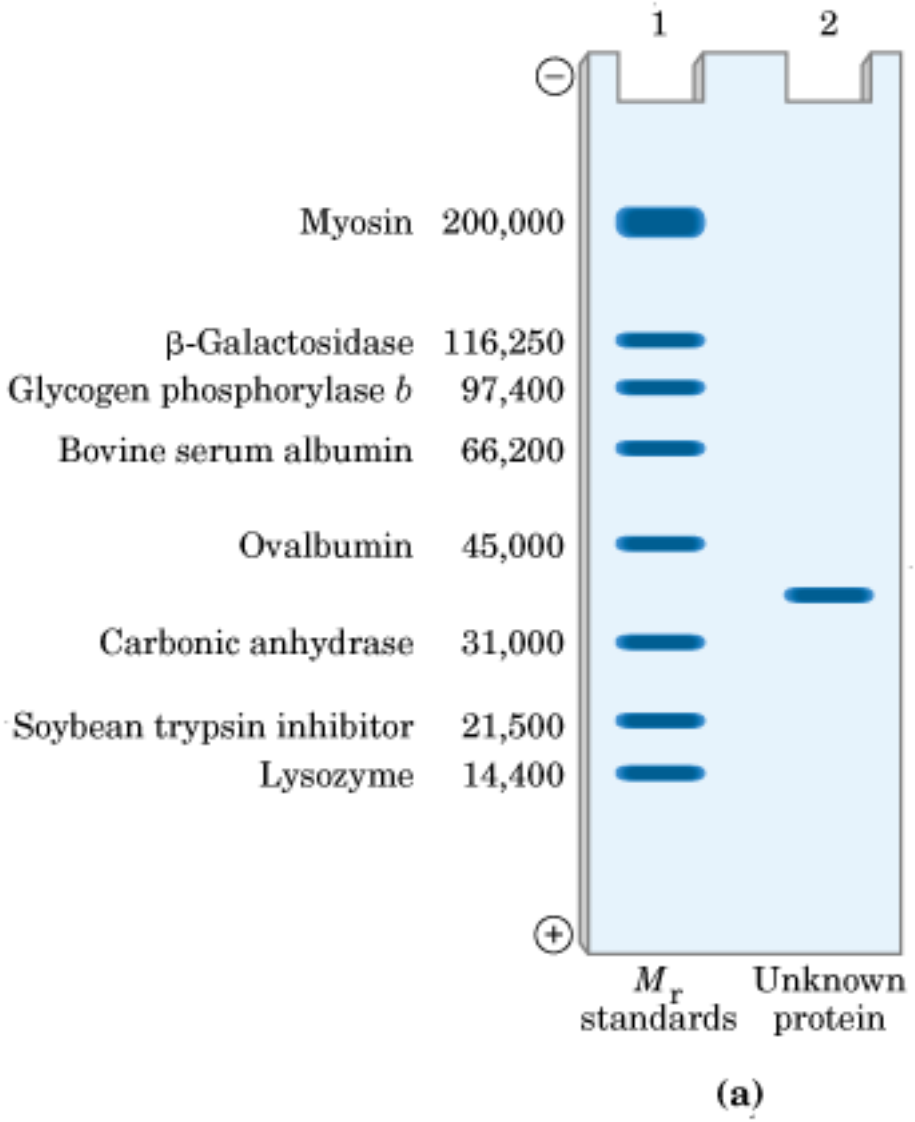
- Detergent one molecule per two residues.

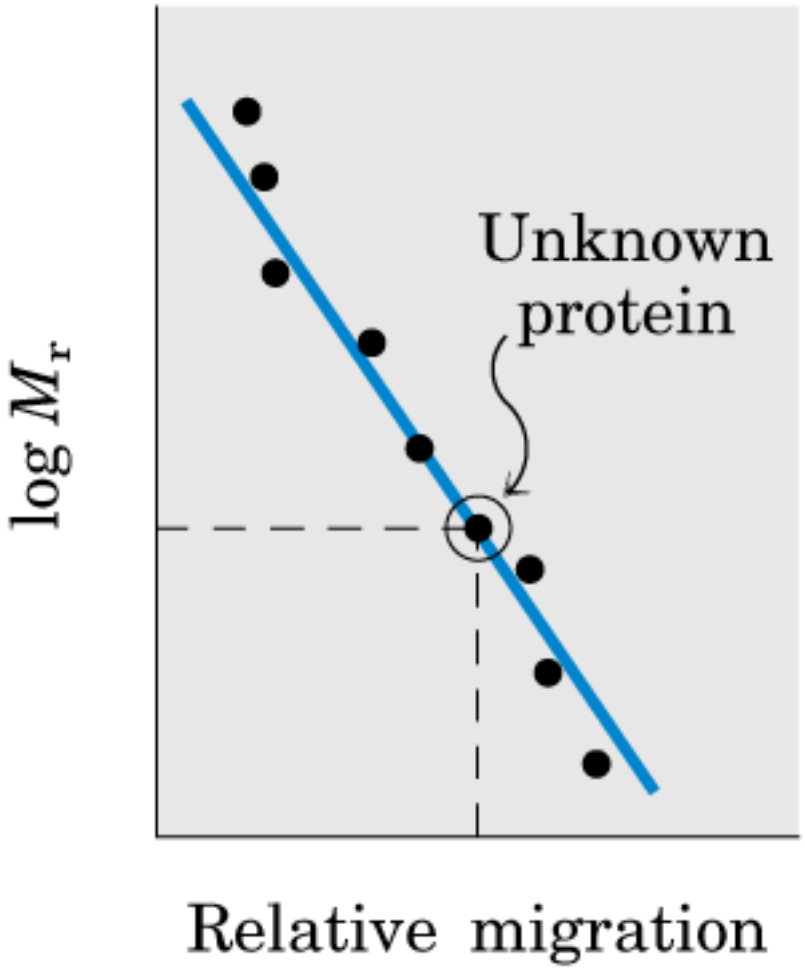


Sodium dodecyl sulfate
(SDS)


1
2
Myosin 200,000
β-Galactosidase 116,250
Glycogen phosphorylase b 97,400
Bovine serum albumin 66,200
Ovalbumin 45,000
Carbonic anhydrase 31,000
Soybean trypsin inhibitor 21,500
Lysozyme 14,400
M r standards
Unknown protein
(a)
log M r
Unknown protein
Relative migration
(b)

