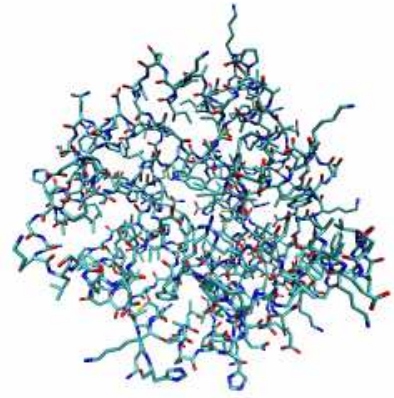
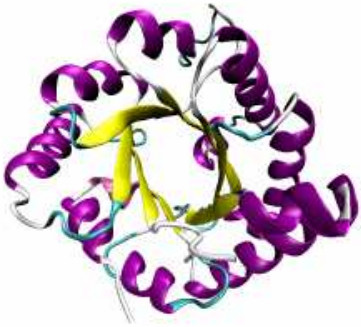
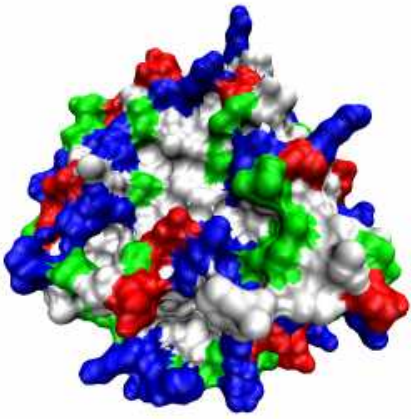
>gi|161984995|gb|CP000896.1| Acholeplasma laidlawii PG-8A, complete genome
TATTTGATTTTTTGCTTAATTTTTATGTAATTACTTCCCCACAATTTTTGCCCACATATTGTGGATAAAT
TTTCCACATTTTATTCACAATGTTGATAAGTTGTGTAAGTCGACTTGTTGGCTTTATAAAGCAAATGACA
CACTGAAAGTTATCCACAAACAAATTTTTCAAAAGTGTTTAAATCAAAAAGTTATCCACAAATAATGTGG
AAAACTTTTTAATAAATTTGTCGTTTCTTATGCTATCATAGTTTTACATAAATTATTAACTACTATAGGG
AGGCAGTCATGAGTCCAAACAGCACATTATGGCAGACAATATTACAGGATTTAGAAAAACTATACAACGA
GGAGACTTACAACGAGCTATTTCTACCAGTGACTTCTACTTTTAAAGATCAAAACGGATTACTTACAATG
GTTGTAGCTAATGAGTTCTTAAAGAATCGTATCAATAAACTATACATTGCAAAAATTAACGAACTTGCTA
CAAAATATTCAAGTACTCCAGTTAGATTGAAATTCGTATCACAAGAAGAAGTTATTGAAGAACCAGTAGC
TGATCGTAAATTAACCATTGATTATCGTCAAGGGAACTTAAACTCTACCTATACCTTTGACTCTTTTGTT
GTTGGAAAATCTAACATGTTTGCTTTTCGTATGGCGATGAAGGTTGCTGATCAACCTGGAGCAGTAGCAA
ACCCTTTCTACATATTTGGTGATGTAGGTTTAGGTAAAACCCATCTTATGCAAGCAATAGGTAACTATAT
ATTAGATAATGATGTCGAAAAACGTATCTTATATGTTAAAGCTGATAACTTTATTGAAGACTTTGTATCT
