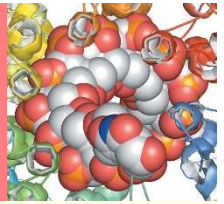
5.4a Homologous proteins from different organism have homologus amino acid sequence



Homologus

proteins sharing a significant degree of sequence similarity and structural resemblance.

Orthologous and **Paralogus** proteins

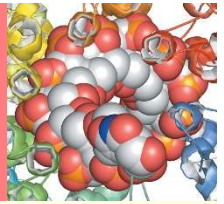
Orthologous

*from **different species** that have **homologus** amino acid sequence with a similar function*

Paralogus

*protein found within **a single species** that have **homologous** aa sequence arose through **gene duplication***
*α **and** β **chains of hemoglobin***

5.4b Computer Programs Can Align Sequences and Discover Homology Between Proteins



S. acidocaldarius FPIAKGGTAAIPGPFGSGKTVTLQSLAKWSAAK---VVIYVGCGERGNEMTD
E. coli CPFAKGGKVGLFGGAGVGKTVNMMELIRNIAIEHSGYSVFAGVGERTREGND

Shown are parts of the amino acid sequences of the catalytic subunits from the major ATP-synthesizing enzyme (**ATP synthase**) in a representative **archaea** and a **bacterium**.

These protein segments encompass the **nucleotide-binding site** of these enzymes

Alignment of the amino acid sequences of two protein homologs using **gaps**.

Identical residues in the two sequences are shown in red.
Introduction of a **three-residue-long gap** in the archaeal sequence **optimizes** the alignment of the two sequences.

Blocks **S**ubstitution **M**atrix (BLOSUM)

- Methods for alignment and comparison of protein sequences depend upon some quantitative measure of **how similar two sequences** are.
- One way to measure similarity is to use a matrix that assigns scores for **all possible substitutions** of one amino acid for another.
- **BLOSUM62** is the substitution matrix most often used with **BLAST**.
- **BLOSUM62** assigns a **probability score** for each position in an alignment based on the **frequency** with which that **substitution** occurs in the consensus sequences of related proteins.

Blocks **S**ubstitution **M**atrix (BLOSUM)

The substitution between W⟹N ,D or P " -4" is not very likely

The substitution between V & I " 3" is very likely

*Unique qualities residues, **C, H, P or W** have **high** BLOSUM62 score (9/8/7/11);*
*May **change the protein significantly***

*Similar residues, **(R & K), (D & E)** have low BLOSUM62 score (6/5)* and ***(A, V, L and I)*** *have low BLOSUM62 score (4).*

*Can replace with the other with **less change to the protein structure***

The score are derived using sequence sharing no more than 62% identity

***-4**: the lowest probability of substitution*
***11**: the highest probability of substitution*

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	−1	5																		
N	−2	0	6																	
D	−2	−2	1	6																
C	0	−3	−3	−3	9															
Q	−1	1	0	0	−3	5														
E	−1	0	0	2	−4	2	5													
G	0	−2	0	−1	−3	−2	−2	6												
H	−2	0	1	−1	−3	0	0	−2	8											
I	−1	−3	−3	−3	−1	−3	−3	−4	−3	4										
L	−1	−2	−3	−4	−1	−2	−3	−4	−3	2	4									
K	−1	2	0	−1	−3	1	1	−2	−1	−3	−2	5								
M	−1	−1	−2	−3	−1	0	−2	−3	−2	1	2	−1	5							
F	−2	−3	−3	−3	−2	−3	−3	−3	−1	0	0	−3	0	6						
P	−1	−2	−2	−1	−3	−1	−1	−2	−2	−3	−3	−1	−2	−4	7					
S	1	−1	1	0	−1	0	0	0	−1	−2	−2	0	−1	−2	−1	4				
T	0	−1	0	−1	−1	−1	−1	−2	−2	−1	−1	−1	−1	−2	−1	1	5			
W	−3	−3	−4	−4	−2	−2	−3	−2	−2	−3	−2	−3	−1	1	−4	−3	−2	11		
Y	−2	−2	−2	−3	−2	−1	−2	−3	2	−1	−1	−2	−1	3	−3	−2	−2	2	7	
V	0	−3	−3	−3	−1	−2	−2	−3	−3	3	1	−2	1	−1	−2	−2	0	−3	−1	4

Figure 5.18 The BLOSUM62 substitution matrix provides scores for **all possible exchanges** of one amino acid with another.

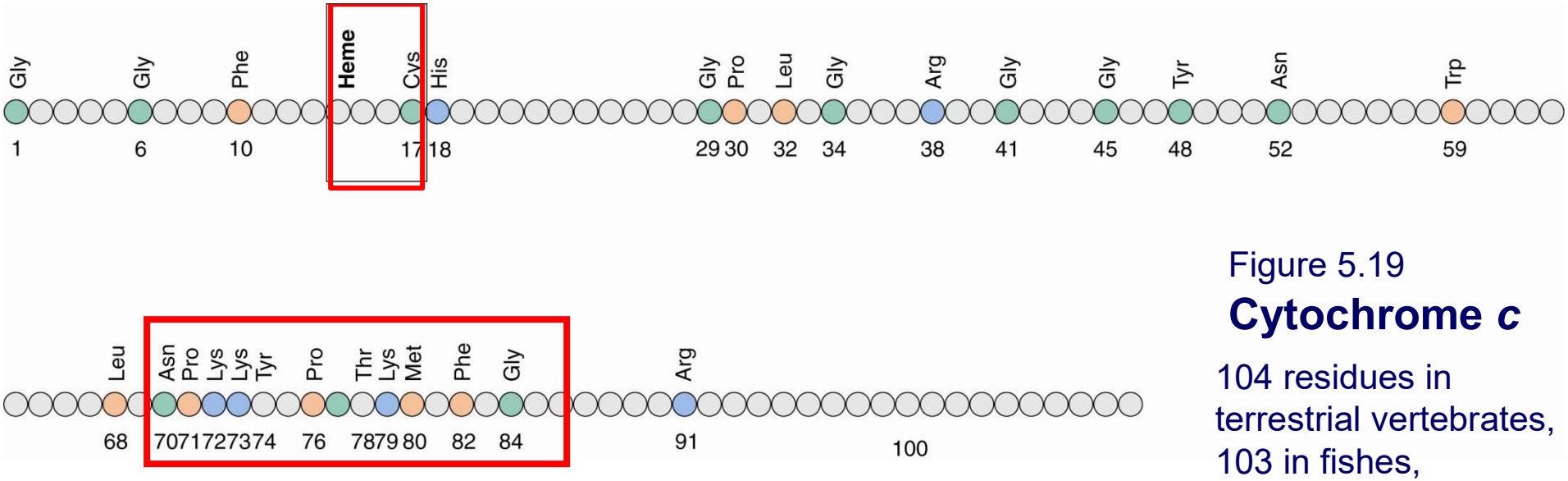


Figure 5.19

Cytochrome *c*

104 residues in terrestrial vertebrates,
103 in fishes,
107 in insects,
107 to 109 in fungi & yeasts,
111 or 112 in green plants

They are scattered **irregularly** along the polypeptide chain, except for a **cluster** between residues **70- 80**.