❖ Molecular Weight Markers





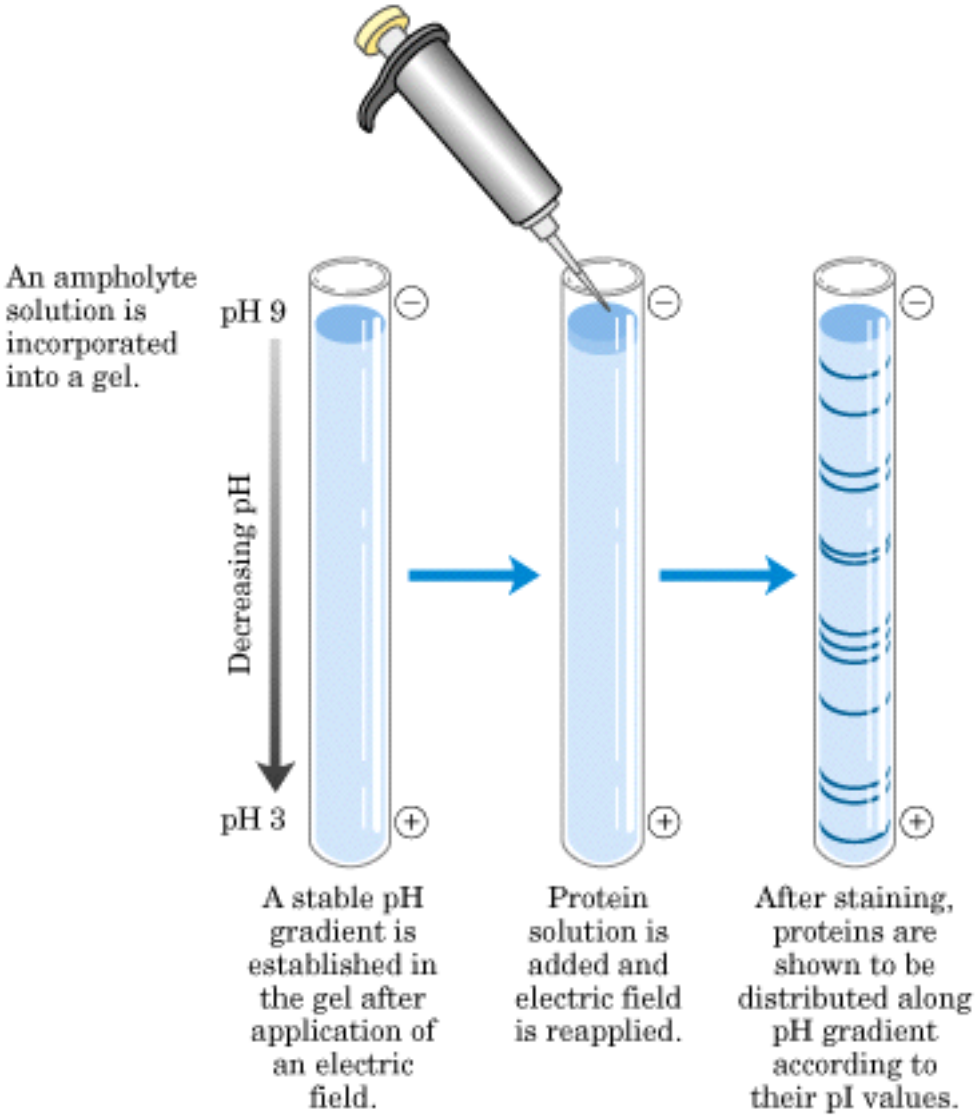
(b)

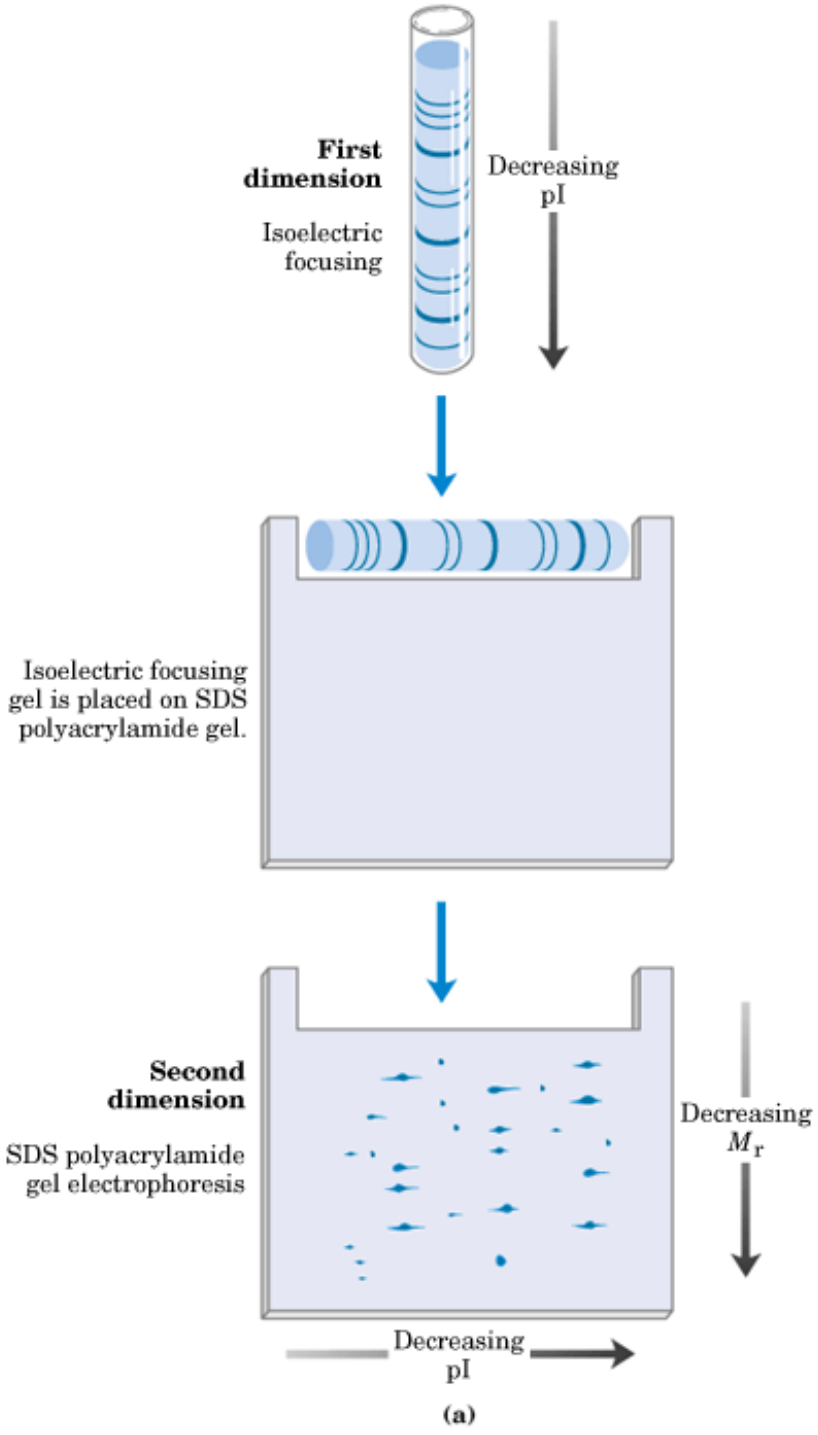
table 5–6

The Isoelectric Points of Some Proteins

Protein	pI
Pepsin	~1.0
Egg albumin	4.6
Serum albumin	4.9
Urease	5.0
β-Lactoglobulin	5.2
Hemoglobin	6.8
Myoglobin	7.0
Chymotrypsinogen	9.5
Cytochrome *c*	10.7
Lysozyme	11.0

