

3 Amino Acids, Peptides, and Proteins

© 2021 Macmillan
Learning



Principle 1 (1 of 2)

In every living organism, proteins are constructed from a common set of 20 amino acids. Each amino acid has a side chain with distinctive chemical properties. Amino acids may be regarded as the alphabet in which the language of protein structure is written.

Principle 2 (1 of 3)

In proteins, amino acids are joined in characteristic linear sequences through a common amide linkage, the peptide bond. The amino acid sequence of a protein constitutes its primary structure, a first level we will introduce within the broader complexities of protein structure.

Principle 3 (1 of 2)

For study, individual proteins can be separated from the thousands of other proteins present in a cell, based on differences in their chemical and functional properties arising from their distinct amino acid sequences. As proteins are central to biochemistry, the purification of individual proteins for study is a quintessential biochemical endeavor.

Principle 4 (1 of 2)

Shaped by evolution, amino acid sequences are a key resource for understanding the function of individual proteins and for tracing broader functional and evolutionary relationships.

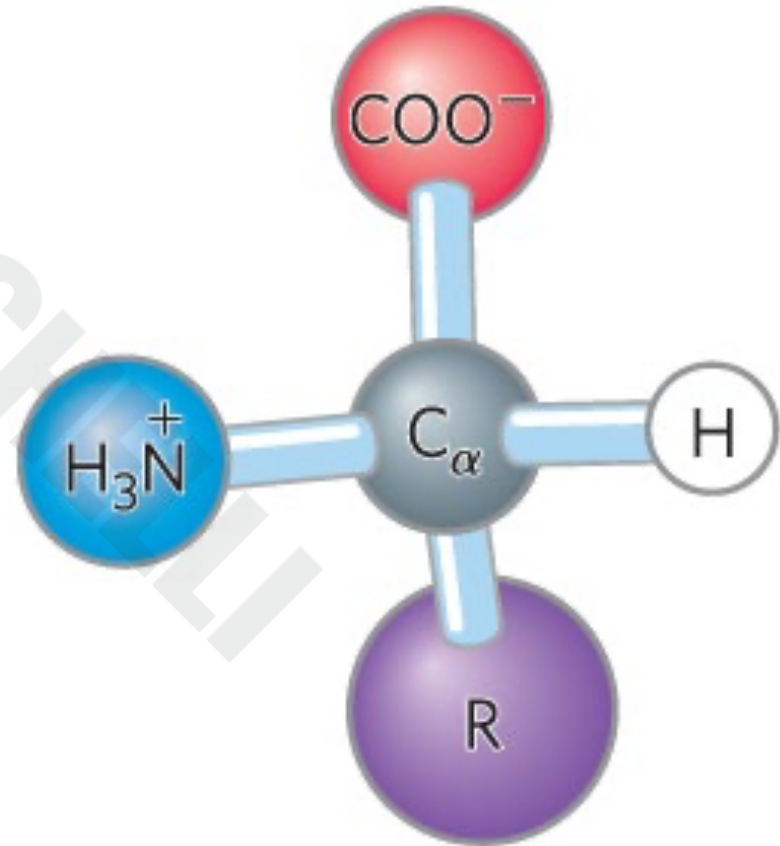
3.1 Amino Acids

Principle 1 (2 of 2)

In every living organism, proteins are constructed from a common set of 20 amino acids. Each amino acid has a side chain with distinctive chemical properties. Amino acids may be regarded as the alphabet in which the language of protein structure is written.

Amino Acids Share Common Structural Features

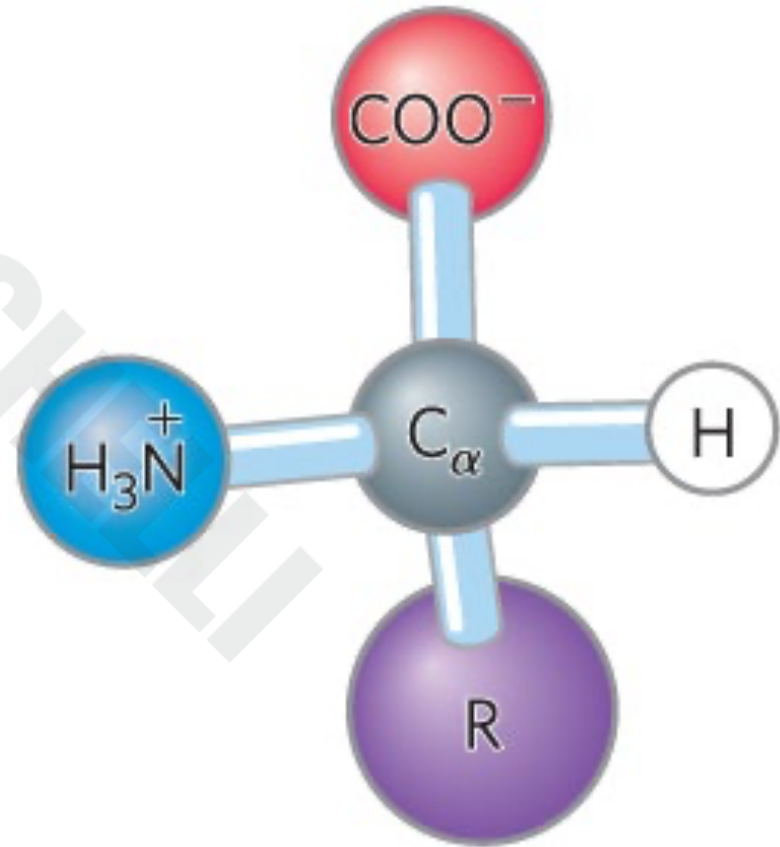
- α carbon and four substituents
- α carbon is the **chiral center**
- tetrahedral



Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021
W. H. Freeman and Company

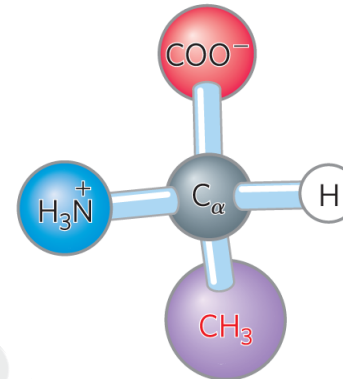
Amino Acid Substituents

- four substituents:
 - a carboxyl group
 - an amino group
 - a hydrogen atom
 - an **R group** (a side chain unique to each amino acid)
 - glycine has a second hydrogen atom instead of an R group

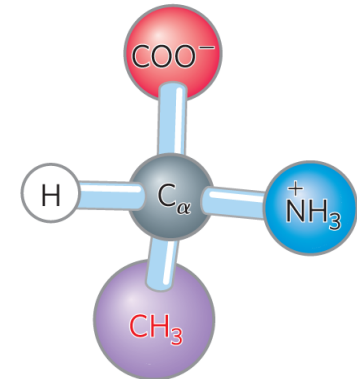


The Amino Acid Residues in Proteins are L Stereoisomers

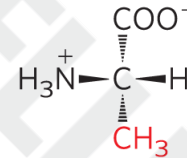
- two possible stereoisomers = **enantiomers**
- **optically active**
- D, L system specifies **absolute configuration**



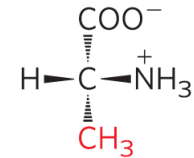
(a) L-Alanine



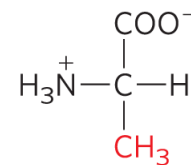
D-Alanine



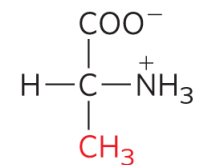
(b) L-Alanine



D-Alanine



(c) L-Alanine



D-Alanine

Amino Acids Can Be Classified by R Group

- five main classes:
 - nonpolar, aliphatic (7)
 - aromatic (3)
 - polar, uncharged (5)
 - positively charged (3)
 - negatively charged (2)