TTATTATCAAGAAACAAAAATAAGACTGAAGAATTCAATGCTAAATATAAAGATATTGACGTTATATTAG
TAGACGACATCCAAATTATGGCAAATGCTAGTAAAACTCAAATGGAATTCTTTAAGCTCTTTGACTACCT
ATATTTAAATAATAAACAAATCGTTATAACATCTGATAAACCAGCTTCACAATTAACAAATATCATGCCA
CGTTTAACGACACGTTTTGAAGCTGGTCTCTCTGTAGACATACAAATACCTGAATTAGAACATAGAATAA
GTATTTTAAAGAGAAAAACAGCTACATTAGATGCCAACTTAGAGGTAAGTGAGGATATTTTAACCTTTAT
AGCATCTCAATTTGCAGCAAACATTAGAGAAATGGAAGGTGCACTCATTCGTTTAATTAGTTATGCACAG
ACCTTTAATCTAGAAATTACAATGAATGTTGTTGAAGAAGCACTTGGTGCTGTATTAAAAACTAAGAAGA
AAACAAATGATTTAAACGAAAATAACTACGATAAGATCCAAAGTATTGTAGCAGATTACTTCCAAGTTTC
ATTACCAGACTTAATTGGTAAGAAAAGACATGCTAAATTCACATTACCTAGACATATAGCGATGTATCTT
ATTAAACTTAAATACAATATTCCTTATAAAACAATTGGATCTTTATTTAATGATAGAGACCACTCCACAG
TATTATCTGCTTGTGAAAAAGTAGAACGCGATATGAGGATGGATTCGAACTTAAAGTTTGCTGTTGACTC
GATTGTCAAAAAAATAGATTCACCATTATTAAAGTGATAAATGTTTATAAAAATGATTAAATGTGGTAAA
ATAAATGGTAGATGAAGCGATTATTGCTAGTTTCCCACTTTCCCACAGACACTAACAATAACAAAGAAGA
AATAATAATTAATAAAAGGGGTAAAATATGAATTTTACAATTGAAAGAGATATCTTACTAAACCATCTAA