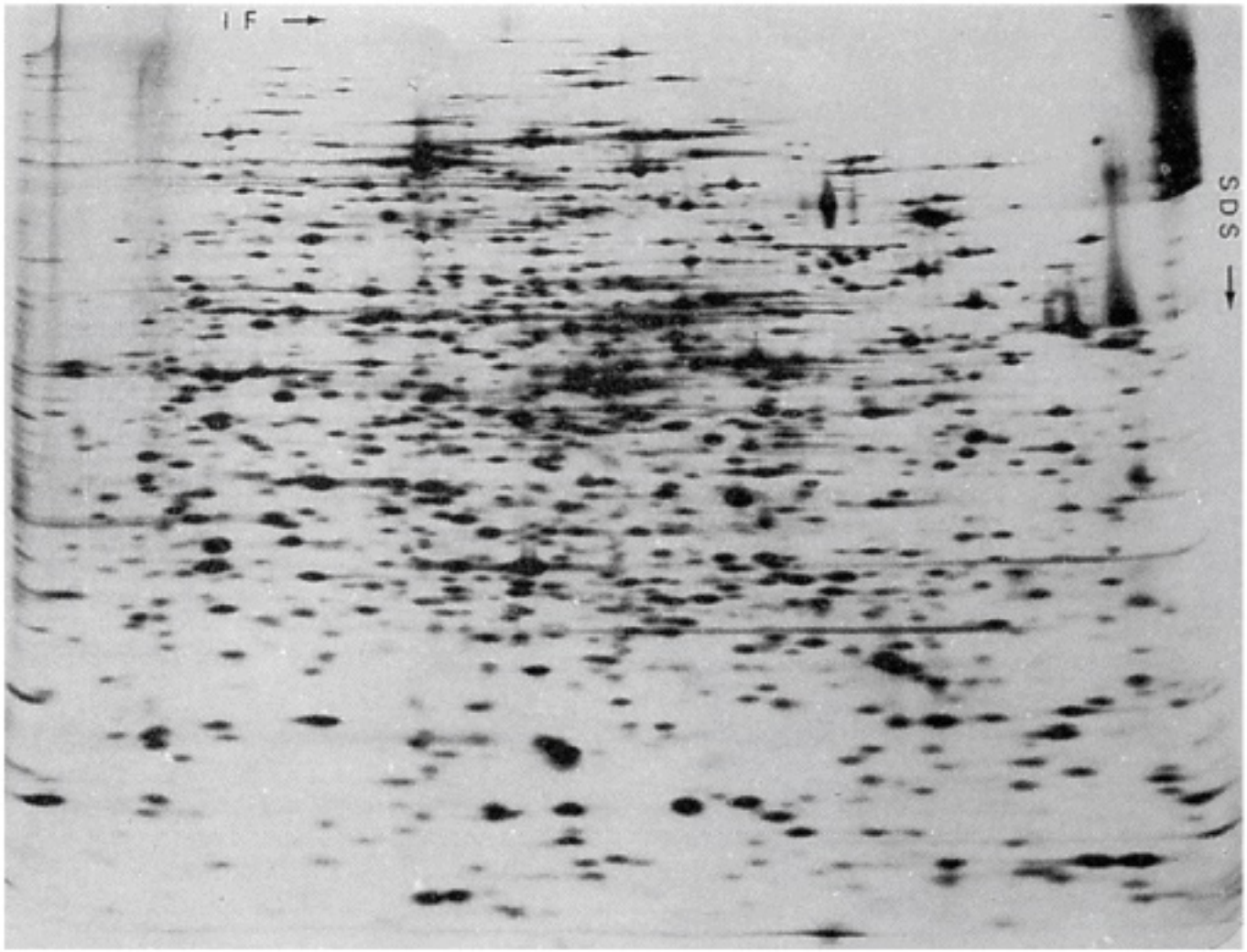
❖ Isoelectrofocusing




First dimension
Isoelectric focusing
Decreasing pI
Isoelectric focusing gel is placed on SDS polyacrylamide gel.
Second dimension
SDS polyacrylamide gel electrophoresis
Decreasing M_r
Decreasing pI

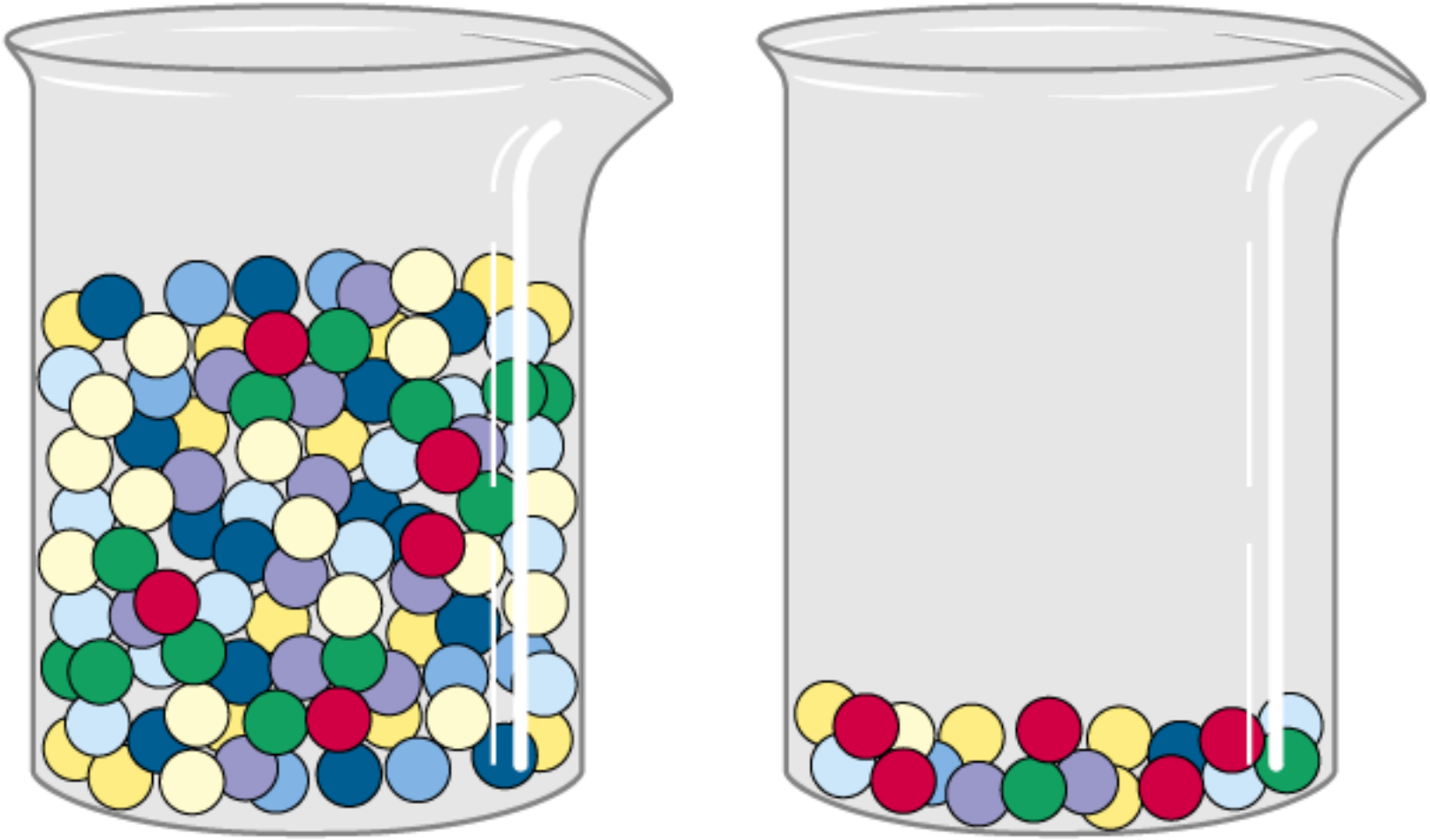
(a)

- Two Dimensions Electrophoresis

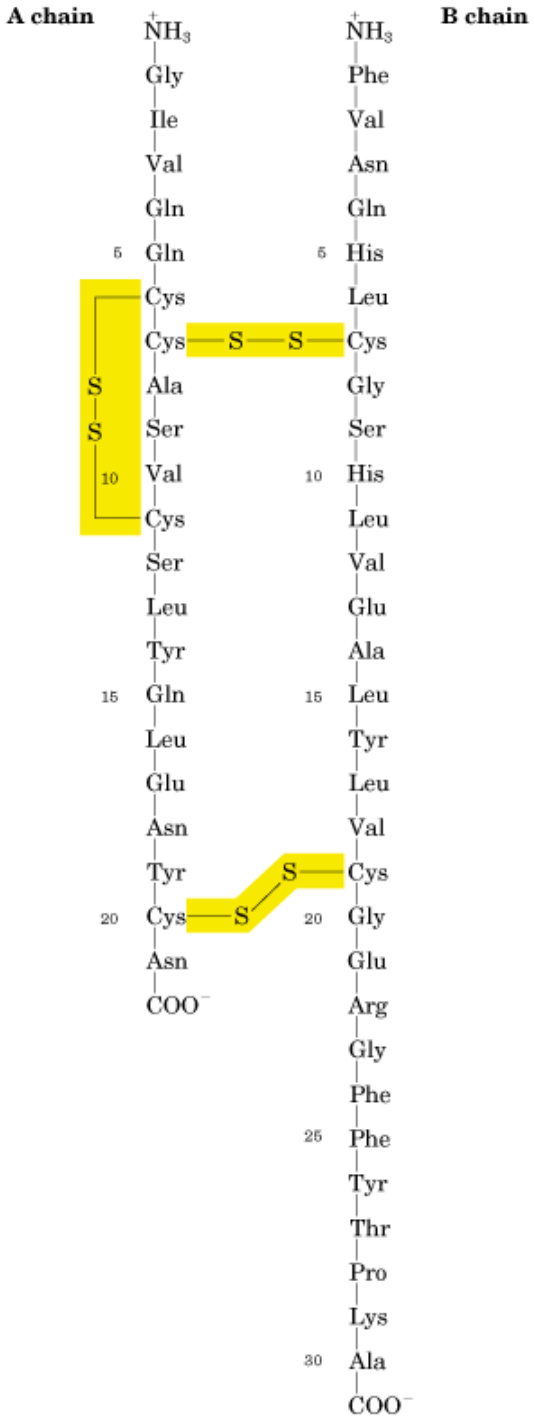


(b)

