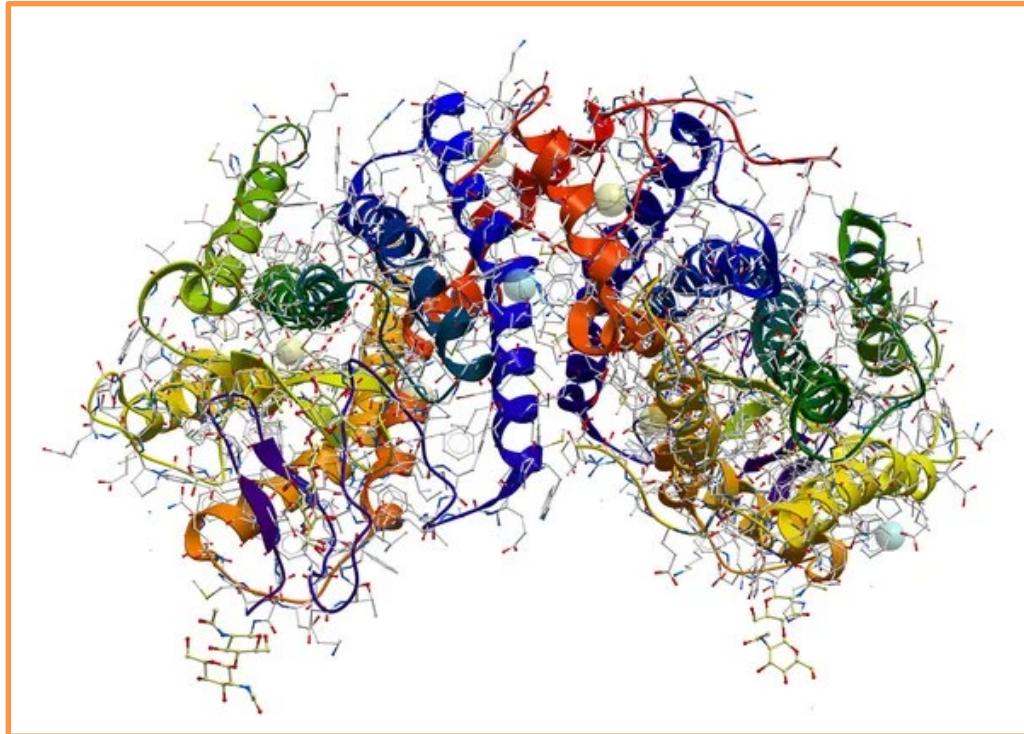
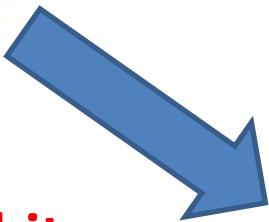
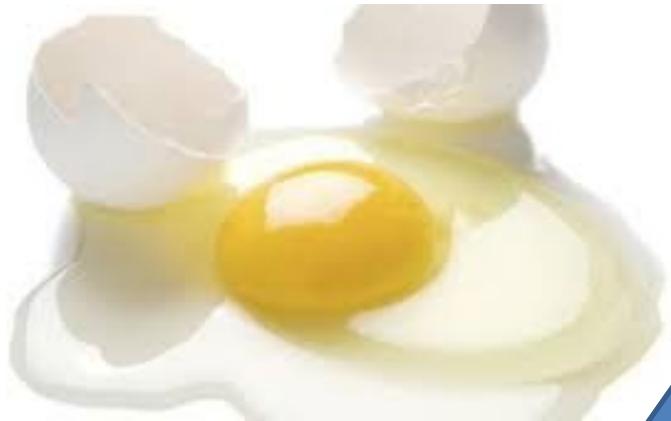


Lecture 4: Structures and Functions of Proteins I



Kwok-On LAI
Department of Neuroscience



**Why egg white turns white
when it is cooked?**



Learning outcomes

- Define the structure and properties of amino acids
- Recognize the principles underlying how proteins fold
- Distinguish between primary, secondary, tertiary and quaternary structures of proteins
- Describe the relationship between structures and functions of selected proteins
- Recognize the link between misfolded protein and disease

What is amino acid?

General structure of amino acid

- Carboxyl group
- Amino group
- R group

- α carbon ($C\alpha$)
(bond to four different groups)

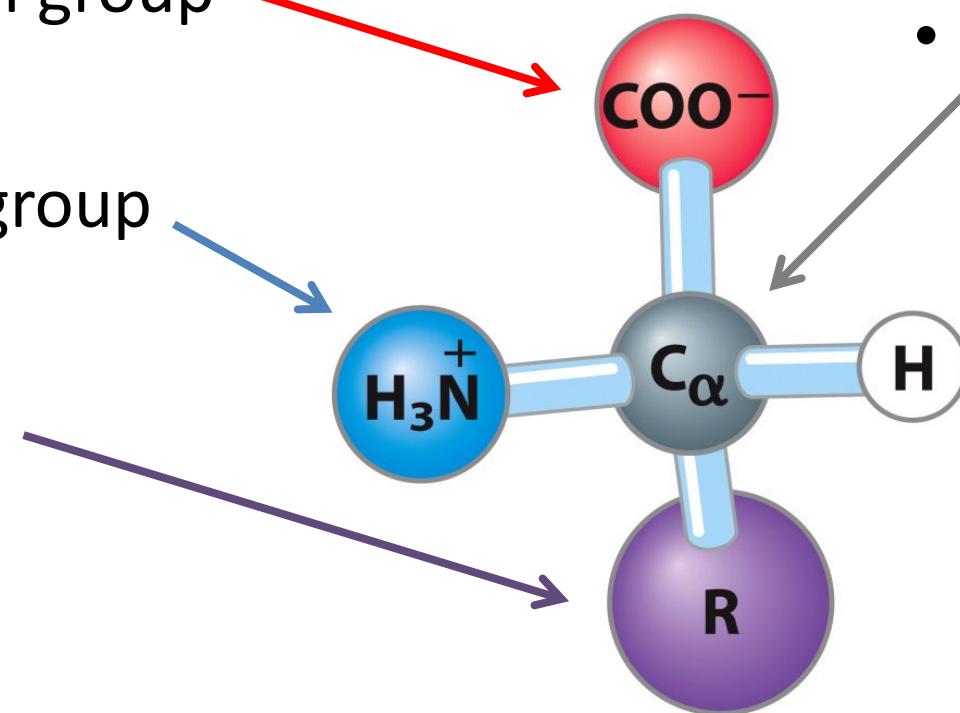
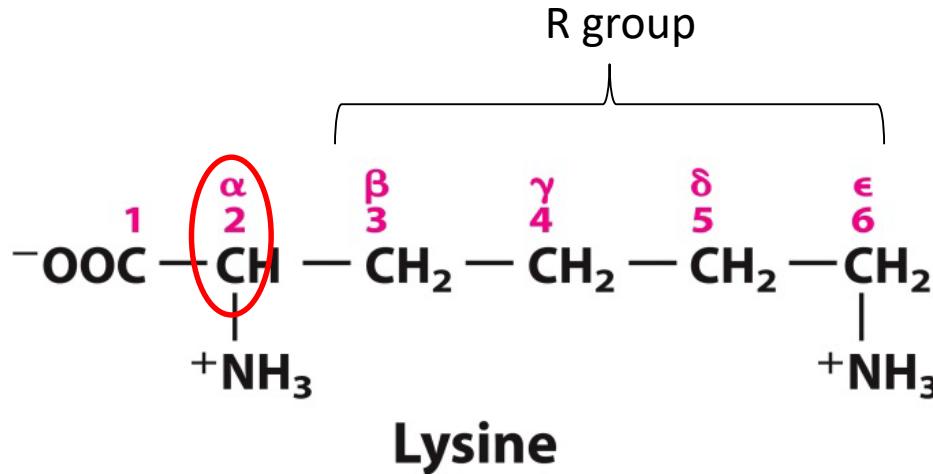


Figure 3-2
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

- There are twenty common amino acids

Structure of amino acid



- Labelling of carbon atom: Carboxyl carbon can be labeled as C-1, C α as C-2, then number along the R chain

C α as a chiral center

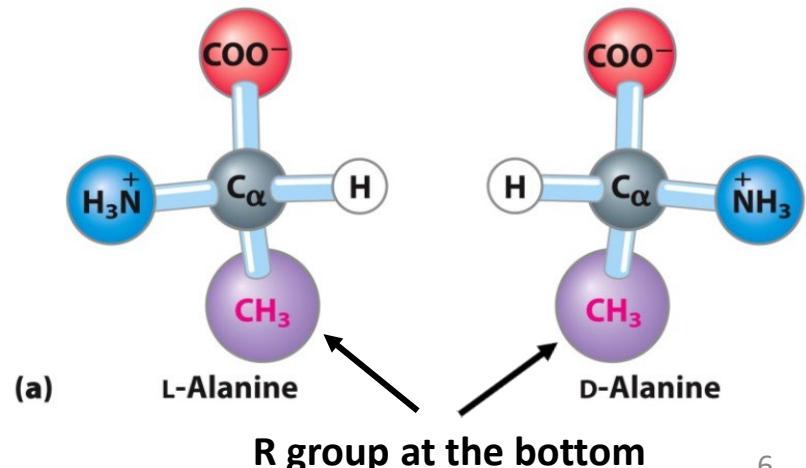
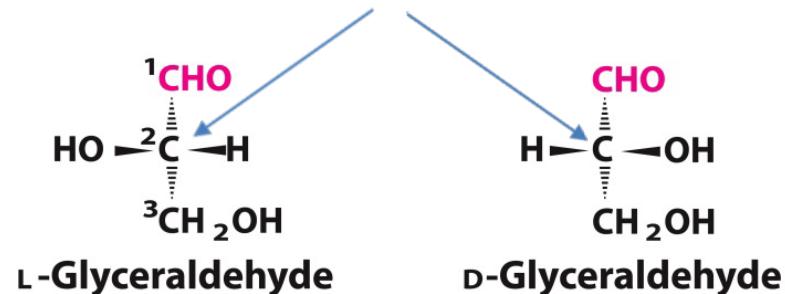
Structure of amino acid

- Two mirror configurations of amino acid: “L-” and “D-”

What's the difference?

- All amino acids in proteins are L-stereoisomers

C_α as a chiral center



List of amino acids



“Pain is inevitable.
Suffering is optional.”

— Haruki Murakami,
What I Talk About When I
Talk About Running

TABLE 3–1 Properties and Conventions Associated with the Common Amino Acids Found in Proteins

Amino acid	Abbreviation/ symbol	M_r^*	pK_a values				Hydropathy index [†]	Occurrence in proteins (%) [‡]	
			pK_1 (—COOH)	pK_2 (—NH ₃ ⁺)	pK_R (R group)	pI			
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	
Proline	Pro P	115	1.99	10.96		6.48	-1.6	5.2	
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	
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	
	Gln Q	146	2.17	9.13		5.65	-3.5	4.2	
Lys K		146	2.18	8.95	10.53	9.74	-3.9	5.9	
His H		155	1.82	9.17	6.00	7.59	-3.2	2.3	
Arg R		174	2.17	9.04	12.48	10.76	-4.5	5.1	
Asp D		133	1.88	9.60	3.65	2.77	-3.5	5.3	
Glu E		147	2.19	9.67	4.25	3.22	-3.5	6.3	

ures as shown in Figure 3–5. The elements of water (M_r , 18) are deleted when the amino acid is incorporated into a polypeptide. hydrophobicity and hydrophilicity of R groups. The values reflect the free energy (ΔG) of transfer of the amino acid side chain from a polar solvent to a nonpolar solvent. This transfer is favorable ($\Delta G < 0$; negative value in the index) for charged or polar amino acid side chains, and unfavorable ($\Delta G > 0$) for amino acids with nonpolar or more hydrophobic side chains. See Chapter 11. From Kyte, J. & Doolittle, R.F. (1982) A simple method for displaying hydrophobic character of a protein. *J. Mol. Biol.* 157, 105–132.

than 1,150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Molecular Recognition of Proteins* (Fasman, G.D., ed.), pp. 599–623, Plenum Press, New York.

acted as polar despite having a positive hydropathy index. This reflects the ability of the sulphydryl group to act as a weak acid and to form covalent bonds with oxygen or nitrogen.

Table 3-1

Lehninger Principles of Biochemistry, Sixth Edition

© 2013 W. H. Freeman and Company

Classification of amino acids by R group

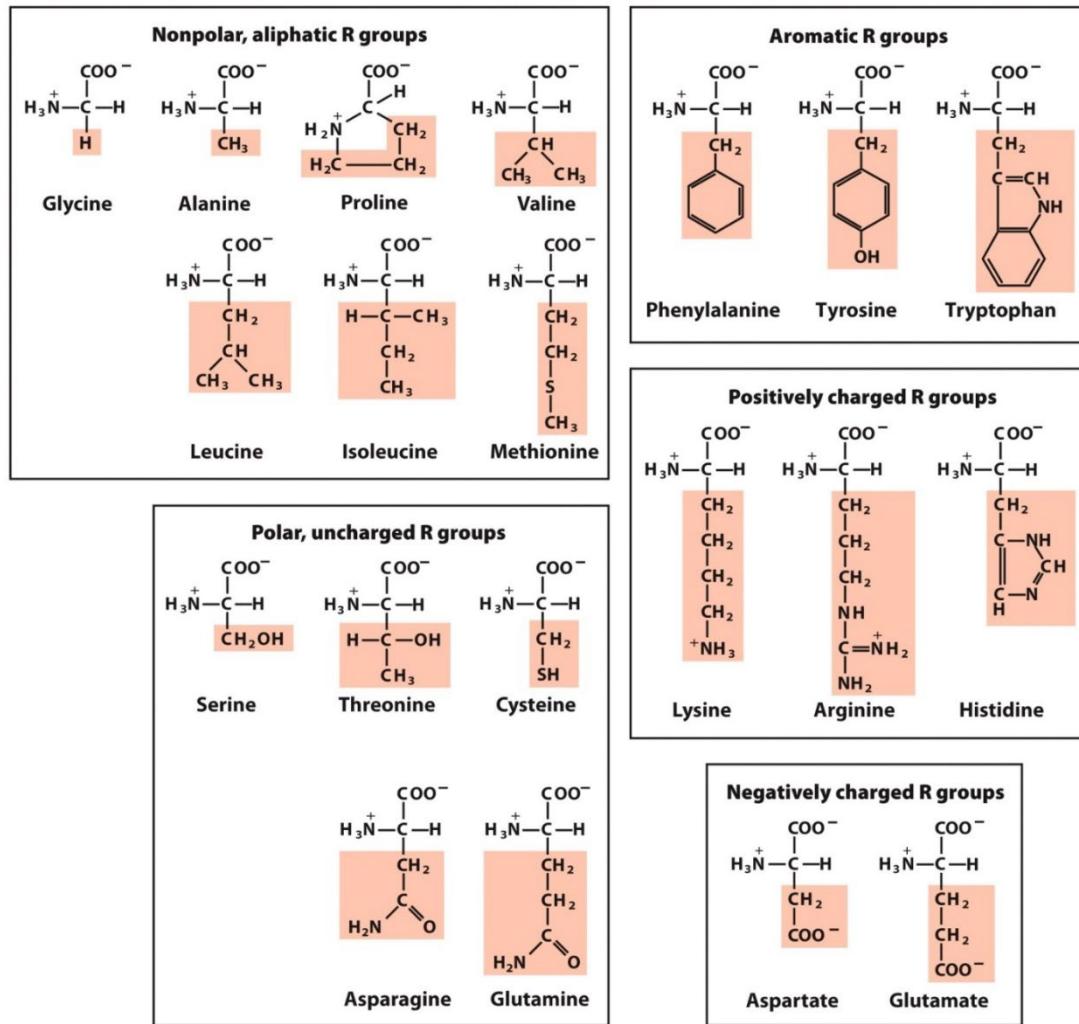
1. **Polarity** (tendency to interact with water at pH 7.0):
 - Nonpolar (hydrophobic)
 - Polar (hydrophilic)

Nonpolar amino acids tend to cluster together within proteins and stabilize proteins through hydrophobic interactions

2. Containing aromatic group (hydrophobic)

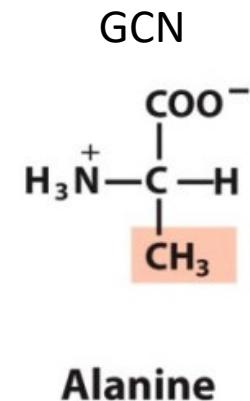
3. Ionizable (charged) at pH 7.0 (hydrophilic):
 - Positively charged
 - Negatively charged

Based on the properties of R groups (5 general types)