Table 3-1

TABLE 3-1 Properties and Conventions Associated with the Common Amino Acids Found in Proteins										
Amino acid	Abbreviation/ symbol	M_r^a	pK_a values			pI	Hydropathy index ^b	Occurrence in proteins (%) ^c		
			pK_1 (—COOH)	pK_2 (—NH ₃ ⁺)	pK_R (R group)					
Nonpolar, aliphatic R groups										
Glycine	Gly G	75	2.34	9.60		5.97	-0.4	7.2	7.3	7.3
Alanine	Ala A	89	2.34	9.69		6.01	1.8	7.8	9.4	7.2
Proline	Pro P	115	1.99	10.96		6.48	-1.6 ^d	5.2	4.4	4.2
Valine	Val V	117	2.32	9.62		5.97	4.2	6.6	7.1	8.2
Leucine	Leu L	131	2.36	9.60		5.98	3.8	9.1	10.6	9.9
Isoleucine	Ile I	131	2.36	9.68		6.02	4.5	5.3	6.0	7.6
Methionine	Met M	149	2.28	9.21		5.74	1.9	2.3	2.2	2.2
Aromatic R groups										
Phenylalanine	Phe F	165	1.83	9.13		5.48	2.8	3.9	4.0	4.5
Tyrosine	Tyr Y	181	2.20	9.11	10.07	5.66	-1.3	3.2	3.0	3.9
Tryptophan	Trp W	204	2.38	9.39		5.89	-0.9	1.4	1.3	1.1
Polar, uncharged R groups										
Serine	Ser S	105	2.21	9.15		5.68	-0.8	6.8	6.1	5.7
Threonine	Thr T	119	2.11	9.62		5.87	-0.7	5.9	5.4	4.5
Cysteine ^e	Cys C	121	1.96	10.28	8.18	5.07	2.5	1.9	1.2	0.8
Asparagine	Asn N	132	2.02	8.80		5.41	-3.5	4.3	3.7	3.4
Glutamine	Gln Q	146	2.17	9.13		5.65	-3.5	4.2	4.5	2.0
Positively charged R groups										
Lysine	Lys K	146	2.18	8.95	10.53	9.74	-3.9	5.9	4.7	6.8
Histidine	His H	155	1.82	9.17	6.00	7.59	-3.2	2.3	2.4	1.6
Arginine	Arg R	174	2.17	9.04	12.48	10.76	-4.5	5.1	5.6	5.9
Negatively charged R groups										
Aspartate	Asp D	133	1.88	9.60	3.65	2.77	-3.5	5.3	5.1	5.0
Glutamate	Glu E	147	2.19	9.67	4.25	3.22	-3.5	6.3	6.0	8.2

^a M_r values reflect the structures as shown in Figure 3-5. The elements of water (M_r 18) are deleted when the amino acid is incorporated into a polypeptide.

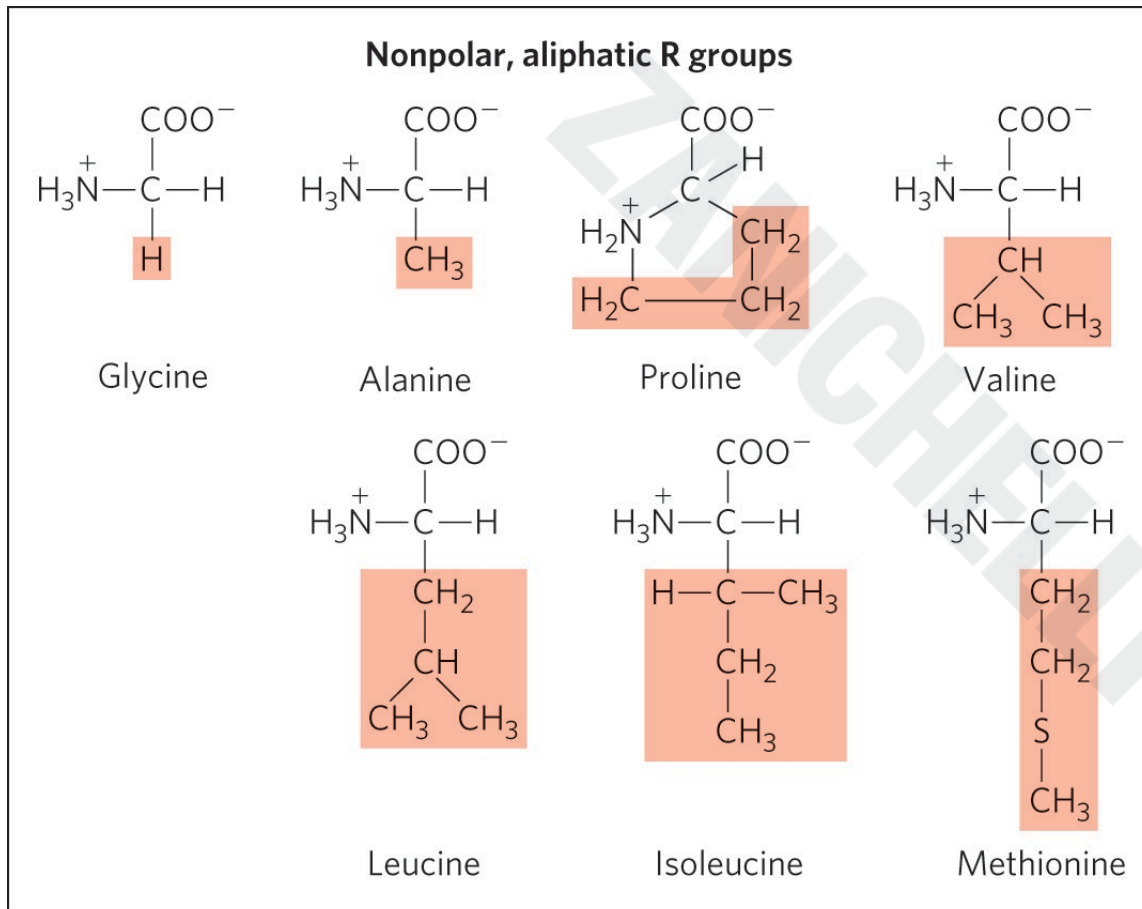
^bA scale combining hydrophobicity and hydrophilicity of R groups. The values reflect the free energy (ΔG) of transfer of the amino acid side chain from a hydrophobic environment to water. This transfer is favorable ($\Delta G < 0$; negative value in the index) for charged or polar amino acid side chains, and it is unfavorable ($\Delta G > 0$; positive value in the index) for amino acids with nonpolar or more hydrophobic side chains. See Chapter 11. Source: Data from J. Kyte and R. F. Doolittle, *J. Mol. Biol.* 157:105, 1982.

^cThe first value in each row is the average occurrence in more than 1,150 proteins. Source: Data from R. F. Doolittle, in *Prediction of Protein Structure and the Principles of Protein Conformation* (G. D. Fasman, ed.), p. 599, Plenum Press, 1989. The second and third values are, respectively, from the complete proteomes of nine mesophilic bacterial species and seven thermophilic bacterial species. Mesophiles grow at commonly encountered temperatures, whereas thermophiles grow at elevated temperatures up to and beyond the boiling point of water. The decline in glutamine occurrence in thermophiles may reflect a tendency of this amino acid to deaminate at high temperatures. Source: Data from A. C. Singer and D. A. Hickey, *Gene* 317:39, 2003.

^dAs originally composed, the hydropathy index takes into account the frequency with which an amino acid residue appears on the surface of a protein. As proline often appears on the surface in β turns, it has a lower score than its chain of methylene groups would suggest.

^eCysteine is generally classified as polar, despite having a positive hydropathy index. This reflects the ability of the sulfhydryl group to act as a weak acid and to form a weak hydrogen bond with oxygen or nitrogen.