- Enzymatic activity

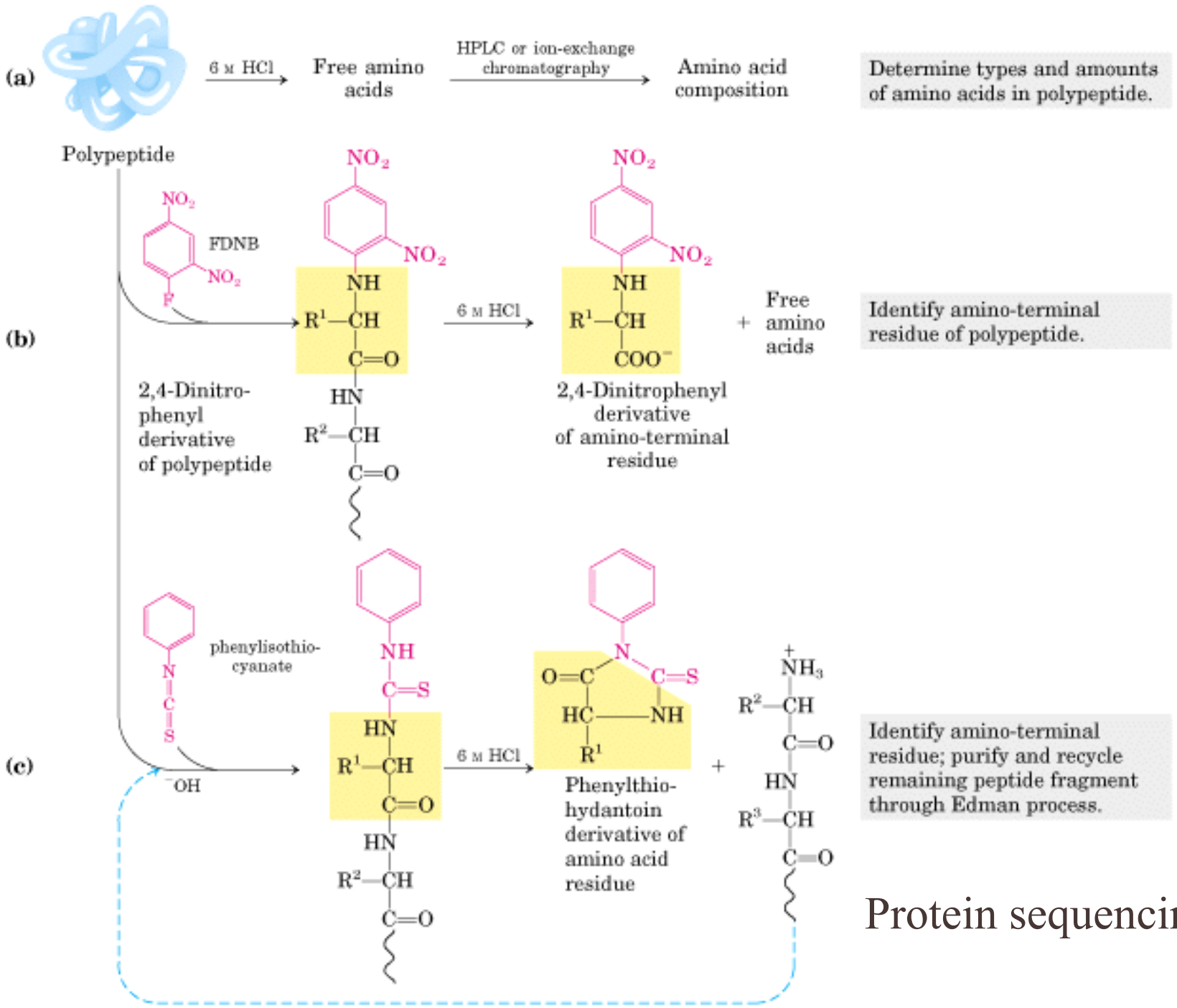


Amino acids sequence of INSULIN

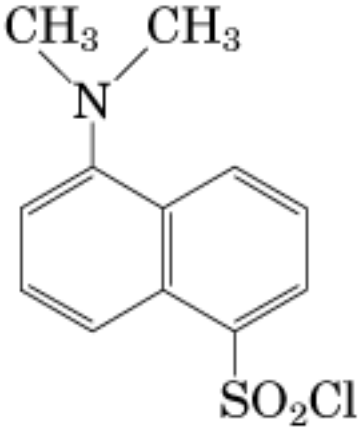



1 5 10 15 20 25 30 35 40 45
G D N S V I A P S E D A K T K A N S T G K A E K A T N I V L T F I V T K M Q T K R C S A E Q L E C H T G V I C E E D G K N A G G N G L A K G P T H K Q T Q I V G P N A L H N S W Y G L I F F Y I G S R K H Q T S G Q S T A V T P V A E D Q G Y F S T A Y S T S
50 55 60 65 70 75 80 85 90 95 100
A D E N T A N K I N S Q K R A K R M A G A N I V L V T N Q E W G A N Q D E K E D N Q E D E N P K T L M M S F Y R E D I A E Y L E T L E N P K K Y I P G T K M I V A I F V A G P T G I L I K S X K K K T A P E D E E D T N S Q A G K E E D R A G V T Q N E D N L I V A T A Y F L M L K K D E Q S T K A K T S T C S A A T N S S K A A S E K A S

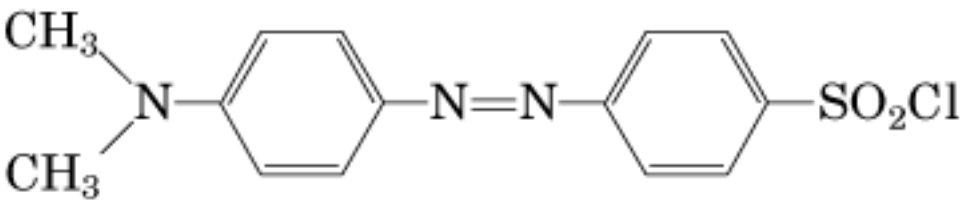

Mammals
Monkey
Homo sapiens, chimp
Mouse
Kangaroo
Rabbit
Horse, pig, sheep, cow
Dog, seal, bat
Hippopotamus
Birds and reptiles
Ostrich
Emu
Chicken, turkey
Pigeon
Penguin
Turtle
Duck
Rattlesnake
Bony fishes
Carp
Tuna
Cartilaginous fishes
Dogfish
Lamprey
Amphibians
Bullfrog
Annelids
Earthworm
Echinoderms
Starfish
Insects
Moth, hawkmoth
Honey bee
Fly
Locust
Fungi
Yeast
Candida
Neurospora
Humicola
Plants
Wheat
Sesame
Sunflower
Rice
Spinach
Ginkgo



