CEM 882

Lecture Notes 1

Welcky

Chemistry 882: Thermodynamics with Applications to Biomolecules

Spring 2024 TuTh, 10:20 a.m. – 11:40 a.m.
http://www.chemistry.msu.edu/courses/CEM882/cem882.htm

Instructor

Dr. David Weliky
Room 41 Chemistry
353-1177
weliky@chemistry.msu.edu

Textbooks (on reserve in Main Library)

1. K. E. van Holde, W. C. Johnson, and P. S. Ho, Principles of Physical Biochemistry, second edition (2006).
2. K. A. Dill and S. Bromberg, Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology, (2003) – there are also electronic copies available through the library website
3. D. Voet and J. G. Voet, Biochemistry 3rd edition, 2004

Reference Books (on reserve in Main Library)

1. Donald A. McQuarrie and John D. Simon, Physical Chemistry: A Molecular Approach
2. James R. Barrante, Applied Mathematics for Physical Chemistry
3. Charles R. Cantor and Paul R. Schimmel, Biophysical Chemistry (volumes 1-3)
4. Michel Daune, Molecular Biophysics, Structures in Motion
5. Rodney Cotterill, Biophysics: An Introduction
6. Charles Kittel and Herbert Kroemer, Thermal Physics

Course Homepage: http://www.chemistry.msu.edu/courses/CEM882/cem882.htm

Lecture and Course notes, homework problem sets, homework and exam solution sets, and course announcements can be found on the course homepage.

Syllabus (approximate)

Week	Topic	Reading
January 8 – 19	Survey of Biomolecular Structure	VV, Chapter 4, 5.1-5.3, 8.1, 8.3-8.5 VJH, 1.5
January 22 – 26	Probability and Entropy	DB, Chapters 1-3 VJH, Chapter 2.1,2.2
Jan. 29 – Feb. 2	Boltzmann Law and Partition Function	DB, Chapters 6,7,10
February 5 – 16	Thermodynamic Functions and Laws	DB, Chapters 8,10 VJH Chapters 2.3,2.4
Feb. 19 – 23	Calorimetry	VJH Chapter 2.5
March 4 – 22	Ideal Gas and Chemical Equilibrium	DB Chapter 11,13

March 25 – April 5	Solutions and Chemical Potential	DB Chapters 9,15,16
April 8 – 12	Intermolecular Interactions and Electrostatics	DB Chapters 22,24
April 15 – 19	Water and the Hydrophobic Effect	DB Chapters 29,30

Homework

There will be about five homework assignments in the course. You can work with your classmates on solving the problems but what you hand in should not be a direct copy of another student's work.

Exams

There will be two exams in the course.

I will provide a review sheet prior to each exam. In addition, fundamental constants will be provided on the exam. You can also bring in one page to each exam with any information that you think is important.

You should show all of your work and explain your reasoning on homework and exams. Most of the grading credit will be based on your reasoning.

Grading

The final grade will be a combination of exam grades (~2/3 contribution) and homework grades (~1/3 contribution).

The approximate final grading scale will likely be:

4.0 80% – 100%
3.5 70% – 80%
3.0 60% – 70%
2.5 50% – 60%
2.0 40% – 50%
1.5 30% – 40%
1.0 20% – 30%
0.0 0% – 20%

Many "small" organic molecules (< 1kDa) are important for life.

There are also four classes of larger molecules: carbohydrates (polymers of monomer units like glucose, fructose, mannose), proteins (polymers of amino acids), nucleic acids (polymers of nucleotides) and lipids (self-assemble to form bilayers). There are also conjugates among the classes

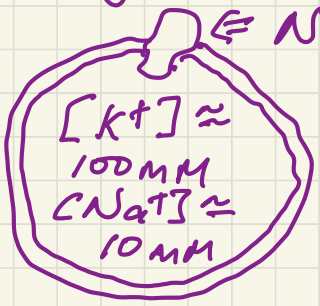
protein + carbohydrate (glycoprotein)
lipid + carbohydrate (glycolipid)
protein + lipid (lipoprotein)

Typical MW of a protein is in the range of $5 \times 10^3 - 10^6$ Da

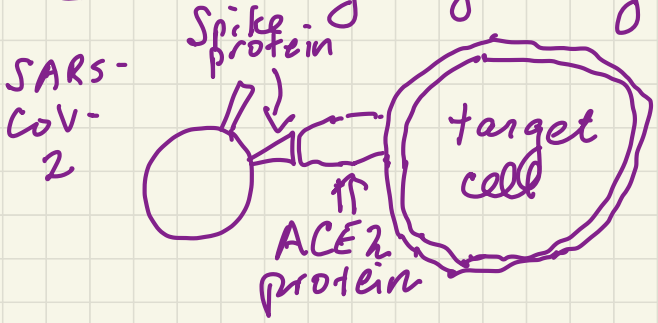
Proteins have diverse functions in life

① Structural integrity + motion
muscle protein like titin

② transport of molecules
hemoglobin protein transports O_2



③ Binding/Signaling

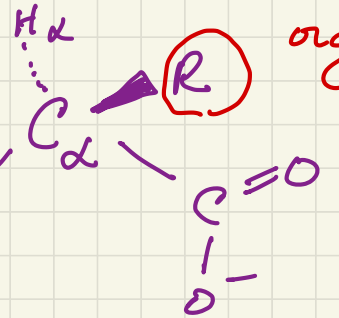


(4) catalysis of chemical reactions
enzymes

SARS-CoV-2 protease (hydrolytically cleaves protein) => inhibitory target of Pfizer drug which is highly effective at mitigating COVID if taken within 5 days of onset of symptoms => protease inhibitors were first developed in 1990's to treat HIV which is a virus similar to SARS-CoV-2 => part of COVID treatment regimen is a HIV protease inhibitor Ritonavir through its "off-target" => inhibition of cytochrome P450, a human enzyme that hydroxylates organic molecules

Many evolved proteins in an organism have well-defined 3D geometric structure

Proteins are linear polymers of amino acids



Note "L-enantiomer" → stereochemistry

organic group that distinguishes one amino acid from another

20 common amino acids in proteins

Chemical (primary) structure of proteins

residue: $R_1 R_2 R_3 R_4 R_5 \ldots\ldots$

AVGIG

polymer bond is amide ("peptide bond")

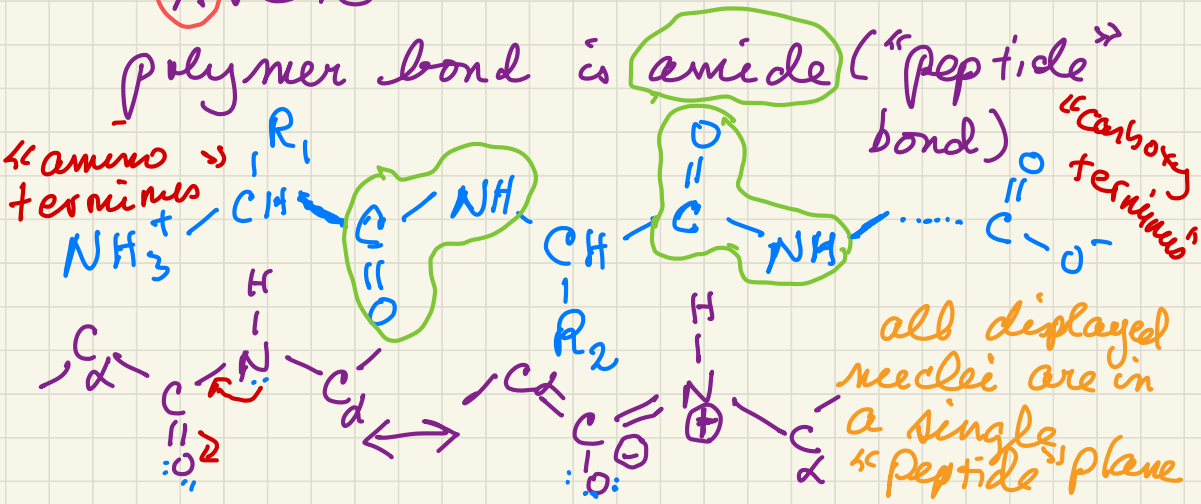
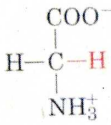
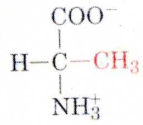
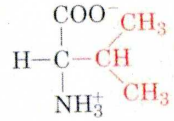
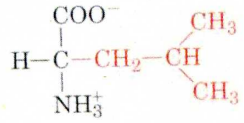
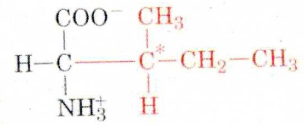
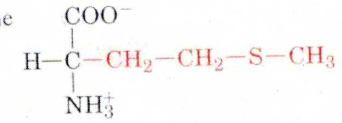
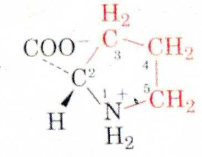
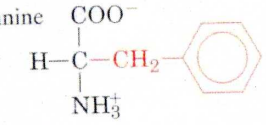
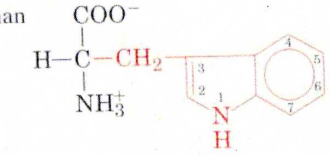


TABLE 4-1 Covalent Structures and Abbreviations of the "Standard" Amino Acids of Proteins, Their Occurrence, and the p*K* Values of Their Ionizable Groups

Name, Three-letter Symbol, and One-letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurrence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+[d]	pK_R Side Chain[d]
Amino acids with nonpolar side chains						
Glycine Gly G		57.0	6.8	2.35	9.78	
Alanine Ala A		71.1	7.6	2.35	9.87	
Valine Val V		99.1	6.6	2.29	9.74	
Leucine Leu L		113.2	9.5	2.33	9.74	
Isoleucine Ile I		113.2	5.8	2.32	9.76	
Methionine Met M		131.2	2.4	2.13	9.28	
Proline Pro P		97.1	5.0	1.95	10.64	
Phenylalanine Phe F		147.2	4.1	2.20	9.31	
Tryptophan Trp W		186.2	1.2	2.46	9.41	

(*continued*)

[a]The ionic forms shown are those predominating at pH 7.0 (except for that of histidine[e]), although residue mass is given for the neutral compound. The C_α atoms, as well as those atoms marked with an asterisk, are chiral centers with configurations as indicated according to Fischer projection formulas. The standard organic numbering system is provided for heterocycles.

[b]The residue masses are given for the neutral residues. For molecular masses of the parent amino acids, add 18.0 D, the molecular mass of H_2O, to the residue masses. For side chain masses, subtract 56.0 D, the formula mass of a peptide group, from the residue masses.

[c]The average amino acid composition in the complete SWISS-PROT database (http://www.expasy.ch/sprot), Release 40.7.

[d]From Dawson, R.M.C., Elliott, D.C., Elliott, W.H., and Jones, K.M., *Data for Biochemical Research* (3rd ed.), pp. 1–31, Oxford Science Publications (1986).

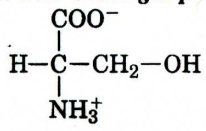
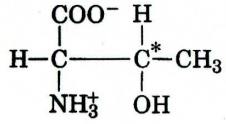
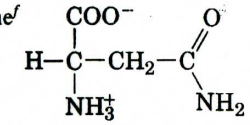
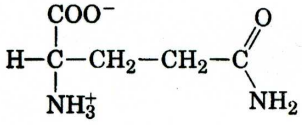
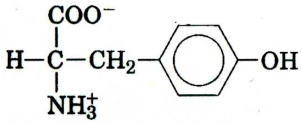
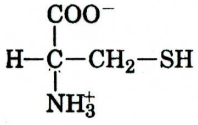
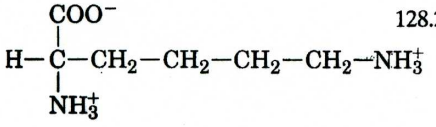
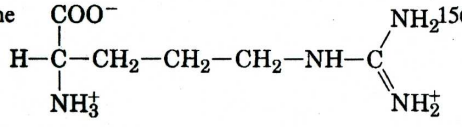
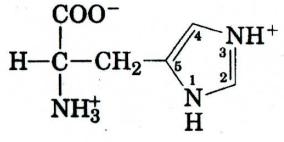
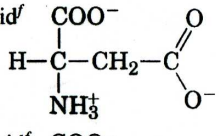
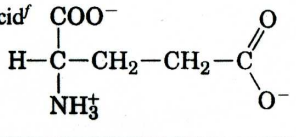
[e]Both the neutral and protonated forms of histidine are present at pH 7.0 because its pK_R is close to 7.0. The imidazole ring of histidine is numbered here according to the biochemistry convention. In the IUPAC convention, N3 of the biochemistry convention is designated N1 and the numbering increases clockwise around the ring.