All cytochrome *c* polypeptide chains have a **cysteine residue** at position **17**, and all but one have another Cys at position **14**.

These **Cys** residues serve to link the **heme** prosthetic group of cytochrome *c* to the protein, a role explaining their invariable presence.

These invariant residues sever roles **crucial to the biological function** of this protein, and thus substitutions of other aa at these positions **cannot be tolerated**

Analysis of the sequence of cytochrome *c* from more than **40** different species reveals that **28** residues are **invarian**t. (marked)

The Phylogenetic Tree for Cytochrome c

Infer potential **ancestral sequences** represented by **nodes** , or branch points.

The minimum number of **mutational changes** connecting the branch

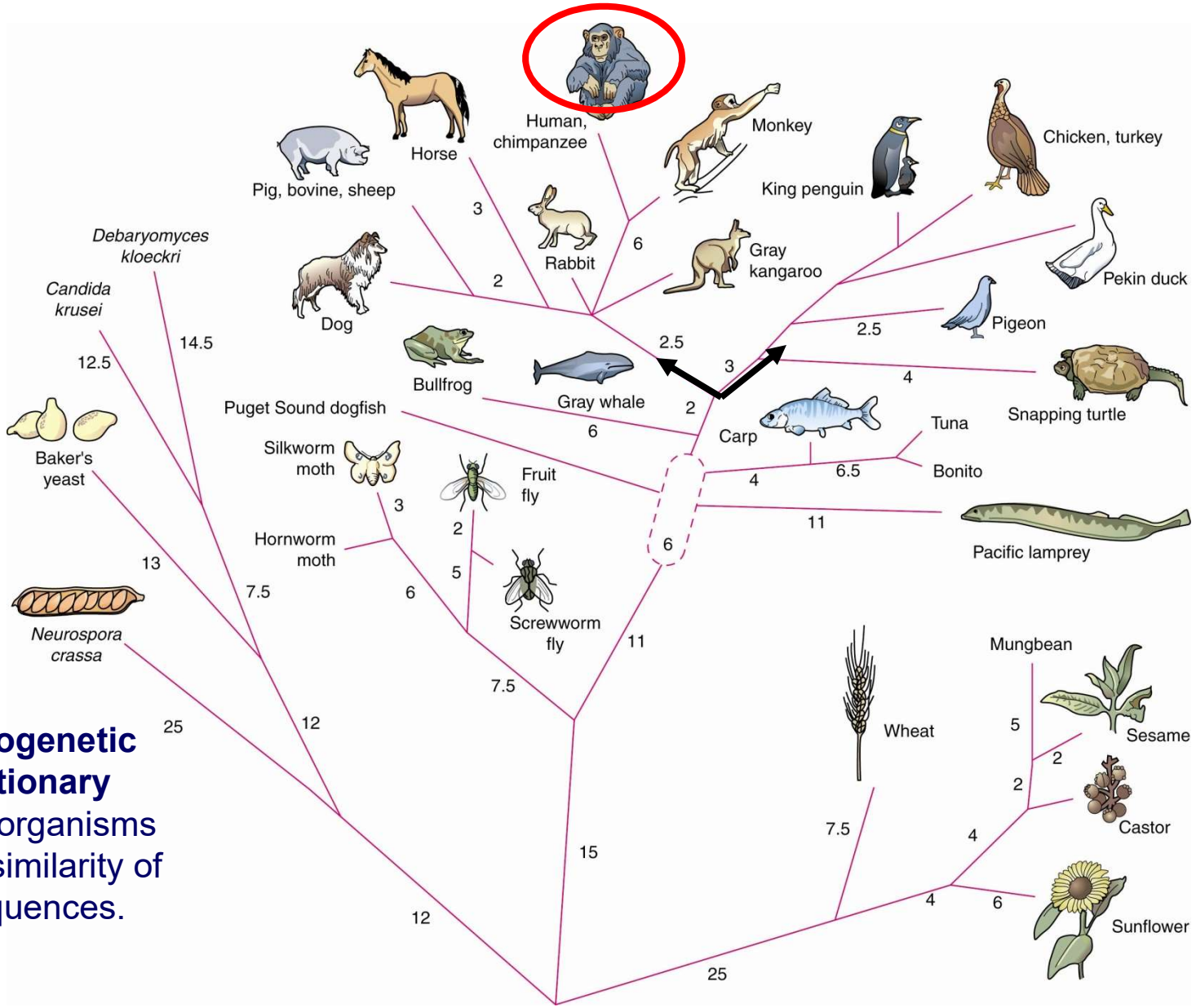
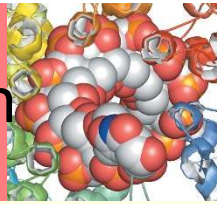


Figure 5.20 This **phylogenetic tree** depicts the **evolutionary relationships** among organisms as determined by the similarity of their cytochrome c sequences.

5.4c Related Proteins Show a Common Evolutionary Origin



The **α-globin and β-globin** chains share **64** residues of their approximately **140** residues in common. **Myoglobin and the α-globin** chain have **38** amino acid sequence identities.