Protein sequencing



Dansyl chloride



Dabsyl chloride

Breaking the sulfur bridges

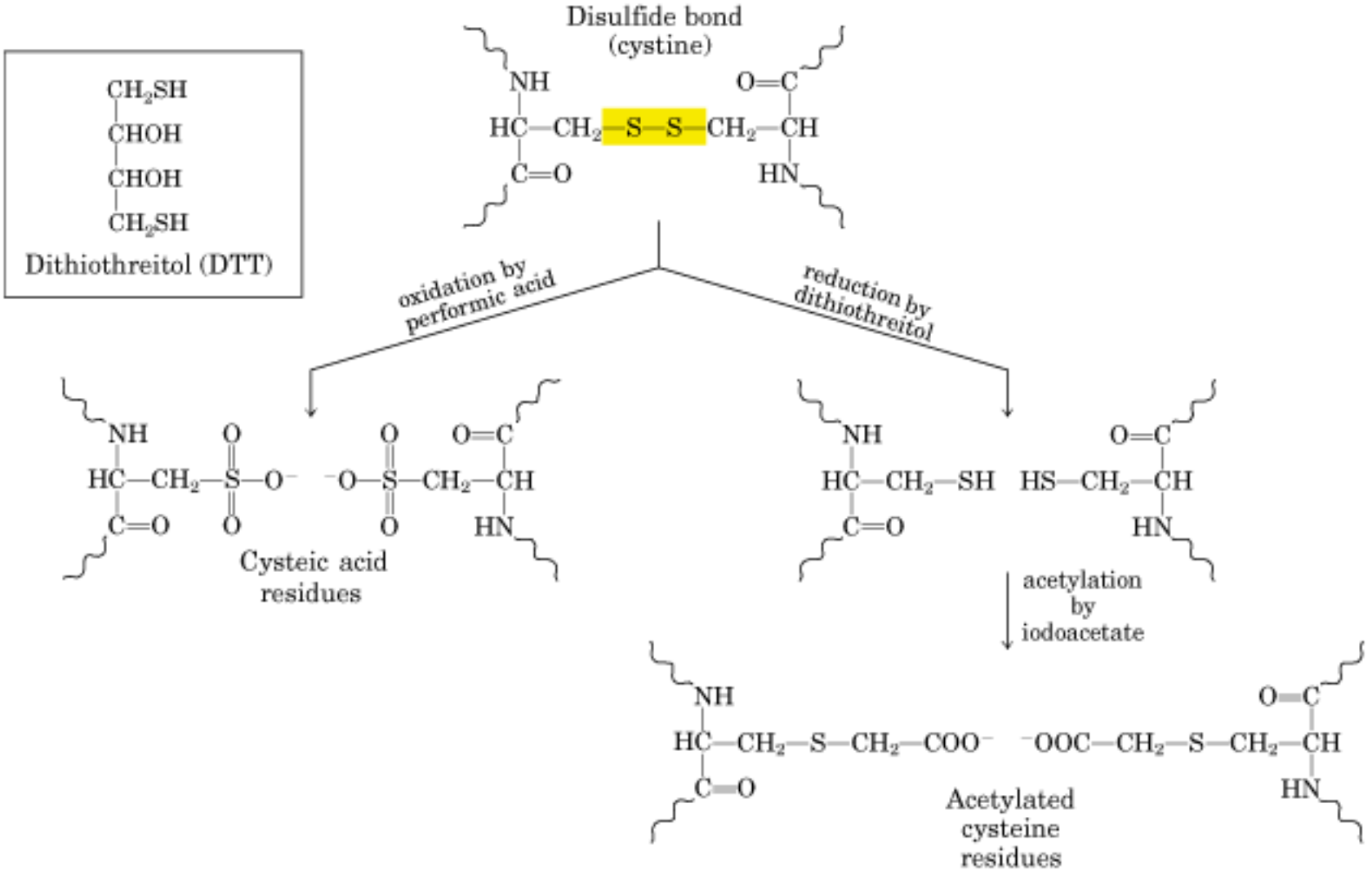


table 5–7

The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

Treatment*	Cleavage points†
Trypsin	Lys, Arg (C)
Submaxillarus protease	Arg (C)
Chymotrypsin	Phe, Trp, Tyr (C)
Staphylococcus aureus V8 protease	Asp, Glu (C)
Asp-*N*-protease	Asp, Glu (N)
Pepsin	Phe, Trp, Tyr (N)
Endoproteinase Lys C	Lys (C)
Cyanogen bromide	Met (C)

*All except cyanogen bromide are proteases. All are available from commercial sources.

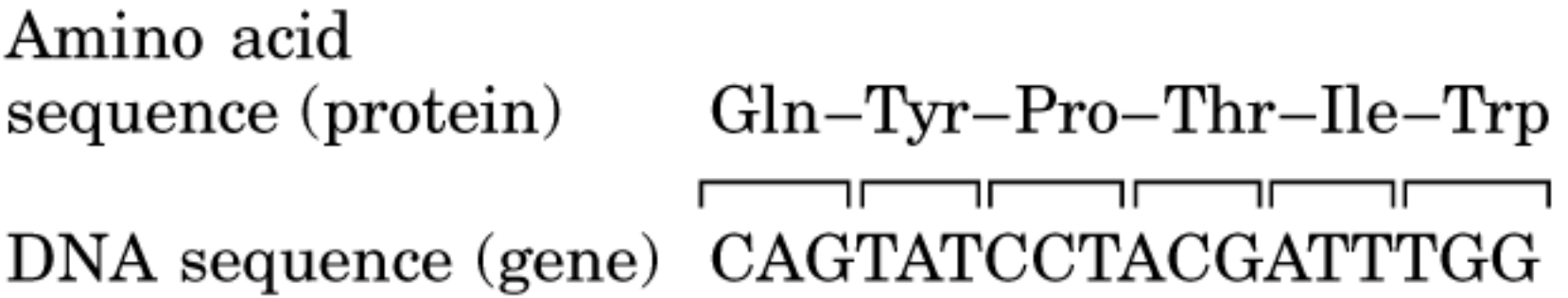
†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