[f]The three- and one-letter symbols for asparagine *or* aspartic acid are Asx and B, whereas for glutamine *or* glutamic acid they are Glx and Z. The one-letter symbol for an undetermined or "nonstandard" amino acid is X.

R groups in red

(1)

TABLE 4-1 *(continued)*

Name, Three-letter Symbol, and One-letter Symbol	Structural Formula[a]	Residue Mass (D)[b]	Average Occurrence in Proteins (%)[c]	pK_1 α-COOH[d]	pK_2 α-NH_3^+[d]	pK_R Side Chain[d]
Amino acids with uncharged polar side chains						
Serine Ser S		87.1	7.1	2.19	9.21	
Threonine Thr T		101.1	5.6	2.09	9.10	
Asparagine[f] Asn N		114.1	4.3	2.14	8.72	
Glutamine[f] Gln Q		128.1	3.9	2.17	9.13	
Tyrosine Tyr Y		163.2	3.2	2.20	9.21	10.46 (phenol)
Cysteine Cys C		103.1	1.6	1.92	10.70	8.37 (sulfhydryl)
Amino acids with charged polar side chains						
Lysine Lys K		128.2	6.0	2.16	9.06	10.54 (ε-NH_3^+)
Arginine Arg R		156.2	5.2	1.82	8.99	12.48 (guanidino)
Histidine[e] His H		137.1	2.2	1.80	9.33	6.04 (imidazole)
Aspartic acid[f] Asp D		115.1	5.2	1.99	9.90	3.90 (β-COOH)
Glutamic acid[f] Glu E		129.1	6.5	2.10	9.47	4.07 (γ-COOH)

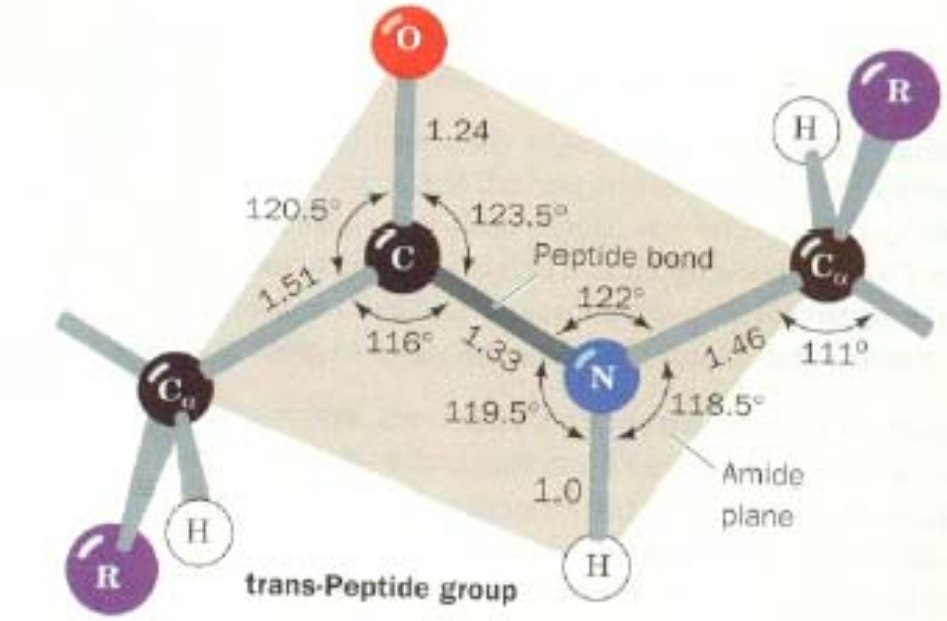


FIGURE 7-1. The standard dimensions (in angstroms, Å, and degrees,°) of the planar trans-peptide group derived by averaging the results of X-ray crystal structure determinations of amino acids and peptides. [After Marsh, R.E. and Donohue, J., *Adv. Protein Chem.* **22**, 249 (1967).]

less common

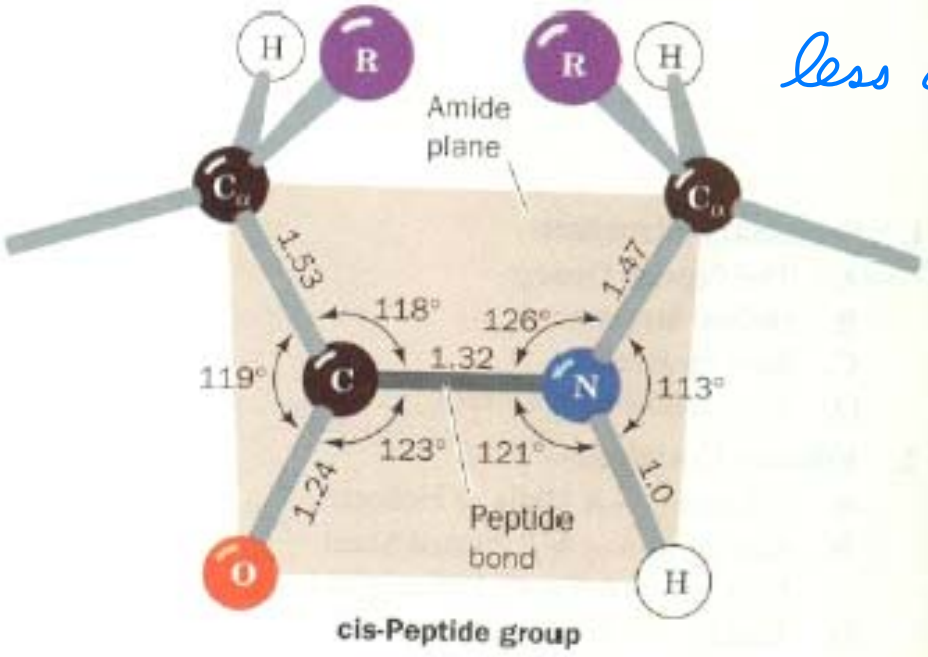
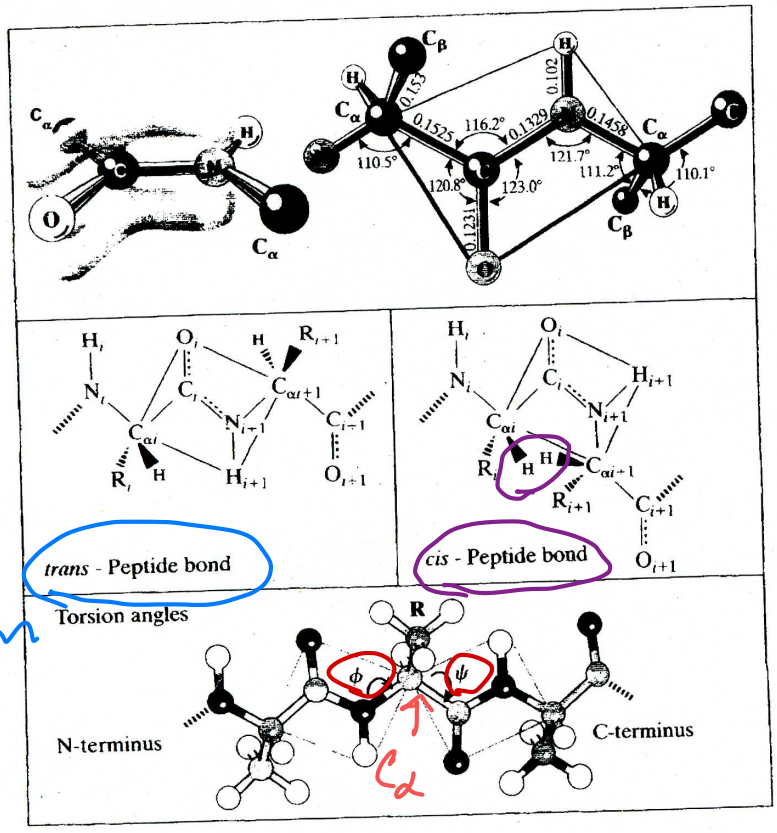


FIGURE 7-2. The cis-peptide group.



much more common

much less common

Figure 1.21 The peptide bond. The peptide bond that chemically links two amino acid residues along a polypeptide chain is a C—N bond of an amide linkage. The bond shows partial double-bond character, with the π-electrons distributed between the O—C—N atoms (top panel), and is not freely rotating [bond lengths and angles from Engh and Huber (1991), *Acta Crystallogr. A* **47**: 392–400]. The bond is found predominantly in the *trans*-configuration, but could also adopt the *cis*-form (middle panel). The two freely rotating bonds on either side of the C_α-carbon thus define the torsion angles ϕ (for rotation about the N—C_α bond) and ψ (for rotation about the Ca—C bond).

$\varphi, \psi \equiv$ twist (dihedral) angles about N-C_α (φ) and C_α-CO(ψ) single bonds

Set of φ, ψ along protein polymer backbone $\equiv$ secondary structure

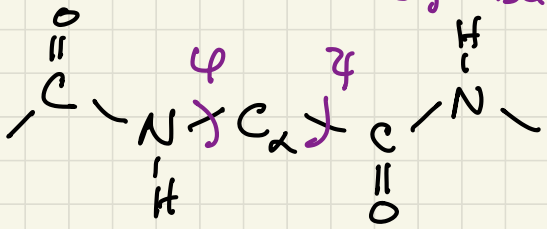
sequence of amino acids (N→C) $\equiv$ primary structure

VH

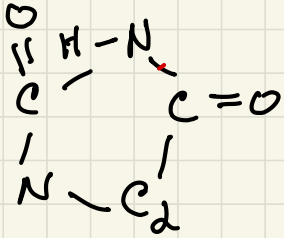
φ : $-180 \rightarrow 180^\circ$

ψ : $-180 \rightarrow 180^\circ$

$\varphi = 180^\circ, \psi = 180^\circ \Rightarrow$ fully-extended co-planar geometry of backbone



$\varphi = 0^\circ, \psi = 0^\circ \Rightarrow$ least-extended co-planar geometry



Ramachandran Plot

green regions have favorable steric energies

Experimental ϕ, ψ are typically in sterically-favored regions

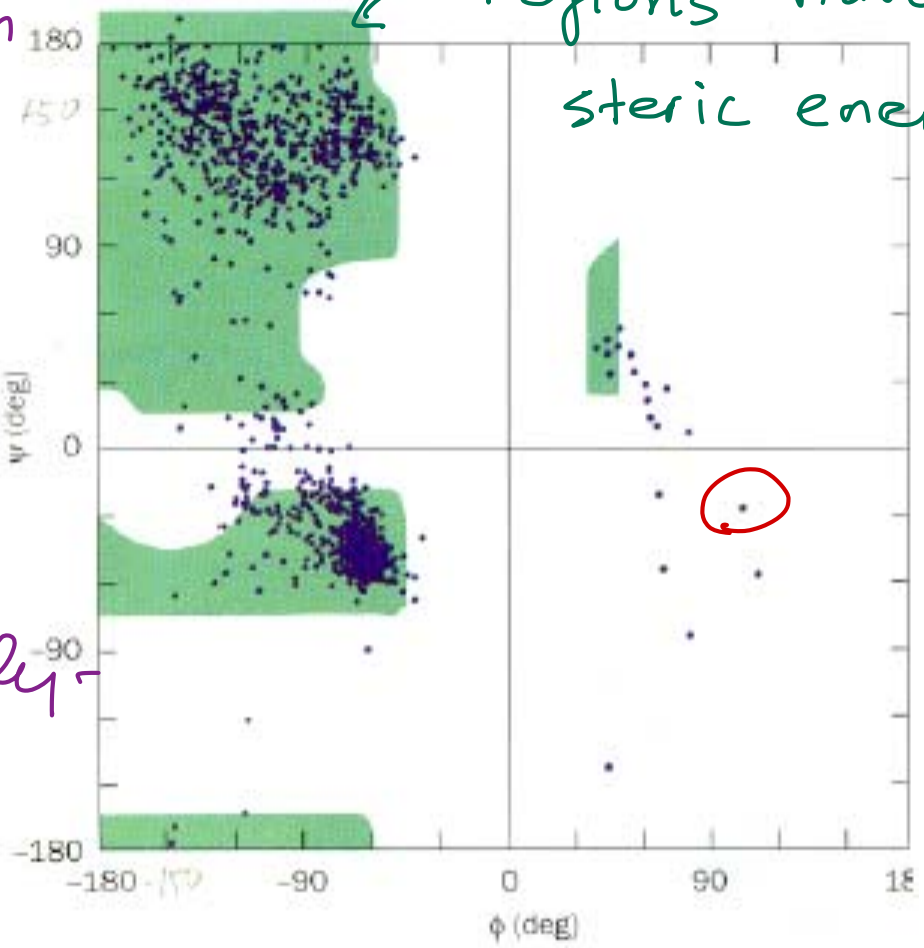
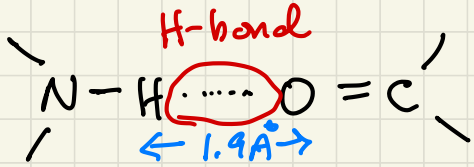


FIGURE 7-8. The conformation angle distribution of all residues but Gly and Pro in 12 precisely determined high-resolution X-ray structures with a superimposed Ramachandran diagram. [After Richardson, J.S. and Richardson, D.C., *in* Fasman, G.D. (Ed.), *Prediction of Protein Structure and the Principles of Protein Conformation*, p. 6, Plenum Press (1989).]