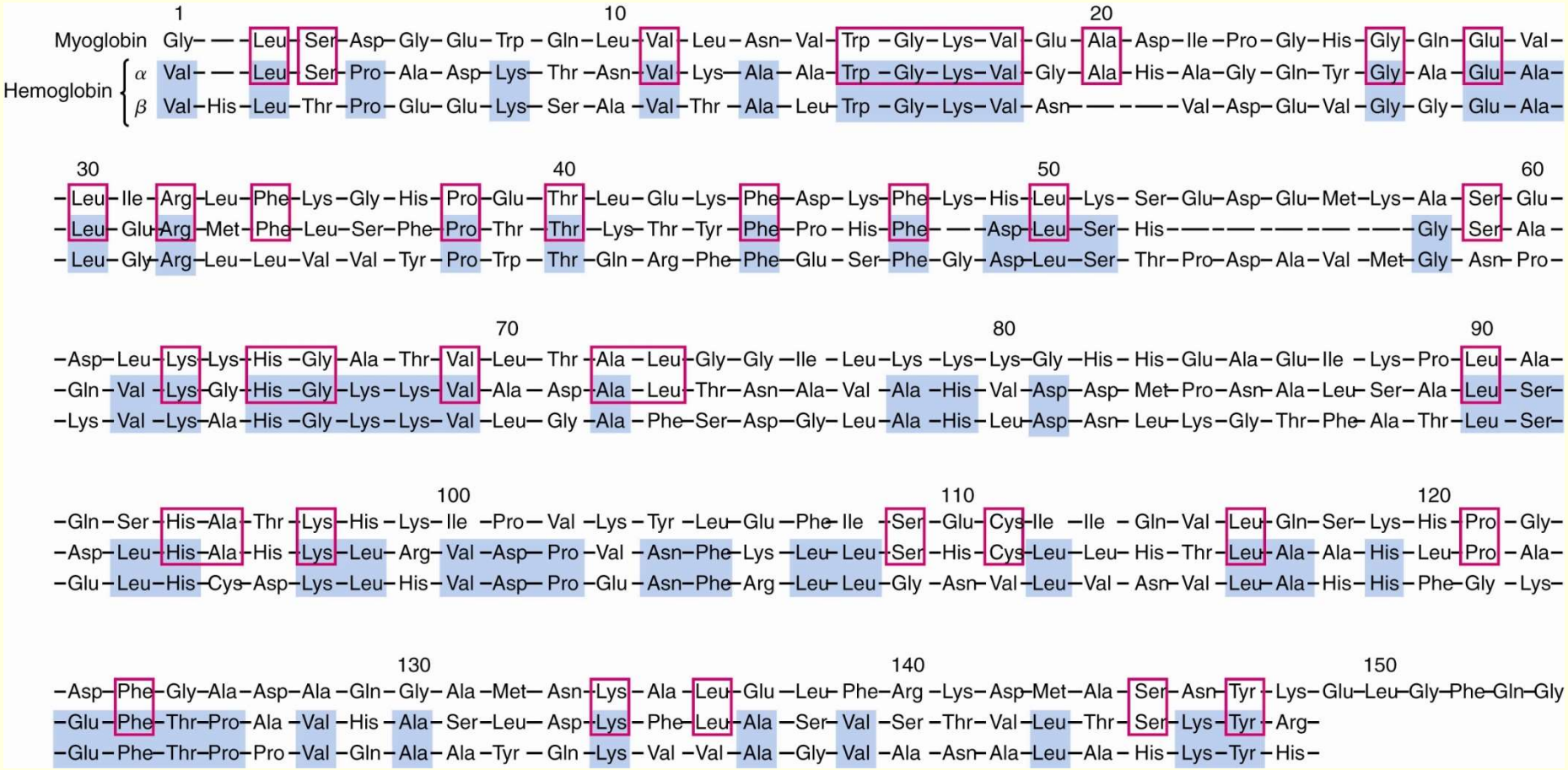
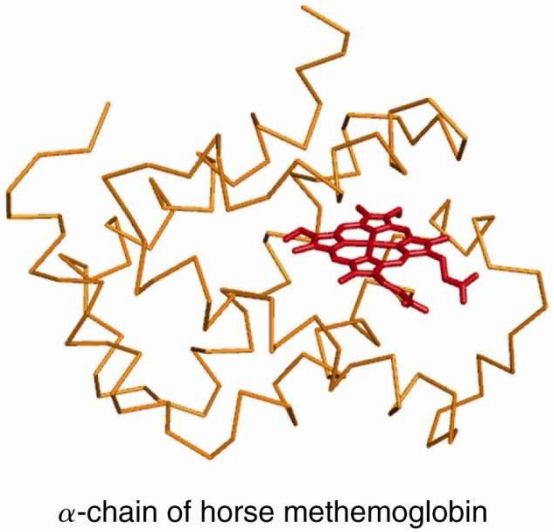


Figure 5.21 The amino acid sequences of the globin chains of human **hemoglobin** and **myoglobin** show a strong degree of homology.

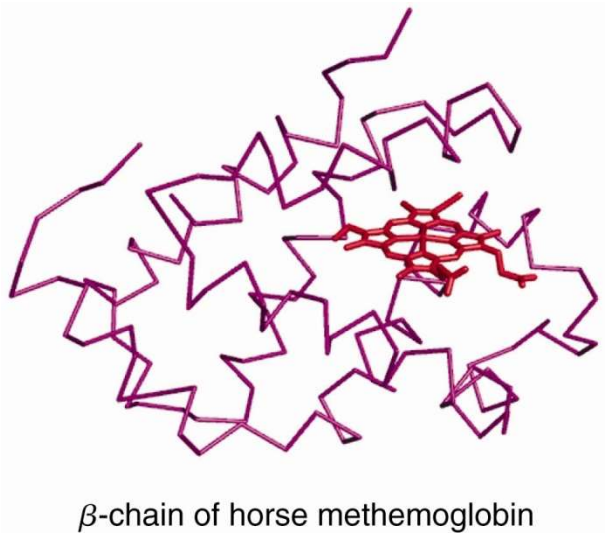
Oxygen-Binding Heme protein

Fig 5.21



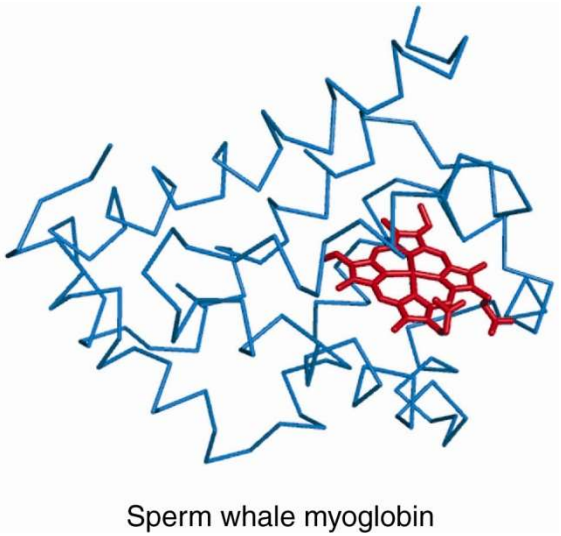
α-chain of horse methemoglobin

141 aa



β-chain of horse methemoglobin

146 aa



Sperm whale myoglobin

153 aa

The ability to **bind O2** via a heme prosthetic group is retained by all three of these polypeptides