Amino acid mutation of R group

No mutation	Point mutations			Designed mutation for protein research	
	Silent	Nonsense	Missense		
			conservative	non-conservative	
DNA level	TTC	TTT	ATC	TCC	TGC
mRNA level	AAG	AAA	UAG	AGG	ACG
protein level	Lys	Lys	STOP	Arg	Thr
				 basic polar	
			No change		
			No change of charge		
			Change of charge		
			Commonly used loss-of-function mutation		



Changing just one amino acid can dramatically change the protein property or function



RESEARCH ARTICLE

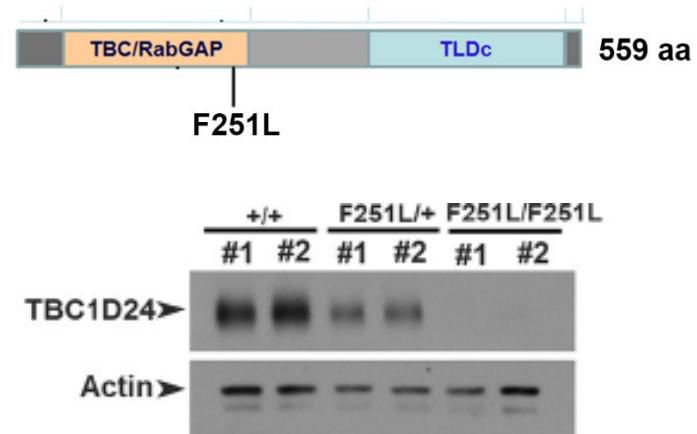
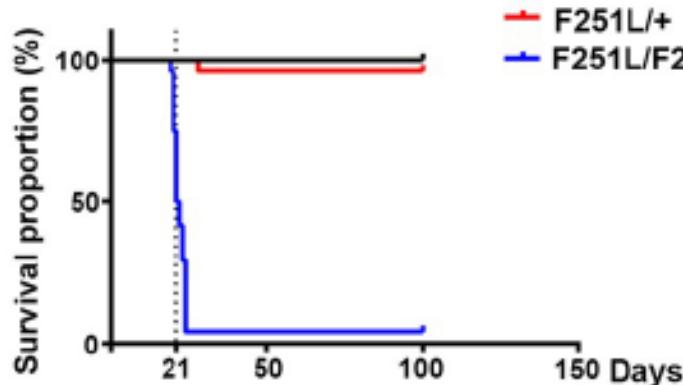
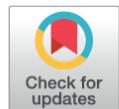
The epilepsy and intellectual disability-associated protein TBC1D24 regulates the maintenance of excitatory synapses and animal behaviors

Lianfeng Lin¹, Quanwei Lyu^{1*}, Pui-Yi Kwan¹, Junjun Zhao¹, Ruolin Fan¹, Anping Chai¹, Cora Sau Wan Lai^{1,2}, Ying-Shing Chan^{1,2}, Xuting Shen^{1*}, Kwok-On Lai^{1,2*}

1 School of Biomedical Sciences, The University of Hong Kong, Hong Kong, China, 2 State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China

* These authors contributed equally to this work.

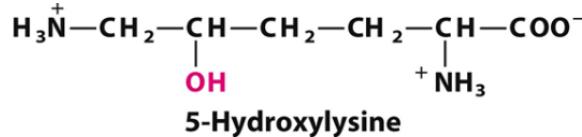
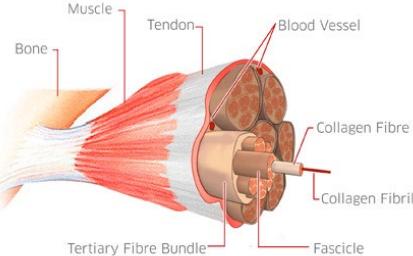
laiko@hku.hk



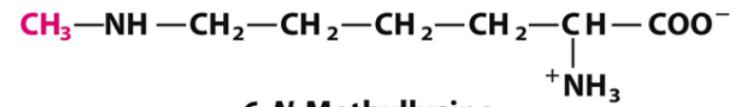
Uncommon amino acids that are present in proteins

(1) Constituents of special proteins

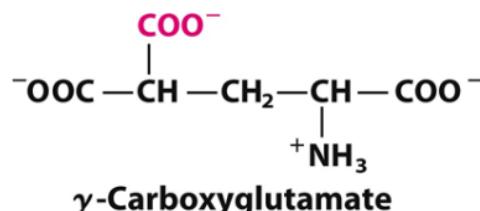
4-Hydroxyproline, 5-Hydroxylysine (in collagen)



6-N-Methyllysine (in myosin)

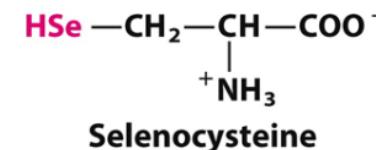


Carboxyglutamate (in prothrombin)



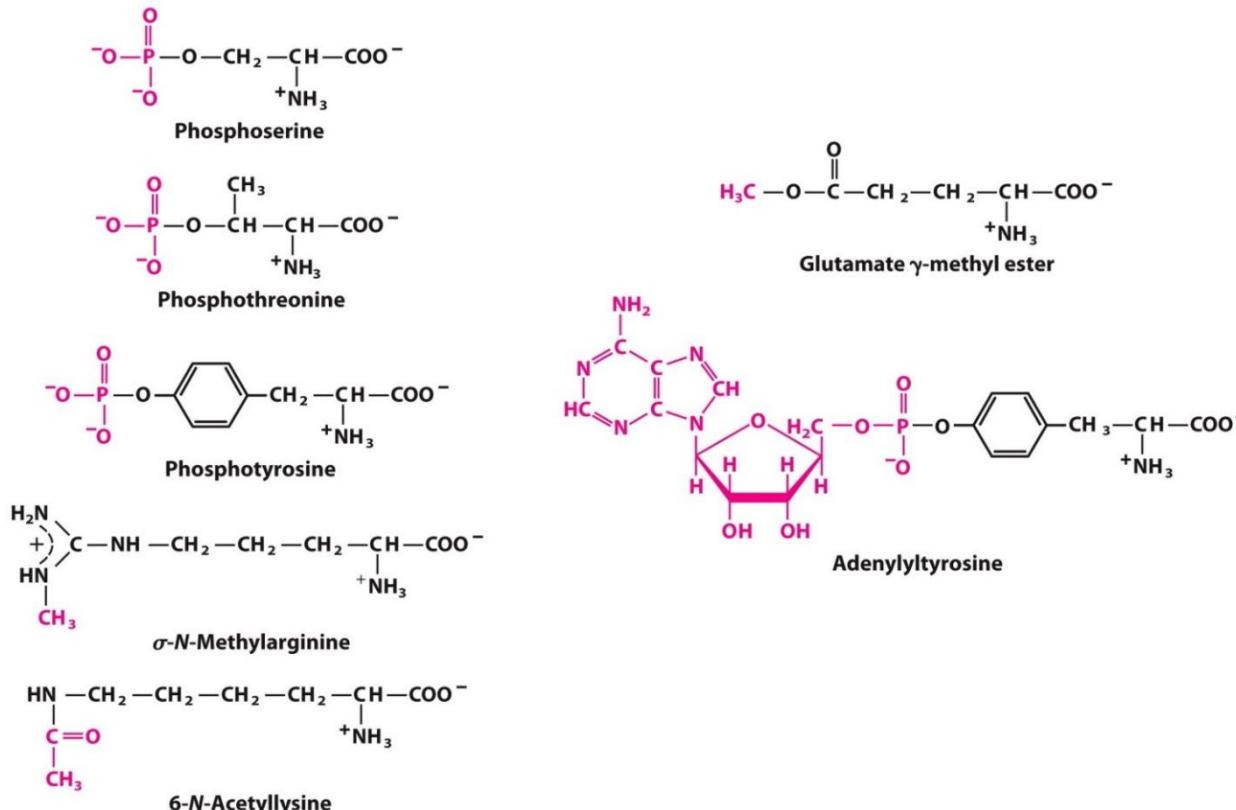
Selenocysteine

Modified before incorporation into a protein



(2) Transient, reversible modifications of amino acids in a protein; alter the protein's function in response to a stimulus (signal transduction)

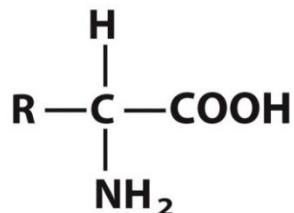
- e.g. Phosphoryl (PO_3^-), methyl (CH_3), acetyl (COCH_3), etc....



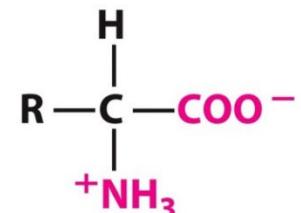
(3) Another 300 additional amino acids which have a variety of functions but are not all constituents of proteins

Acid-base properties of amino acids

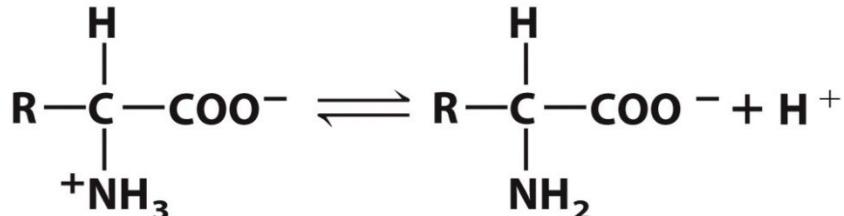
- Amino acids are typical zwitterion, a neutral molecule with both a positive and a negative electrical charge



Nonionic form

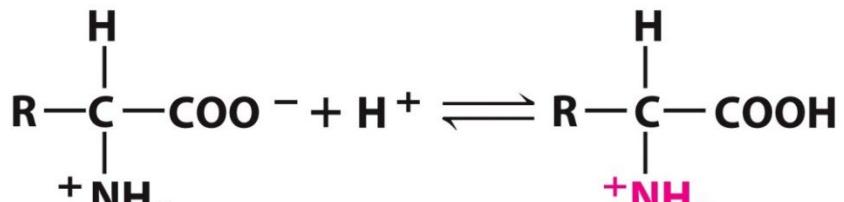


Zwitterionic form



Zwitterion
as acid

Anion



Zwitterion
as base

Cation

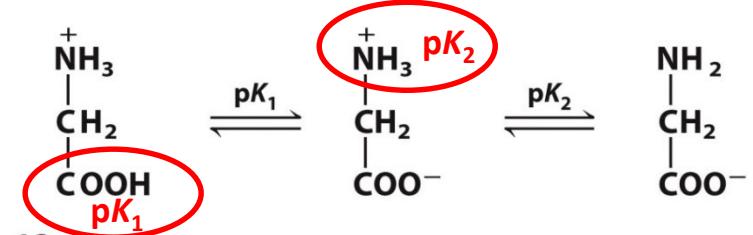
Figure 3-9

Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

Net change: Positive

- pK_a is equal to the pH at which half of the protons are titrated away by NaOH (the lower a pK_a , the stronger the acid)

Neutral

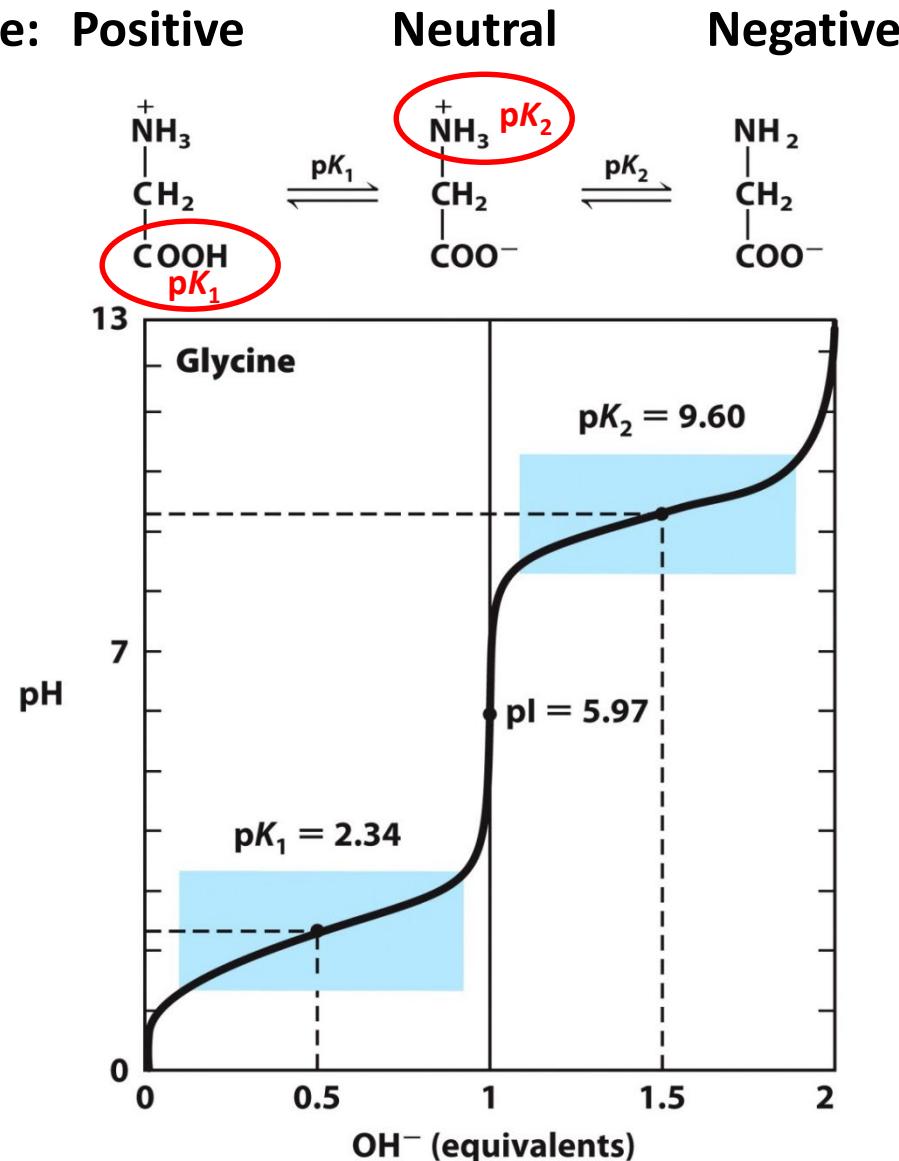


- For glycine, the simplest amino acid with no R group, the isoelectric point (pl) is:

$$pl = \frac{1}{2} (pK_1 + pK_2) = 5.97$$

- When environmental pH equals to pl, the net charge of the molecule is 0 (neutral)

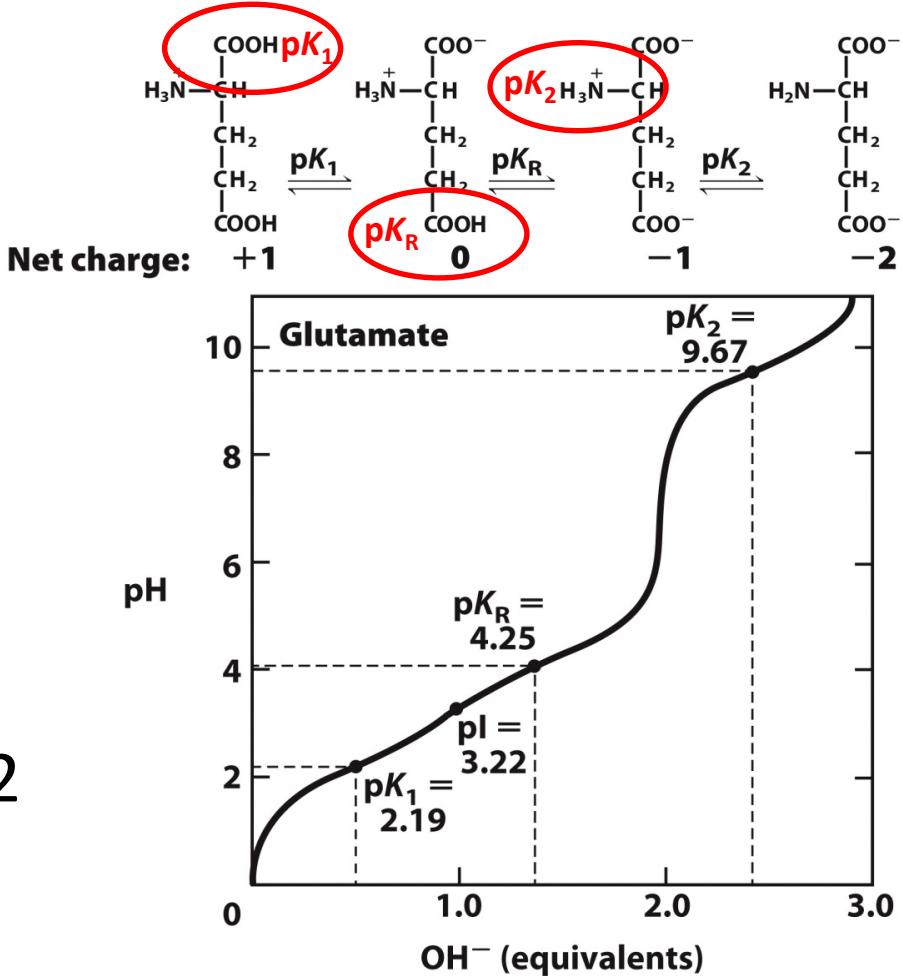
- The titration curves allows prediction of the charge of the amino acid at a given pH