ϕ, ψ values from a single residue in a single protein

Specific ϕ, ψ regions result in regular (periodic) H-bonding for backbone

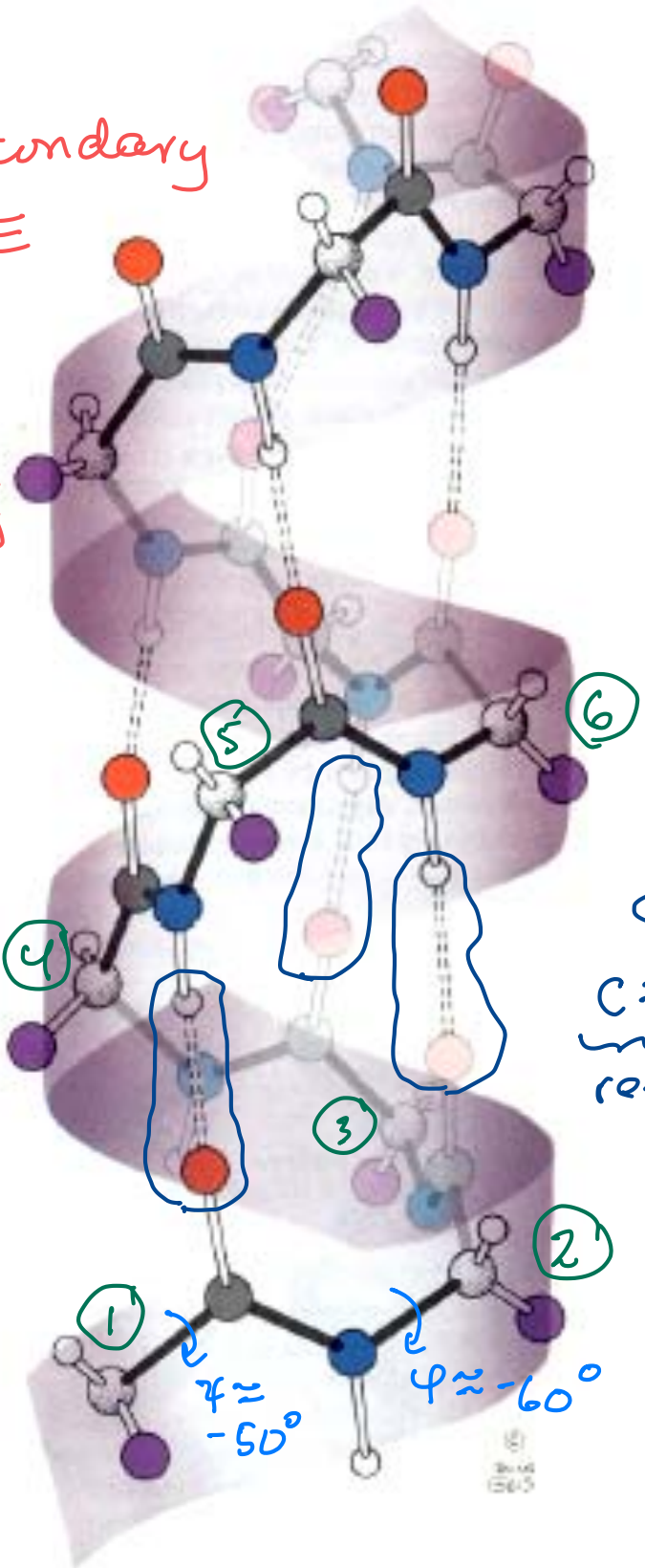
H-bond ≡ partial sharing of lone pair e^-'s of N, O, or F with H



Sum of vanderWaals radii of non-bonded H and O ≈ 2.7 Å

right-handed α helix

(regular secondary structure ≡ periodic pattern of H-bonding



≡

C=O ··· H-N

residue i residue i+4

Residue numbers

FIGURE 7-11. The right-handed α helix. Hydrogen bonds between the N—H groups and the C—O groups that are four residues back along the polypeptide chain are indicated by dashed lines. [Figure copyrighted © by Irving Geis.]

β strands

β sheet regular
Secondary structure

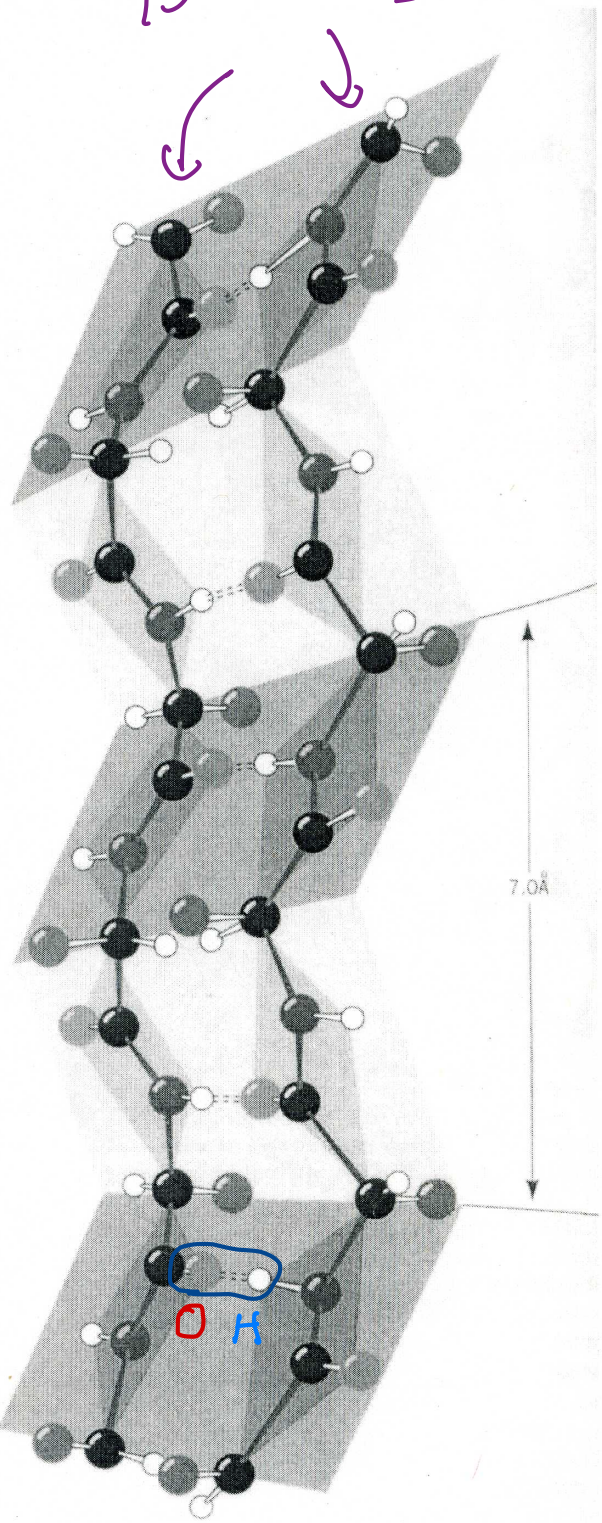
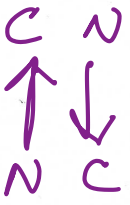


FIGURE 8-17 **A two-stranded β antiparallel pleated sheet drawn to emphasize its pleated appearance.** Dashed lines indicate hydrogen bonds. Note that the R groups (*purple balls*) on each polypeptide chain alternately extend to opposite sides of the sheet and that they are in register on adjacent chains. [Illustration, Irving Geis/Geis Archives Trust. Copyright Howard Hughes Medical Institute. Reproduced with permission.] See Kinemage Exercise 3-3

antiparallel



parallel

● ≡ regular secondary structure

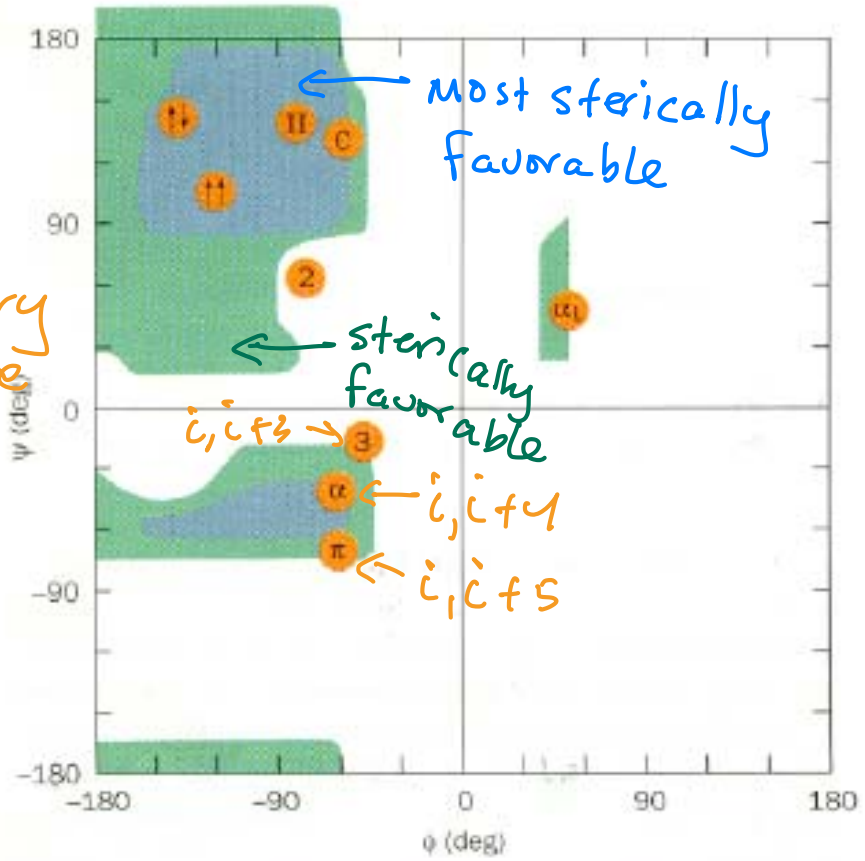


FIGURE 7-7. A Ramachandran diagram (named in honor of its inventor, G.N. Ramachandran) shows the sterically allowed ϕ and ψ angles for poly-L-alanine. The diagram was calculated using the van der Waals distances in Table 7-1. Regions of "normally allowed" ϕ and ψ angles are shaded in blue, whereas green-shaded regions correspond to conformations having "outer limit" van der Waals distances. The conformation angles, ϕ and ψ, of several secondary structures are indicated below:

Secondary Structure	ϕ (deg)	ψ (deg)
Right-handed α helix (α)	−57	−47
Parallel β pleated sheet ($\uparrow\uparrow$)	−119	113
Antiparallel β pleated sheet ($\uparrow\downarrow$)	−139	135
Right-handed 3_{10} helix (3)	−49	−26
Right-handed π helix (π)	−57	−70
2.2_7 ribbon (2)	−78	59
Left-handed polyglycine II and poly-L-proline II helices (II)	−79	150
Collagen (C)	−51	153
Left-handed α helix (α_L)	57	47

[After Flory, P.J., *Statistical Mechanics of Chain Molecules*, p. 253, Interscience (1969); *and* IUPAC–IUB Commission on Biochemical Nomenclature, *Biochemistry* **9**, 3475 (1970).]