TGCCATGTACAAAAAGGATTACCAAATAAAACACCTCTACCTATTTTATATGCTATTAAATTTGAAGTGT
TTTCAGATTATATCCAATTAACAGCTTCTAACTCTGATATAGCCATTCAAGTATTAATTGATGATCCATC
ATTAGTTGTTCATAAGACAGGAAAGATTGCTTTACCCTGAGAGATCCCCTCATAATTTCCCCAAAGCGTA
ACCATGTGTGAATAAATTTTGAGCTAGTAGGGTTGCAGCCACGAGTAAGTCTTCCCTTGTTATTGTGTAG
CCAGAATGCCGCAAAACTTCCATGCCTAAGCGAACTGTTGAGAGTACGTTTCGATTTCTGACTGTGTTAG
CCTGGAAGTGCTTGTCCCAACCTTGTTTCTGAGCATGAACGCCCGCAAGCCAACATGTTAGTTGAAGCAT
CAGGGCGATTAGCAGCATGATATCAAAACGCTCTGAGCTGCTCGTTCGGCTATGGCGTAGGCCTAGTCCG
TAGGCAGGACTTTTCAAGTCTCGGAAGGTTTCTTCAATCTGCATTCGCTTCGAATAGATATTAACAAGTT
GTTTGGGTGTTCGAATTTCAACAGGTAAGTTAGTTGCTAGAACCCATGGCTCCTTTGCCGACGCTGAGTA
GATTTTAGGTGACGGGTGGTGACAATGAGTCCGTGTCGAGCGCTGATTTTTTCGGCCTTTAGAGCGAGAT
TTATACAATAGAATTTGGCATGAGATTGGATTGCTTTTAGTCAGCCTCTTATAGCCTAAAGTCTTTGAGT
GACTAGATGACATATCATGTAAGTTGCTGATAGGTTTCCAGTTTTCCTCTCCTAGGTCTGCATATTGTAC
TTTTCCTCTTACTCGACTTAACCAGTACCAACCCAGCTTCTCAACGGATTTATACCATGGCACTTTAAAG
CCAGCATCACTGACAATGAGCGGTGTGGTGTTACTCGGTAGAATGCTCGCAAGGTCGGCTAGAAATTGGT
CATGAGCTTTCTTTGAACATTGCTCTGAAAGCGGGAACGCTTTCTCATAAAGAGTAACAGAACGACCGTG