Titration curve of glycine

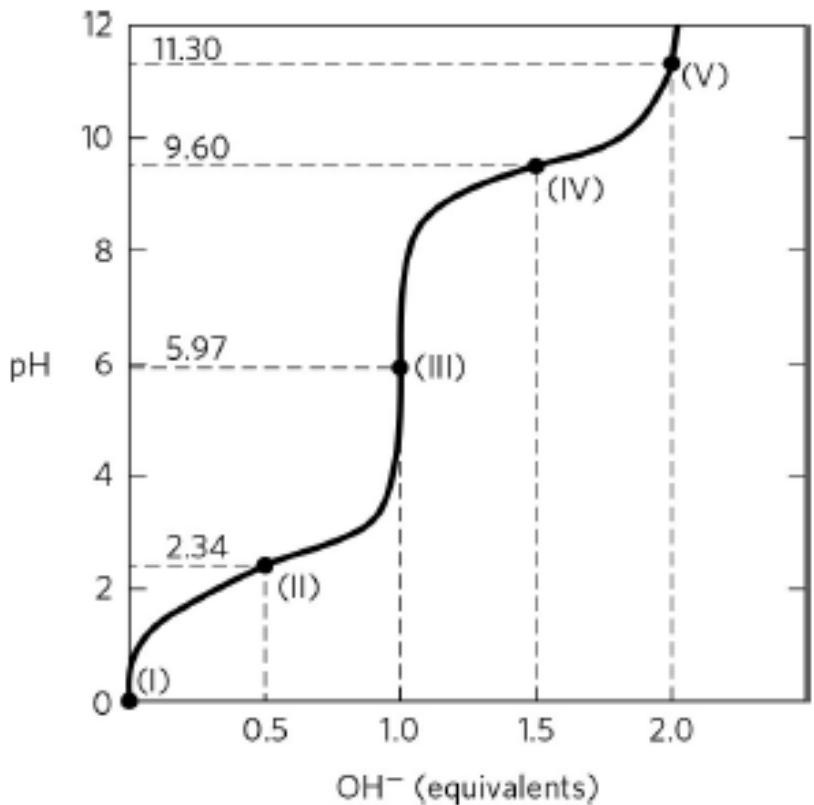
Titration curves of amino acids with ionizable R groups

- For amino acid with ionizable side chain, the R group also has pK_a
- The pI is the overall effects of pK_1 , pK_R and pK_2
- The pI of glutamate is 3.22 (lower than glycine)
What does it mean?



Titration curve of glutamate

Questions



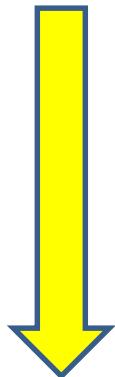
This is the titration curve of glycine.

- (1) At which point(s) the glycine has the least buffering power?
- (2) what would be the change in isoelectric point if the amino acid is lysine?

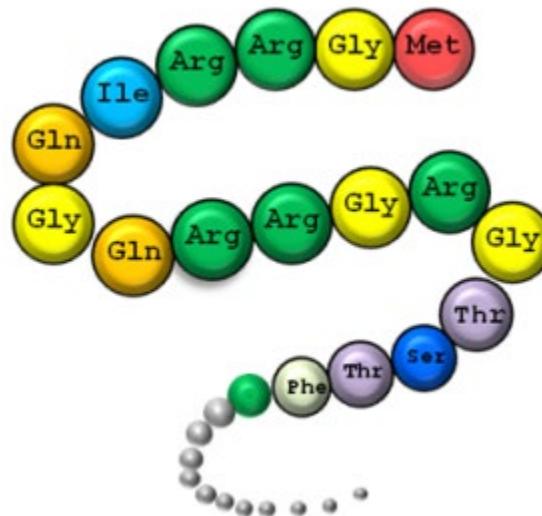
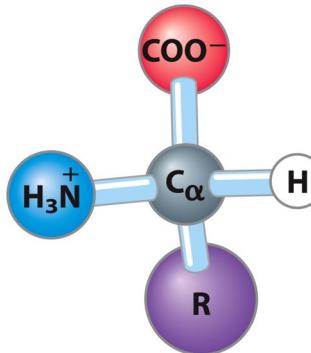
Summary I (amino acids)

- Structure of amino acids
- Twenty common amino acids (different R groups)
- Examples of uncommon amino acids (e.g. modifications to change protein function or activity)
- Acid-base properties (titration curves to predict the charges; different titration curves of different amino acids)

Amino acids

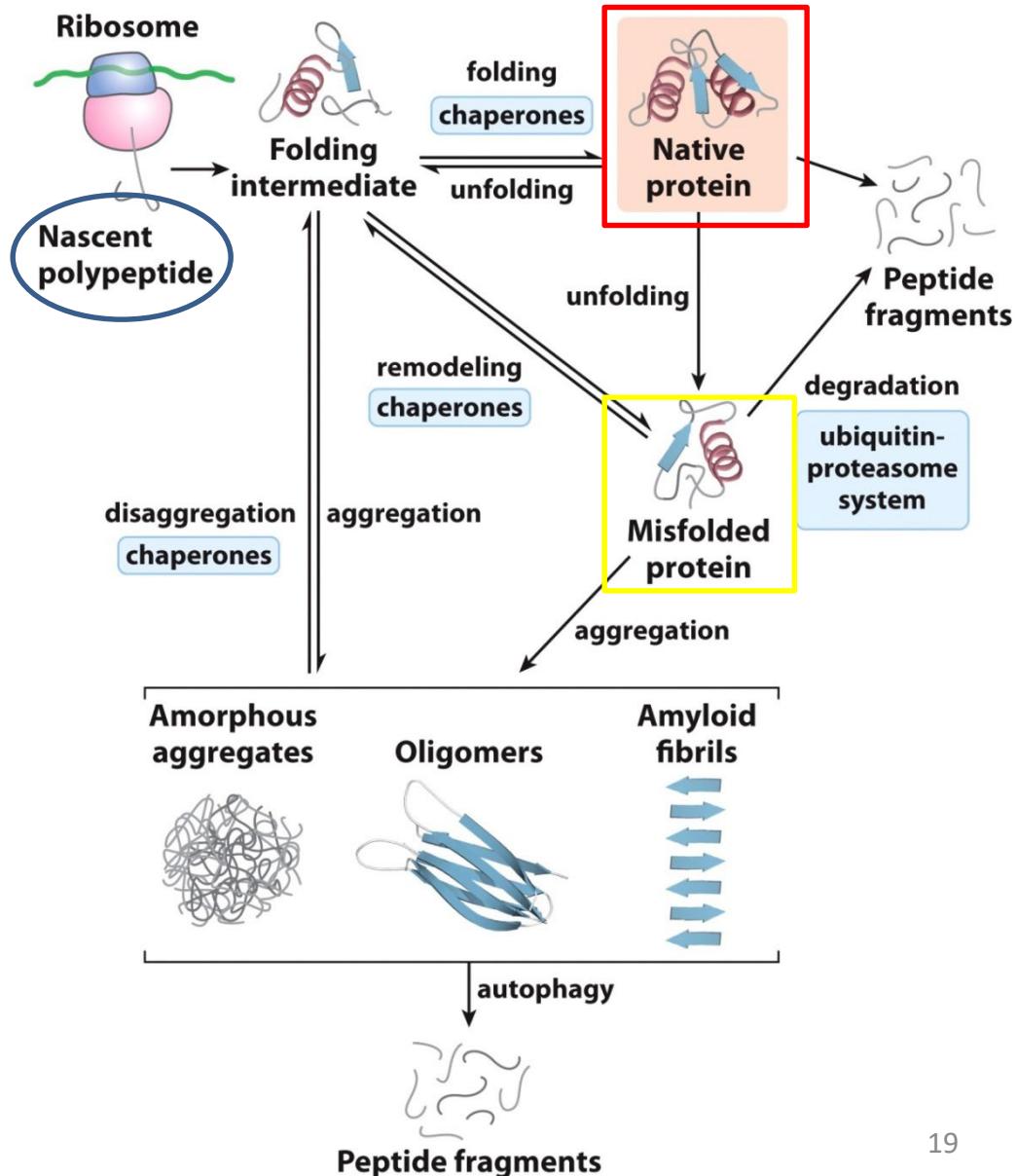


Proteins



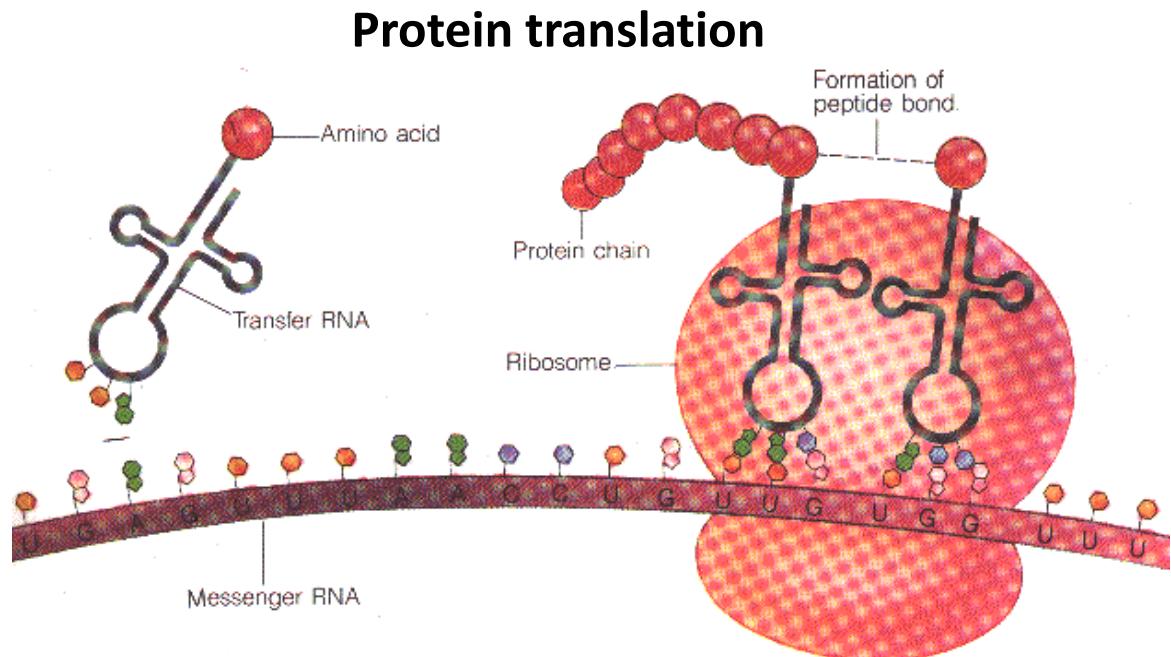
Proteostasis

- The continual maintenance of the active set of cellular proteins required under a given set of conditions is called **proteostasis**
- Requires coordinated functions of protein synthesis and folding, the refolding of misfolded proteins and the degradation of irreversibly unfolded proteins

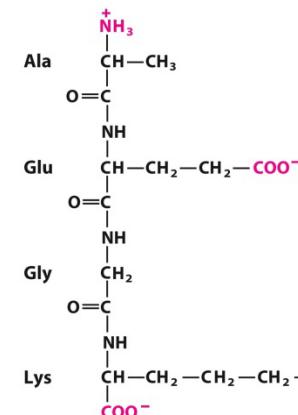
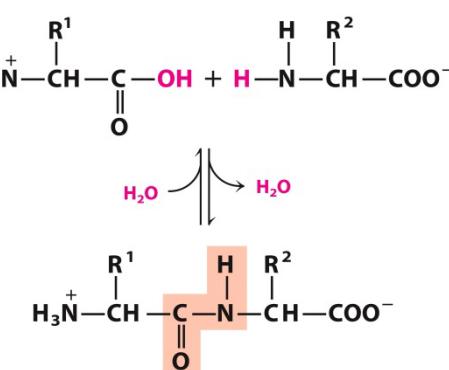


Protein synthesis

- Amino acids are joined together by peptide bond
- Catalyzed by ribosome (a large complex formed by rRNAs and ribosomal proteins)
- Dehydration between α -amino and α -carboxyl forms peptide bond



Peptide bond formation



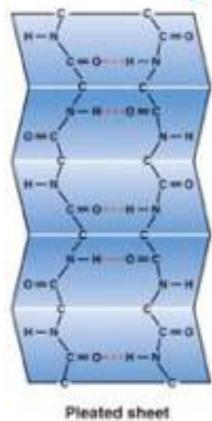
Structures of proteins

Primary sequence

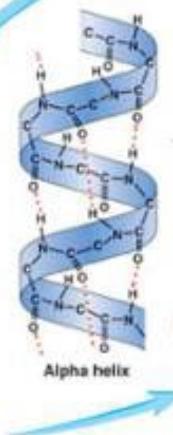
N terminal (N-ter, amino group) ...KLAGGC... Carboxyl terminal (C-ter)

Secondary structure

Local spatial arrangement
of main chain atoms
 α helix, β sheet



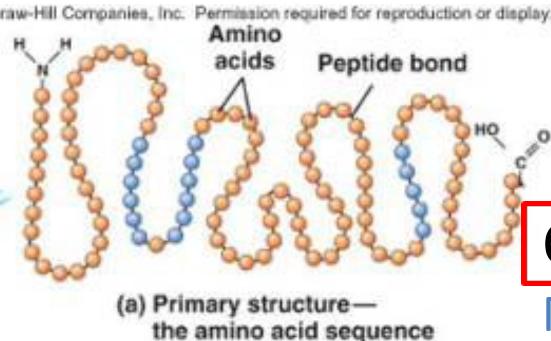
(b) Secondary structure
with folding as a result
of hydrogen bonding
(dotted red lines)



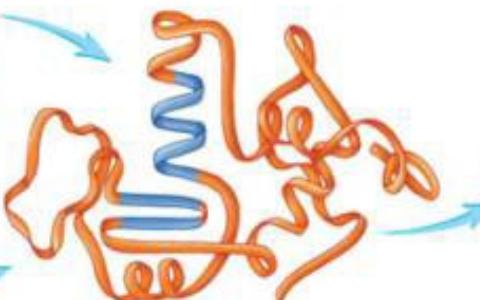
(c) Tertiary structure with
secondary folding caused by
interactions within the
polypeptide and its immediate
environment