Polypeptide

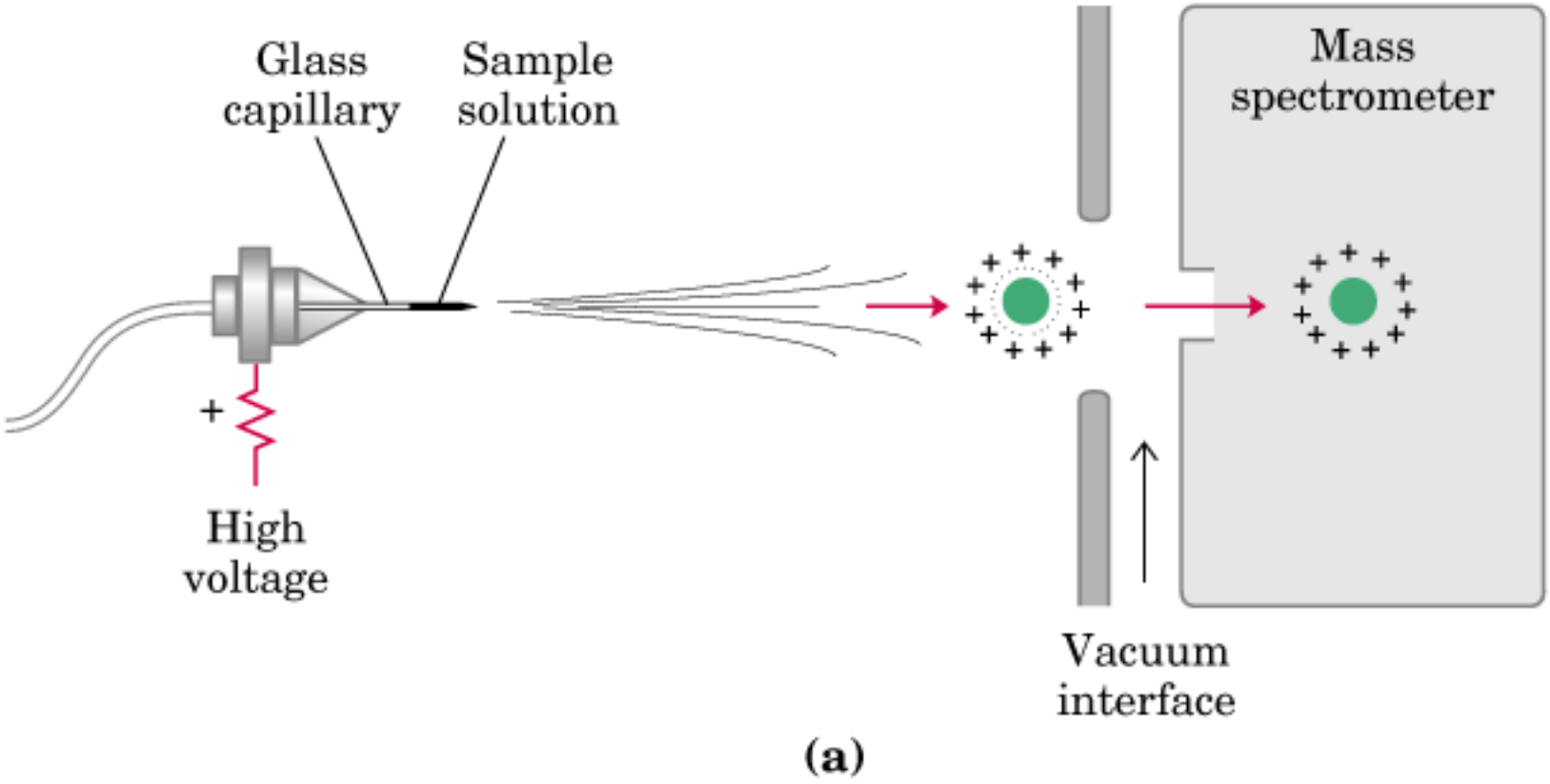
Procedure	Result	Conclusion
hydrolyze; separate amino acids	A 5, C 2, D 4, E 2, F 1, G 3, H 2, I 3, K 2, L 2, M 2, P 3, R 1, S 2, T 1, V 1, Y 2	Polypeptide has 38 amino acid residues. Trypsin will cleave three times (at one R (Arg) and two K (Lys)) to give four fragments. Cyanogen bromide will cleave at two M (Met) to give three fragments.
react with FDNB; hydrolyze; separate amino acids	2,4-Dinitrophenylglutamate detected	E (Glu) is amino-terminal residue.
reduce disulfide bonds (if present)		

cleave with trypsin separate fragments; sequence by Edman degradation	T-1 GASMALIK T-2 EGAAYHDFEPIDPR T-3 DCVHSD T-4 YLIACGPMTK	T-2 placed at amino terminus because it begins with E (Glu). T-3 placed at carboxyl terminus because it does not end with R (Arg) or K (Lys).
cleave with cyanogen bromide; separate fragments; sequence by Edman degradation	C-1 EGAAYHDFEPIDPRGASM C-2 TKDCVHSD C-3 ALIKYLIACGPM	C-3 overlaps with T-1 and T-4, allowing them to be ordered.

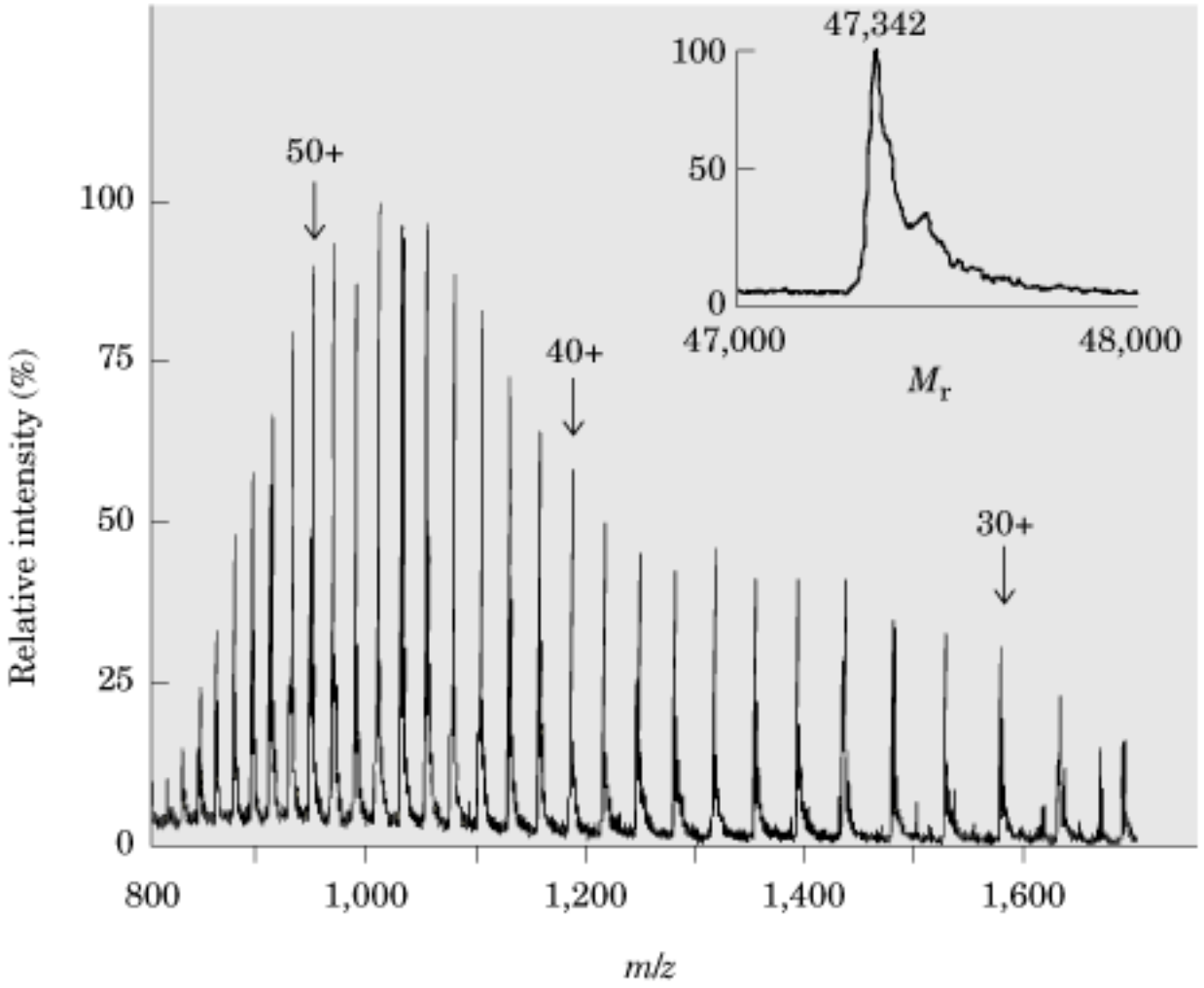
establish sequence


Amino acid
sequence (protein)
Gln–Tyr–Pro–Thr–Ile–Trp
DNA sequence (gene)
CAGTATCCTACGATTTGG