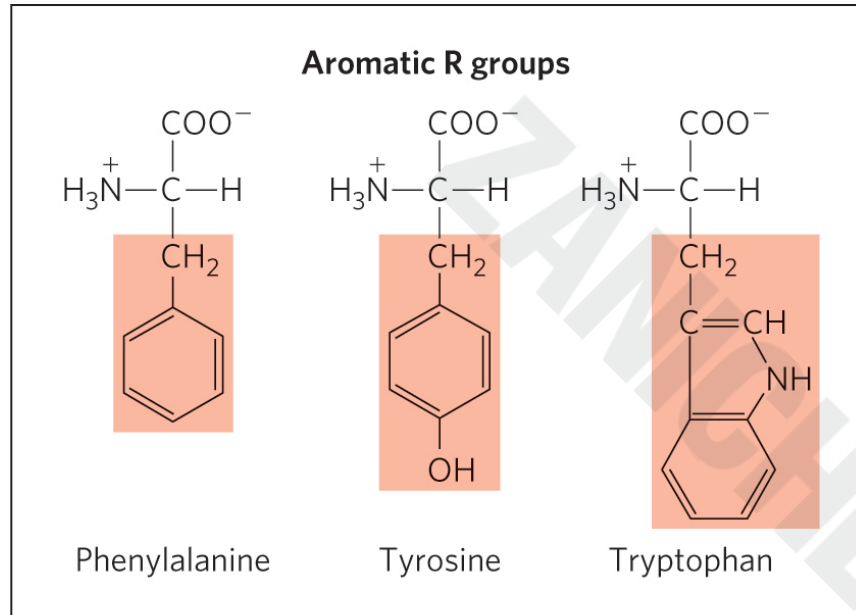
Nonpolar, aliphatic R groups



- the hydrophobic effect stabilizes protein structure

Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company

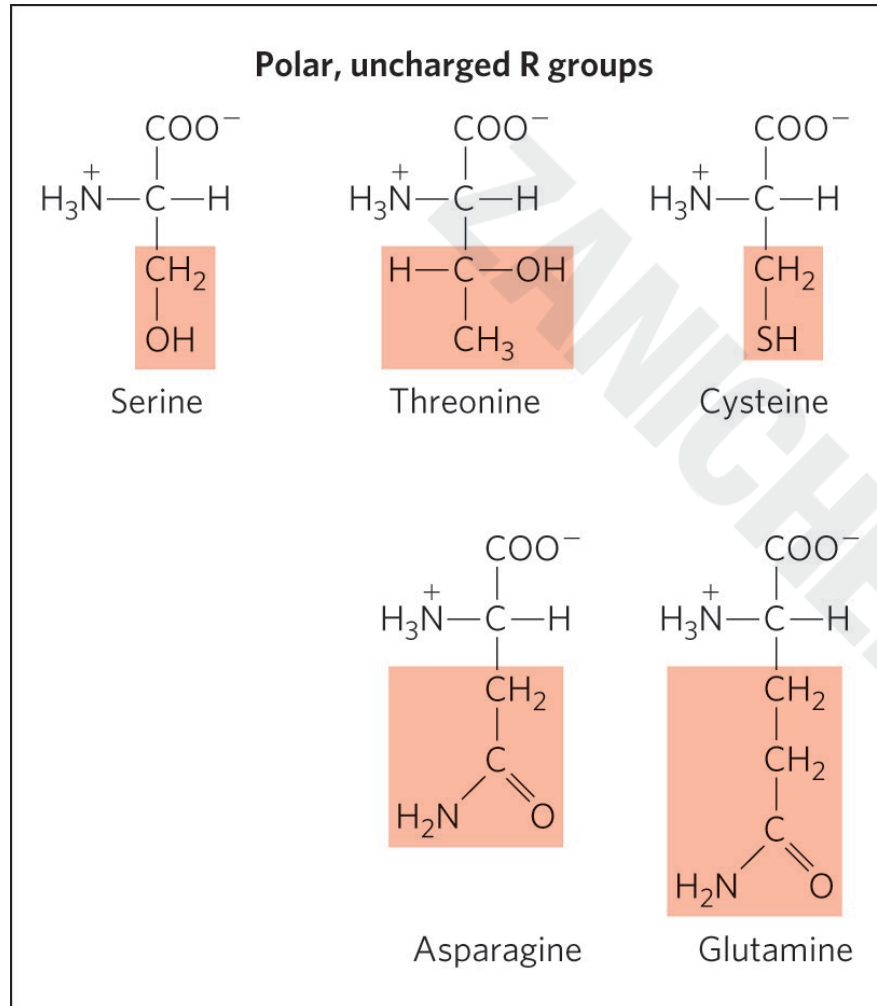
Aromatic R Groups



Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company

- R groups absorb UV light at 270–280 nm
- can contribute to the hydrophobic effect