Tertiary structure

3D structure of one polypeptide/protein



(a) Primary structure—
the amino acid sequence



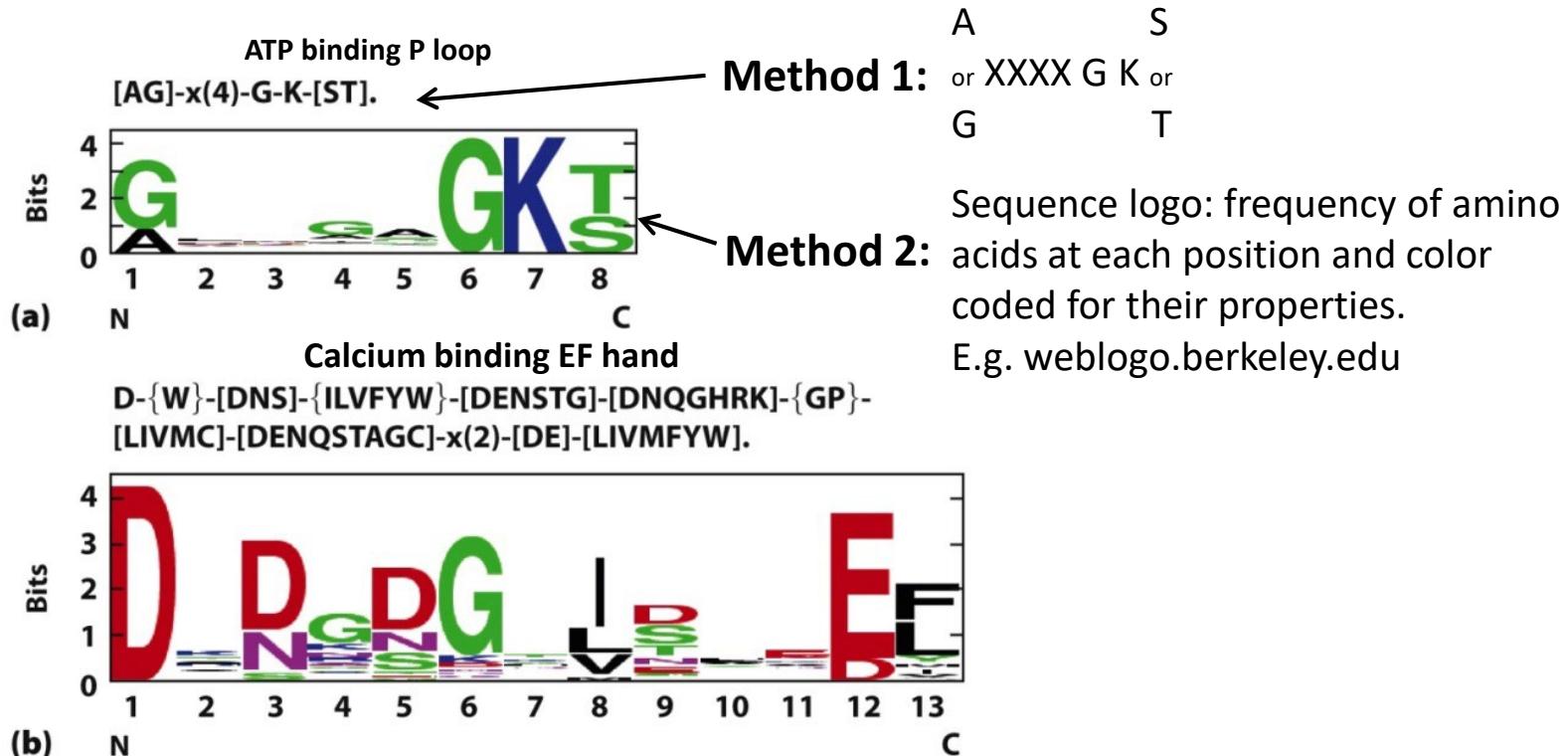
Quaternary structure
Multiple protein subunits
form a large complex



(d) Quaternary structure
— the relationships
between individual
subunits

Primary structure

- Orders of amino acids in a protein (read from amino-terminal to carboxyl-terminal)
- Consensus sequence (motif): parts of the sequence that have particularly good agreement, often represent evolutionarily conserved functions

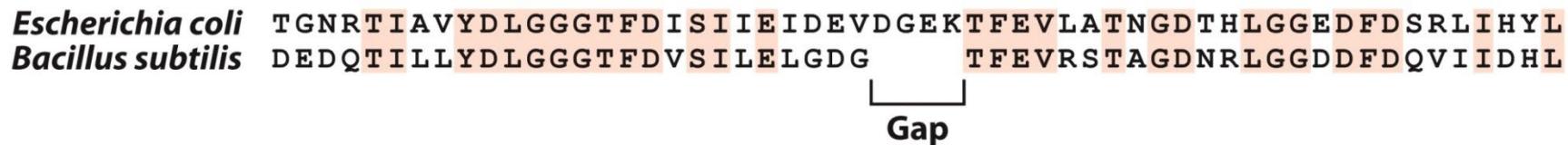


Comparison of sequences by alignment can elucidate the origin of the genes

Homologs: closely related proteins of the same protein family

(1) Paralogs: homologs that are present in the same species

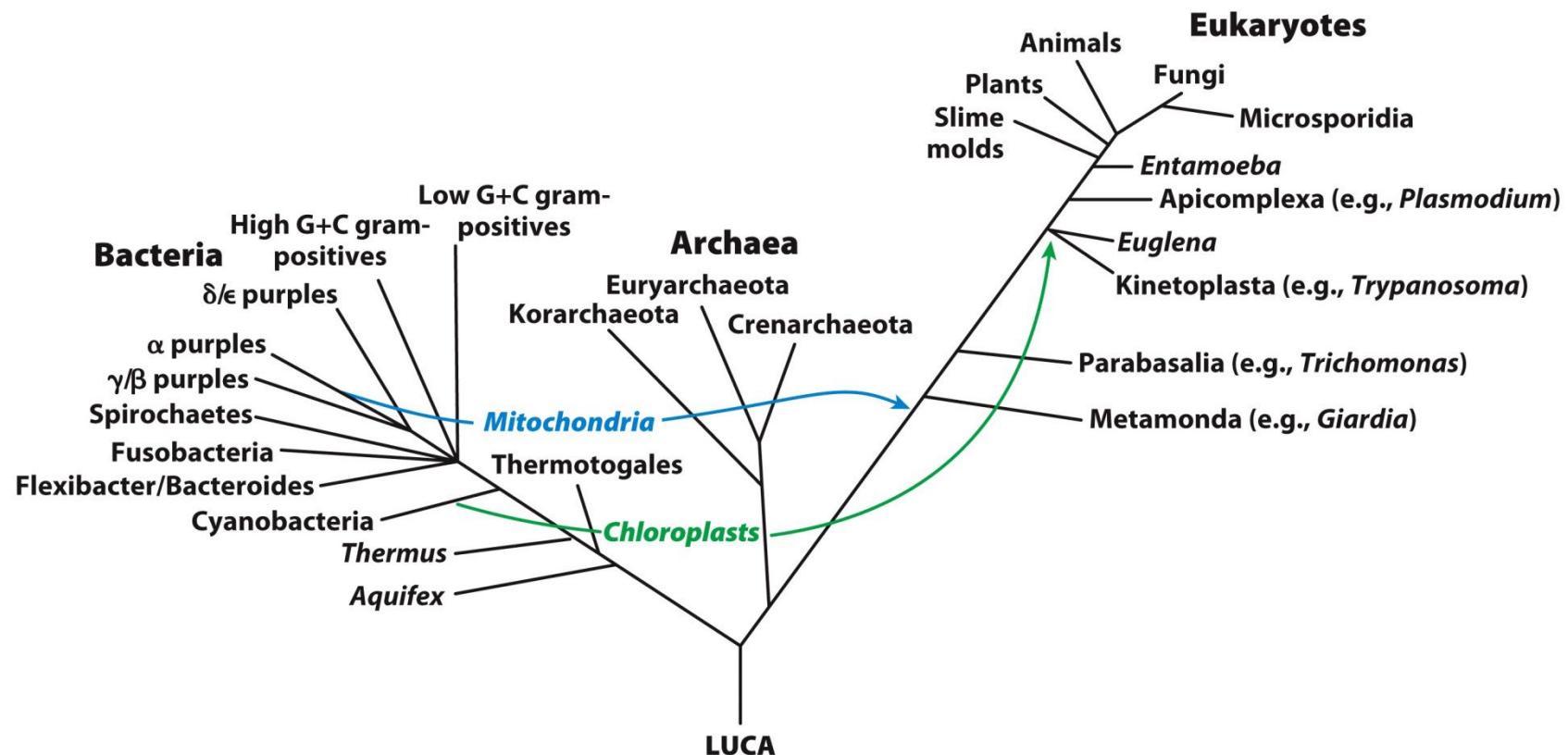
(2) Orthologs: homologs that are present in different species



Signature sequence: certain segment of a protein sequence found in the organisms of one taxon but not others

		Signature sequence
Archaea	{ <i>Halobacterium halobium</i>	I G H V D H G K S T M V G R I L L Y E T G S V P E H V I E Q H
	<i>Sulfolobus solfataricus</i>	I G H V D H G K S T L V G R I L L M D R G F I D E K T V K E A
Eukaryotes	{ <i>Saccharomyces cerevisiae</i>	I G H V D S G K S T T G H L I Y K C G G I D K R T I E K F
	<i>Homo sapiens</i>	I G H V D S G K S T T G H L I Y K C G G I D K R T I E K F
Gram-positive bacterium	<i>Bacillus subtilis</i>	I G H V D H G K S T M V G R
Gram-negative bacterium	<i>Escherichia coli</i>	I G H V D H G K T T L T A A

Phylogenetic tree – “the tree of life”: relationship of species inferred from the similarities of protein or DNA sequences



Unique amino acids in the receptor binding domain of SARS-CoV-2

‘Polybasic sequence’ is unique to SARS-CoV-2:
R (Arg, Arginine) has a positively charged side group

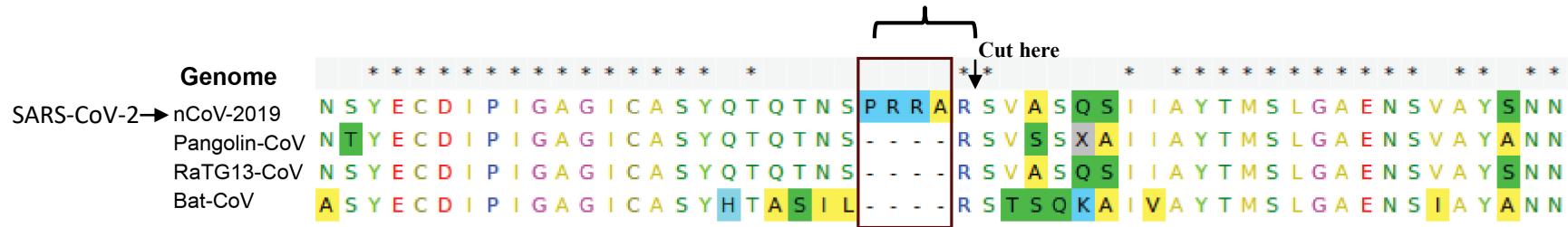


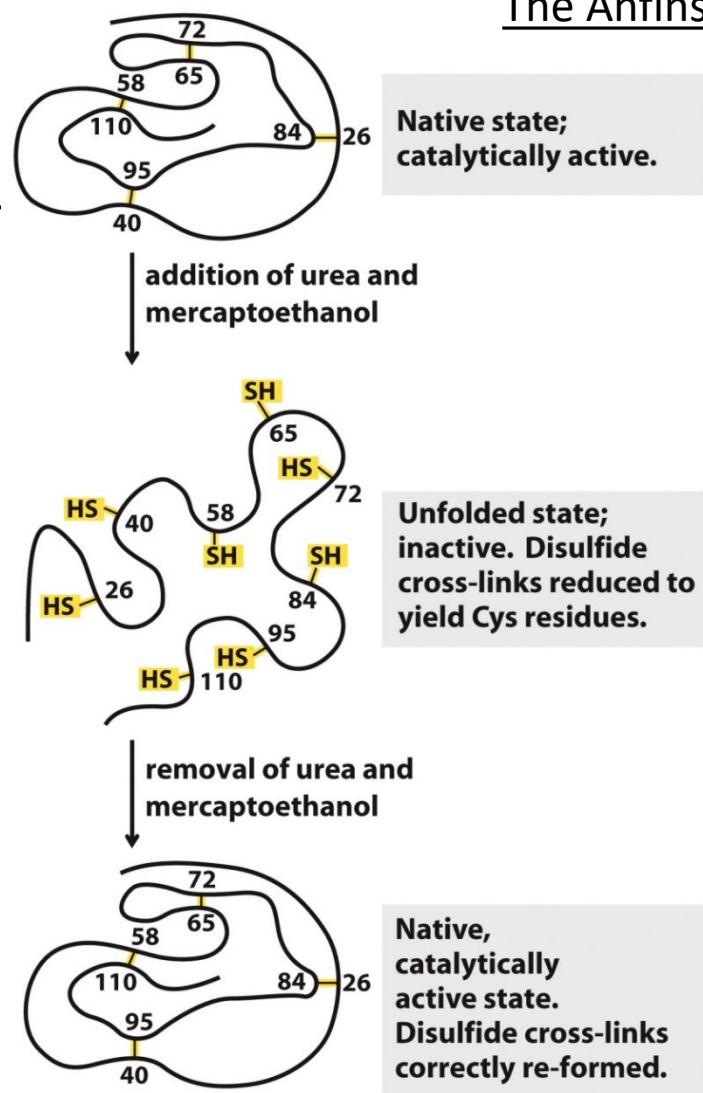
Figure 2: Multiple alignment at the amino acid level of PRRA insertion region and flanking areas across three coronavirus genomes relative to the outbreak strain (nCoV-2019). Box show the PRRA insertion present in nCoV-2019 and missing in the other three CoV genomes.

- RRAR: Polybasic sequence is a furin cleavage motif.
- Furin protease could cleave protein at the sequence: Arg-X-X-Arg
- Evidence of recombination in coronaviruses implicating pangolin origins of nCoV-2019
<https://www.biorxiv.org/content/10.1101/2020.02.07.939207v1>



Primary sequence determines native structure of the protein

- Purified Ribonuclease A is completely denatured by treatment with urea plus 2-mercaptoethanol (2-ME, breaking disulfide bridges)
- After removing urea and 2-ME, Ribonuclease A can regain its biological activity
- Proper refolding of protein could take place *in vitro*, initiated just by its own sequence



The Anfinsen Experiment



Dr Christian Anfinsen

Conformation of a protein

- Conformation: the spatial arrangement of atoms within a protein molecule
- Conformation is essential for the given protein function
- There are limited number of conformations allowed for a given protein



Why?

1. Structural constraints from the peptide backbone

- Six atoms ($\text{C}\alpha\text{-C-N-C}\alpha$, O and H) form a rigid plane because C-N angel (ω , omega) can not rotate.
- Rotation is only allowed at two angles at $\text{C}\alpha$:
 $\text{N-C}\alpha$: φ (phi)
 $\text{C}\alpha\text{-C}$: ψ (psi)
- Protein is a series of those rigid planes connected at $\text{C}\alpha$.

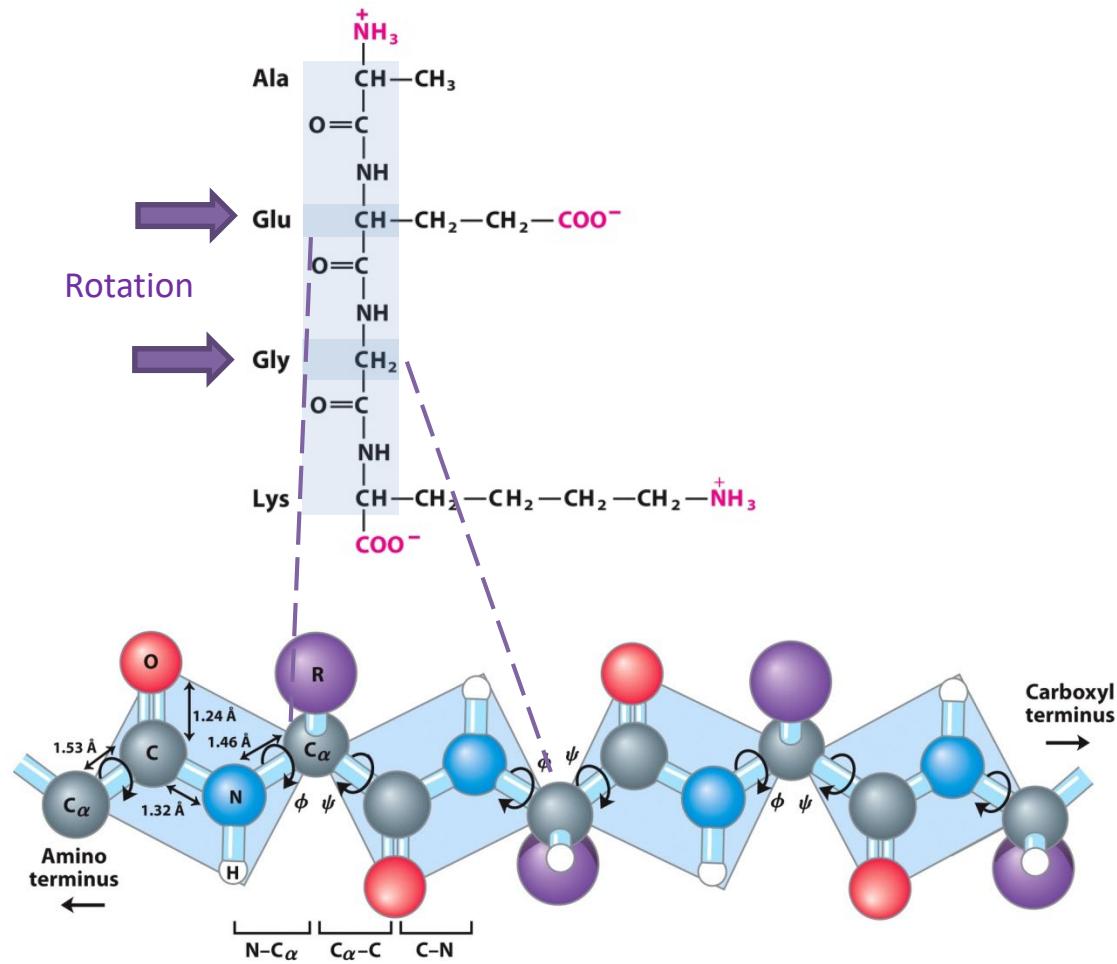


Figure 4-2b
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W.H. Freeman and Company

- Six atoms ($\text{C}\alpha$ -C-N- $\text{C}\alpha$, O and H) forms a rigid plane because of partial double bonds between C-O and C-N.