Tertiary structure ≡ 3D geometric arrangement of secondary structure elements

Many evolved protein sequences (i.e. found in an organism) have well-defined "folded" structures or at least regions with well-defined structures

"Well-defined" ≡ 3D arrangement of atoms is very similar for all molecules

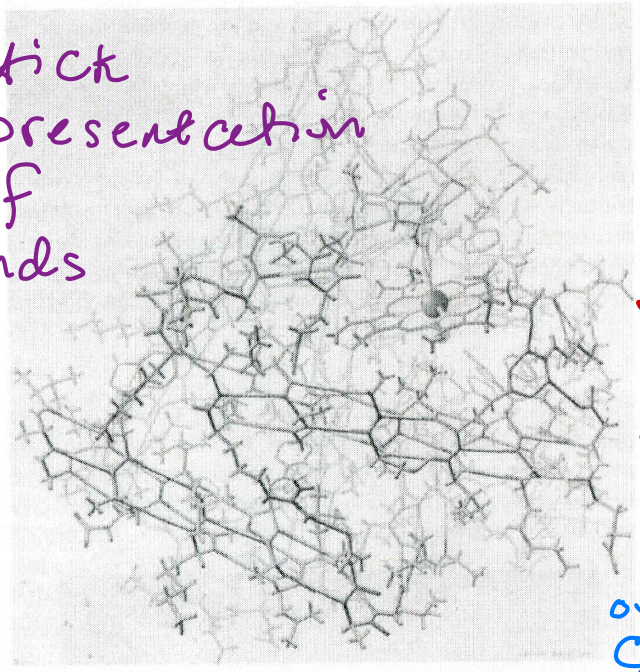
to crystallize. Moreover, since NMR can probe motions over time scales spanning 10 orders of magnitude, it can be used to study protein folding and dynamics (Chapter 9).

d. Protein Molecular Structures Are Most Effectively Illustrated in Simplified Form

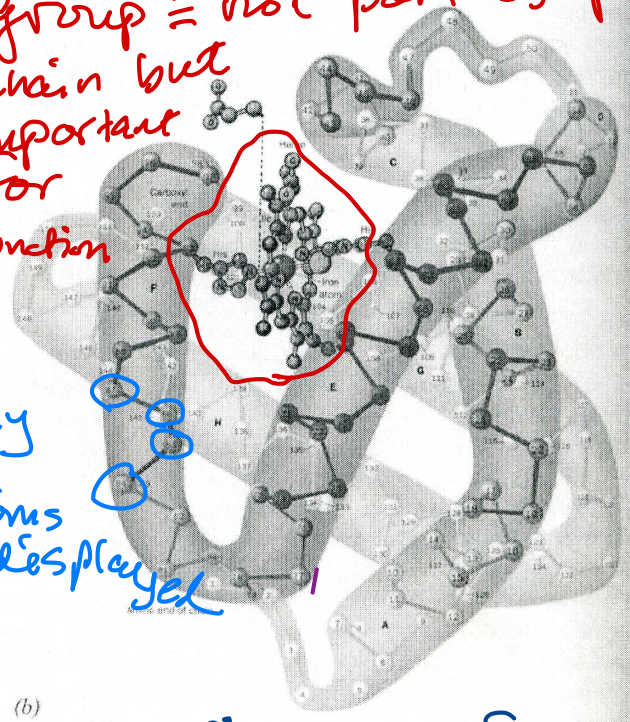
The several hundred nonhydrogen atoms of even a small protein makes understanding a protein's detailed structure a considerable effort. This complexity makes building a skeletal (ball-and-stick) model of a protein such a time-consuming task that such models are rarely available. Moreover, a drawing of a protein showing all its nonhydrogen atoms (e.g., Fig. 8-39*a*) is too complicated to be

Stick representation of bonds

heme porphyrin prosthetic group ≡ not part of polymer chain but important for function



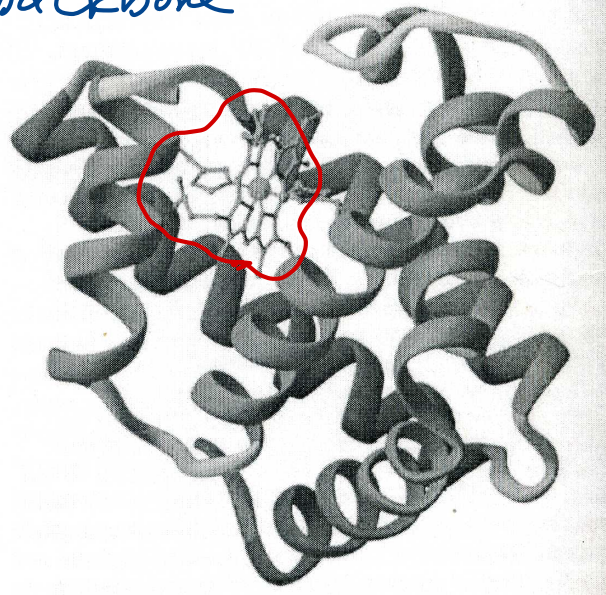
(*a*)



only Cα atoms displayed

(*b*)

Ribbon diagram of backbone



(*c*)

FIGURE 8-39 Representations of the X-ray structure of sperm whale myoglobin. (*a*) The protein and its bound heme are drawn in stick form, with protein C atoms green, heme C atoms red, N atoms blue, and O atoms red. The Fe and its bound water molecule are shown as orange and gray spheres and hydrogen bonds are gray. In this one-of-a-kind painting of the first known protein structure, the artist has employed "creative distortions" to emphasize the protein's structural features, particularly its α helices. (*b*) A diagram in which the protein is represented by its computer-generated C_α backbone, with its C_α atoms, shown as balls, consecutively numbered from the N-terminus. The 153-residue polypeptide chain is folded into eight α helices (highlighted here by hand-drawn envelopes), designated A through H, that are connected by short polypeptide links. The protein's bound heme group (*purple*, with its Fe atom represented by a red sphere), in complex with a water molecule (*orange sphere*), is shown together with its two closely associated His side chains (*blue*). One of the heme group's propionic acid side chains has been displaced for clarity. Hydrogen atoms are not visible in the X-ray structure. (*c*) A computer-generated cartoon drawing in an orientation similar to that of Part *b*, emphasizing the protein's secondary structure. Here helices are green and the intervening coil regions are yellow. The heme group with its bound O_2 molecule and its two associated His side chains are shown in ball-and-stick form with C magenta, N blue, O red, and Fe orange. [Parts *a* and *b* are based on an X-ray structure by John Kendrew, MRC Laboratory of Molecular Biology, Cambridge, U.K. PDBid 1MBN. Illustrations, Irving Geis/Geis Archives Trust. Copyright Howard Hughes Medical Institute. Reproduced with permission. Part *c* is based on an X-ray structure by Simon Phillips, University of Leeds, Leeds, U.K. PDBid 1MBO.]

See Kinemage Exercise 6-1

≈ 150 residues

Creighton (11)

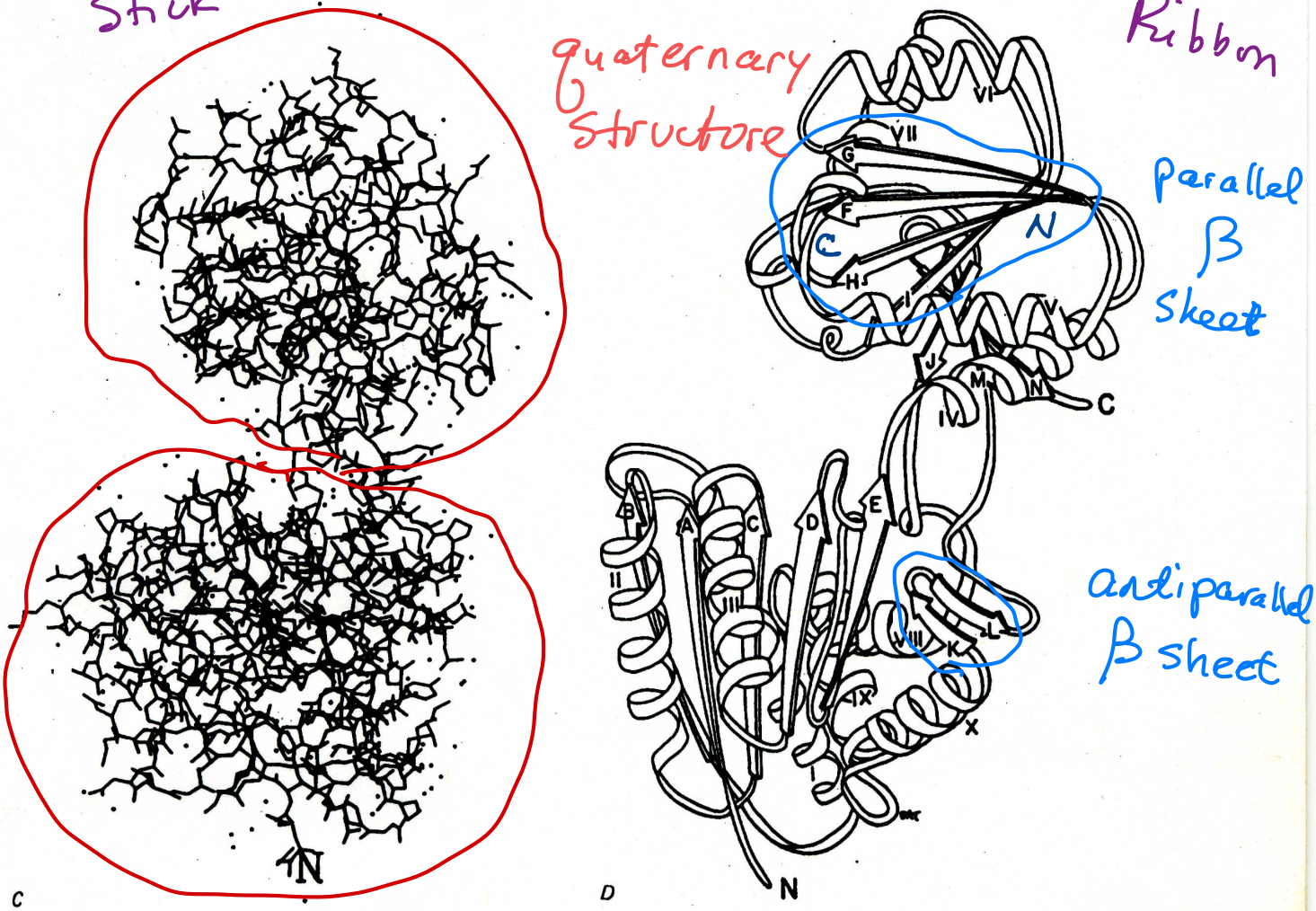


FIGURE 6.12 *(Continued)*

or more structural units, usually referred to as **domains**. The domains of a protein molecule interact to varying extents, but less extensively than do structural elements within domains. Often a single segment of polypeptide chain links the domains, and each domain consists of a single stretch of polypeptide chain. Domains along a polypeptide chain are not always segregated in this way, however; in phosphofructokinase, pyruvate kinase, and arabinose-binding protein, for example, there are two or three polypeptide connections between domains. In some cases, the end of a polypeptide chain in one domain interacts with another domain, appearing to serve as a strap that holds the domains together. The definition of a domain is not rigorous, and the division of a structure into domains is a subjective process that is done in different ways by different people. Other terms and subdivisions, such as *subdomain* and *folding unit*, are also encountered in the literature. Nevertheless, the presence of domains in many protein molecules is clear to all observers (Fig. 6.12C, D). Domains are most evident by their compactness, which can be expressed quantitatively as the ratio of the surface area of a domain to the surface area of a sphere with the same volume: observed values are 1.64 ± 0.08 for obvious domains.