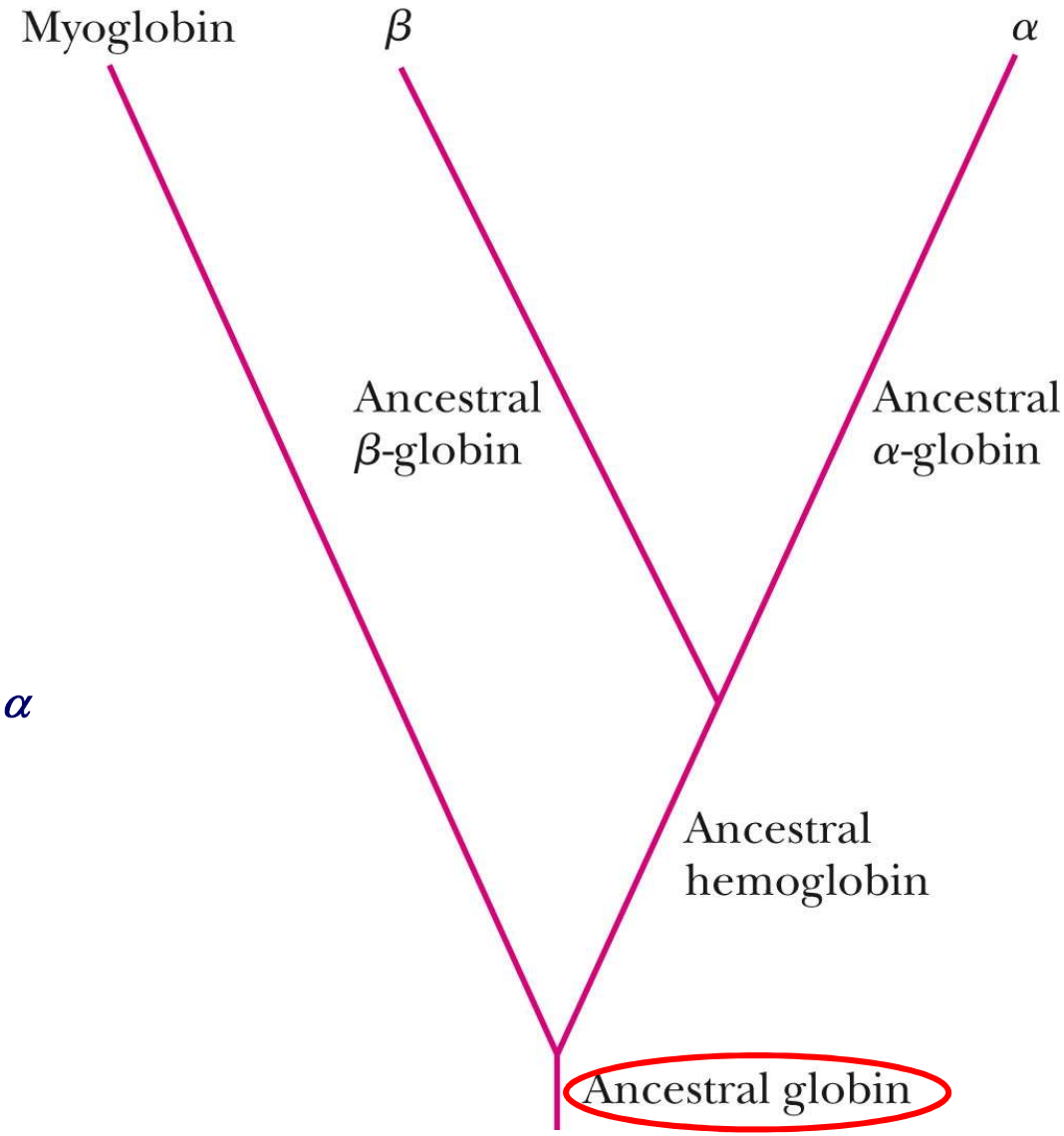
Oxygen-Binding Heme protein

Figure 5.22 This evolutionary tree is inferred from the homology between the amino acid sequences of the **α-globin, β-globin, and myoglobin chains.**

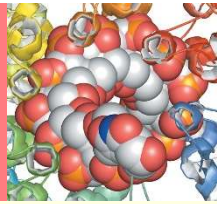
Duplication of an ancestral globin gene allowed the divergence of the **myoglobin** and ancestral **hemoglobin** genes.

Another gene **duplication** event subsequently gave rise to ancestral α and β forms, as indicated.

Gene duplication is an important evolutionary force in creating diversity.



5.4d Apparently Different Proteins May Share a Common Ancestry



- **Evolutionary relatedness** is inferred **from sequence homology**
- **Ex:lysozyme** (129aa) and **human milk α-lactalbumin (**123aa)
- These proteins are identical at **48** positions (**~37%** identity)
- **Functions** of these two are **not related**
- **lysozyme** hydrolyzes the polysaccharide wall of the bacterial cells
- human milk **α-lactalbumin** regulates milk sugar (lactose) synthesis
- **Similar 3D structures**
- It is conceivable that many proteins are related in this way, but **time** and the course of **evolutionary change erased** most evidence of their common ancestry