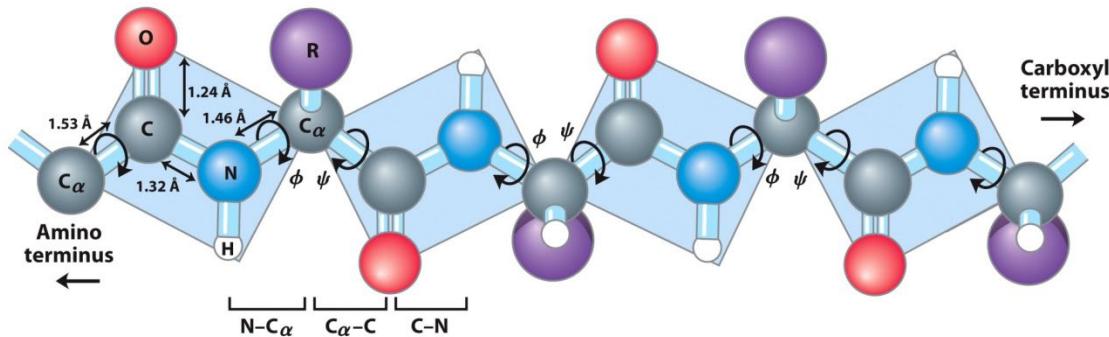
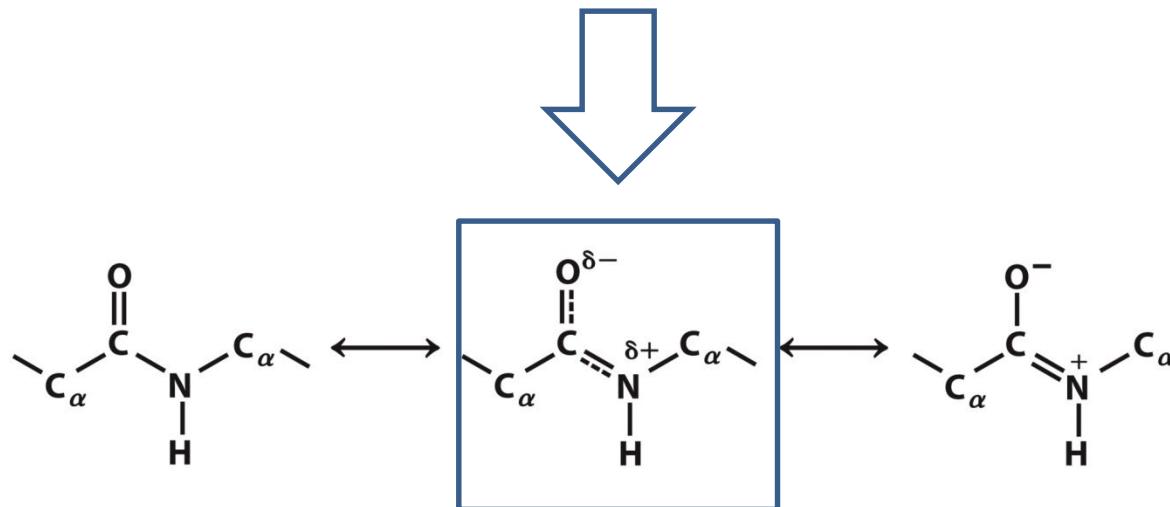
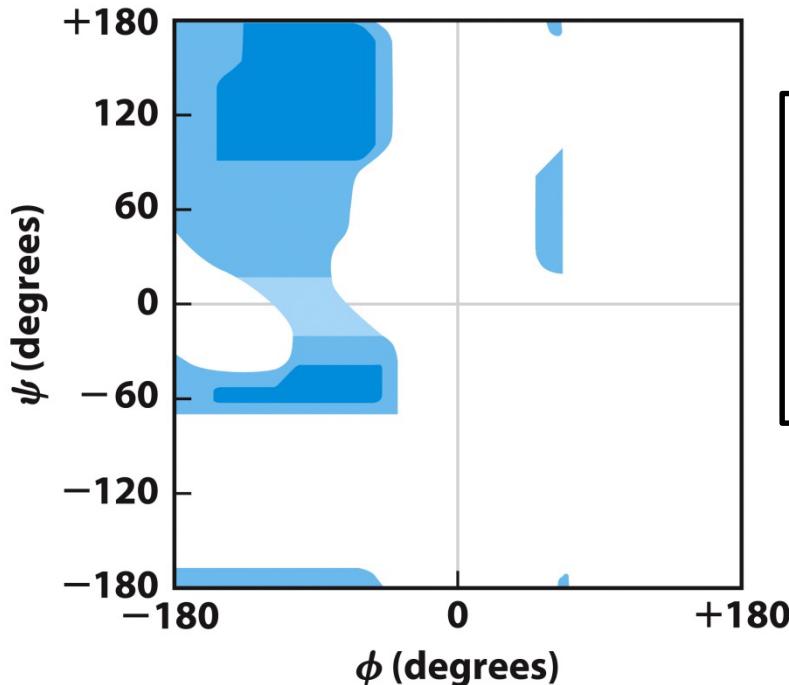
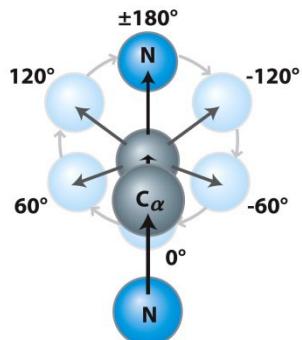
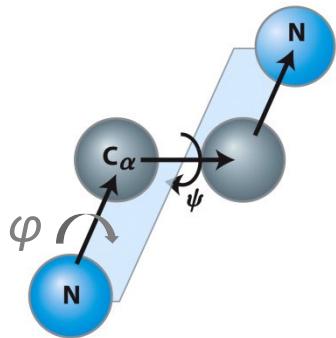


Figure 4-2b
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

The rigid peptide bonds limit the range of conformations possible for a polypeptide chain

Ramachandran plot

Demonstrating ψ



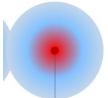
C-N: ω is fixed at mostly $\pm 180^\circ$

N-C α : φ (0 to $\pm 180^\circ$)

C α -C: ψ (0 to $\pm 180^\circ$)

In natural proteins, φ and ψ are only allowed within a certain range due to the clash of van der Waals radii.

Dark blue: stable
Medium to light blue:
less stable
White: not stable



2. Interactions between side-chains

- The conformations existing are usually the ones that are thermodynamically most stable (i.e. having the lowest free energy)
- Protein structure is stabilized largely by multiple weak, non-covalent interactions
 - *The hydrophobic effect (nonpolar groups are clustered in the interior) makes the major contribution to stabilizing most soluble proteins*
 - *The number of hydrogen bonds and ionic interactions within the protein is maximized*
- Disulfide bonds (covalent) play a role in the stabilization of structure in some proteins (especially extracellular proteins)

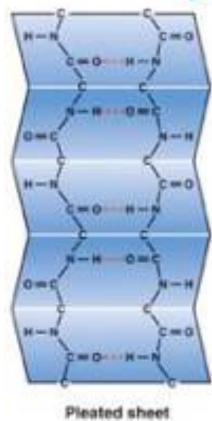
Structures of proteins

Primary sequence

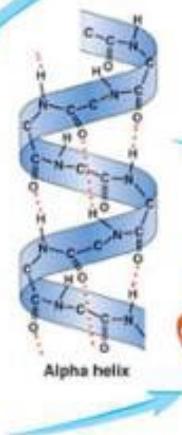
N terminal (N-ter, amino group) ...KLAGGC... Carboxyl terminal (C-ter)

Secondary structure

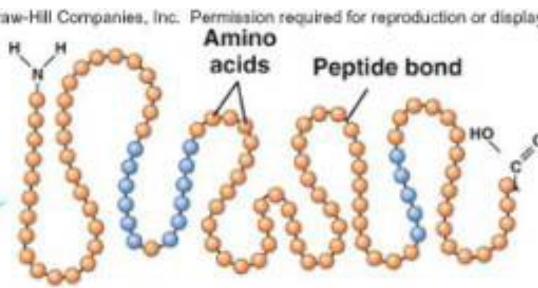
Local spatial arrangement
of main chain atoms
 α helix, β sheet



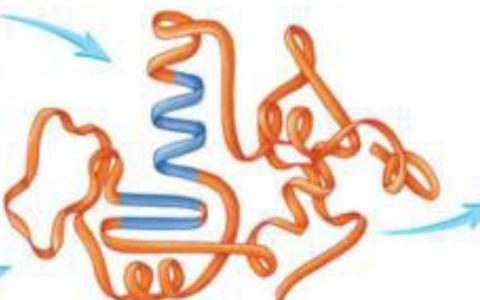
(b) Secondary structure
with folding as a result
of hydrogen bonding
(dotted red lines)



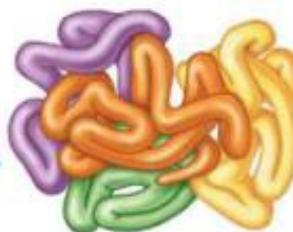
(a) Primary structure—the amino acid sequence



(a) Primary structure—the amino acid sequence



(c) Tertiary structure with
secondary folding caused by
interactions within the
polypeptide and its immediate
environment



(d) Quaternary structure
—the relationships
between individual
subunits

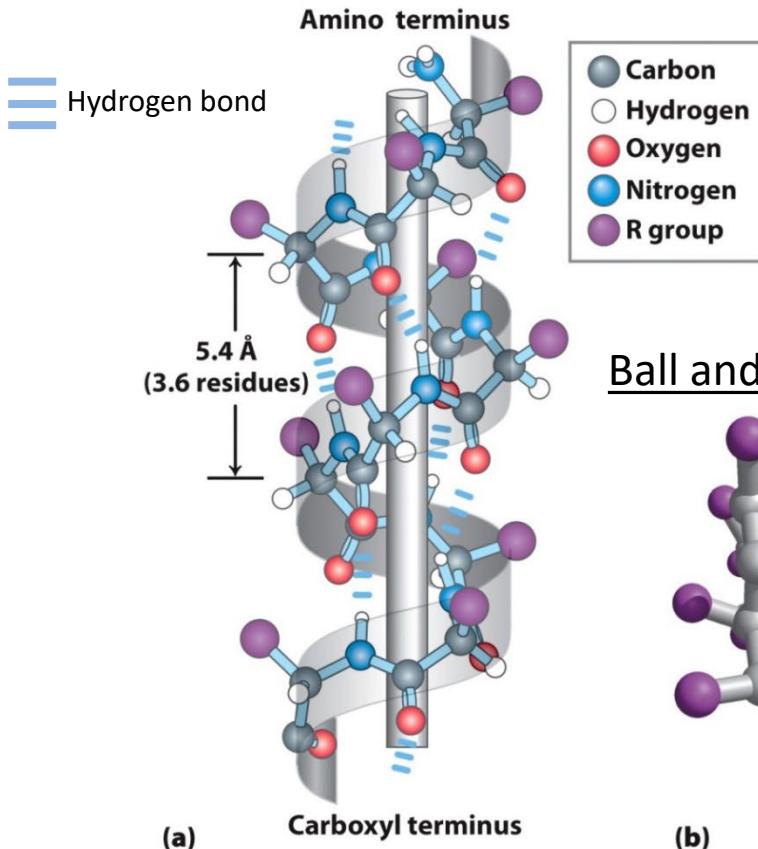
Quaternary structure
Multiple protein subunits
form a large complex

Tertiary structure

3D structure of one polypeptide/protein

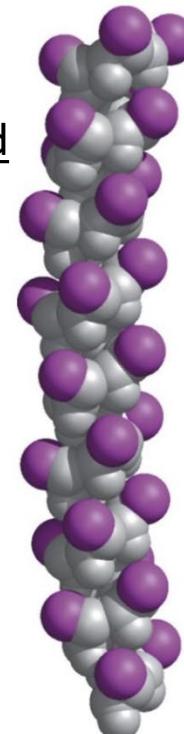
Protein secondary structure - α helix

- Right hand helical structure (no left hand ones in natural proteins); each helical turn includes 3.6 amino acids
- Stabilized by hydrogen bond between the hydrogen atom attached to the nitrogen of the first amino acid and the carbonyl oxygen atom of the fourth amino acid

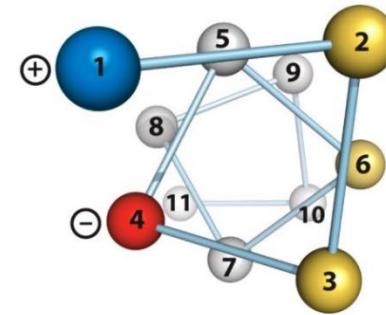


Space-filled

Ball and stick

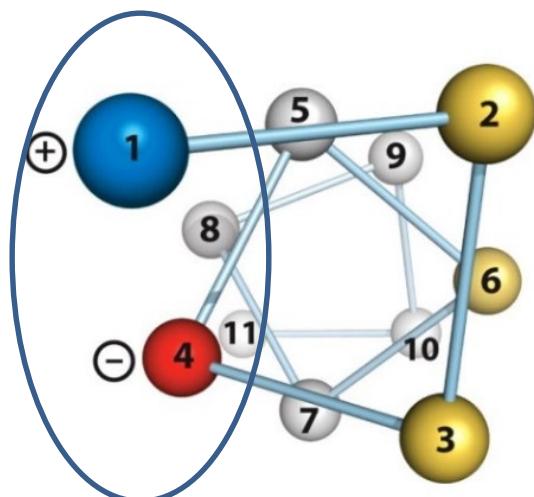


Helical wheel projection
(highlight the properties of the R groups)



Protein secondary structure - α helix

- α helices is the most readily-formed secondary structure because it maximizes the use of internal hydrogen bonding
- Not all polypeptides form α helix
- Alanine shows the greatest tendency to form α helices; Pro and Gly tend to destabilize α helices
- Some combinations of amino acids favour α helices; e.g. oppositely-charged amino acids spaced 3-4 residues apart



Protein secondary structure - β sheet

- Polypeptide forms zigzag structure (also R group placed up and down alternately) – β strand

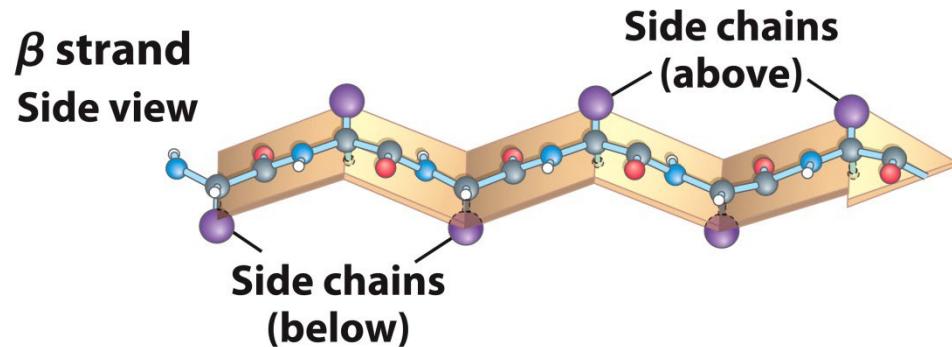


Figure 4-6a
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

- Repeating parallel β strands are held together side-by-side by hydrogen bonds forming β sheet