The course of a polypeptide backbone through a domain is irregular, but it generally follows a moderately straight course across the entire domain and then makes a U-turn and recrosses the domain in a more or less direct but different path to the other side (Fig. 6.13). The observer receives an impression of segments of a somewhat stiff polypeptide chain interspersed with relatively tight turns or bends, which are almost always on the molecule's surface. This general type of structure has been compared to the behavior of a fire hose when dropped in one spot. It can be contrasted with other possible limiting situations: one more irregular, such as

Space-filling model = represent each atom by its VanderWaals e⁻ density

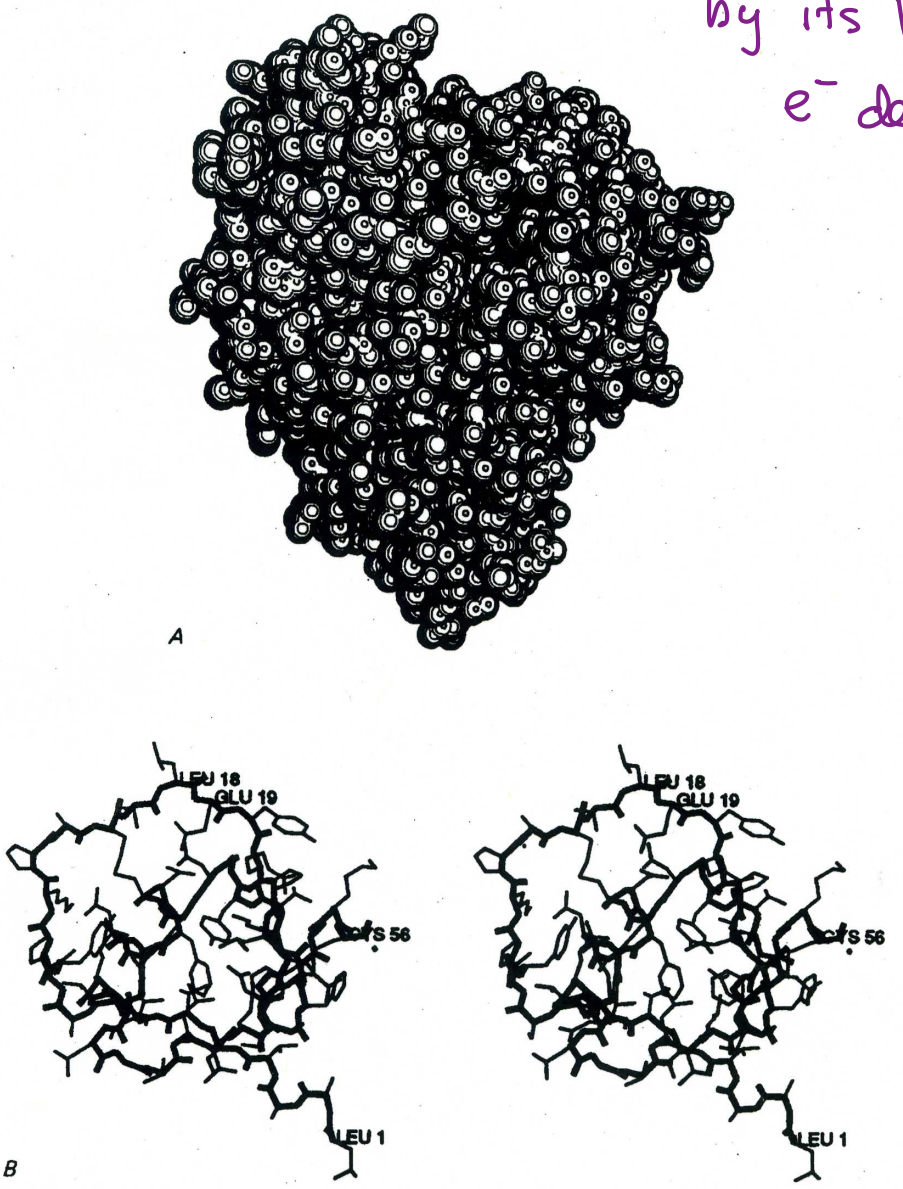


FIGURE 6.12

Various representations of the three-dimensional structures of globular proteins. *A*: Space-filling model of cytochrome P450, with 414 amino acid residues, in which each atom (except hydrogen) is represented by a sphere of the appropriate van der Waals radius. Note that only the surface atoms are visible and that it is impossible to trace the polypeptide chain. (T. L. Poulos et al., *J. Mol. Biol.* 195:687–700, 1987.) *B*: Stereo picture of a skeletal model of the isolated 56-residue third domain of silver pheasant ovomucoid. Bonds between all the atoms (except hydrogen) are illustrated; the polypeptide backbone is indicated by a thicker line. The illustration can be viewed in stereo by using a stereo viewer or by diverging or converging the two eyes. With such a small protein, all of the structure can be viewed, but little information about the packing of the atoms is apparent. The terminal residues are labeled, as are residues 18 and 19 that flank the peptide bond that is cleaved reversibly by proteases. (Kindly provided by W. Bode.) *C*: Skeletal model of the two-domain 344-residue leucine/isoleucine/valine–binding protein. Bonds of the polypeptide backbone are shown as thick lines. The dots are fixed water molecules. Note that it is impossible to trace the chain in this larger protein without stereo. *D*: Same view of leucine/isoleucine/valine–binding protein as in *C*, but a schematic representation of the polypeptide backbone, with arrows for β-strands and coils for α-helices. (Kindly provided by F. Quiocho.)

Creighton

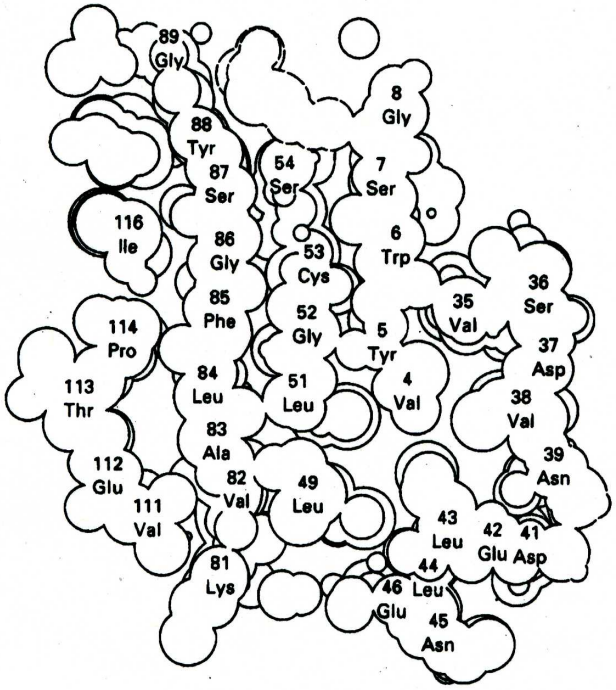
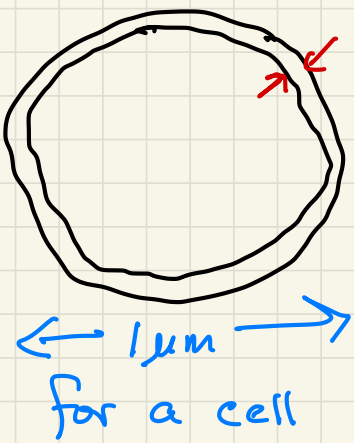


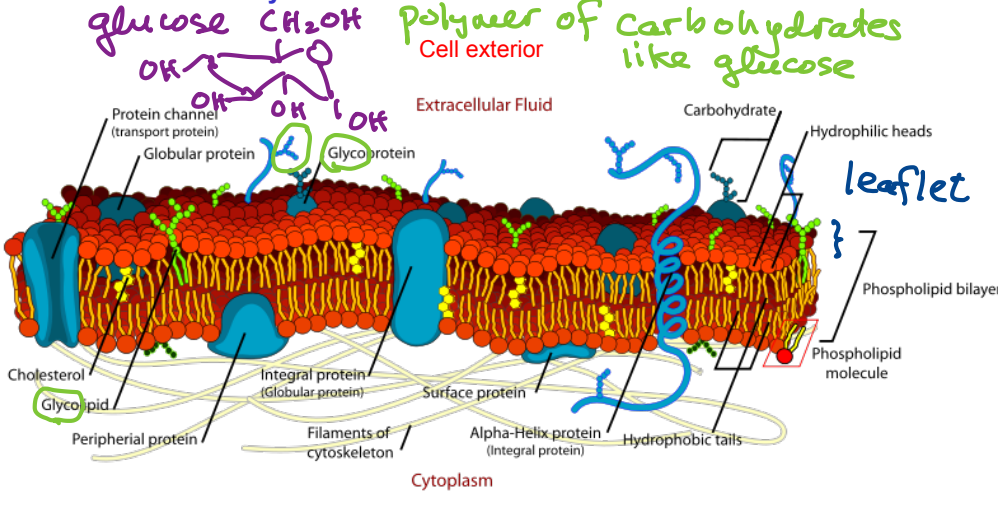
FIGURE 6.21
Serial section through the interior of flavodoxin, demonstrating the dense packing of atoms. Atoms of the protein are drawn with their van der Waals radii as solid lines; the dashed line shows the prosthetic group FMN. Three successive sections cut 1 Å apart are shown. The labels indicate the positions of the C^{α} atoms. Several parallel β-strands are visible. (Kindly provided by A. M. Lesk.)

Creighton
14

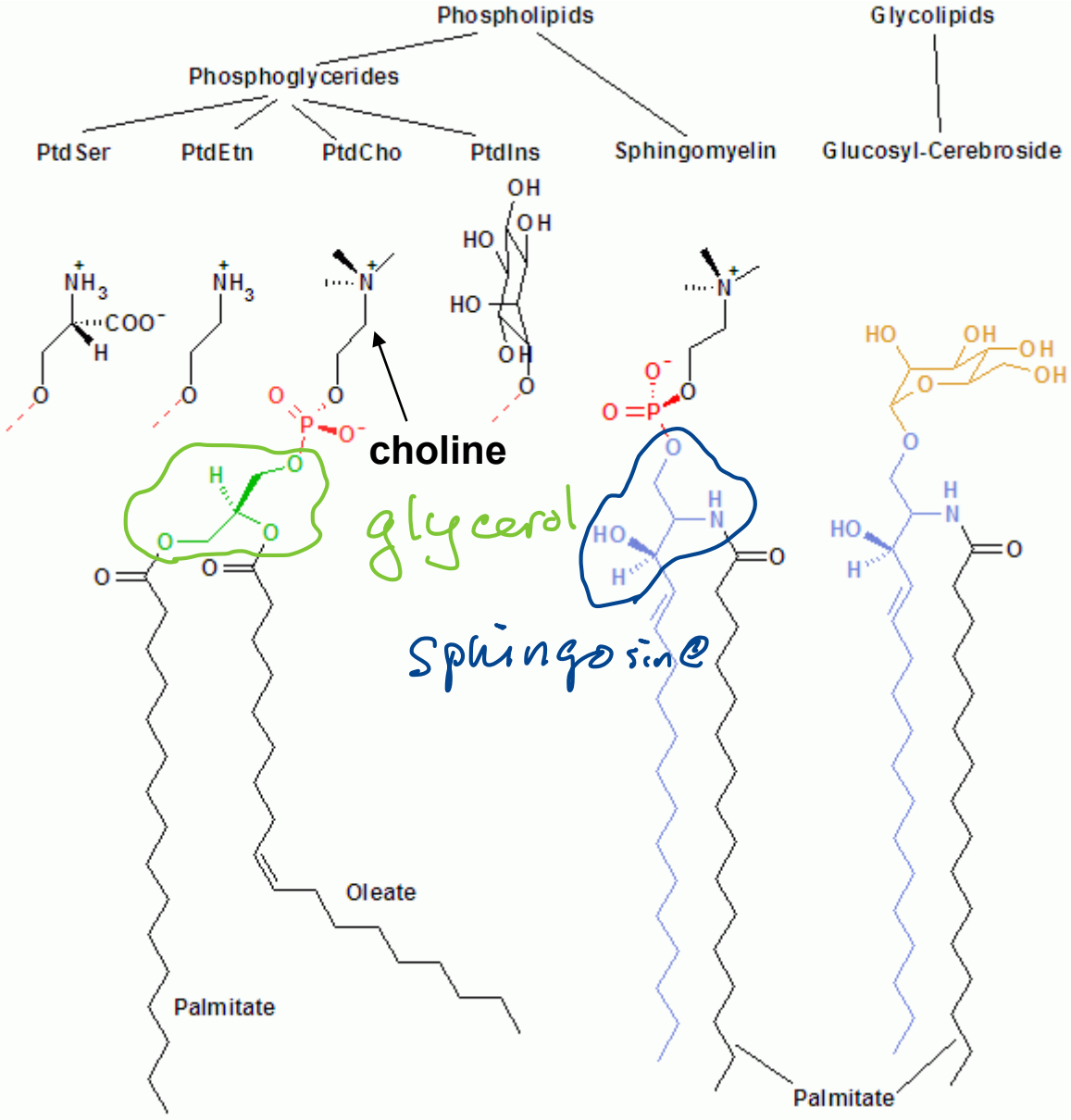
Requirement of Life ≡ control [ ]'s of molecules in a defined space ≡ usually accomplished by a lipid membrane which provides a barrier to diffusion