Similar structures, sequence homology, but different function

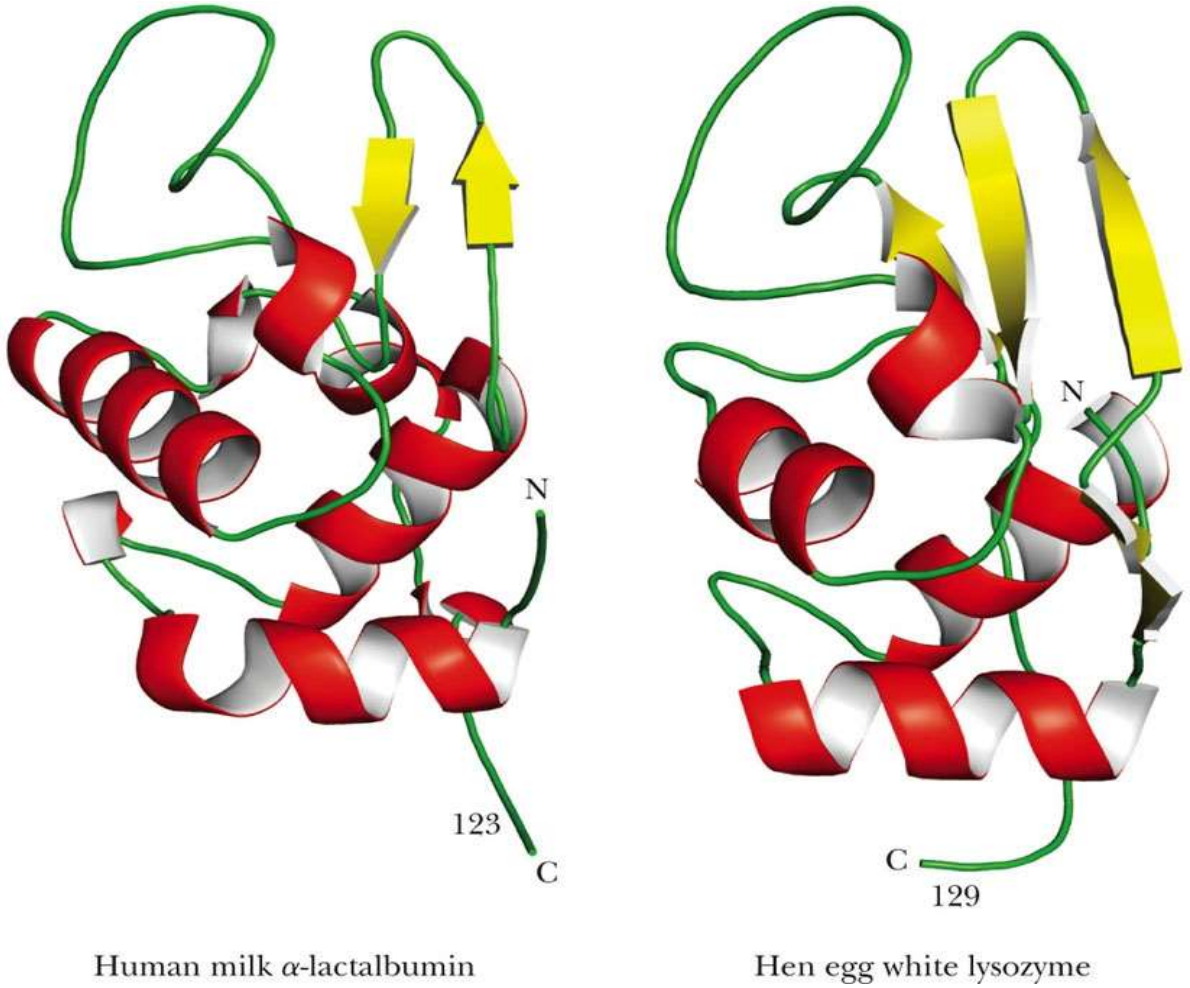


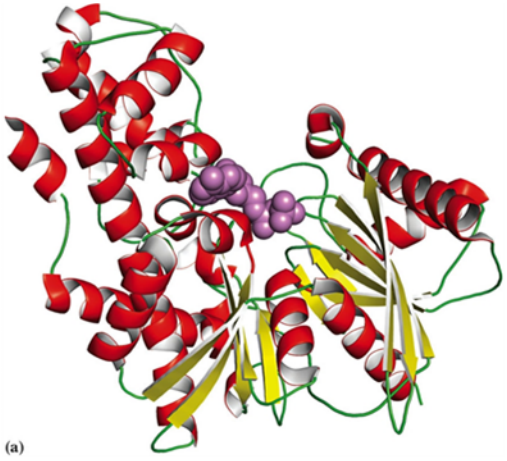
Figure 5.23 The tertiary structures of hen egg white lysozyme and human α-lactalbumin are very similar.

Similar structures, but different sequence and function

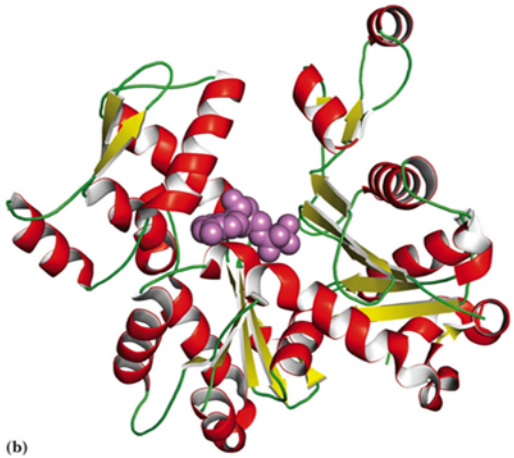
- **Evolutionary relatedness** can be inferred from **sequence homology**

G-actin & hexokinase

- **NO** sequence homology
- **Similar** 3D structure
- **Different** biological roles and physical properties
- **Actin** forms a filamentous polymer in muscle
- **Hexokinase** is a cytysolic enzyme that catalyzes the first reaction in glucose catabolism.

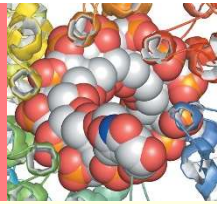


Hexokinase-ADP complex



Actin-ADP complex

5.4e A **Mutant Protein** is a Protein with a slightly Different Amino Acid Sequence



Some mutants: alternations in Oxygen affinity, heme affinity, stability, solubility, and subunit interactions

Some mutants NO apparent changes,

HbS (sickel-cell Hb) result in serious illness

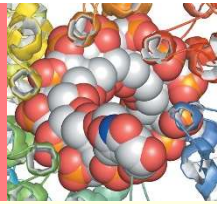
TABLE 5.4 Some Pathological Sequence Variants of Human Hemoglobin

Abnormal Hemoglobin*	Normal Residue and Position	Substitution
α-chain		
Torino	Phenylalanine 43	Valine
M_{Boston}	Histidine 58	Tyrosine
Chesapeake	Arginine 92	Leucine
$G_{Georgia}$	Proline 95	Leucine
Tarrant	Aspartate 126	Asparagine
Suresnes	Arginine 141	Histidine
β-chain		
S	Glutamate 6	Valine
Riverdale–Bronx	Glycine 24	Arginine
Genova	Leucine 28	Proline
Zurich	Histidine 63	Arginine
$M_{Milwaukee}$	Valine 67	Glutamate
$M_{Hyde\ Park}$	Histidine 92	Tyrosine
Yoshizuka	Asparagine 108	Aspartate
Hiroshima	Histidine 146	Aspartate

5.5 How Are Polypeptides Synthesized in the Laboratory?

- Chemical synthesis of peptides and polypeptides of defined sequence can be carried out in a laboratory.
- Strategies are complex because of the need to control side-chain reactions
- Blocking groups must be added and later removed
- du Vigneaud's synthesis of oxytocin in 1953 was a milestone
- Bruce Merrifield's solid phase method was even more significant

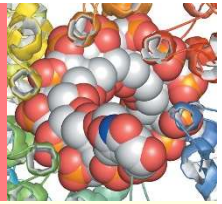
5.5 Can Polypeptides Be Synthesized in the Laboratory?



- Strategies are **complex** because of the need to control **side chain reactions**
- **Blocking groups** must be added and later removed