TAGTGCGACTGAAGCTCGCAATACCATAAGTCGTTTTTGCTCACGAATATCAGACCAGTCAACAAGTACA
ATGGGCATCGTATTGCCCGAACAGATAAAGCTAGCATGCCAACGGTATACAGCGAGTCGCTCTTTGTGGA
GGTGACGATTACCTAACAATCGGTCGATTCGTTTGATGTTATGTTTTGTTCTCGCTTTGGTTGGCAGGTT
ACGGCCAAGTTCGGTAAGAGTGAGAGTTTTACAGTCAAGTAATGCGTGGCAAGCCAACGTTAAGCTGTTG
AGTCGTTTTAAGTGTAATTCGGGGCAGAATTGGTAAAGAGAGTCGTGTAAAATATCGAGTTCGCACATCT
TGTTGTCTGATTATTGATTTTTCGCGAAACCATTTGATCATATGACAAGATGTGTATCCACCTTAACTTA
ATGATTTTTACCAAAATCATTAGGGGATTCATCAGTGCTTTACCTGGCCGTTATTTTATTGATATTATTA
GAAAAGTAGCTGCTAAGCGTATTGAGGTTGCTCTACAAGAAGAAAGACTATTAGTTATTAAAGCAGATCG
TTCAGAATTTAAATTAAAACTTATGGATATCGATGATTATCCAGATGTTGATTTTTTAGACTTAGCAGAA
CCAGTGATTCTATCATCTGATATGGTGAAAGAAATTATTAAAGAAACTAATTTCGCTACAGCTGATAACG
AAAAACGCCCAATTCTTACTGGGGTAAACTTTAAATATCAAGATAACCACTTATTTGCAGTTGCAACAGA
LOCUS   CP000896   **1496992** bp   DNA   **circular** BCT 06-SEP-2011
DEFINITION Acholeplasma laidlawii PG-8A, complete genome.

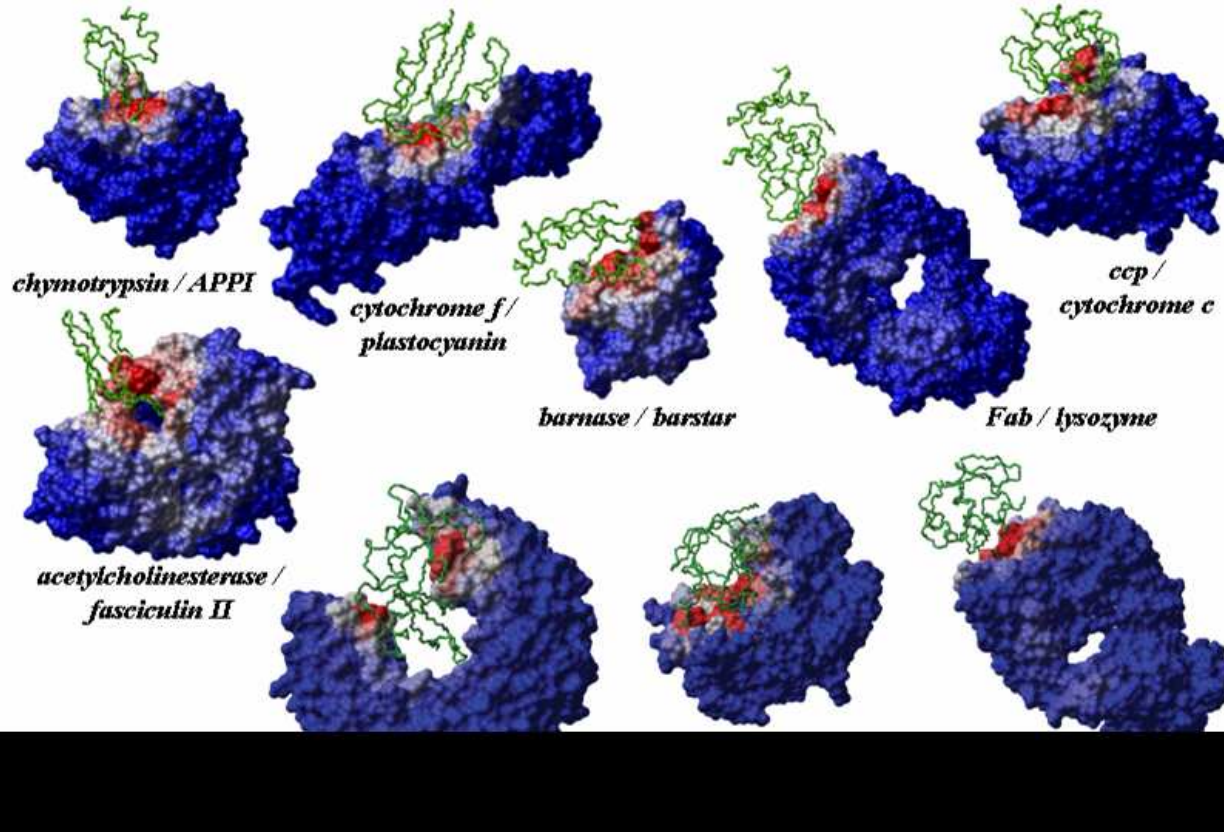
>gi|161985001|gb|ABX80650.1| **DNA gyrase**, subunit A [Acholeplasma laidlawii PG-8A]