❖ Mass Spectrometry

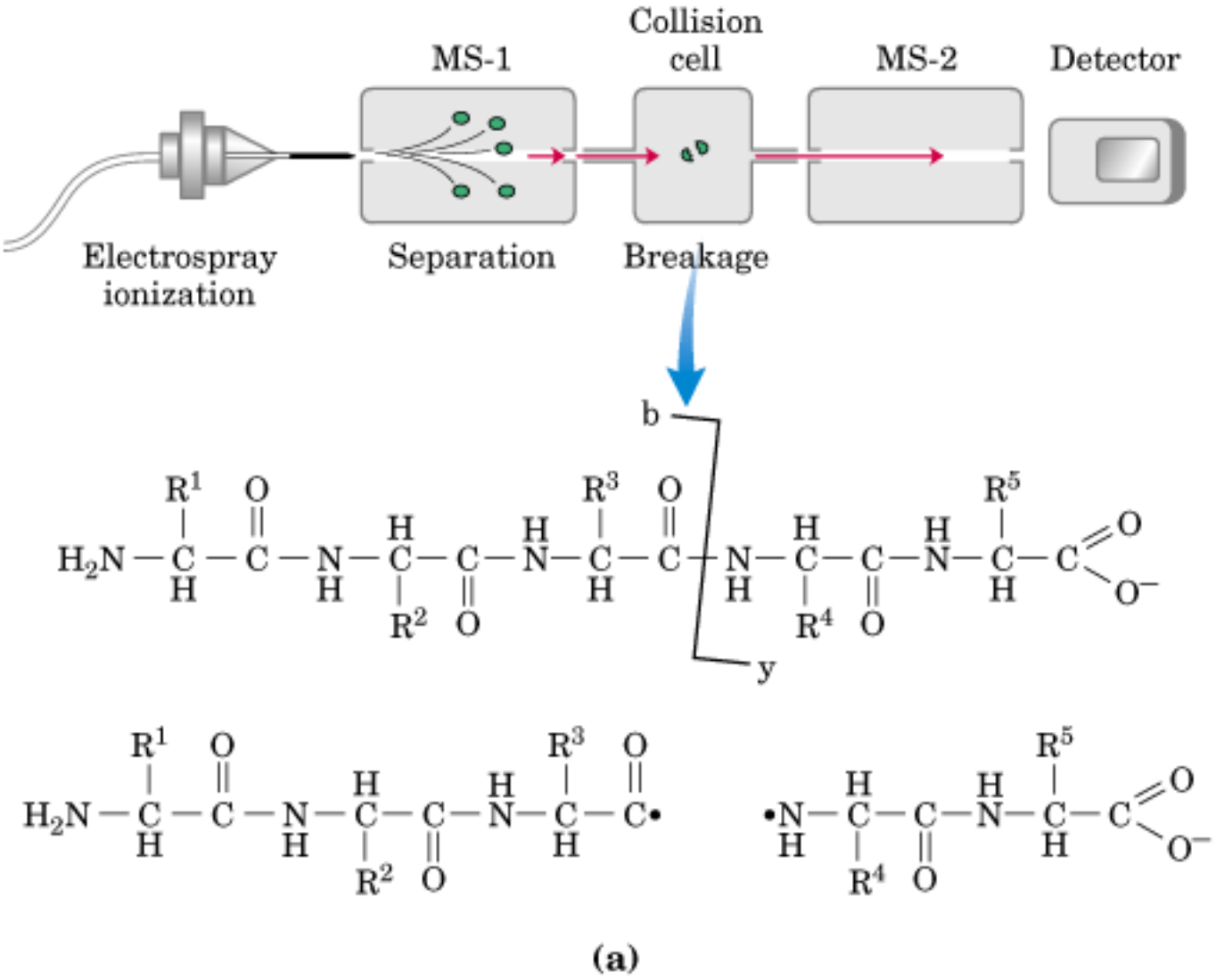


- MS Spectrum by electrospray

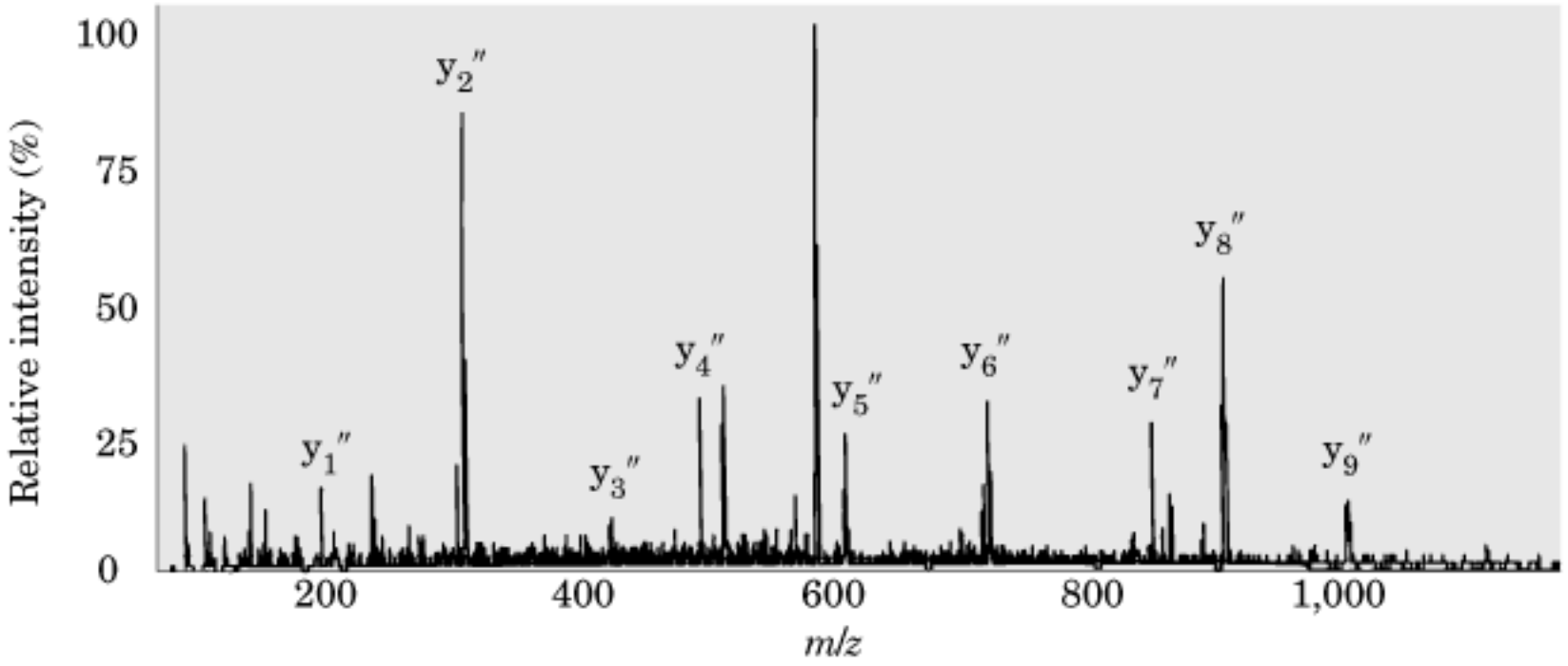


(b)