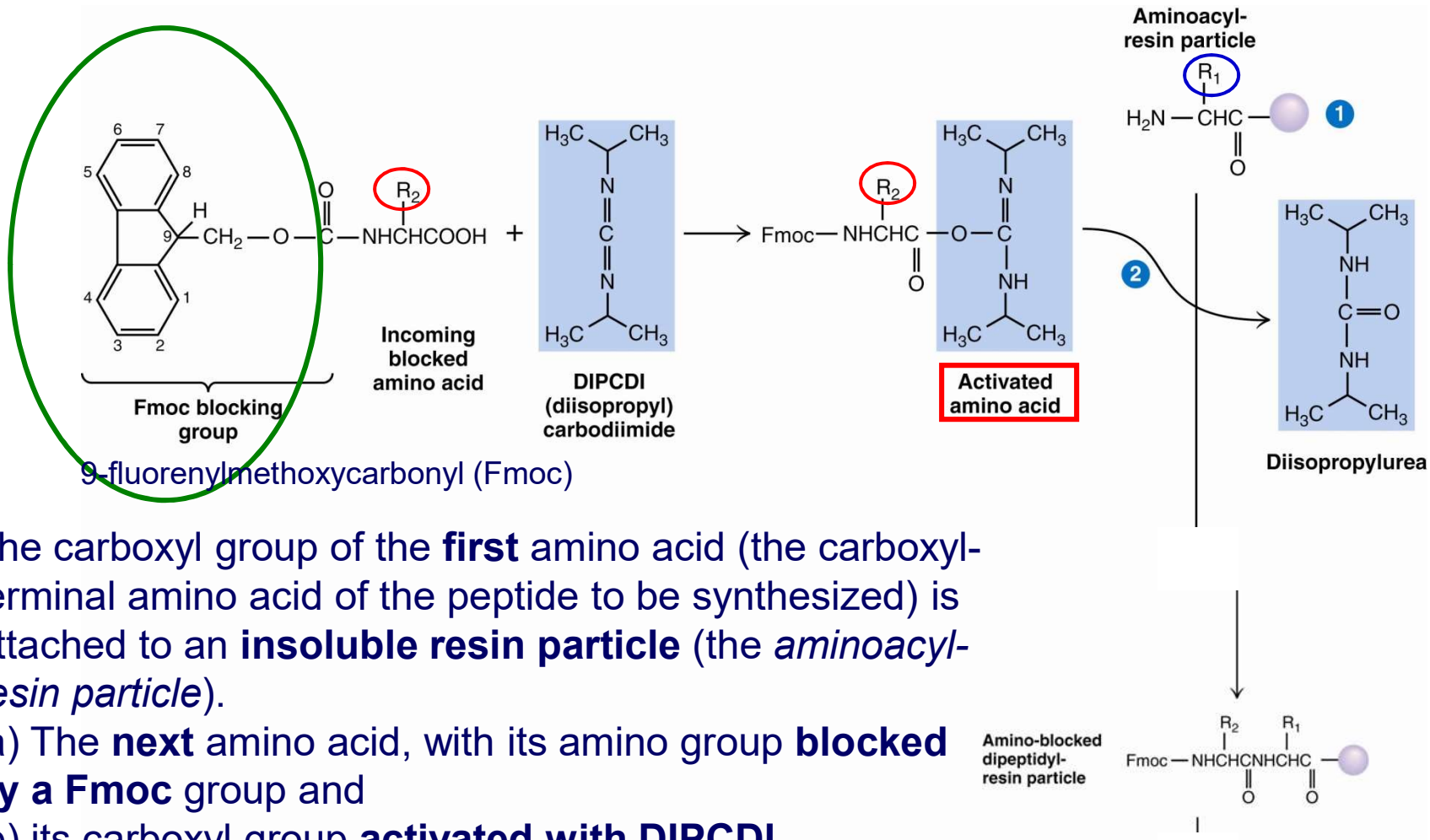
- **Vincent du Vigneaud's** (Nobel Prize in Chemistry 1955) synthesis of **oxytocin** (腦垂體後葉激素) in 1953 was a milestone
- **Bruce Merrifield's** (Nobel Prize in Chemistry 1984) **solid phase method** was even more significant
- **Kary B. Mullis** (Nobel Prize in Chemistry 1993) "for his invention of the **polymerase chain reaction** (PCR) method"

5.5b Solid-Phase Methods Are Very Useful in Peptide Synthesis



- **Carboxy terminus** of a nascent peptide is covalently anchored to an **insoluble resin**
- After each **addition of a residue**, the resin particles are collected by filtration
- Automation and computer control now permit synthesis of peptides of **30 residues or more**

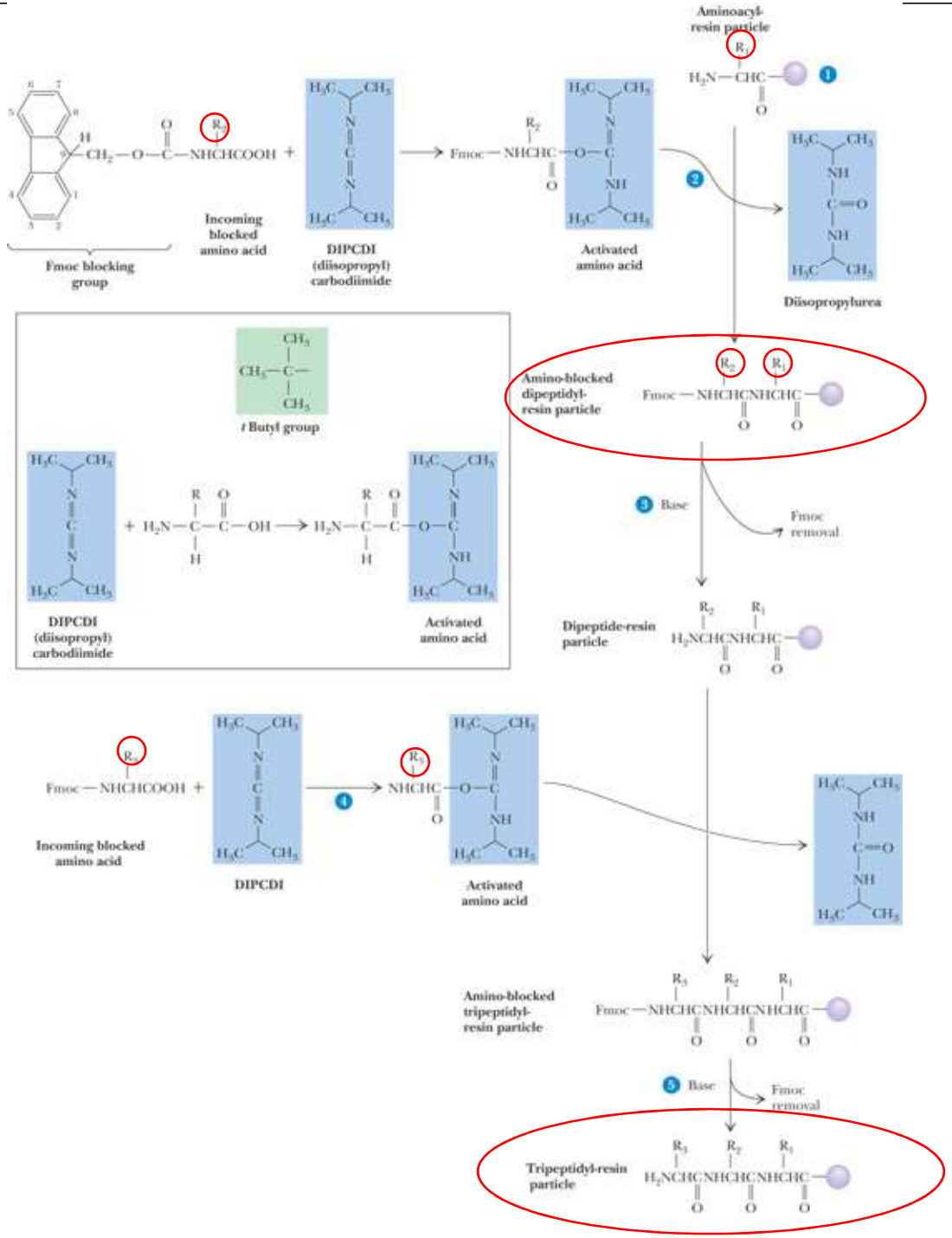
Figure 5.25 **Solid-phase synthesis of a peptide**.