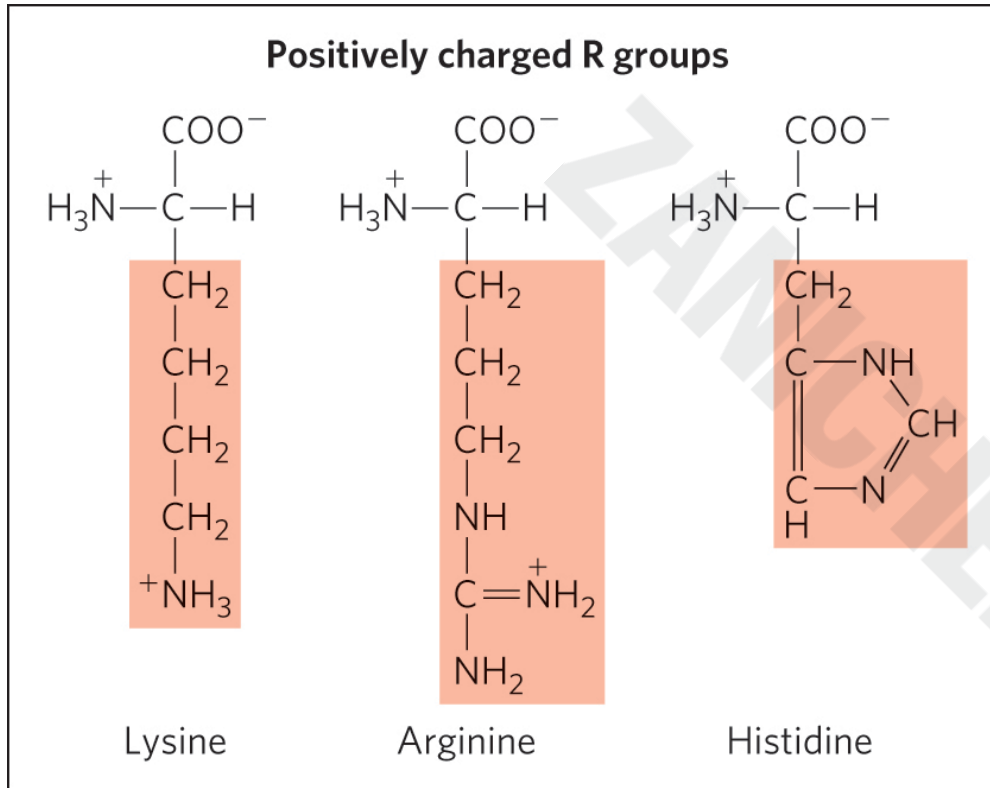
Polar, Uncharged R Groups



- R groups can form hydrogen bonds
- cysteine can form disulfide bonds

Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company

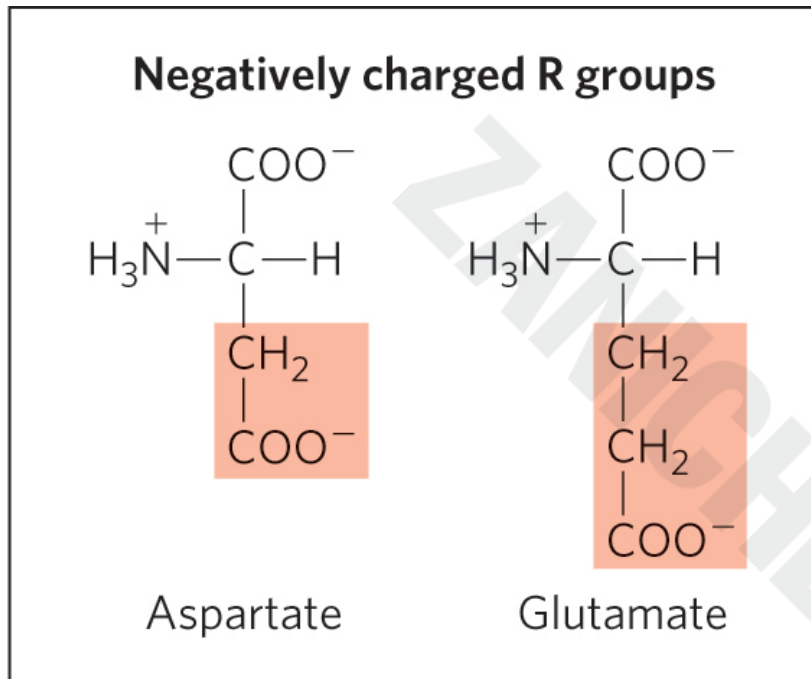
Positively Charged R Groups



- have significant positive charge at pH 7.0

Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company

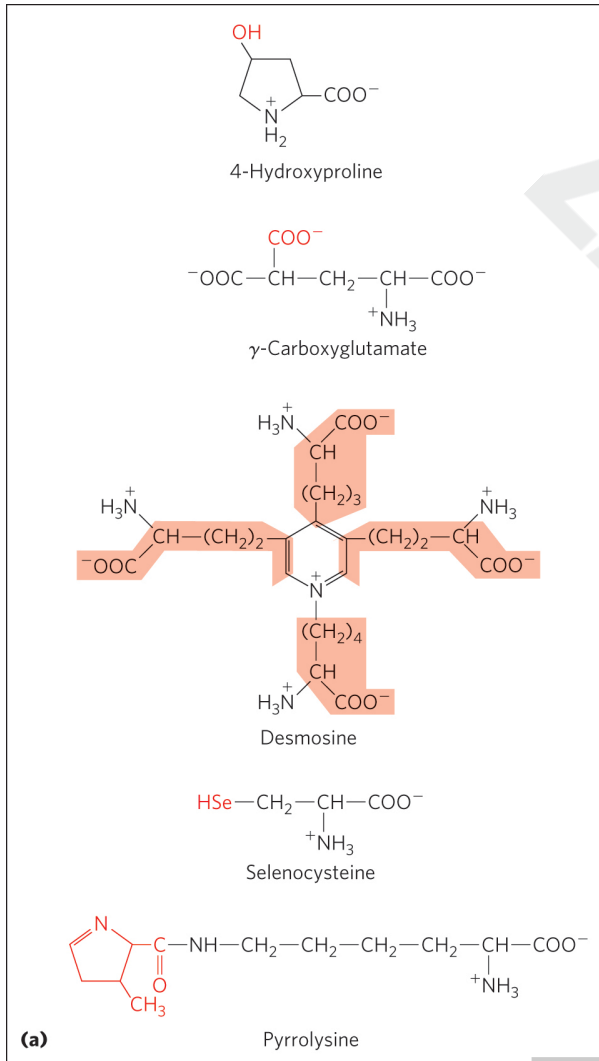
Negatively Charged R Groups



- have a net negative charge at pH 7.0

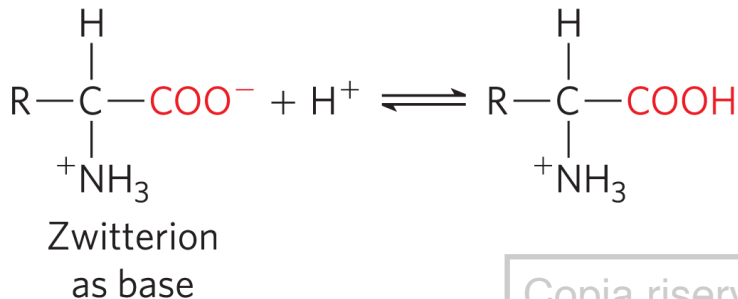
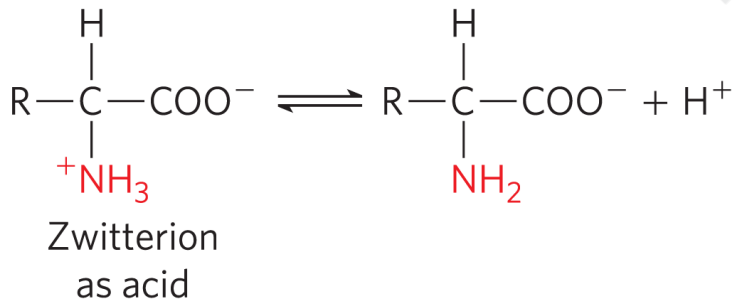
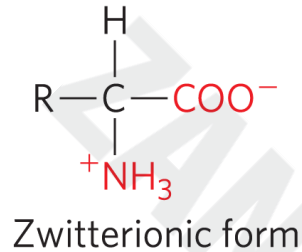
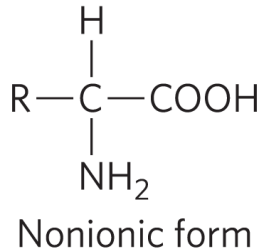
Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e,
© 2021 W. H. Freeman and Company

Uncommon Amino Acids Also Have Important Functions



- modifications of common amino acids:
 - modified after protein synthesis (e.g., 4-hydroxyproline, found in collagen)
 - modified during protein synthesis (e.g., pyrrolysine, contributes to methane biosynthesis)
 - modified transiently to change protein's function (e.g., phosphorylation)
- free metabolites (e.g., ornithine, intermediate in arginine biosynthesis)

Amino Acids Can Act as Acids or Bases

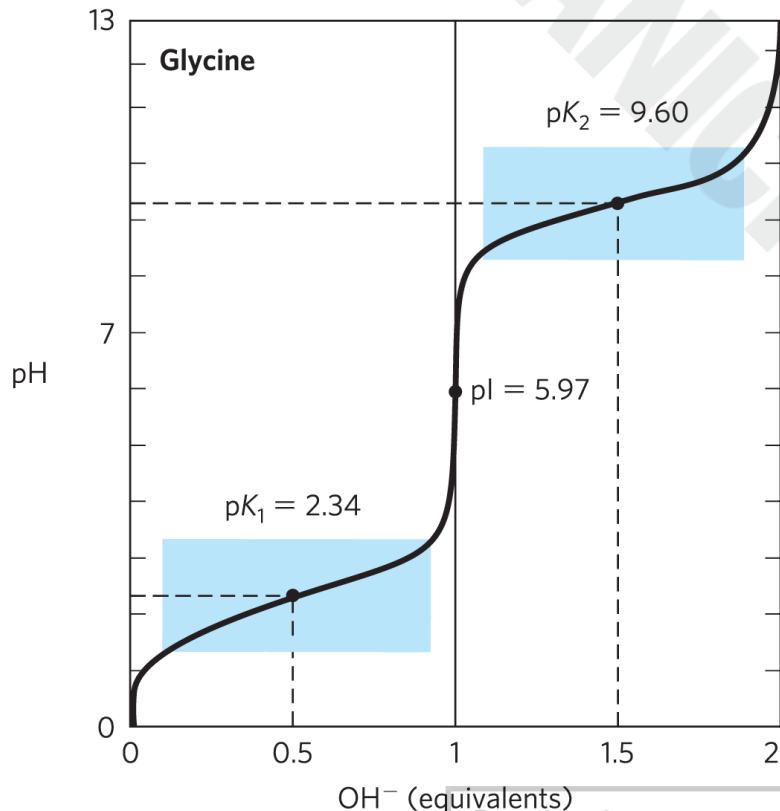
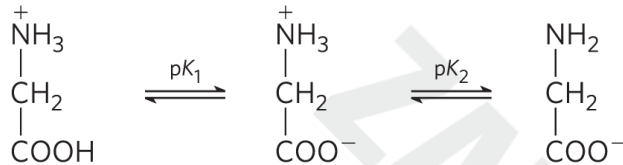


- amino groups, carboxyl groups, and ionizable R groups = weak acids and bases
- **zwitterion** occurs at neutral pH

Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company

Titration of Amino Acids

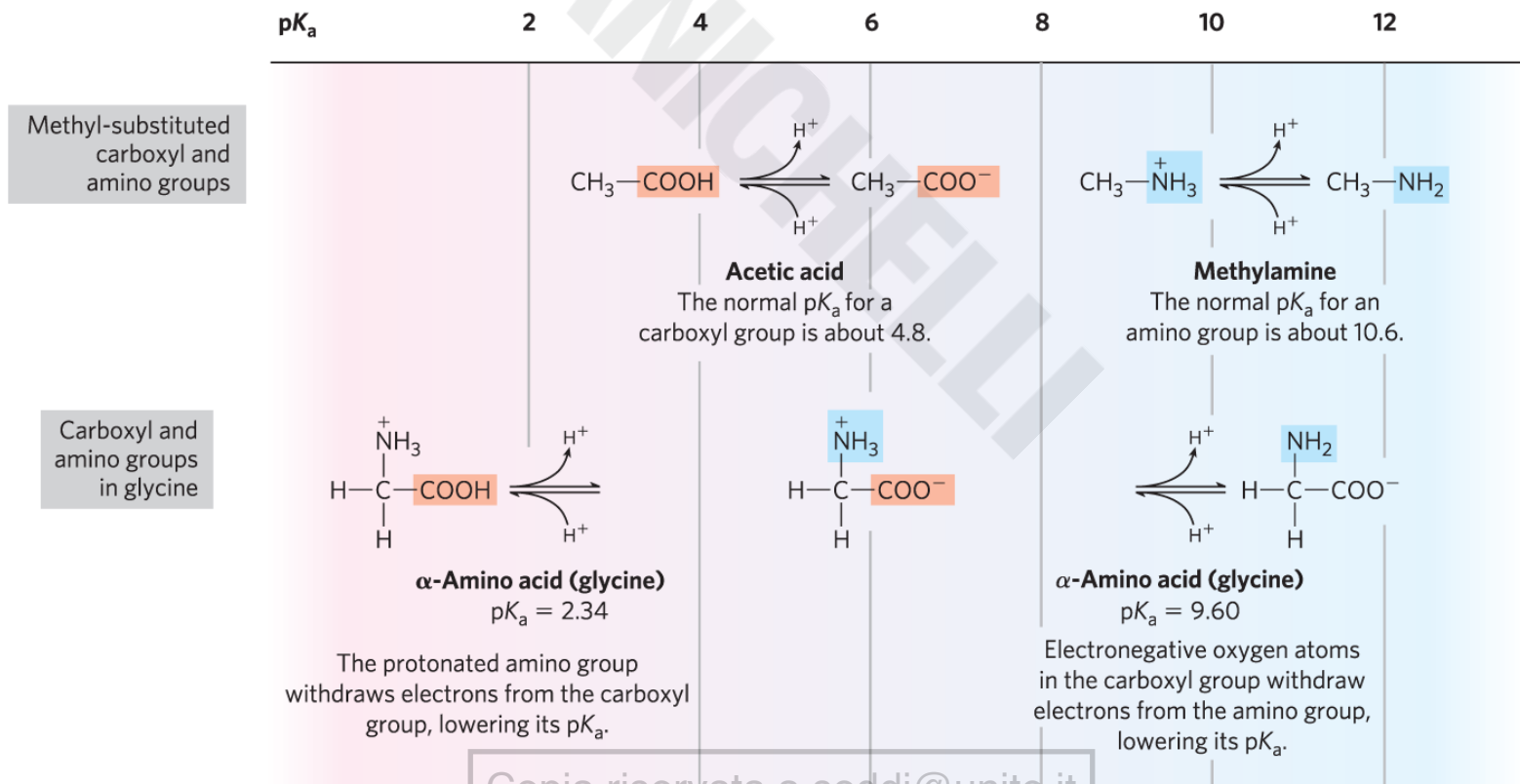
cation \rightleftharpoons zwitterion \rightleftharpoons anion