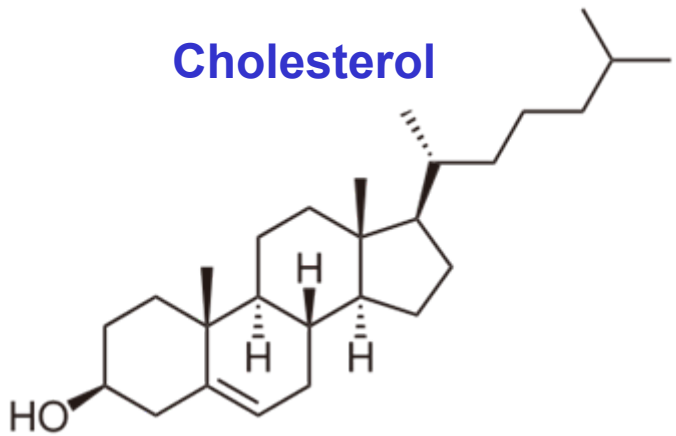
Cell membrane bilayer forms barrier between interior and exterior of cell



Lipid molecules



Cholesterol

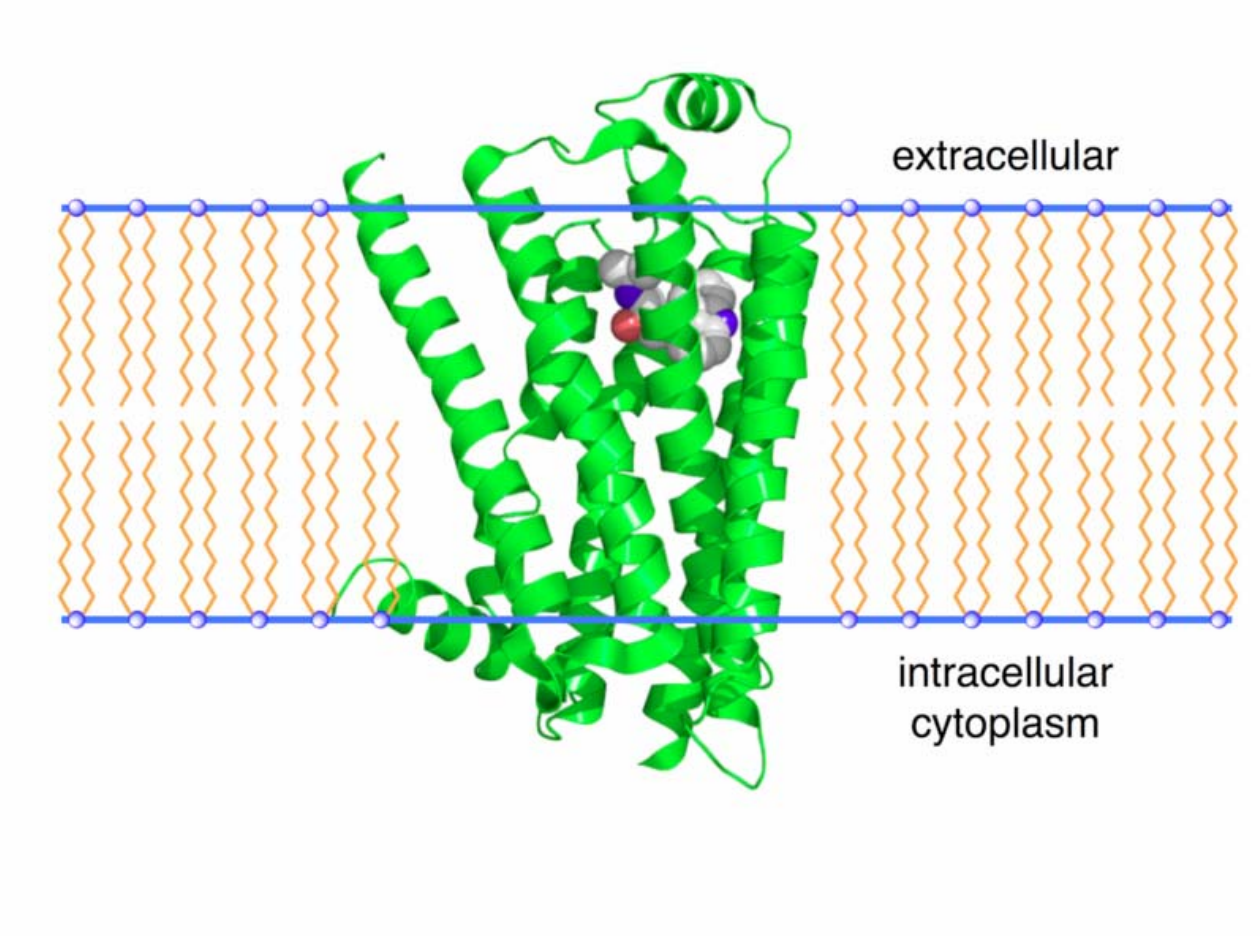


Structure of Beta Adrenergic Receptor (~400 amino acids)

Binds adrenaline

Initiates “fight or flight” response

Target of asthma and cardiac drugs – 40 billion dollars/year



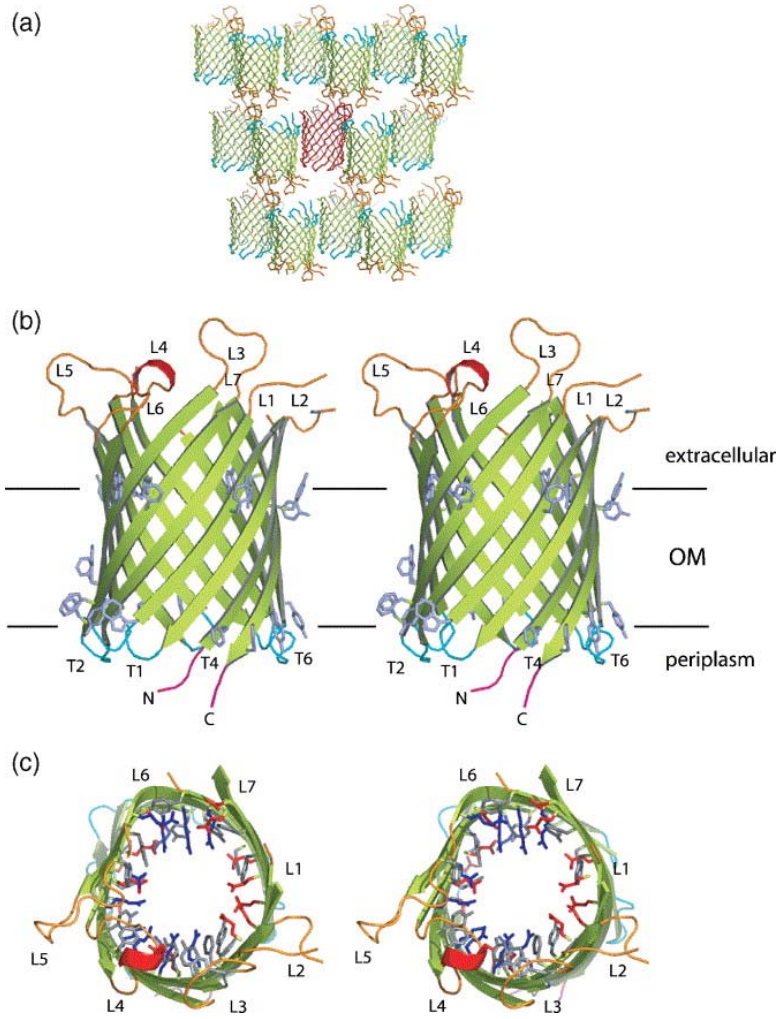
Cherezov, V. et al. (2007) Science 318, 1258-1265

Structure of Outer Membrane Protein G of E. coli bacterium (301 amino acids)

14 antiparallel β strands in a β barrel structure

13 Å diameter pore

channel for larger oligosaccharides (sugars)?



Subbarrao, G. et al. (2010) J. Mol. Biol. 360, 750-759

Nucleic acids: DNA + RNA

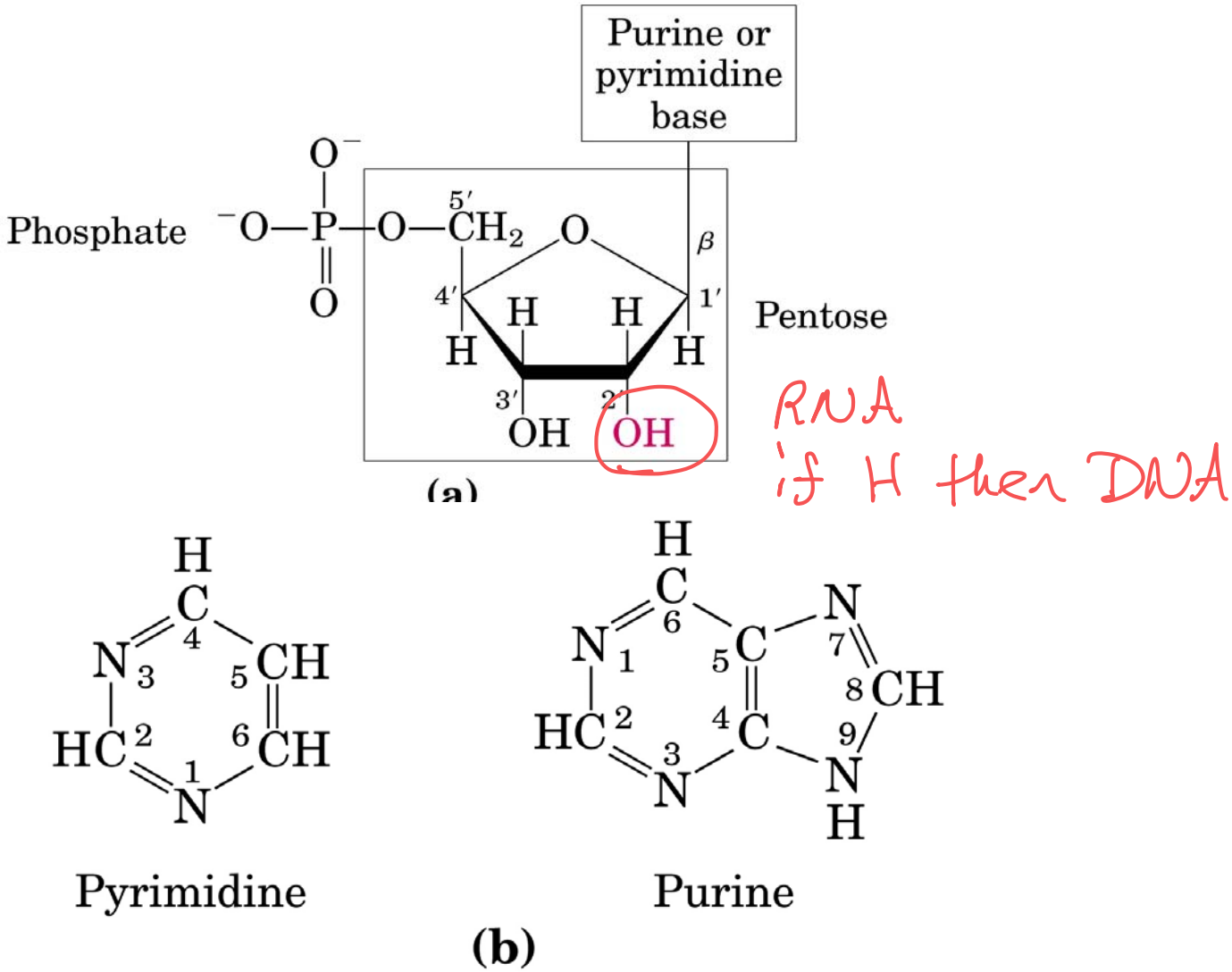
linear polymers of nucleotides

code for the specific amino acid sequences of proteins

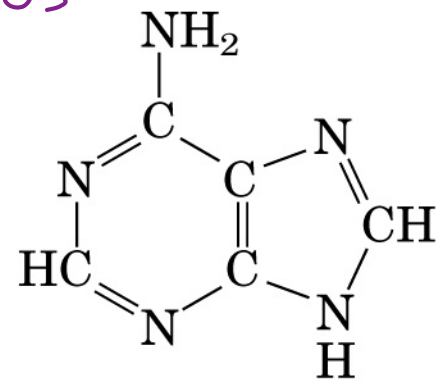
DNA $\Longrightarrow$ RNA $\Longrightarrow$ protein with a specific sequence