9-fluorenylmethoxycarbonyl (Fmoc)

1) The carboxyl group of the **first** amino acid (the carboxyl-terminal amino acid of the peptide to be synthesized) is attached to an **insoluble resin particle** (the *aminoacyl-resin particle*).
   (a) The **next** amino acid, with its amino group **blocked by a Fmoc** group and
   (b) its carboxyl group **activated with DIPCDI**,

2) is reacted with the aminoacyl-resin particle to form a peptide linkage, with elimination of DCCD as diisopropylurea

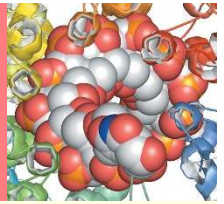

Aminoacyl-resin particle
Fmoc blocking group
Incoming blocked amino acid
DIPCDI (diisopropyl) carbodiimide
Activated amino acid
Diisopropylurea
Amino-blocked dipeptidyl-resin particle
t Butyl group
Base
Fmoc removal
Dipeptide-resin particle
Incoming blocked amino acid
DIPCDI
Activated amino acid
Amino-blocked tripeptidyl-resin particle
Base
Fmoc removal
Tripeptidyl-resin particle

5.6 Do Proteins Have Chemical Groups Other Than Amino Acids?

Proteins may be "conjugated" with other chemical groups

- If the non-amino acid part of the protein is important to its function, it is called a prosthetic group.
- Be familiar with the terms: glycoprotein, lipoprotein, nucleoprotein, phosphoprotein, metalloprotein, hemoprotein, flavoprotein.
- Post-translational modifications are chemical changes made to the protein after synthesis

5.6 Do Proteins Have Chemical Groups Other Than Amino Acids?



Proteins may be "conjugated" with other chemical groups

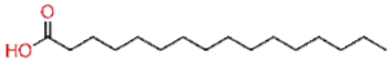
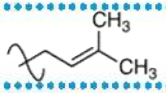
- If the **non**-amino acid part of the protein is important to its **function**, it is called a **prosthetic** group.
- May other protein contain various chemical constituents as as an **integral** part of their structure.
- Through **covalent modification** of aa side chain in proteins after the protein has been synthesized.

- Such alternations are called **post-translational** modification
- **Glycoprotein**,
- **Lipoprotein,**
- **Nucleoprotein,**
- **Phosphoprotein,**
- **Metalloprotein,**
- **Hemoprotein,**
- **Flavoprotein**.

Post-Translational Modification

"**On-off switches**" that regular the function of cellular location on the protein

TABLE 5.5 Some Prominent Post-Translational Modifications Found in Proteins

Name	Nonprotein Part	Amino Acid Side Chain Modified	Examples
Phosphorylation	$-PO_3^{2-}$	S, T, Y	Hormone receptors, regulatory enzymes
Acetylation	$-CH_2COO^-$	K	Histones, metabolic enzymes
Methylation	$-CH_3$	K, R	Histones
Acylation	Palmitic acid 	C	G-protein-coupled receptors (GPCR)
Prenylation	Prenyl group 	C	Ras p21
ADP-ribosylation	ADP-ribose	H, R	G proteins, eukaryotic elongation factors
Adenylylation	AMP	Y	Glutamine synthetase

Proteins containing noprotein constituents is **"Conjugated Proteins"**

TABLE 5.6 Some Common Conjugated Proteins

Name	Nonprotein Part	Association	Examples
Lipoproteins	Lipids	Noncovalent	Blood lipoprotein complexes (HDL, LDL)
Nucleoproteins	RNA, DNA	Noncovalent	Ribosomes, chromosomes
Glycoproteins	Carbohydrate groups	Covalent	Immunoglobulins, LDL receptor
Metalloproteins and metal-activated proteins	Ca^{2+}, K^{+}, Fe^{2+}, Zn^{2+}, Co^{2+}, others	Covalent to noncovalent	Metabolic enzymes, kinases, phosphatases, among others
Hemoproteins	Heme group	Covalent or noncovalent	Hemoglobin, cytochromes
Flavoproteins	FMN, FAD	Covalent or noncovalent	Electron transfer enzymes

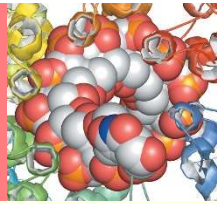
5.7 What Are the Many Biological Functions of Proteins?

- Many proteins are **enzymes**
- **Regulatory proteins** control metabolism and gene expression
- Many **DNA-binding proteins** are gene-regulatory proteins
- **Transport proteins** carry substances from one place to another or across membranes
- **Storage proteins** serve as reservoirs of amino acids or other nutrients

5.7 What Are the Many Biological Functions of Proteins?

- Movement is accomplished by **contractile and motile proteins**
- Many proteins serve **a structural role**
- Proteins of **signaling pathways** include scaffold proteins (adapter proteins)
- **Other proteins** have protective and exploitive functions
- A few proteins have **exotic functions**