≡ Hydrogen bond

Antiparallel β sheet

Top view

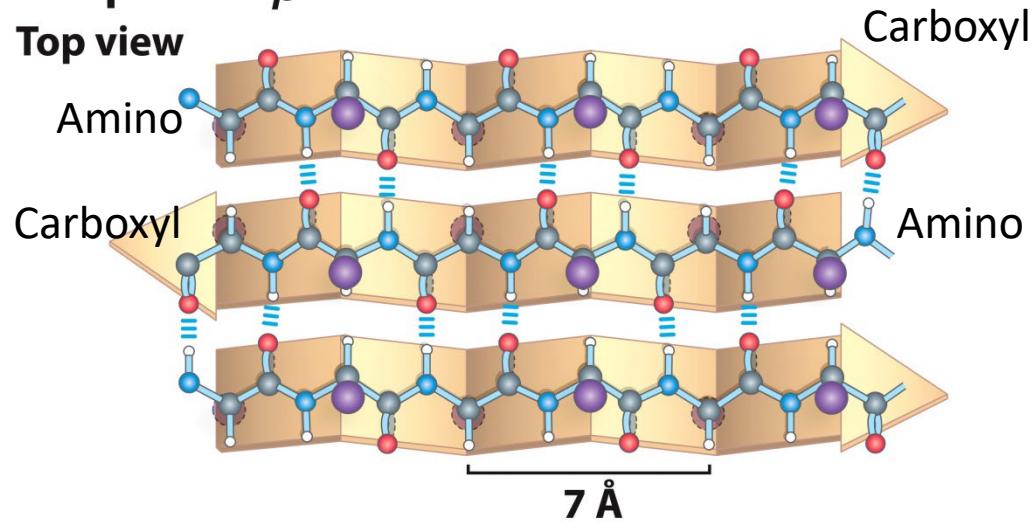


Figure 4-6b
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company

β -sheets can also be parallel

Secondary structures in Ramachandran plot

- A certain range of combination of ϕ and ψ angles is preferred by each type of secondary structure

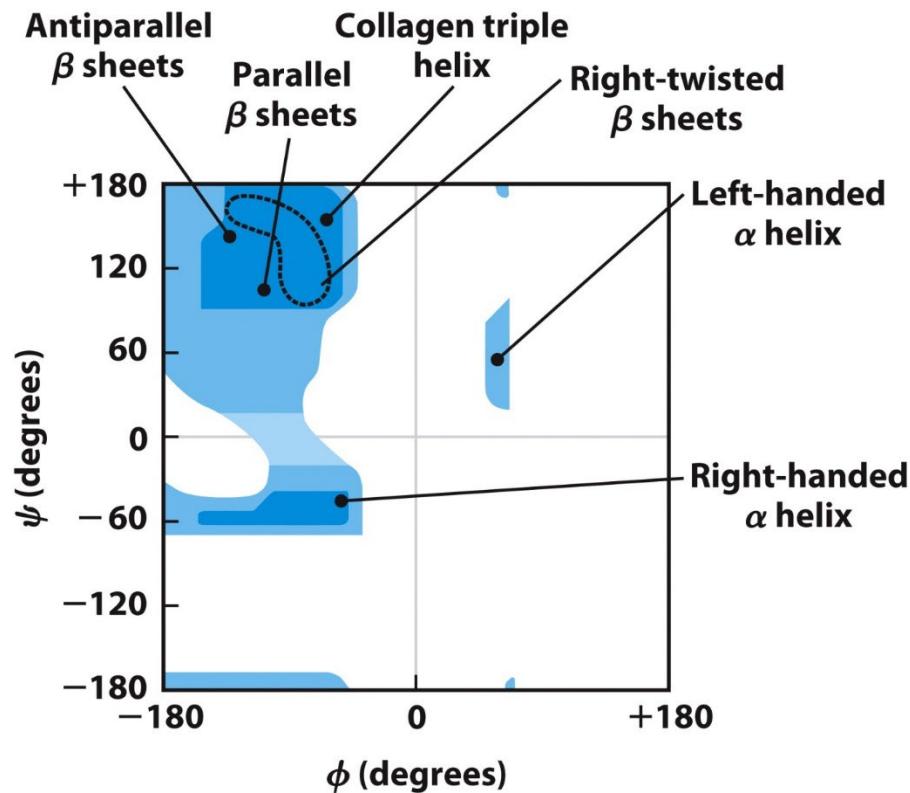
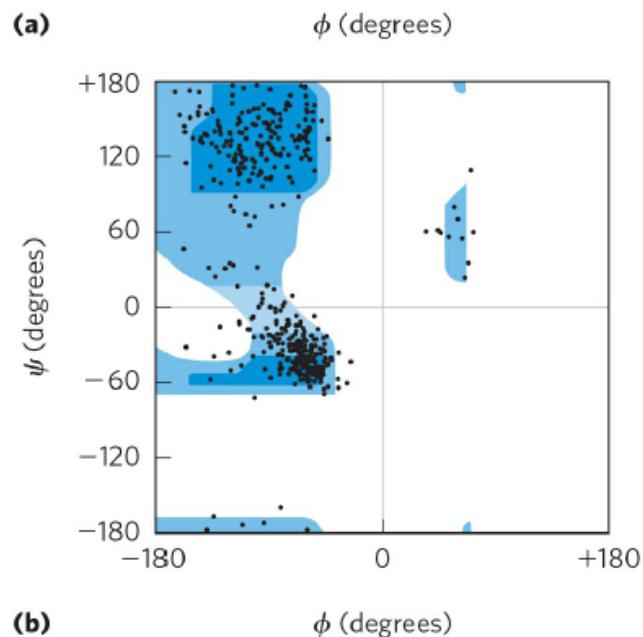


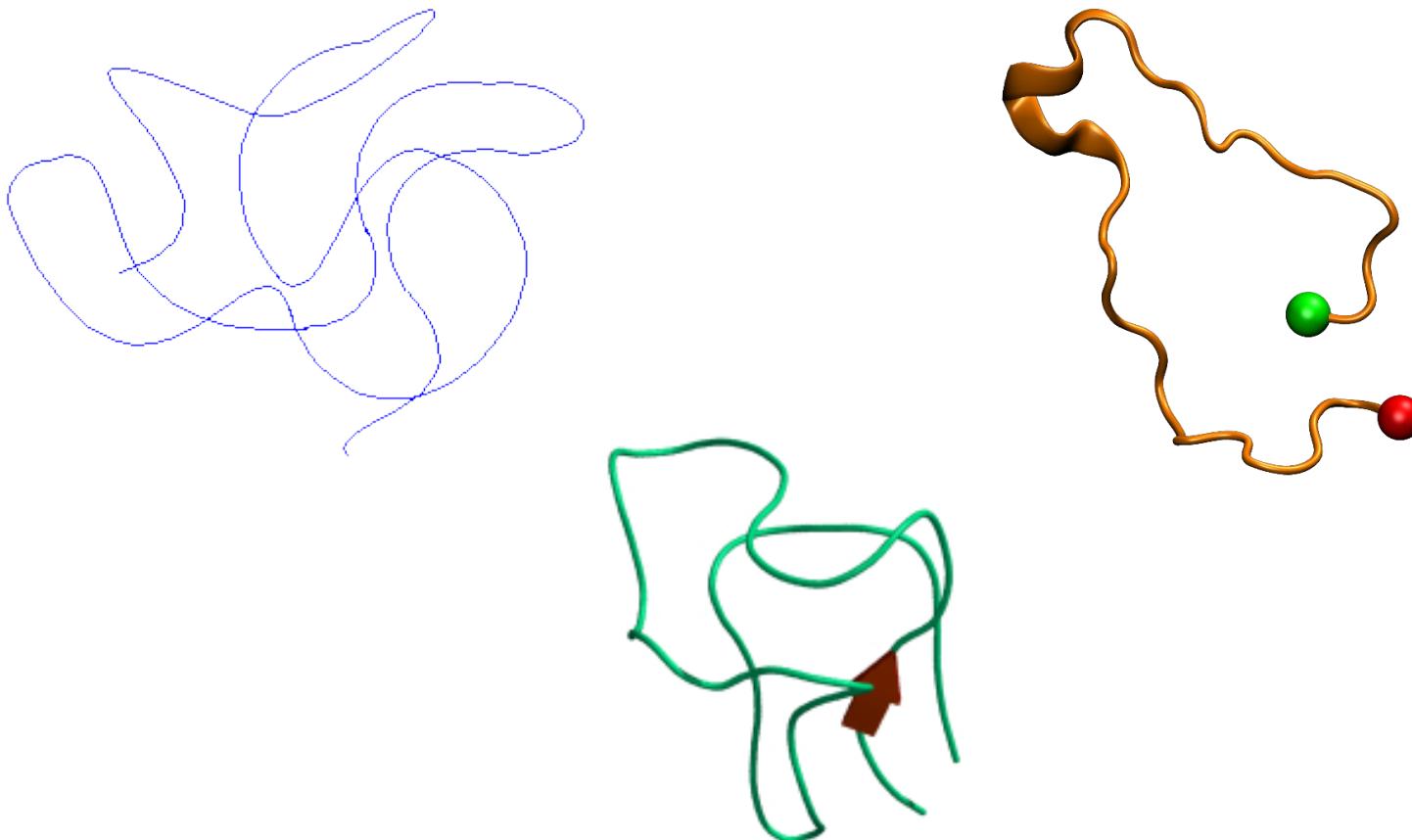
Figure 4-9a
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W. H. Freeman and Company



All the ϕ and ψ values of every amino acid residue (except glycine) of the enzyme pyruvate kinase

Random coil

- Not all proteins contain secondary structures such as α helices and β -sheets that have regular patterns
- Non-repetitive, unstructured proteins

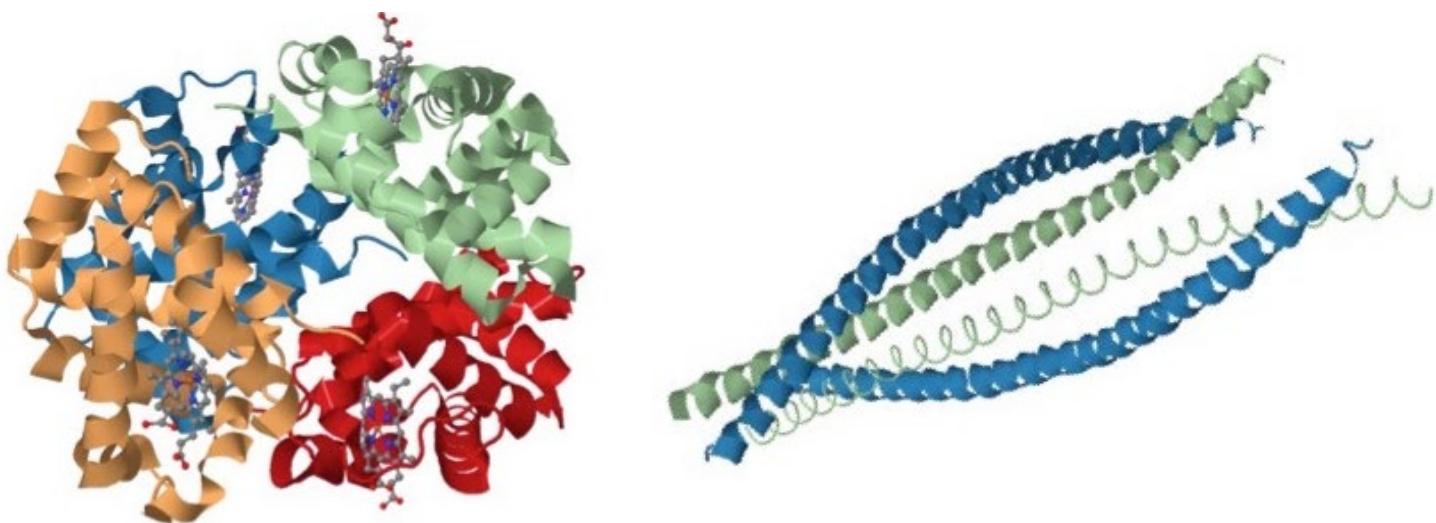


Protein tertiary and quaternary structures

- **Tertiary structure:** The overall three-dimensional arrangement of all atoms in a protein
- Including long-range interactions of distant parts of a protein (amino acids far apart in the primary sequence)
- Combination of many secondary structural elements
- **Quaternary structure:** Multiple subunits (chains) of a protein being arranged in a three-dimensional complex

General classification of proteins based on structure

- Fibrous proteins: polypeptide chains arrange in long strands or sheets
- Globular proteins: polypeptide chains fold into a spherical or globular shape



Example of fibrous protein: α -Keratin

- Major constituent of the outer layer of skin, hair, wool, nail, claws, quills, horns, hooves
- Structural functions; tough
- Connected together by many **disulfide bonds**

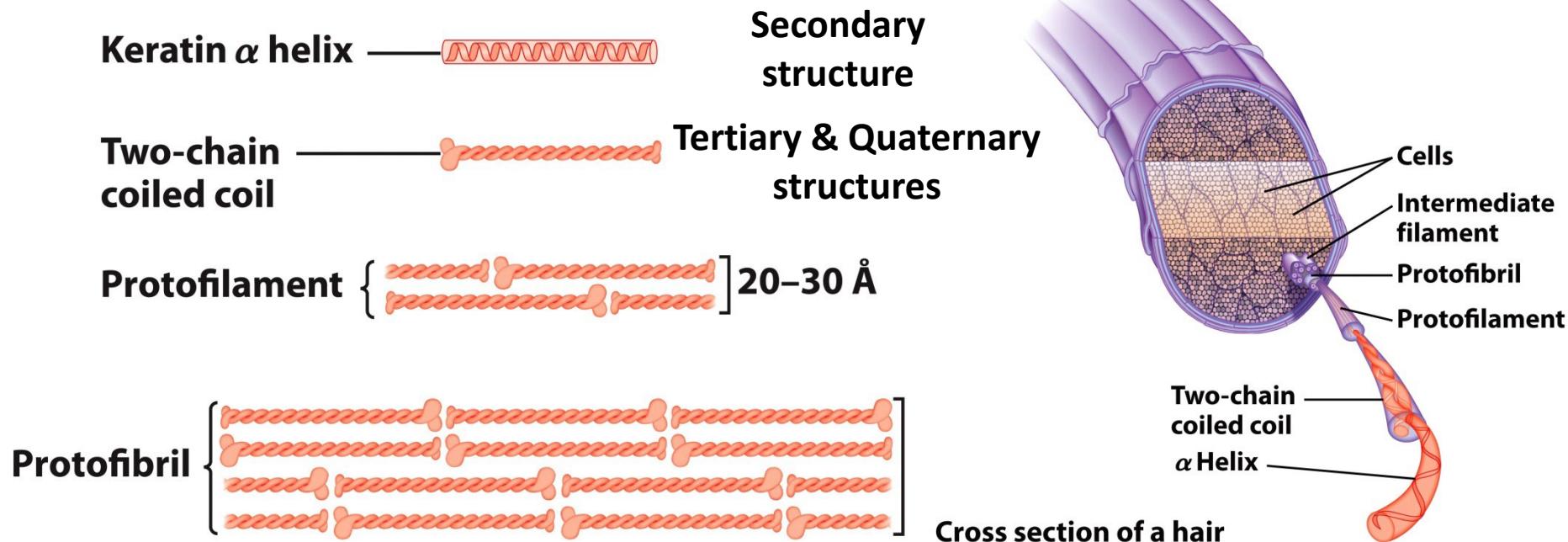


Figure 4-11b
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W.H. Freeman and Company

Disulfide bond

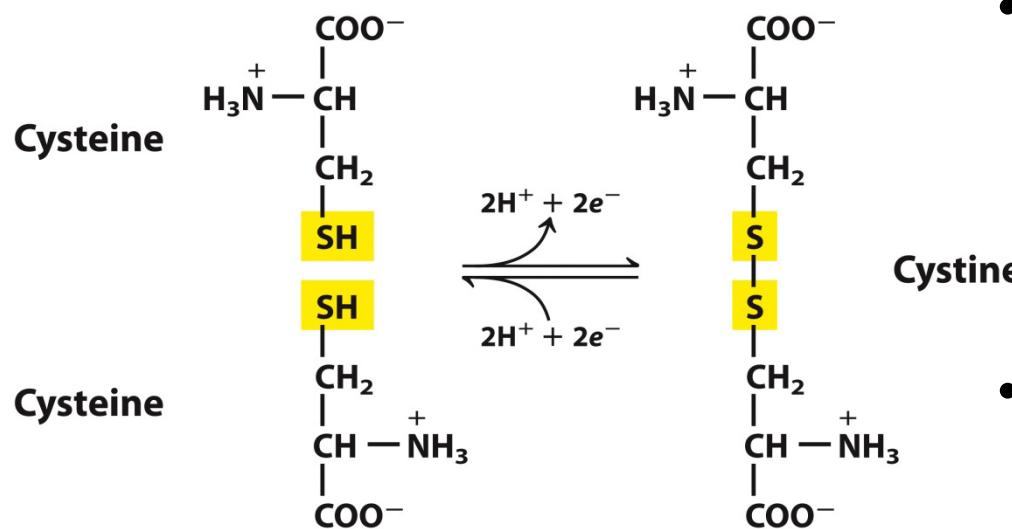


Figure 3-7
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W.H. Freeman and Company