- $-\text{COOH}$ has an acidic pK_a (pK_1)
- $-\text{NH}_3^+$ has a basic pK_a (pK_2)
- the pH at which the net electric charge is zero is the **isoelectric point (pI)**

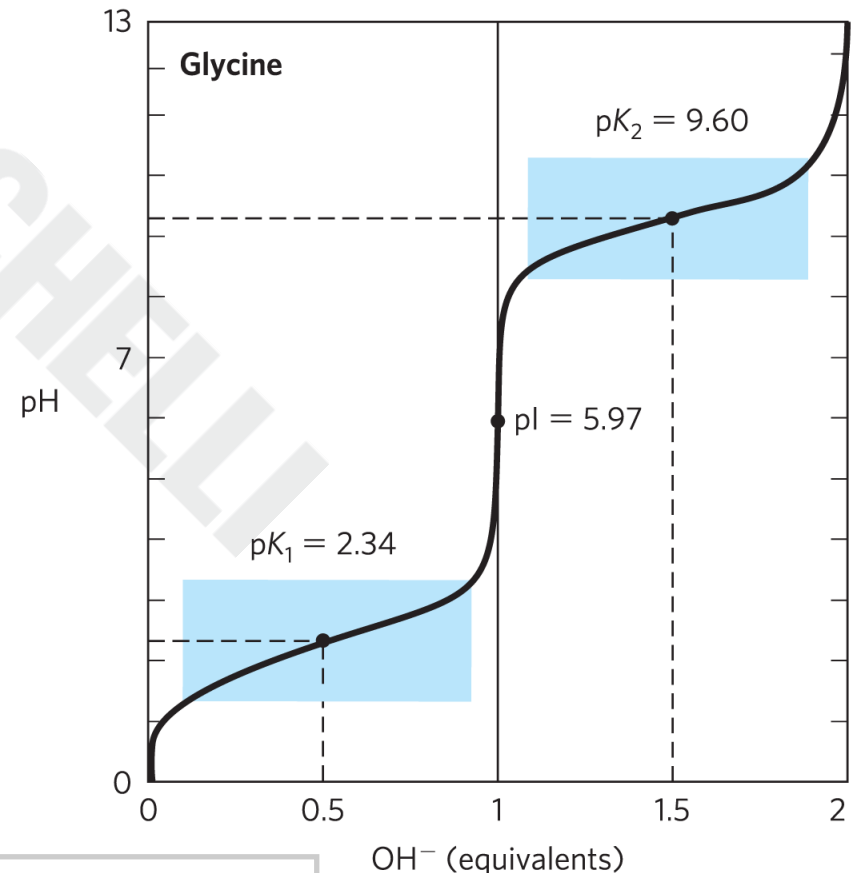
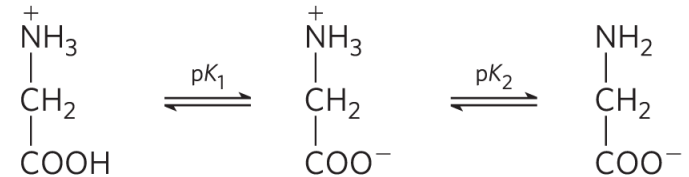
Effect of the Chemical Environment on pK_a

- α -carboxyl group is more acidic than in carboxylic acids
- α -amino group is less basic than in amines



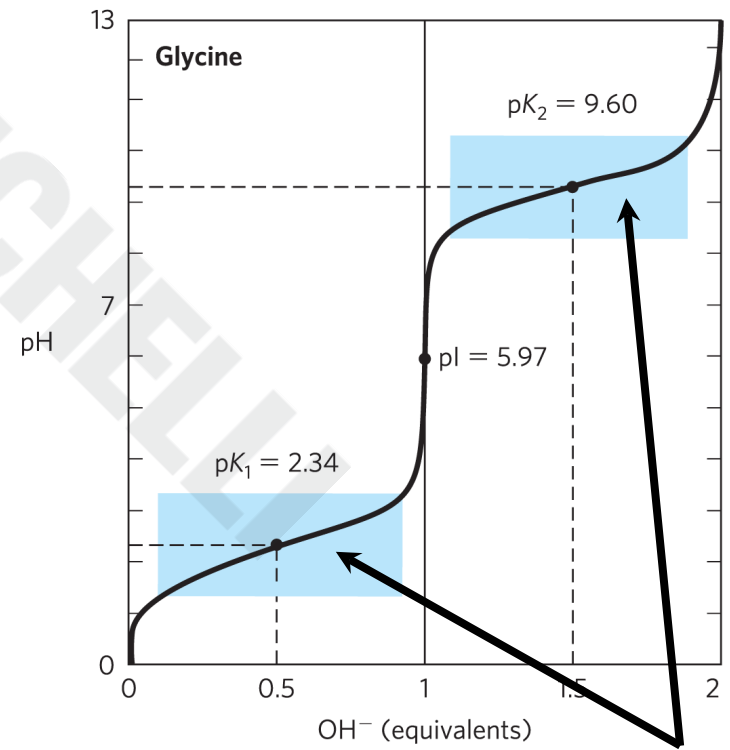
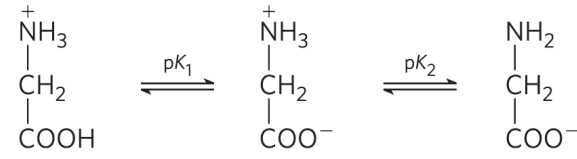
Information from a Titration Curve

- quantitative measure of the pK_a of each ionizing group
- regions of buffering power
- relationship between its net charge and the pH of the solution
 - **isoelectric point, pI**



Amino Acids as Buffers

- buffers prevent changes in pH close to the pK_a
- glycine has two buffer regions:
 - centered around the pK_a of the α -carboxyl group ($pK_1 = 2.34$)
 - centered around the pK_a of the α -amino group ($pK_2 = 9.6$)



Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company

Isoelectric Point, pI

- for amino acids without ionizable side chains, the **isoelectric point (pI)** is:

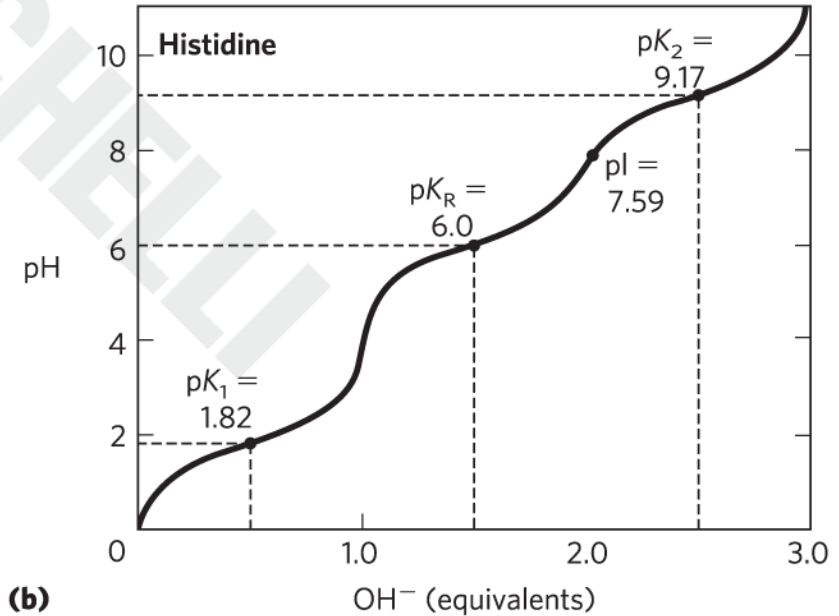
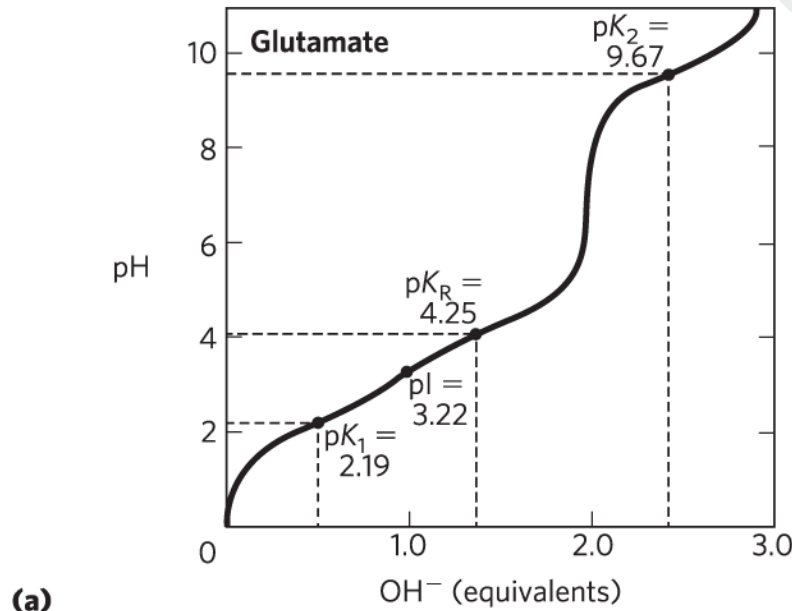
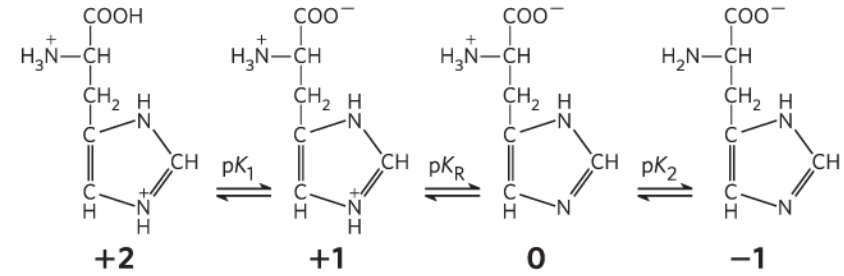
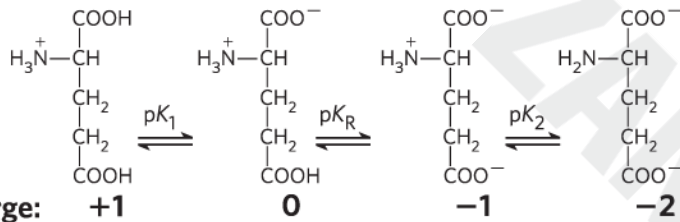
$$pI = \frac{pK_1 + pK_2}{2}$$

- pH = pI = net charge is zero (amino acid least soluble in water, does not migrate in electric field)
- pH > pI = net negative charge
- pH < pI = net positive charge

Amino Acids Differ in Their Acid-Base Properties

- ionizable side chains:
 - have a pK_a value
 - act as buffers
 - influence the pI of the amino acid
 - can be titrated (titration curve has 3 ionization steps)

Titration of Amino Acids with an Ionizable R Group



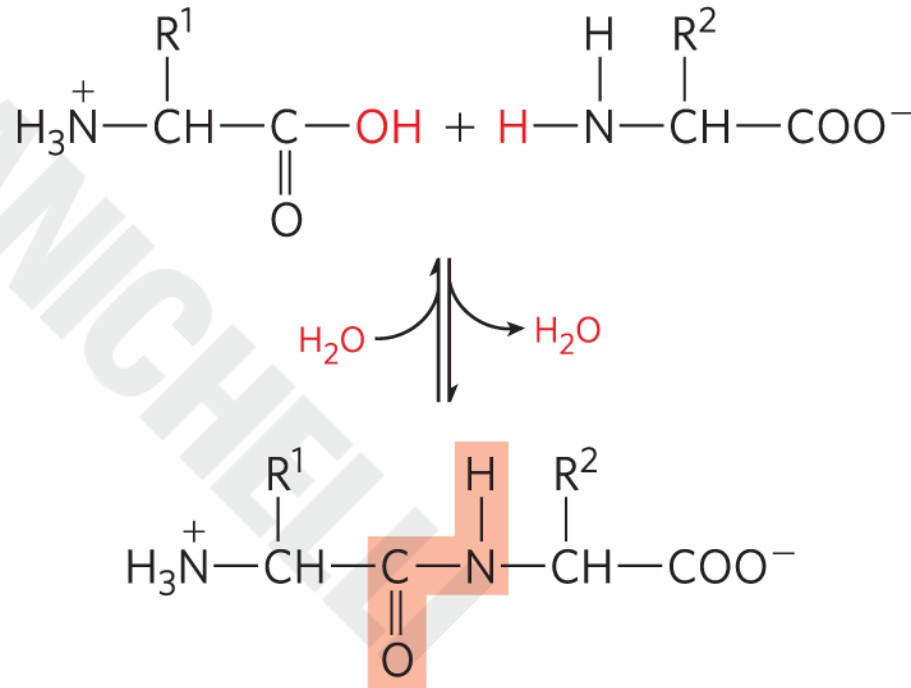
3.2 Peptides and Proteins

Principle 2 (2 of 3)

In proteins, amino acids are joined in characteristic linear sequences through a common amide linkage, the peptide bond. The amino acid sequence of a protein constitutes its primary structure, a first level we will introduce within the broader complexities of protein structure.

Peptides Are Chains of Amino Acids

- **peptide bond:**
 - covalent
 - formed through **condensation**
 - broken through **hydrolysis**



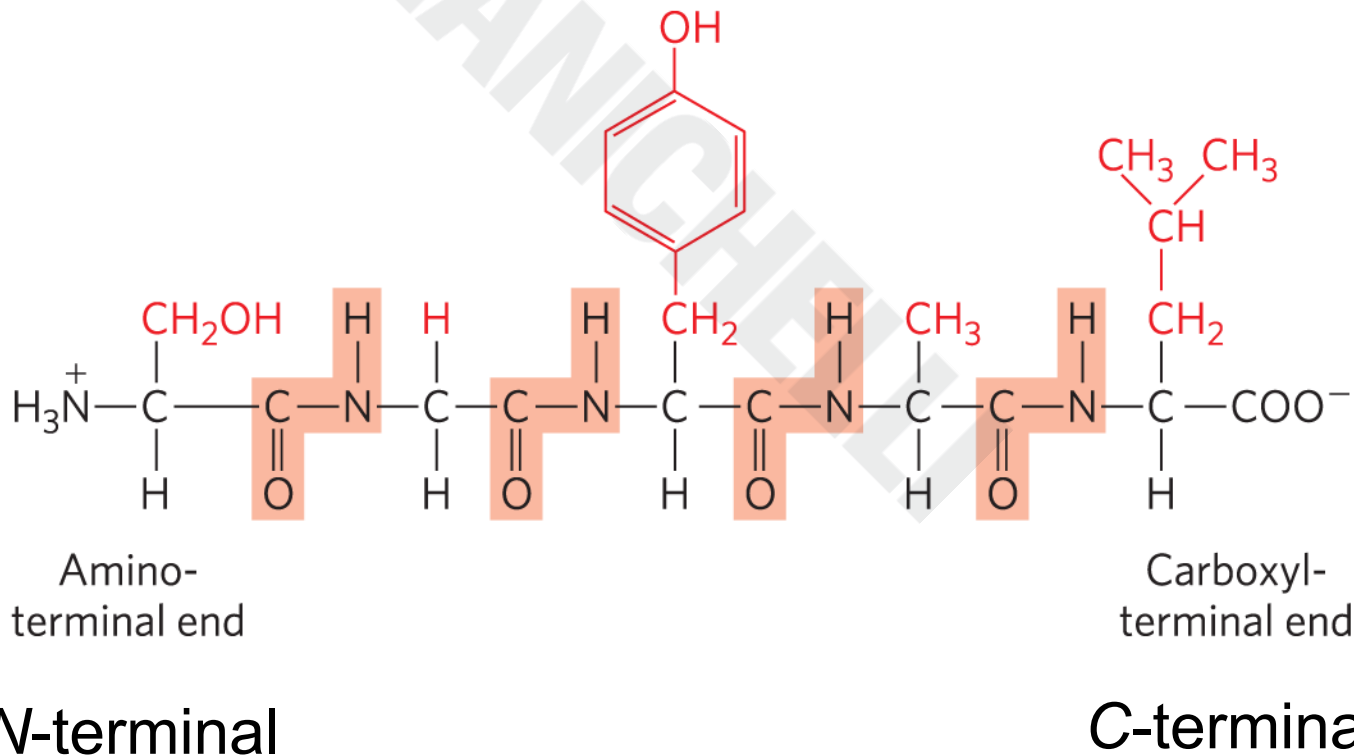
Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021
W. H. Freeman and Company