(DNA → RNA: transcription; RNA → protein: translation)

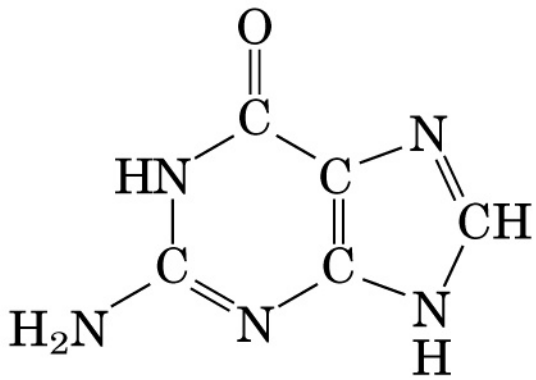
The Building blocks of nucleic acids



nitrogenous bases

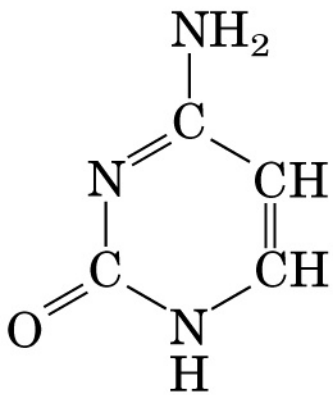


Adenine

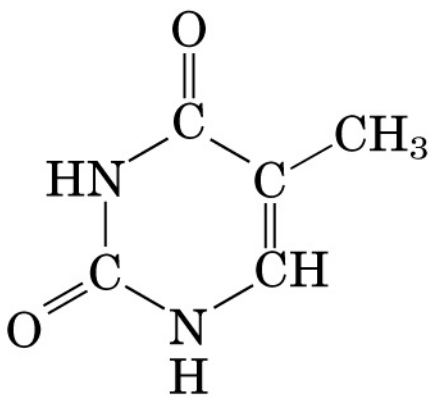


Guanine

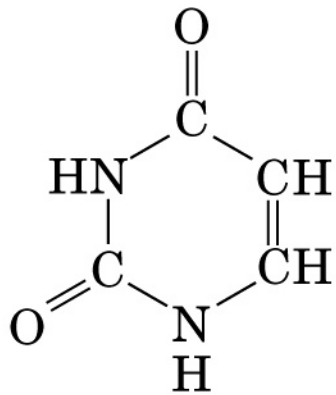
Purines



Cytosine



Thymine
(DNA)

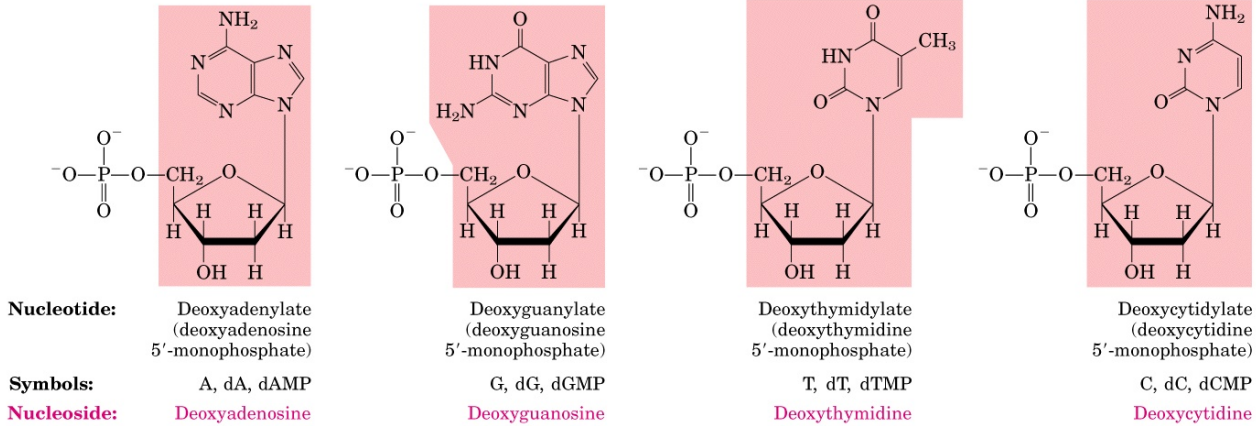


Uracil
(RNA)