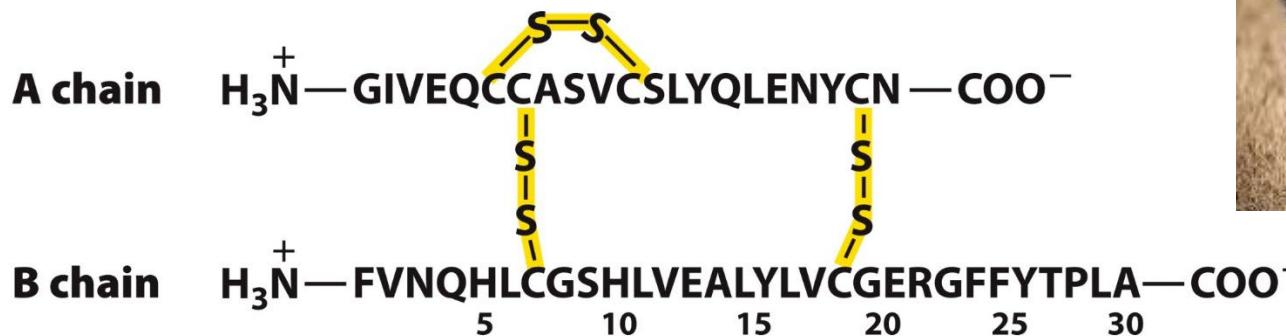


Figure 3-24
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W.H. Freeman and Company

- Disulfide bonds are covalent bonds formed post-translationally by the oxidation of a pair of cysteines
- Can be formed between cysteines of the same or different protein chains



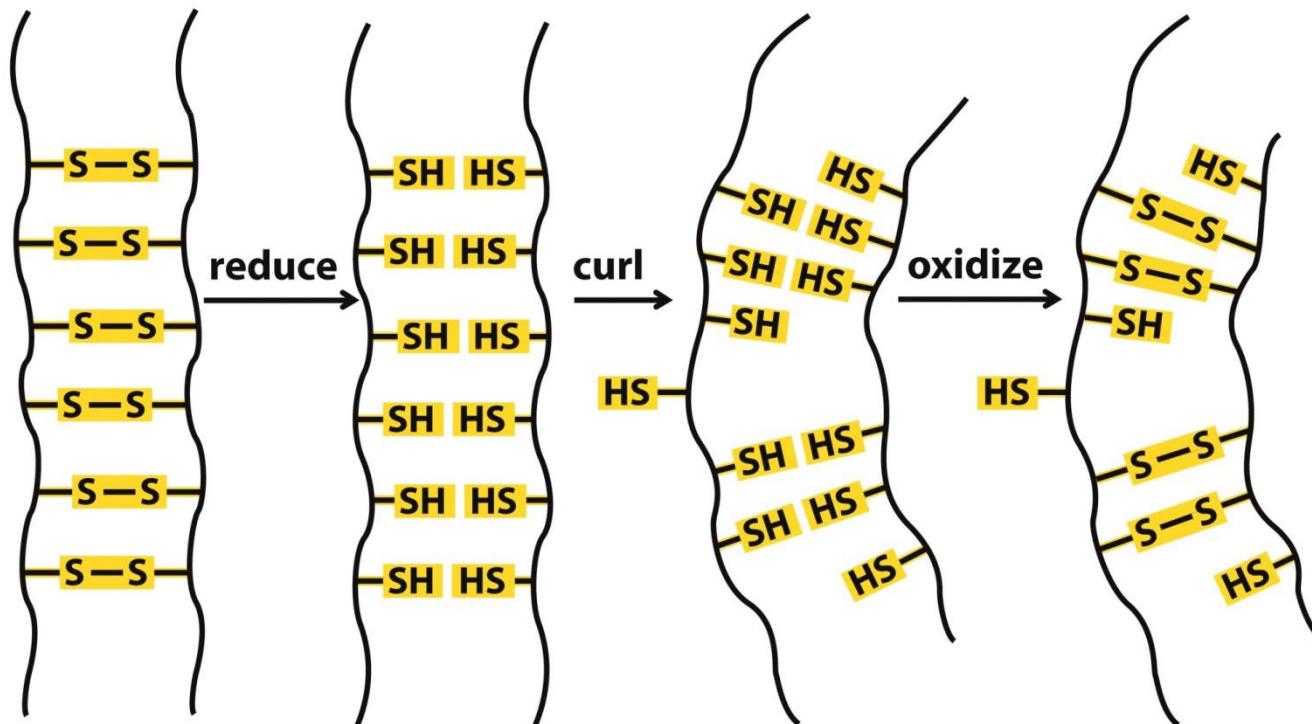
Up to 18% of the keratin in rhino horn are made up of Cys

Permanent waving is biochemical engineering



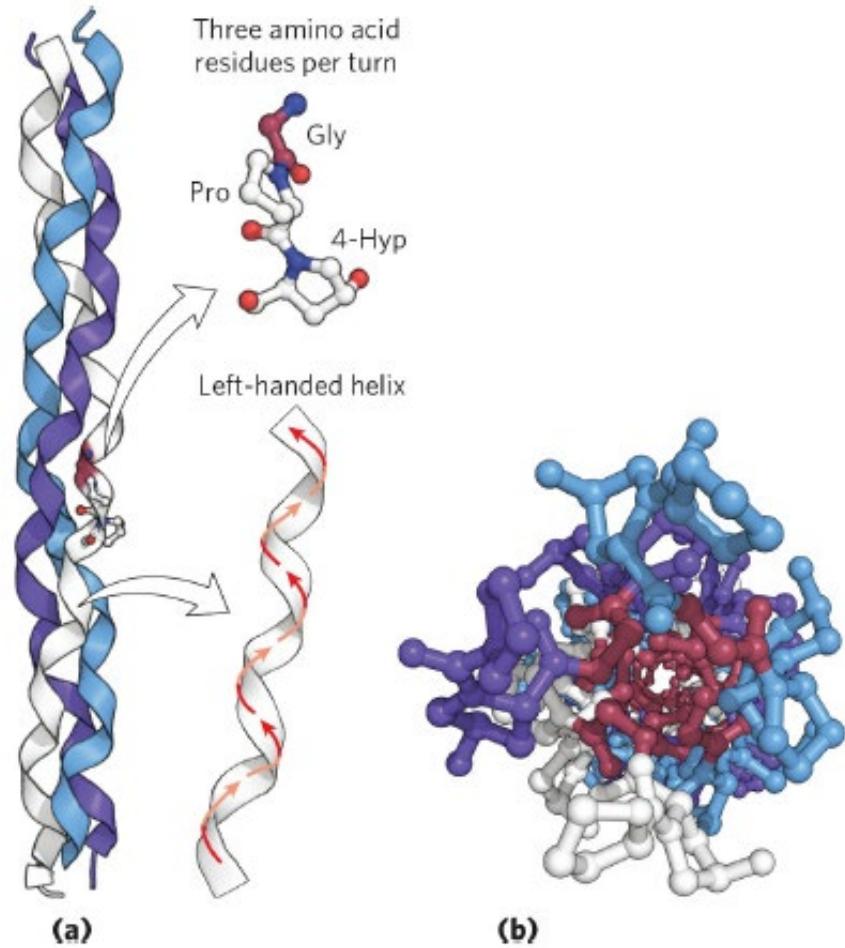
- Under moist heat α -Keratin can be stretched out from α helix to extended β conformation
- Breaking (reduce) and rebuilding disulfide (oxidize) bonds can be used to curl or straighten hairs

<https://louisdietvorst.wordpress.com/2012/09/22/nothing-is-permanent/>



Example of fibrous protein: Collagen

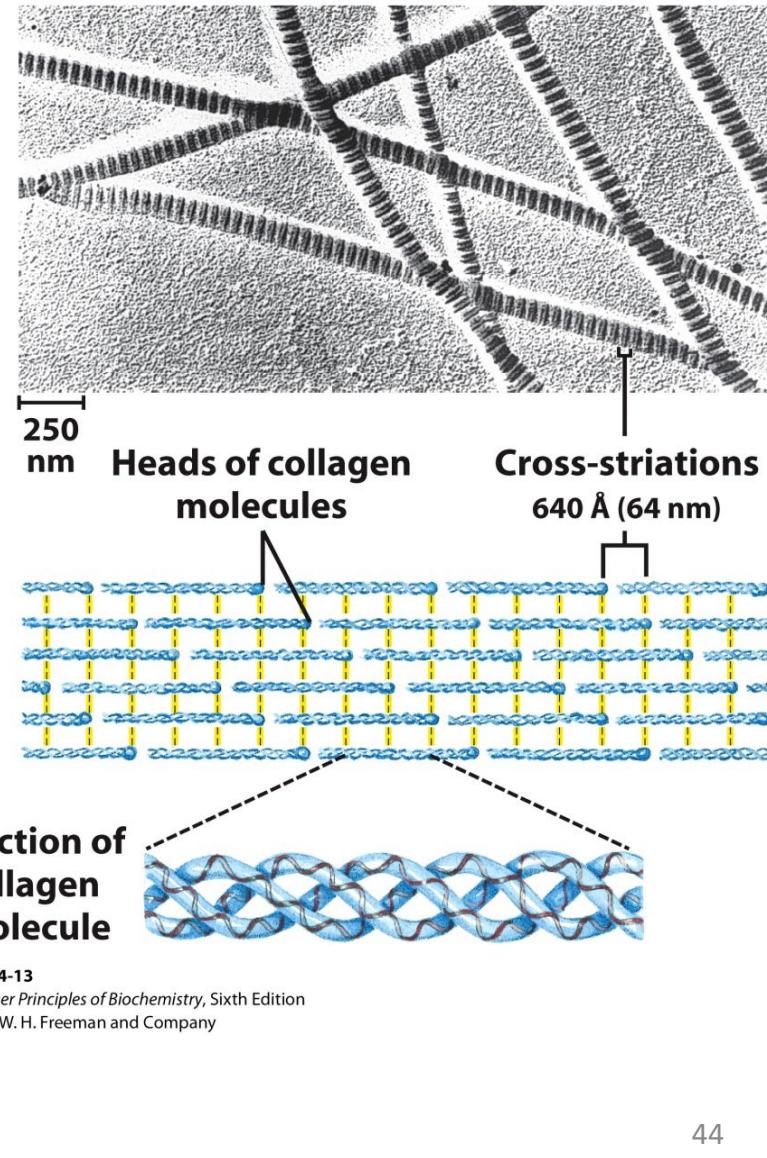
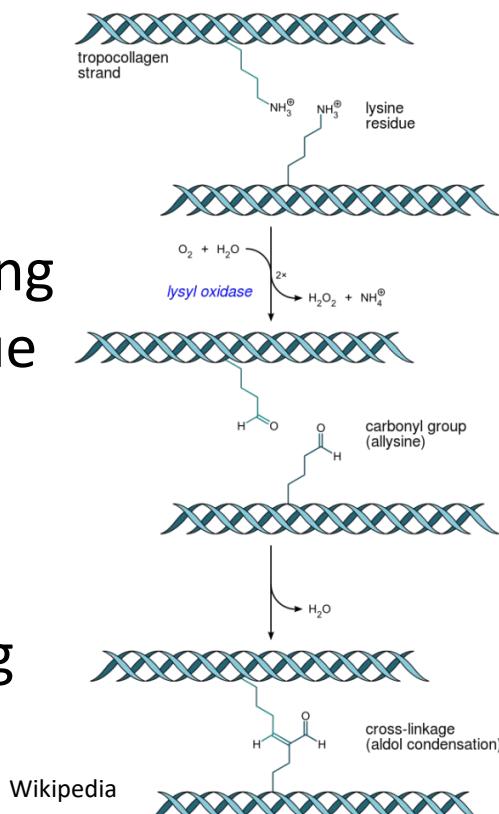
- Found in connected tissues such as tendons, cartilage, the organic matrix of bones and cornea of the eye
- Provide strength to the tissues
- The sequence is generally a repeating tripeptide Gly-X-Y: X is often Pro and Y is often 4-Hyp (4-Hydroxyproline), an uncommon amino acid
- Formed by α chains which is different from α helix
- Three α chains form coiled coil structure



Structure of collagen fibrils

- The α chains are cross-linked by unusual types of covalent bonds involving Lys, HyLys (5-Hydroxylysine) or His at a few of the X and Y positions

- Increasing rigid and brittle character of aging connective tissue results from accumulated covalent bonds along with aging



Secondary structures and properties of fibrous proteins

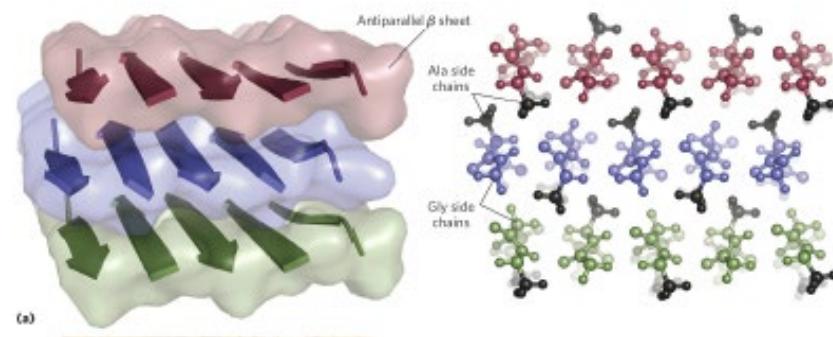
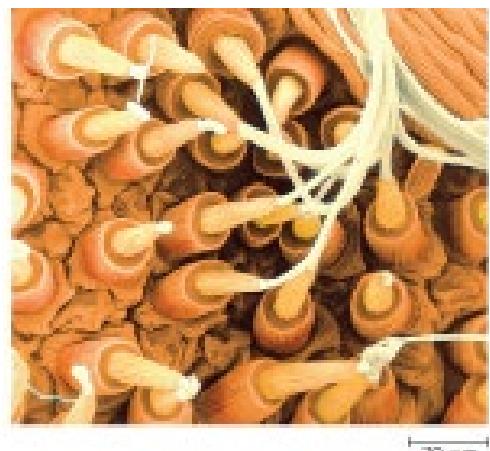
TABLE 4–3 Secondary Structures and Properties of Some Fibrous Proteins

Structure	Characteristics	Examples of occurrence
α Helix, cross-linked by disulfide bonds	Tough, insoluble protective structures of varying hardness and flexibility	α -Keratin of hair, feathers, nails
β Conformation	Soft, flexible filaments	Silk fibroin
Collagen triple helix	High tensile strength, without stretch	Collagen of tendons, bone matrix

Table 4-3

Lehninger Principles of Biochemistry, Sixth Edition

© 2013 W. H. Freeman and Company



Silk (the protein fibroin)

Example of globular protein: Myoglobin

- Oxygen-binding proteins of muscle cells
- 3D structure was solved in 1950s (1962 Nobel Prize, John Kendrew and Max Perutz)
- Key features:
 - Binding pocket for **Heme**;
 - Hydrophobic residues** are buried in the interior

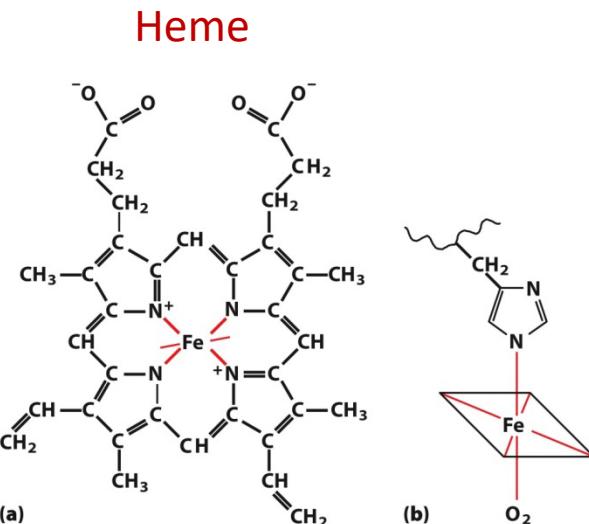


Figure 4-17
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W.H. Freeman and Company