Peptide Types by the Number

- dipeptide = 2 amino acids, 1 peptide bond
- tripeptide = 3 amino acids, 2 peptide bonds
- **oligopeptide** = a few amino acids
- **polypeptide** = many amino acids, molecular weight < 10 kDa
- **protein** = thousands of amino acids, molecular weight > 10 kDa

Peptide Terminals

numbering (and naming) starts from the **amino-terminal residue** (*N*-terminal)



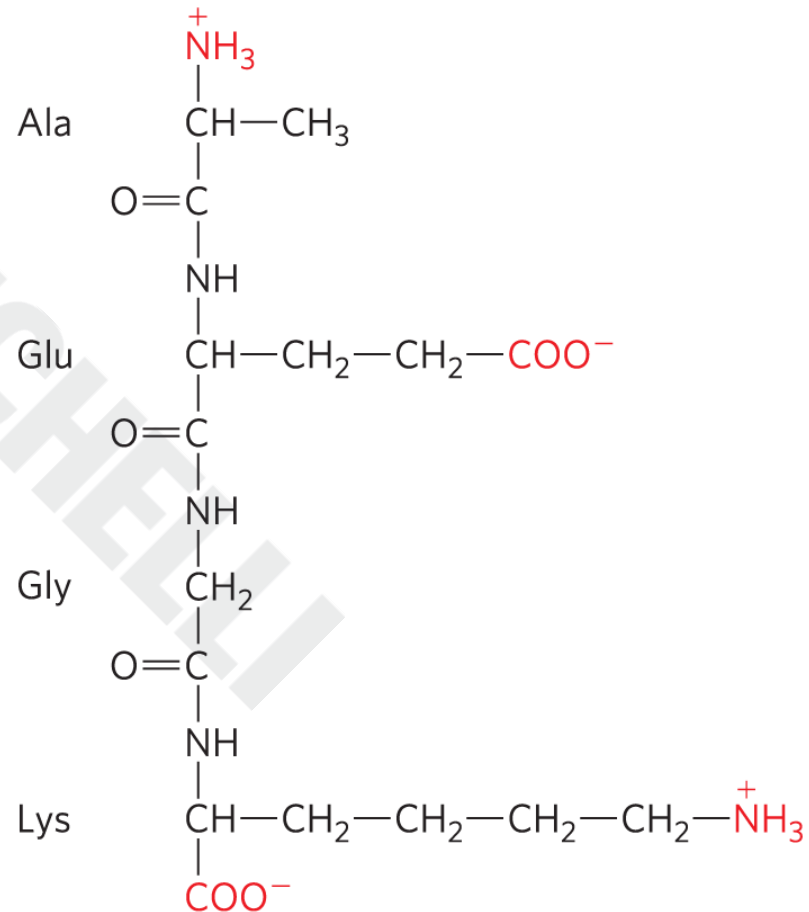
Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021
W. H. Freeman and Company

Naming Peptides

- full amino acid names:
serylglycyltyrosylalanylleucine
- three-letter code abbreviations:
Ser–Gly–Tyr–Ala–Leu
- one-letter code abbreviation:
SGYAL

Peptides Can Be Distinguished by Their Ionization Behavior

- ionizable groups in peptides:
 - one free α -amino group
 - one free α -carboxyl group
 - some R groups



Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company

Biologically Active Peptides and Polypeptides Occur in a Vast Range of Sizes and Compositions

- length of naturally occurring peptides = 2 to many thousands of amino acid residues

Table 3-2 Molecular Data on Some Proteins

Protein	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome c (human)	12,400	104	1
Myoglobin (equine heart)	16,700	153	1
Chymotrypsin (bovine pancreas)	25,200	241	3
Hemoglobin (human)	64,500	574	4
Hexokinase (yeast)	107,900	972	2
RNS polymerase (<i>E. coli</i>)	450,000	4,158	5
Glutamine synthetase (<i>E. coli</i>)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

Peptide Subunits

- **multisubunit** protein = 2+ polypeptides associated noncovalently
- **oligomeric** protein = at least 2 identical subunits
 - identical units = **protomers**

Amino Acid Composition of Proteins

- amino acid composition is highly variable

Table 3-3 Amino Acid Composition of Two Proteins

Amino Acid	Bovine cytochrome c: Number of residues per molecule	Bovine cytochrome c: percentage of total	Bovine chymotrypsinogen: Number of residues per molecule	Bovine chymotrypsinogen: Percentage of total
Ala	6	6	22	9
Arg	2	2	4	1.6
Asn	5	5	14	5.7
Asp	3	3	9	3.7
Cys	2	2	10	4
Gln	3	3	10	4
Glu	9	9	5	2
Gly	14	13	23	9.4
His	3	3	2	0.8
Ile	6	6	10	4
Leu	6	6	19	7.8
Lys	18	17	14	5.7
Met	2	2	2	0.8
Phe	4	4	6	2.4
Pro	4	4	9	3.7
Ser	1	1	28	11.4
Thr	8	8	23	9.4
Trp	1	1	8	3.3
Tyr	4	4	4	1.6
Val	3	3	23	9.4
Total	104	102	245	99.7

Estimating the Number of Amino Acid Residues

- number of residues = molecular weight/110
- average molecular weight of amino acid = ~128
- molecule of water removed to form peptide bond = 18

$$128 - 18 = 110$$

Some Proteins Contain Chemical Groups Other Than Amino Acids

- **conjugated** proteins = contain permanently associated chemical components
 - non–amino acid part = **prosthetic group**
- **lipoproteins** contain lipids
- **glycoproteins** contain sugars
- **metalloproteins** contain specific metals

Class	Prosthetic group	Example
Lipoproteins	Lipids	β_1 -Lipoprotein of blood (Fig. 17-2)
Glycoproteins	Carbohydrates	Immunoglobulin G (Fig. 5-20)
Phosphoproteins	Phosphate groups	Glycogen phosphorylase (Fig. 6-39)
Hemoproteins	Heme (iron porphyrin)	Hemoglobin (Figs 5-8 to 5-11)
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase (Fig. 19-9)
Metalloproteins	Iron Zinc Calcium Molybdenum Copper	Ferritin (Box 16-1) Alcohol dehydrogenase (Fig. 14-12) Calmodulin (Fig. 12-17) Dinitrogenase (Fig. 22-3) Complex IV (Fig. 19-12)

3.3 Working with Proteins

Principle 3 (2 of 2)

For study, individual proteins can be separated from the thousands of other proteins present in a cell, based on differences in their chemical and functional properties arising from their distinct amino acid sequences. As proteins are central to biochemistry, the purification of individual proteins for study is a quintessential biochemical endeavor.