5.7 What Are the Many Biological Functions of Proteins?



“Protein” are the agents of biological function

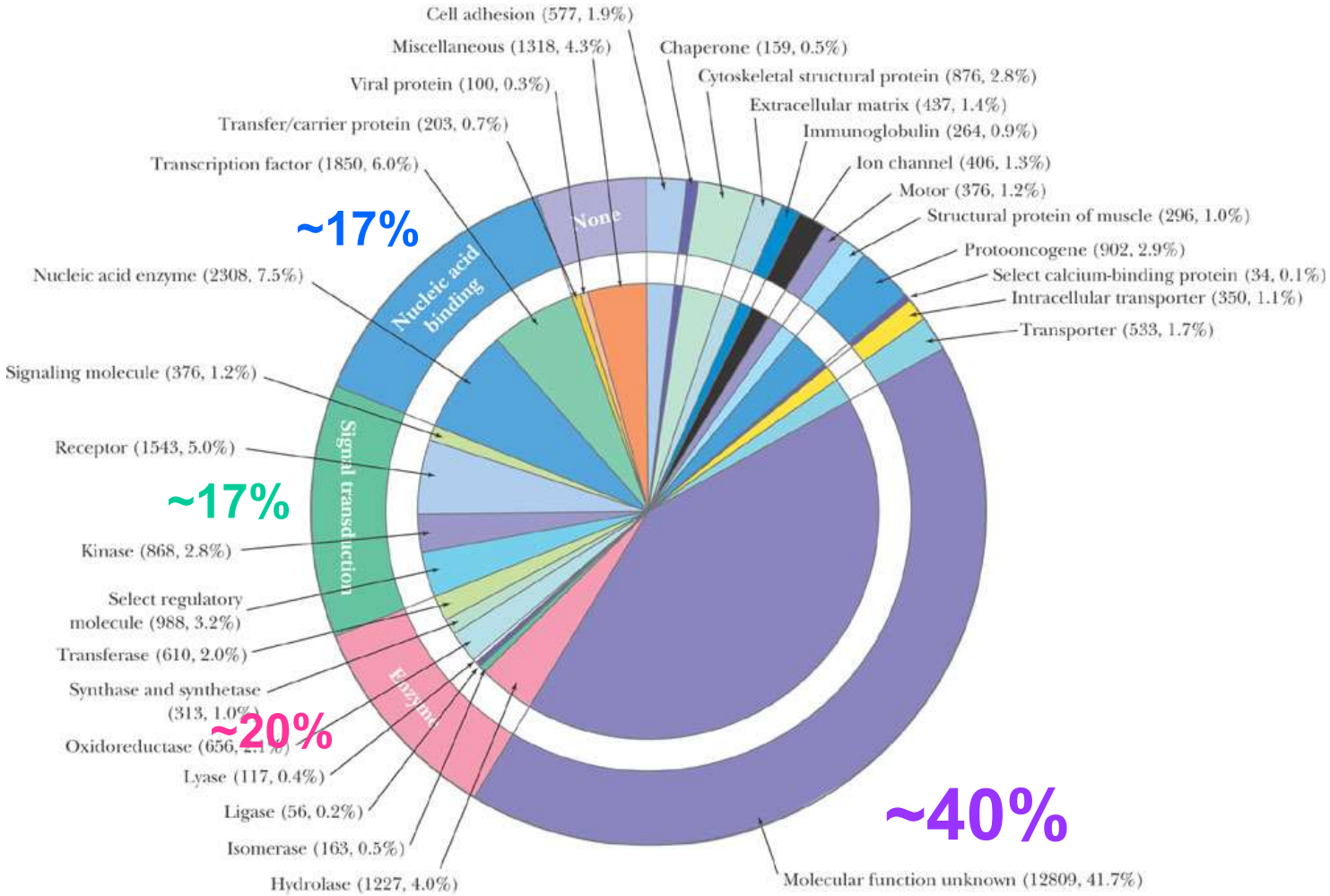


Figure 5.26 **Proteins of the human genome grouped according to their molecular function**.

Protein fill essentially every biological role, with the exception of information storage

- **Binding proteins**: interact nocovalently with their specific ligands
- **Transport proteins**
  Membrane proteins: transport substrate across membrane
  Soluble proteins: deliver nutrients or waste throughout the body
- **Catalytic proteins (Enzymes**) mediate most every metabolic reaction
- **Regulatory proteins**: bind to specific nucleotide sequences within DNA control gene expression
- **Homones**: deliver this information to cells when they bind to specific receptors
- **Switch proteins**: G-protein can switch between 2 conformational states "ON & Off"
- **Structure proteins**: give form to cells and subcellular structure

- **All proteins Function through Specific Recognition and Binding of Some Target Molecule**

Protein Binding

The simplest situation: A protein has a single binding site

$$PL \rightleftharpoons P + L$$

When [L]>> [P],

[PL] /([P]+[PL]) versus [L] plot,

Hyperbolic curve

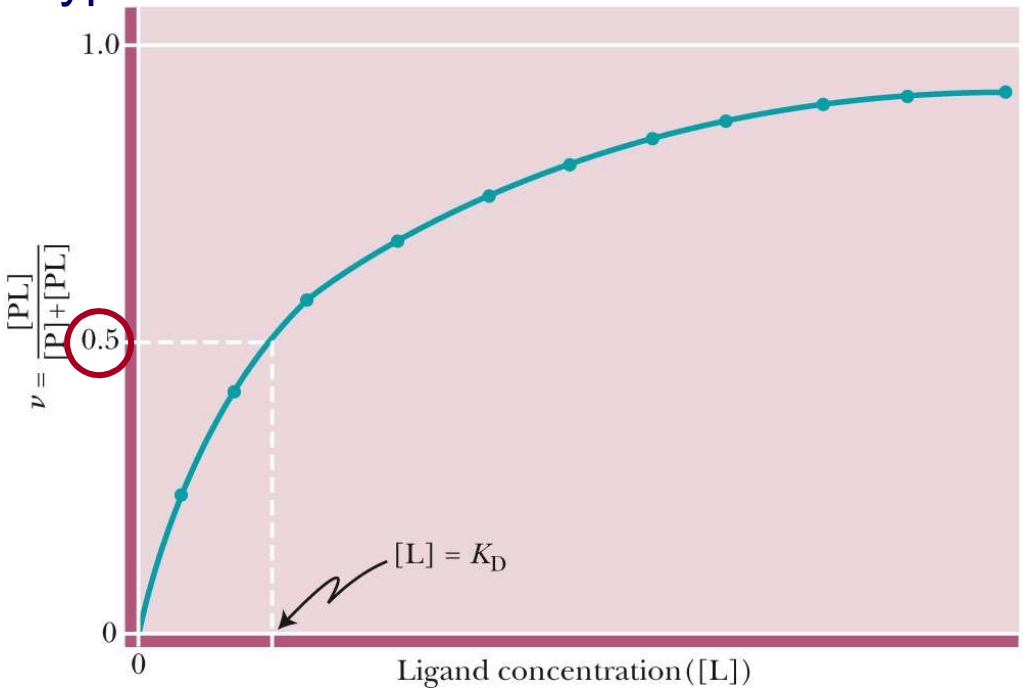


Figure 5.27 Saturation curve or binding isotherm.

Equilibrium constant

$K_{eq} = [P][L]/[PL]$

K_D : Dissociation constants

υ : Fractional saturation of P with L

= bound / free + bound

=[PL] / ([P]+[PL])

[P]=[PL] (1- υ / υ)

$\upsilon = [L] / (K_D + [L])$

υ : =0.5, $[L] = K_D$

half protein were bound with ligand

υ smaller, the better binding

A small K_D means that the protein binds the ligand strongly

K_D 10^{-3} to 10^{-12} M

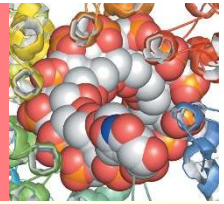
- **Ligand-Binding Site**

- Ligand binding occurs through **noncovalent** interaction between the protein and ligand.
- Binding is readily **reversible**
- Protein display **specificity** in ligand binding with a specific site
- Undergo a **conformational change**
- Provide a **better fit**
- **Ligand-induced conformational changes,** and the result is an even stable interaction between the protein and its ligand.

5.8 What is the Proteome and What Does It Tell Us?

- The full genetic potential of a cell is contained within its genome
- A more accurate reflection of what a cell is doing at any moment is found in the **proteome**
- The proteome is dynamic and may consist of hundreds of thousands of proteins, as the result of post-translational modification, alternative RNA splicing, and RNA editing
- The tools of proteomics are global purification strategies to separate complex mixtures, followed by sequence determination using mass spectroscopy to both identify and quantify each of the different proteins present

5.8 What is the Proteome and What Does It Tell Us?



- The **full genetic potential of a cell** is contained within its genome
- A more **accurate** reflection of what a cell is doing at any moment is found in the **proteome**
- The **proteome is dynamic** and may consist of hundreds of thousands of proteins, as the result of **post-translational modification**, **alternative RNA splicing**, and **RNA editing**
- The tools of proteomics are **global** purification strategies to **separate** complex mixtures, followed by **sequence determination** using mass spectroscopy to both **identify** and **quantify** each of the **different proteins present**

End of Chapter 5

2023/10/2, 10th hr