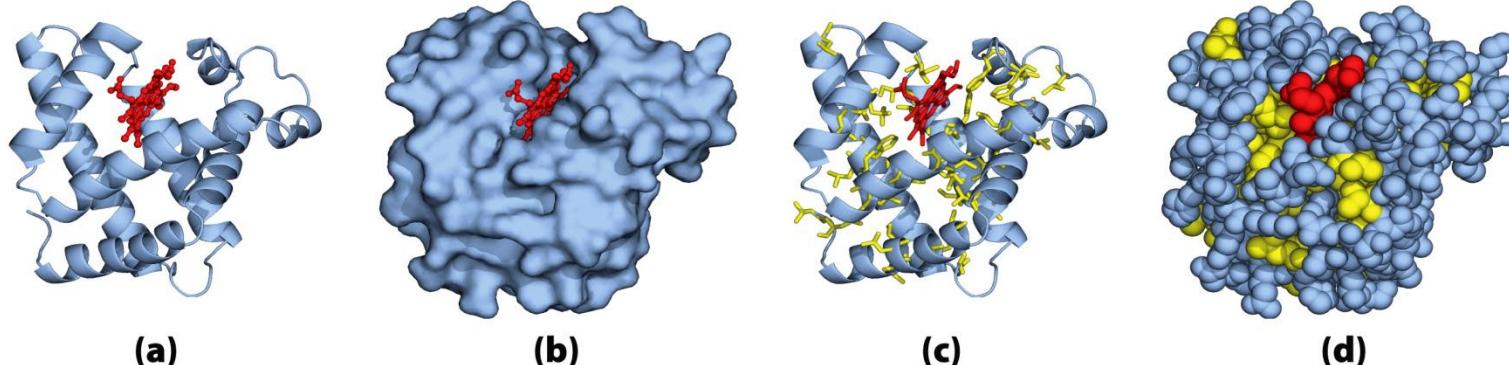
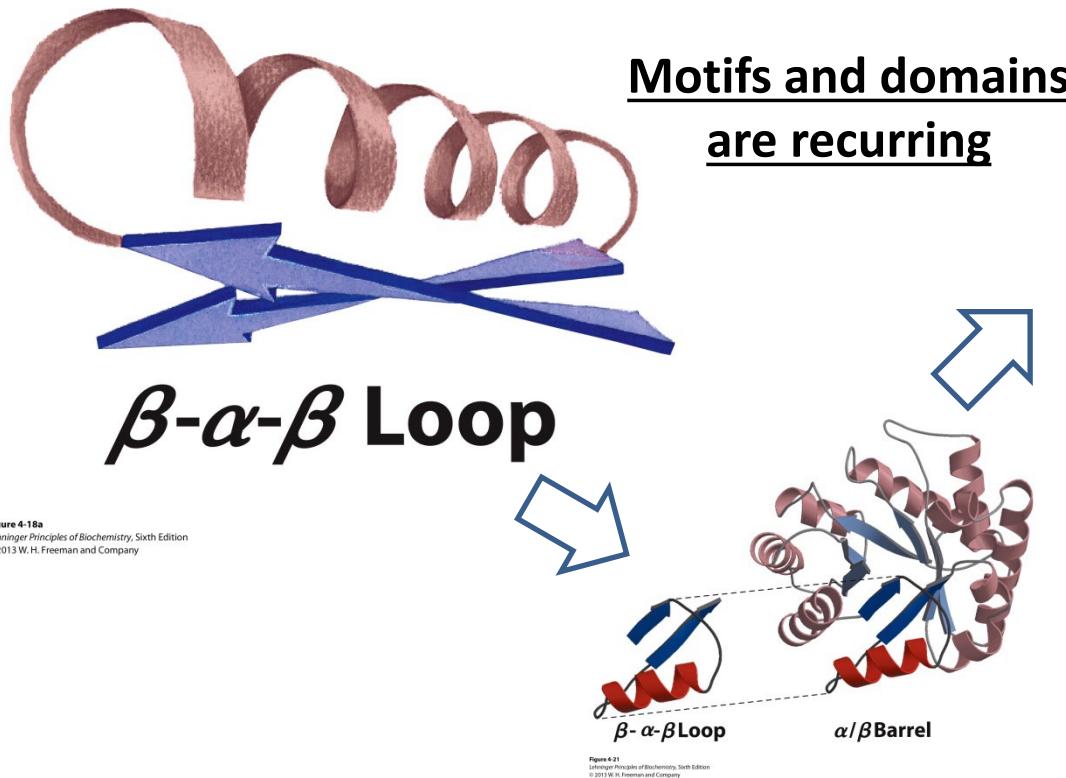


Figure 4-16
Lehninger Principles of Biochemistry, Sixth Edition
© 2013 W.H. Freeman and Company

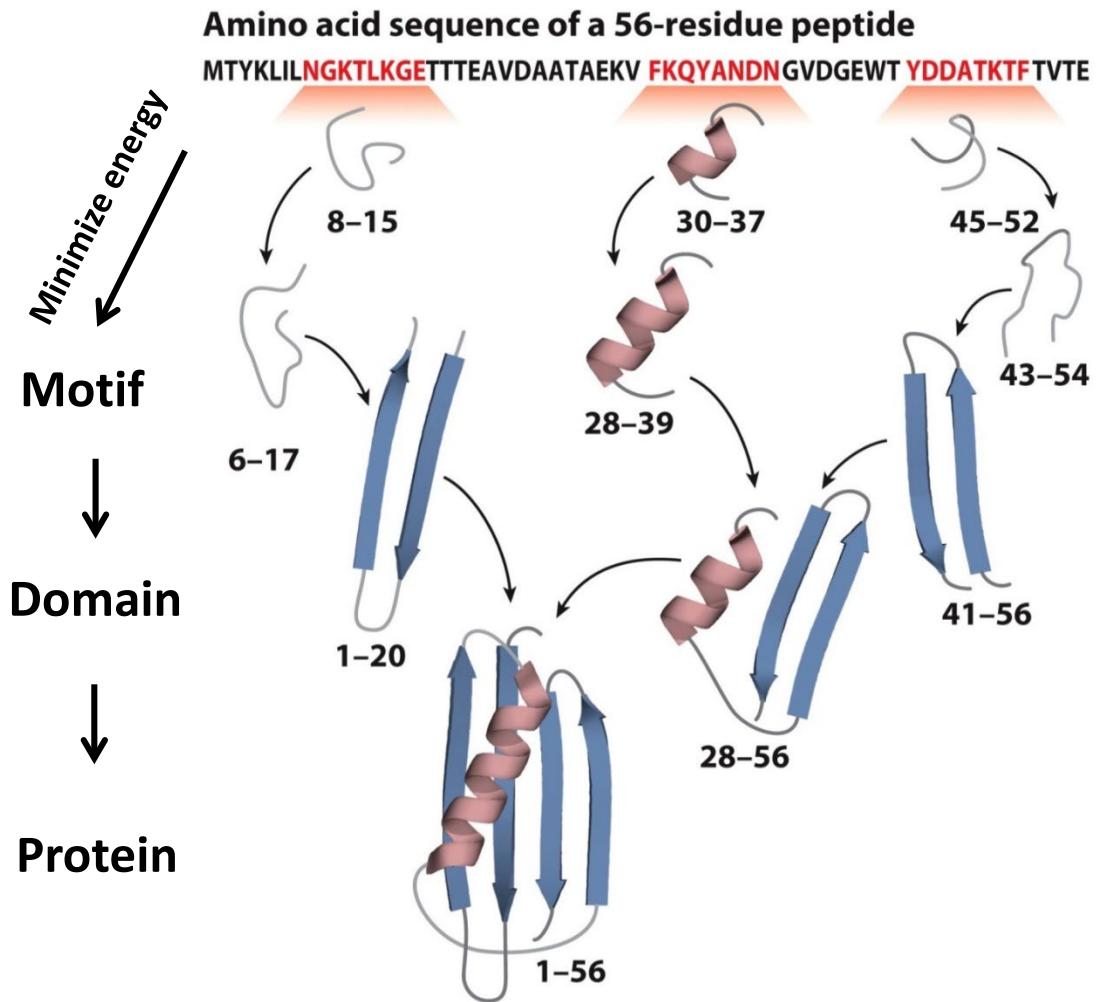
Protein folding: basic folding blocks

- Motif (fold, super-secondary structure): recognizable folding pattern involving two or more elements of secondary structure and the connection
- Domain: a part of a polypeptide chain that is independently stable or could undergo movements of a single entity with respect to the entire protein



Protein folding: a stepwise process

- Protein folding is not a completely random, trial and- error process
- From high energy state (unfolded) to the lowest energy state (native structure): to achieve thermodynamic stability
- From peptide to motif then to domain then to the complex protein
- Protein folding can also be assisted by **chaperone** proteins

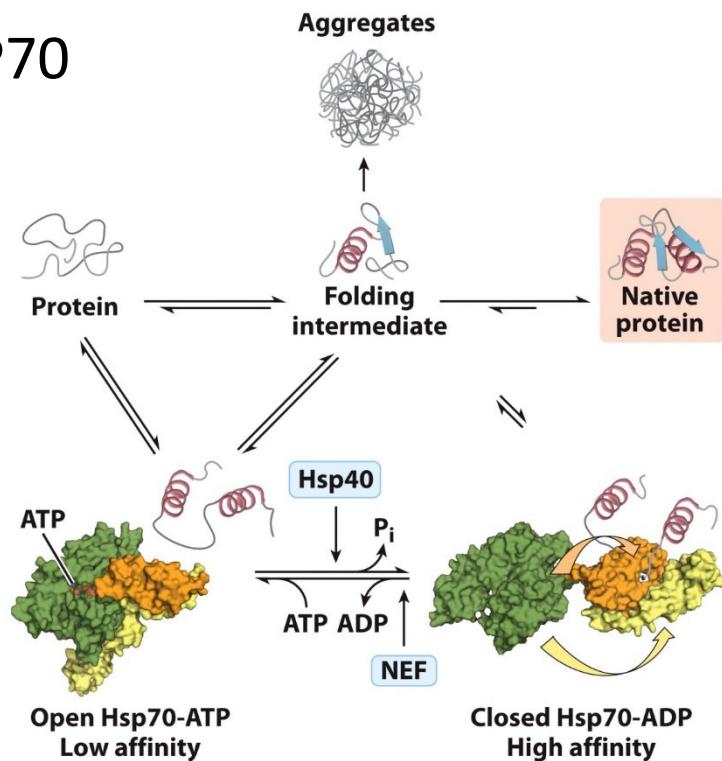


Assisted protein folding by chaperones

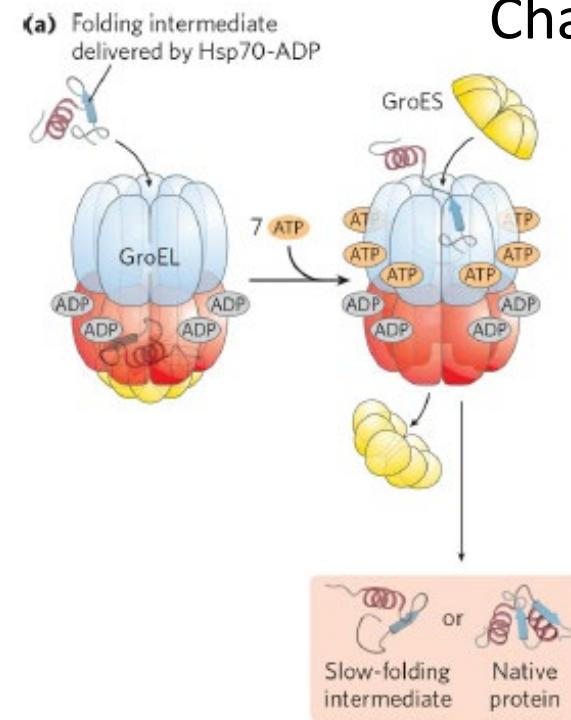
- Chaperones are proteins that assist the folding of other proteins
- Interact with partially folded or improperly folded polypeptides
- Do not actively promote the folding of the substrate protein, but instead prevent aggregation of unfolded peptides

Two major families of chaperones

HSP70

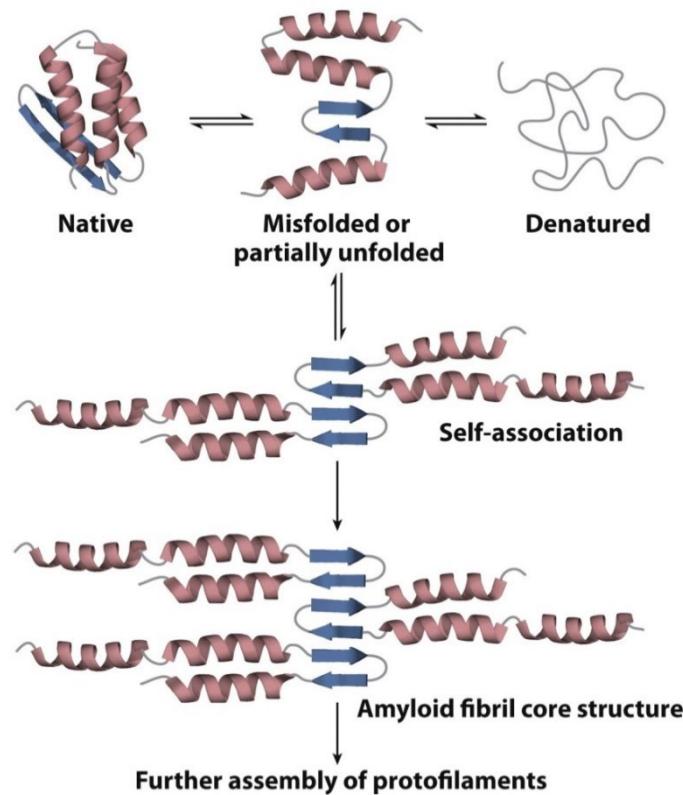
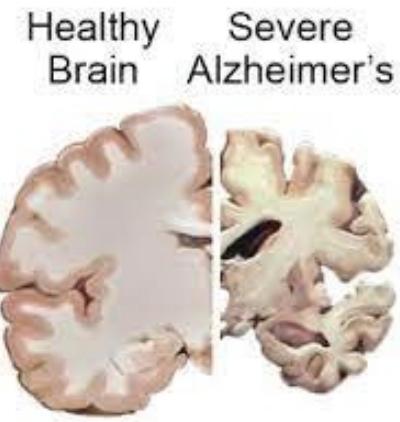
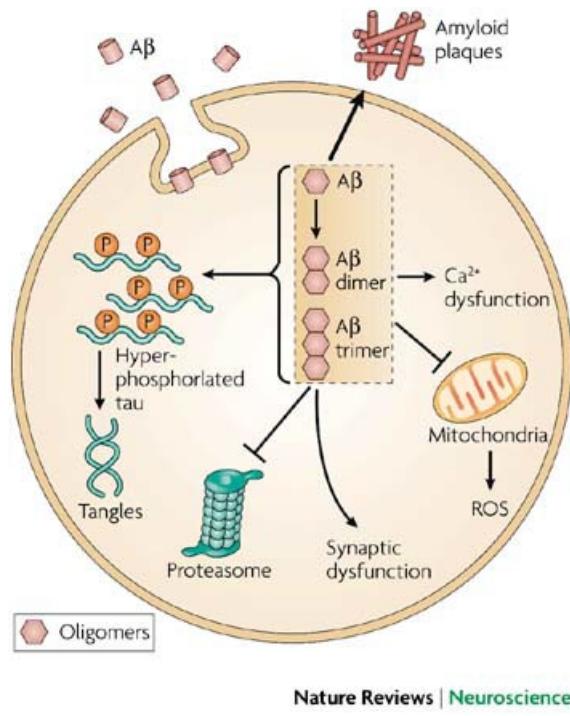


Chaperonins



Alzheimer's disease: misfolded β -amyloid proteins

- Normally intracellular amyloid- β exists as a monomeric form
- Misfolded amyloid- β can aggregate into oligomers and form amyloid fibrils
- The condition called **amyloidosis** might mediate pathological events within a neuron and could be the cause the Alzheimer's disease



Summary (II)

- Primary, secondary, tertiary and quaternary structures of proteins:
 - Consensus sequence, Ramachandran plot, types of secondary structures, disulfide bond, fibrous and globular proteins
- How protein folds:
 - Motif, domain
 - Chaperones
- Protein structure and diseases:
 - Formation of amyloid fibrils and the Alzheimer's disease

Study Question

**Collagen is regarded as a protein with low nutritional value.
Why?**

- (A) Cross-linking between fibrils makes collagen difficult to be digested
- (B) It contains too little diversity of amino acids
- (C) It is too large to be absorbed
- (D) Our intestine does not have the enzyme to digest collagen

Study Question

A protein can be phosphorylated at a specific Tyrosine residue. If you want to study the importance of this phosphorylation, which amino acid you will use to substitute the Tyrosine residue?

- (A) Alanine
- (B) Glycine
- (C) Phenylalanine
- (D) Serine