Proteins Can Be Separated and Purified

- separated based on:
 - size
 - charge
 - binding properties
 - protein solubility

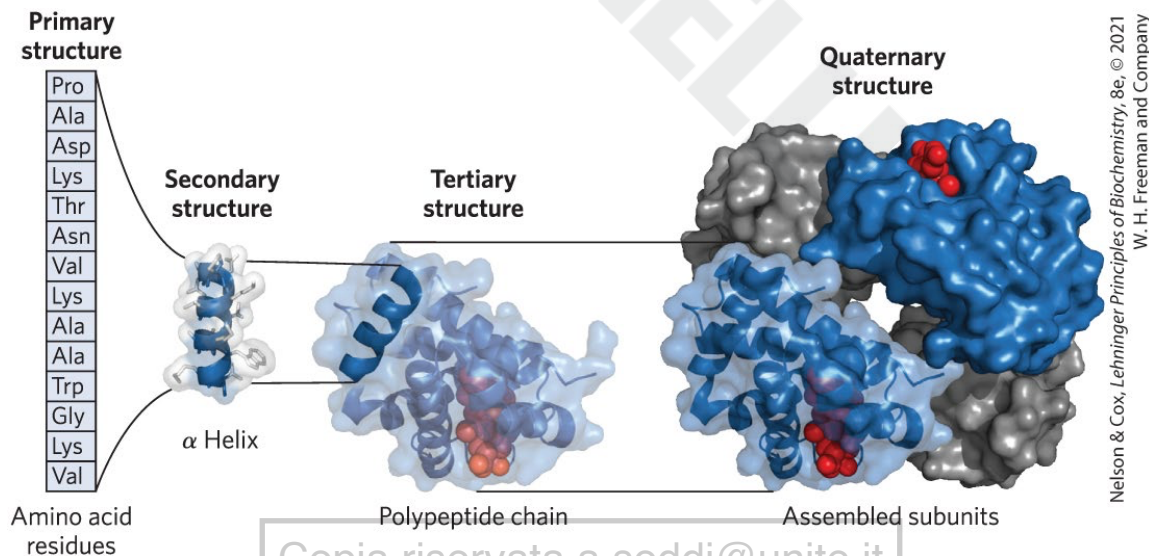
3.4 The Structure of Proteins: Primary Structure

Principle 2 (3 of 3)

In proteins, amino acids are joined in characteristic linear sequences through a common amide linkage, the peptide bond. The amino acid sequence of a protein constitutes its primary structure, a first level we will introduce within the broader complexities of protein structure.

Levels of Structure in Proteins

- four levels:
 - **primary structure** = covalent bonds linking amino acid residues in a polypeptide chain
 - **secondary structure** = recurring structural patterns
 - **tertiary structure** = 3D folding of polypeptide
 - **quaternary structure** = 2+ polypeptide subunits



Principle 4 (2 of 2)

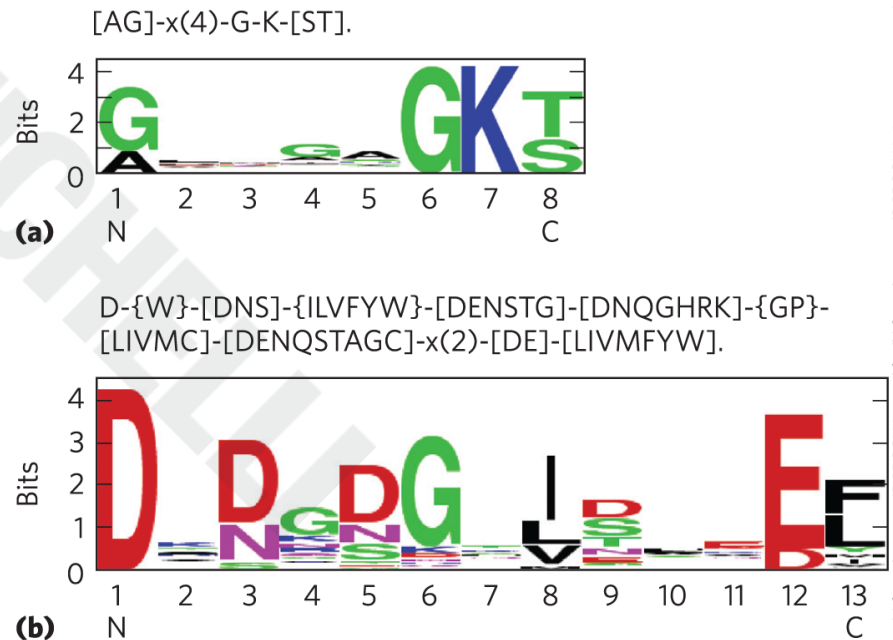
Shaped by evolution, amino acid sequences are a key resource for understanding the function of individual proteins and for tracing broader functional and evolutionary relationships.

The Function of a Protein Depends on Its Amino Acid Sequence

- amino acid sequence confers 3D structure
- 3D structure confers function
- most human proteins = **polymorphic** = have amino acid sequence variants
- **Edman degradation** = classic method of sequencing amino acids

Amino Acid Sequences Provide Important Biochemical Information

- amino acid sequence can inform:
 - 3D structure
 - function
 - cellular location
 - evolution
- **consensus sequence** = reflects most common amino acid at each position



Protein Sequences Help Elucidate the History of Life on Earth

- **bioinformatics:**
 - identifies functional segments in new proteins
 - establishes sequence and structural relationships to known proteins
- essential amino acid residues = conserved over evolutionary time
- less important amino acid residues = vary over evolutionary time

Transfer of Genes from Organism to Organism

- **horizontal gene transfer** = transfer of a gene or group of genes from one organism to another
 - proteins derived from transferred genes are not good candidates for bacterial evolution studies
 - for example, rapid spread of antibiotic-resistance genes in bacterial populations

Defining Members of Protein Families

- **homologs = homologous proteins = members of protein families**
 - **paralogs** = homologs in same species
 - **orthologs** = homologs in different species
 - identified by comparing protein sequences to a database of protein sequences

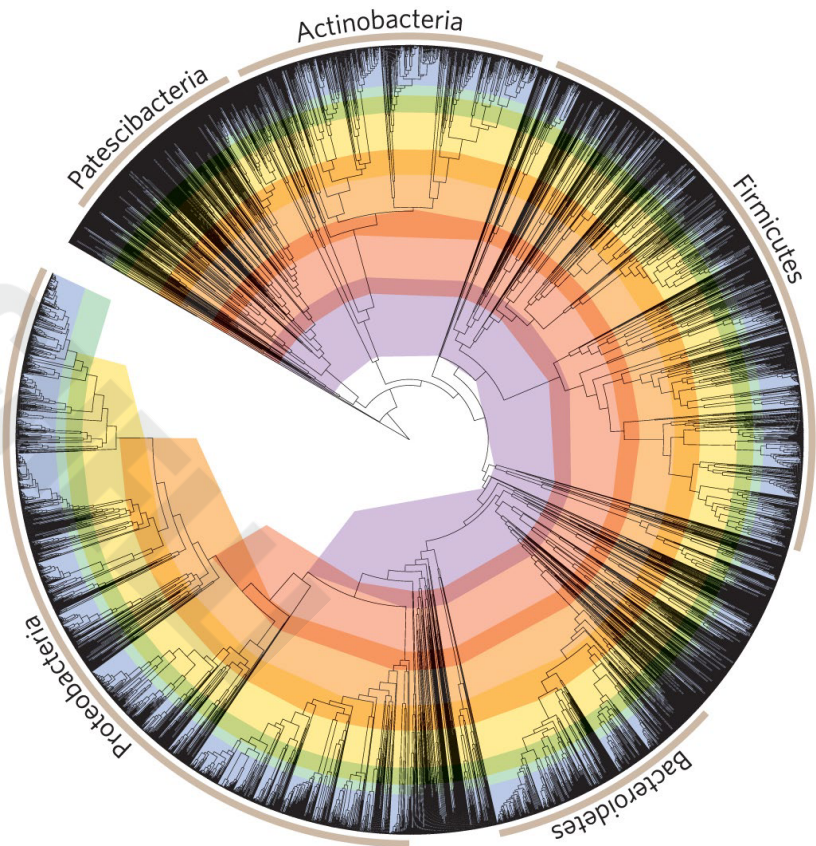
Escherichia coli T GNR T I AV YDLGGGTFD I S I I E IDE VDGE KT F E VLA T N G D T H L G G E D F D S R L I N Y L
Bacillus subtilis DE D Q T I L L YDLGGGTFD V S I L E L G D G V F E V R S T A G D N R L G G D D F D Q V I D H L

└───┘
Gap

Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company

Constructing Evolutionary Trees

- segregate organisms into classes based on sequence divergence in protein families
- **signature sequences** = certain protein segments specific to a taxonomic group



Nelson & Cox, *Lehninger Principles of Biochemistry*, 8e, © 2021 W. H. Freeman and Company