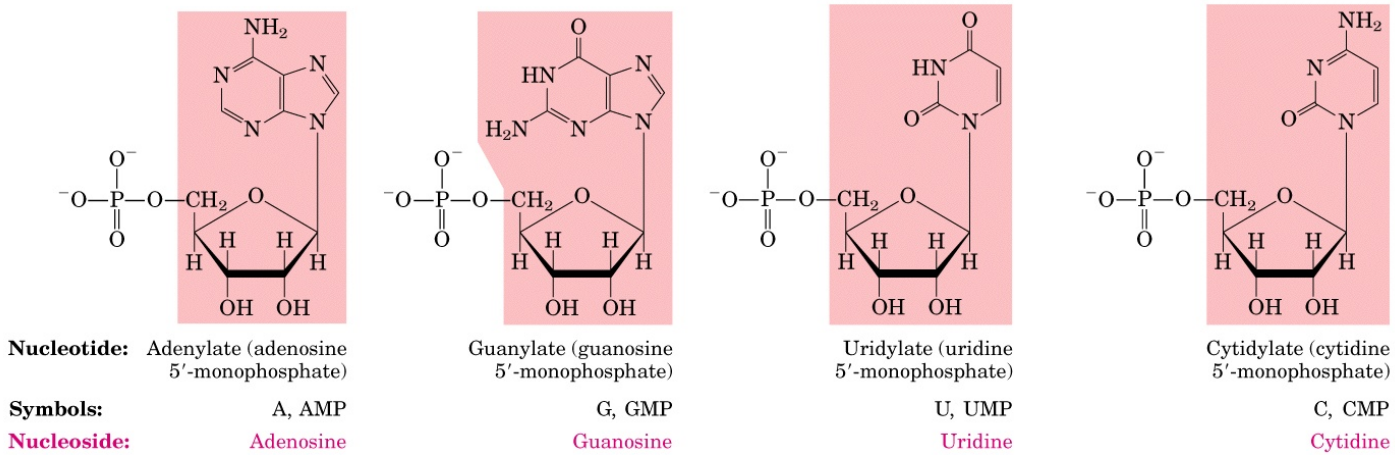
Pyrimidines

Nomenclature: Nucleotide, base+sugar+phosphate

Nucleoside: base+sugar

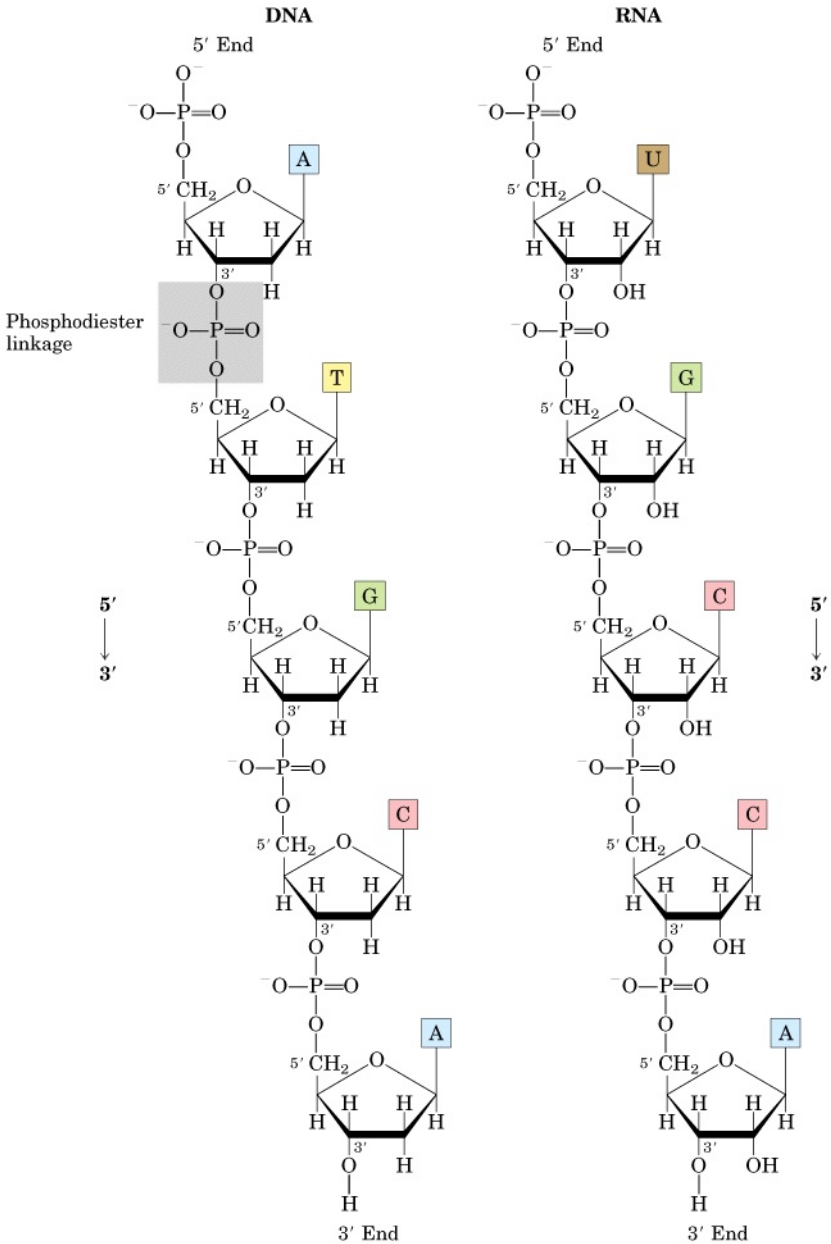


(a) Deoxyribonucleotides

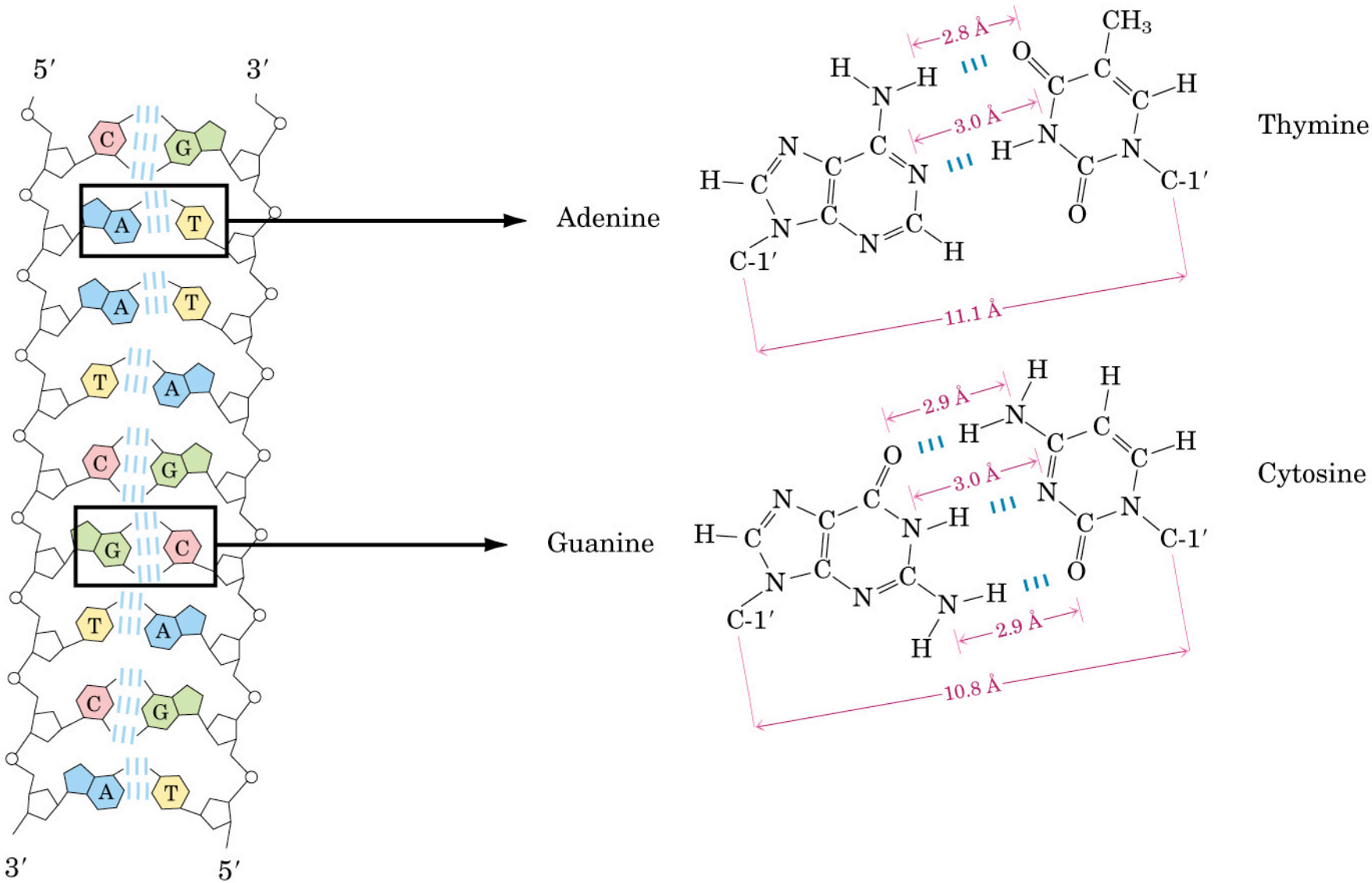


(b) Ribonucleotides

The covalent backbone of DNA and RNA

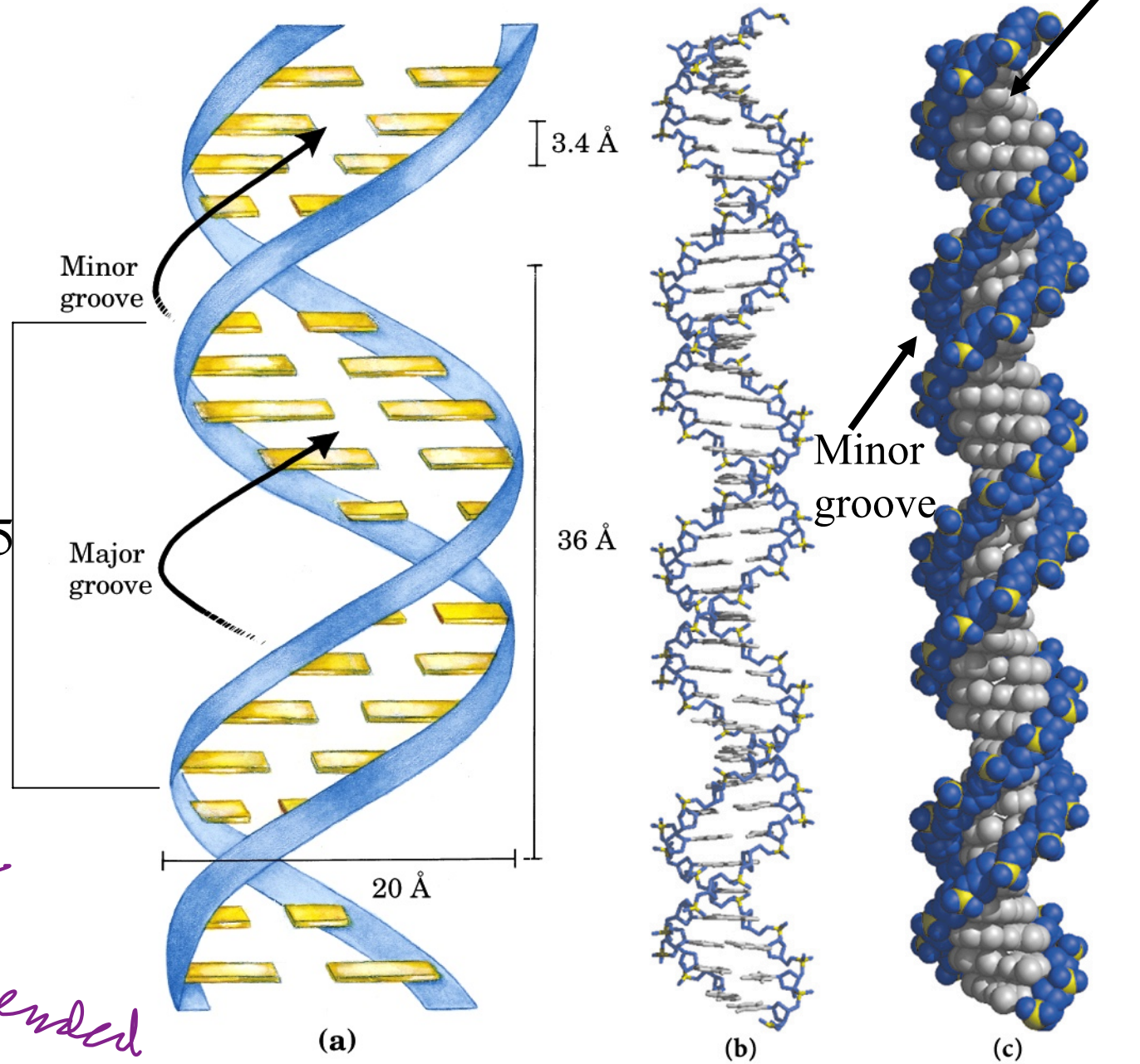


Watson/Crick basepairing holds double helix together.



Structure of double helical B DNA

Rise, h: 3.4 Å
Helical repeat, c: 10-10.5 bases



fully extended

Protein "transcription factor"

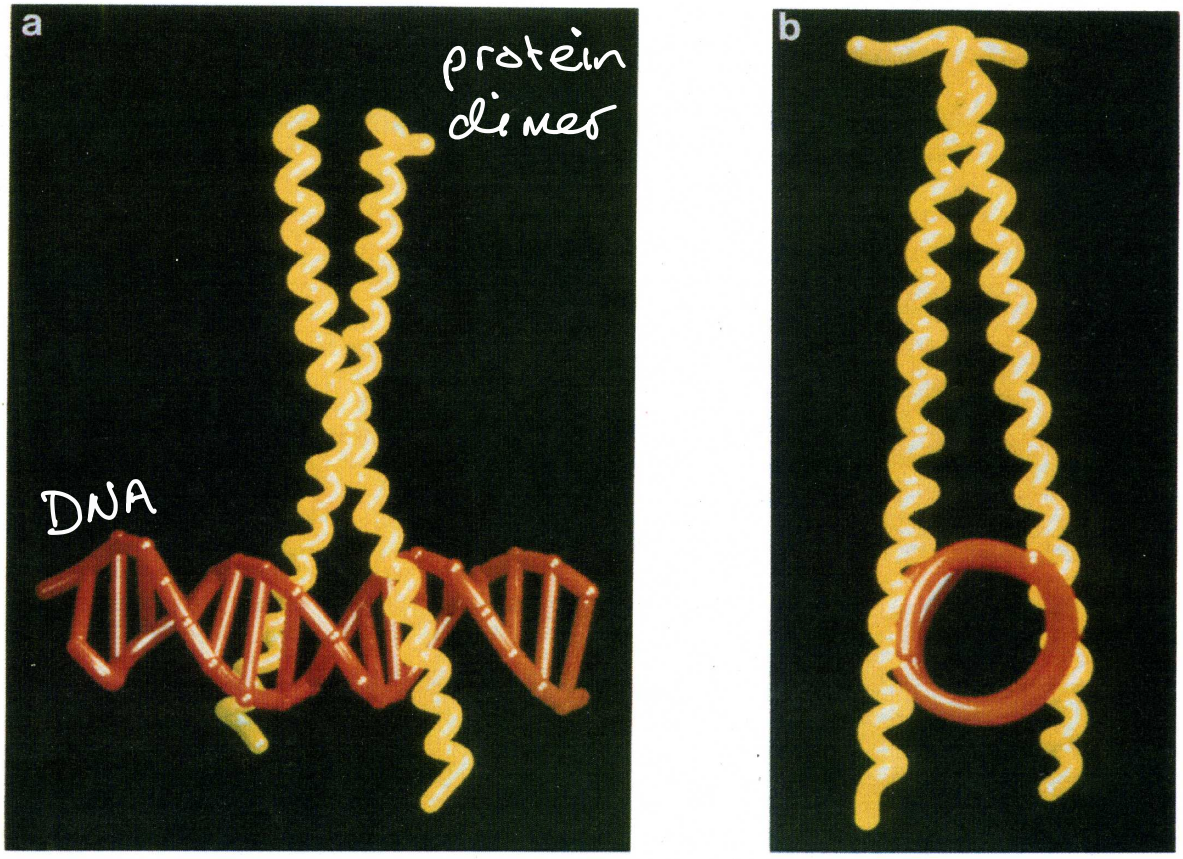


Figure 3. Structure of the GCN4 bZIP Complex

(a) The bZIP dimer (yellow) binds in the major groove of the DNA (red). Each bZIP protomer is a smoothly curved, continuous α helix. The carboxy-terminal residues of the monomers pack together as a coiled coil, which gradually diverges to allow the basic region residues to follow the major groove of either DNA half site. This divergence of the bZIP monomers corresponds to an unwinding of the coiled-coil super helix, with a slight righthanded rotation of basic region residues about the α-helical axis of each chain and a lateral displacement of each monomer along the helical axis of the DNA. The DNA in the complex is straight, and its conformation is in the B form across the region contacted by the protein.

(b) View down the DNA axis, with the DNA backbone represented as a red coil. The basic region residues amino-terminal to the point of DNA contact are in a straight, α-helical conformation. The amino-terminal residues of the basic region do not wrap around the back side of the binding site.

unesterified phosphate oxygen of Cyt-0 is contacted by Arg-243 of the right monomer, with the phosphate of Gua-0′ contacted by Arg-240 of the left monomer. The central asymmetry imposed by contacts to bases at position 0 is corrected by local contacts, which position both monomers over their respective half sites in an analogous way. This adjustment implies that the α-helical fork at the junction between the basic region and the leucine zipper coiled coil is somewhat flexible. Flexibility of GCN4 is likely to be a key component in its specific recognition of the activating transcription factor/cAMP response element–binding protein (ATF/CREB)-binding site (Sellers et al., 1990). The ATF/CREB site is a palindromic sequence homologous to the AP-1 sequence, but it contains an additional central guanine–cytosine base pair.

Structure of the DNA in the Complex

The AP-1 site in the GCN4 complex is straight, B-form DNA across the region contacted by the protein (see Figure 3). The central 15 bp of the AP-1 site DNA have an average helical rise and twist of 3.2 Å and 34°. No systematic variation in phosphodiester backbone or base pair geometry is evident across the binding site. DNAs in adjacent unit cells of the crystals pack end to end, forming a Watson–Crick base pair between the 5′ thymine of one complex and the 5′ adenine of the adjacent DNA (Figure 9a). The DNA is aligned along crystallographic unit cell axis c, which has a length (65.3 Å) commensurate with that of a B-form 20-mer oligonucleotide.

The DNA packing arrangement in the crystals changes markedly in response to crystal harvest conditions. Prior to collecting diffraction data from crystals at −160°C, the crystals were soaked in cryoprotectant to minimize crystalline ice formation during freezing (see Experimental Procedures). Following equilibration of the crystals in a cryoprotectant solution containing an elevated concentration of polyethylene glycol (molecular weight = 400) (PEG 400), the length of the crystallographic c axis is reduced by 6.1 Å (c = 59.2 Å). This shrinkage is due to a local rearrangement in base-stacking interactions that decreases