❖ Tandem of MS

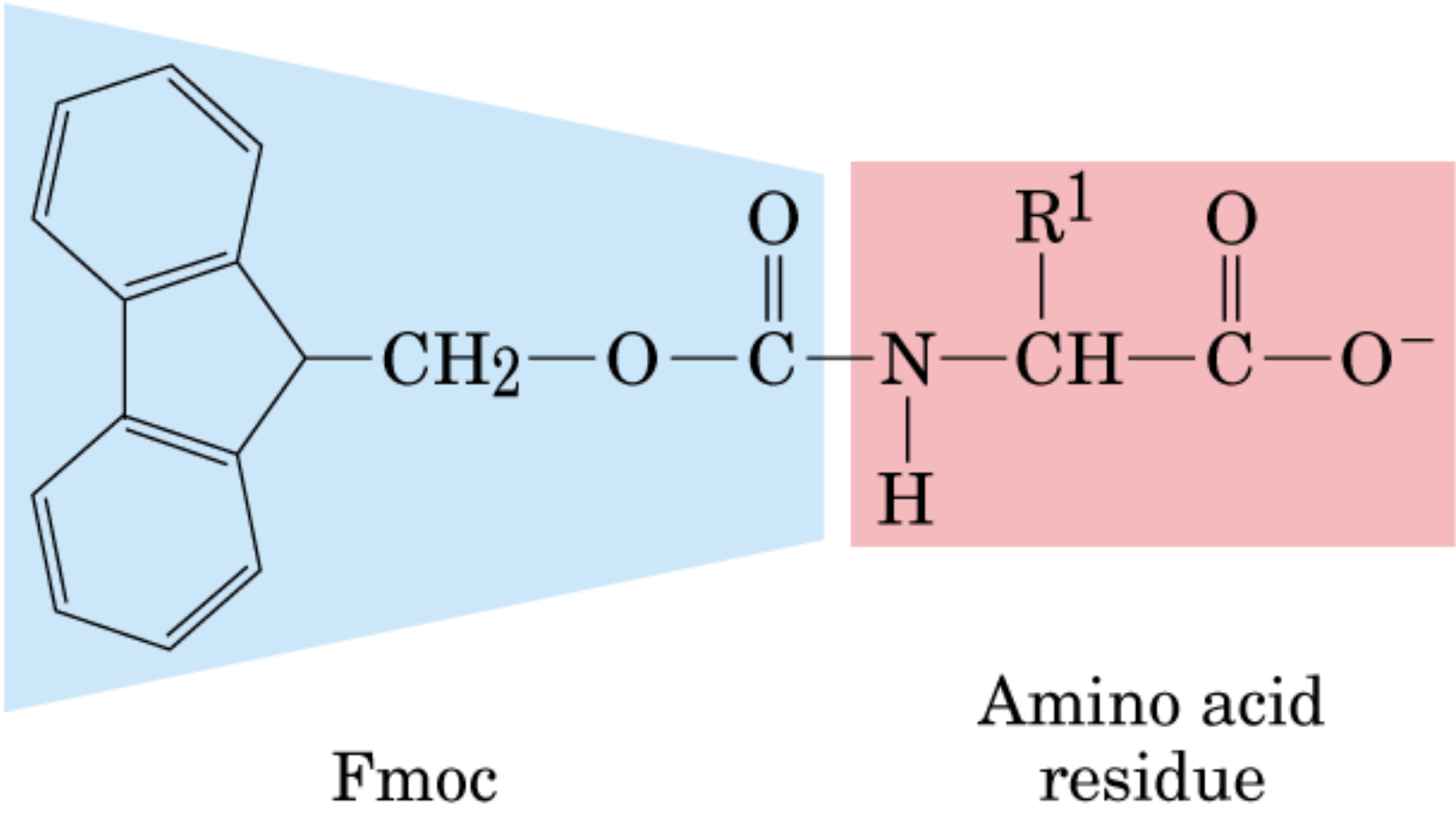

Collision
cell
MS-1
MS-2
Detector
Electrospray
ionization
Separation
Breakage
b
y
(a)

- Typical mass spectrum



(b)

- Peptide synthesis on a insoluble polymer



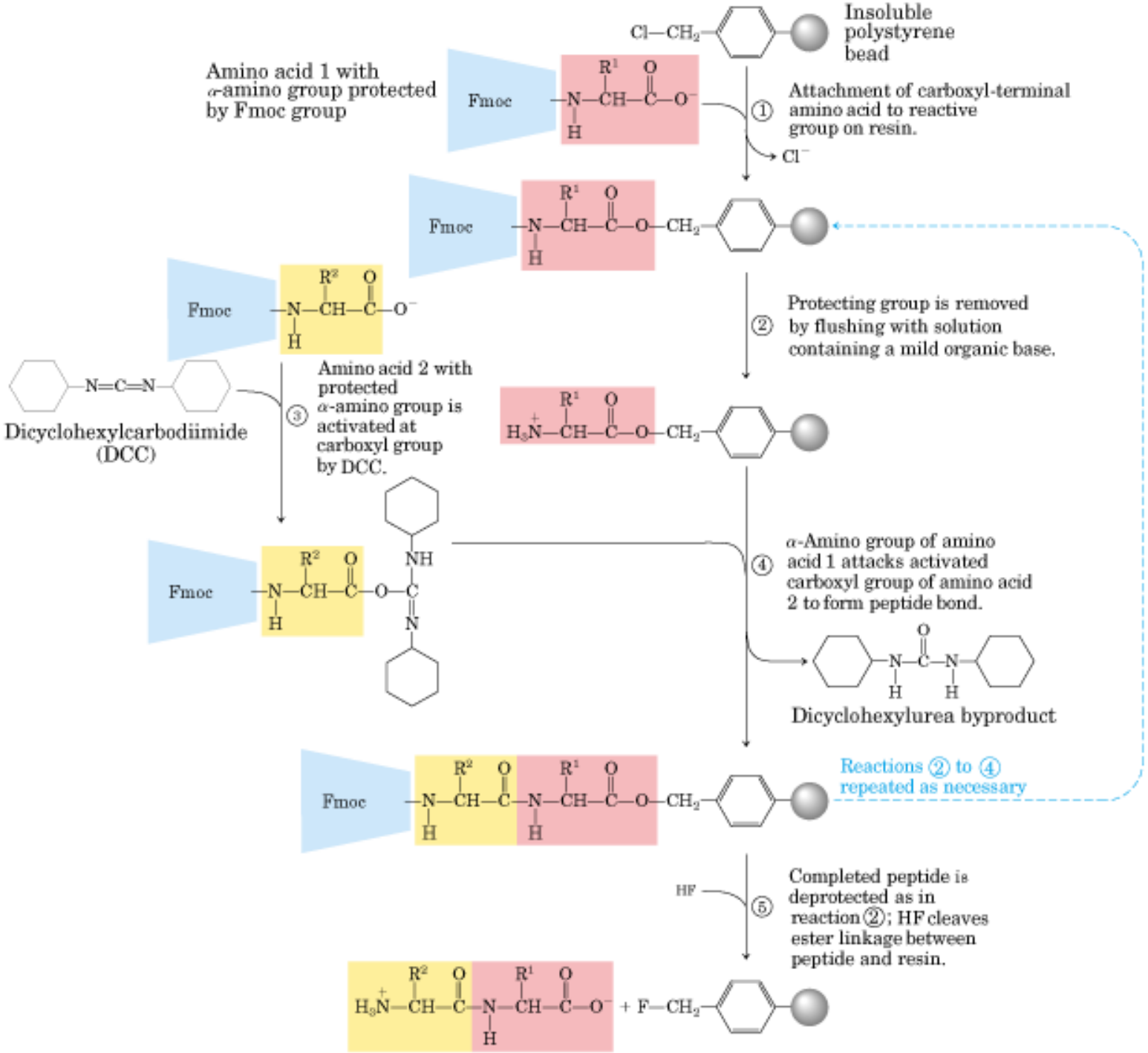

Insoluble polystyrene bead
Amino acid 1 with α-amino group protected by Fmoc group
Attachment of carboxyl-terminal amino acid to reactive group on resin.
Protecting group is removed by flushing with solution containing a mild organic base.
Amino acid 2 with protected α-amino group is activated at carboxyl group by DCC.
Dicyclohexylcarbodiimide (DCC)
α-Amino group of amino acid 1 attacks activated carboxyl group of amino acid 2 to form peptide bond.
Dicyclohexylurea byproduct
Reactions ② to ④ repeated as necessary
Completed peptide is deprotected as in reaction ②; HF cleaves ester linkage between peptide and resin.

table 5–8

Effect of Stepwise Yield on Overall Yield in Peptide Synthesis

Number of residues in the final polypeptide	Overall yield of final peptide (%) when the yield of each step is:	
	96.0%	99.8%
11	66	98
21	44	96
31	29	94
51	13	90
100	1.7	82