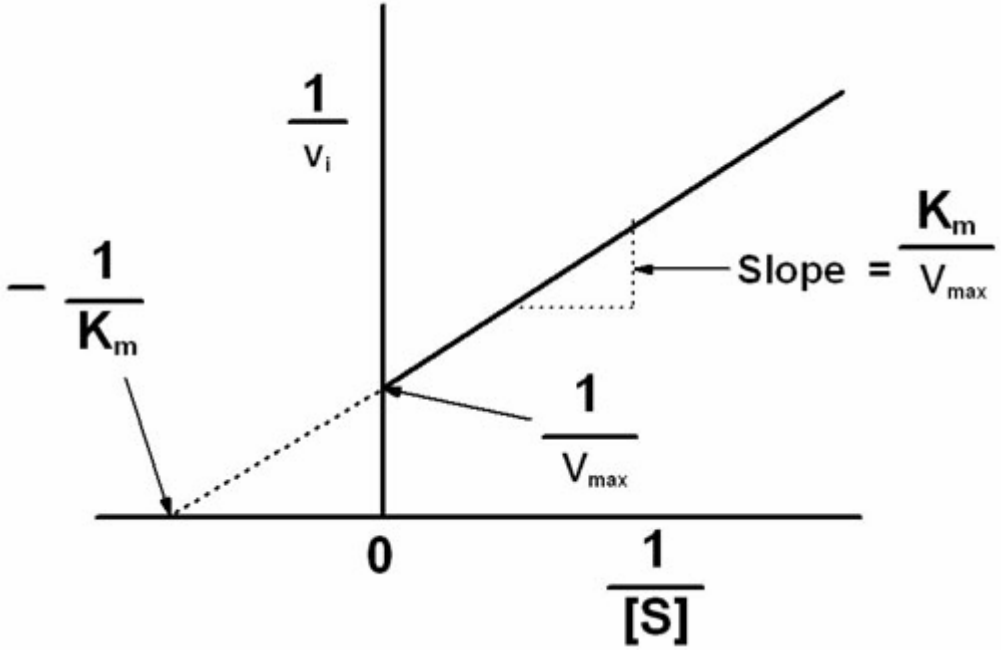
MSENTDKLHDKIKEVNITSEMKTSFLNYAMSVIVSRALPNAKDGLKPVQRRILYGMHENNMYANTAHKKS
ARIVGDVMGKYHPHGDSSIYDAMVRMAQPFSYRKPLVDGHGNFGSVDGDGAAAMRYTEARLSKIAGEMVR
DIDKNTVNFVDNYDGSEREPQVLPSRYPNLLVNGSTGIAVGMATNIPPHNLGEVIDAIFAYLDNPDITVL
ELMQYIKGPDFPTGGQILGITGLRQAYETGKGIIAVRAQHEIVETKNRTELIITEIPYQVNKTTLIERIA
EVVKNKIIEGISDLRDESNRKGMRIVIELRKDTNPSVTLNNLYKHTQLQTSFGINMIALVDGQPRLLSLK
DALTHYIAHQIEIITRRTQFDLEKAEARMHILEALVKALNDVDNVIKIIKDAKNPEDARNNLIATYDFTE
IQARSILTMQLQRLTGLEIDKIHEEADNLAHKITEYKQILGDEDLKIEIIKSELQEIKDKYDDPRMSEVN
LYEDLDIDNEDLIPVEDIVVTVTNNGYIKRMNLDEYKLQNRGGVGMSGVKLHDDDFVEHIAMTSTHDFHL
FFTNLGRVFKLKGYTLPSGSRQSKGVPIVNYLNFQPGEKLASFTTIKDFDAEDHYLFFVTKRGIVKRTHI
NNYKNIRQNGIIAISLREDDELLVTKSTNGHKHILLGASNGKAIHFLETDIREIGRTATGVRGMDIPDDE
EIVGFAVVDGIEDDILVITENGYGKRTSITEYRLQNRGGKGVKTLNVTDKNGKLKTLRKVTNDEDIIVVS
DRGITIRMPIDQISQTKRATQGVRIITLKGDQKVSTIAIVDHQEPEEEVVSNEAVVTQEQMALSNQKKVL
EVVELEALEESVDEEDTKNAKQTKLDL











http://infolific.com/images/fun/metaphor-for-complexity.gif

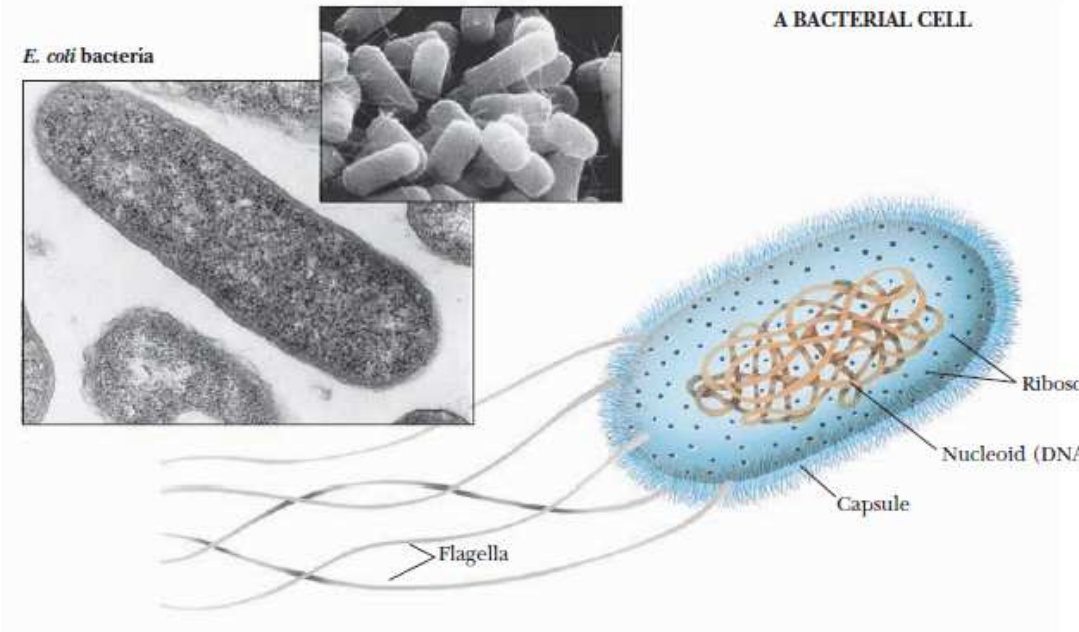
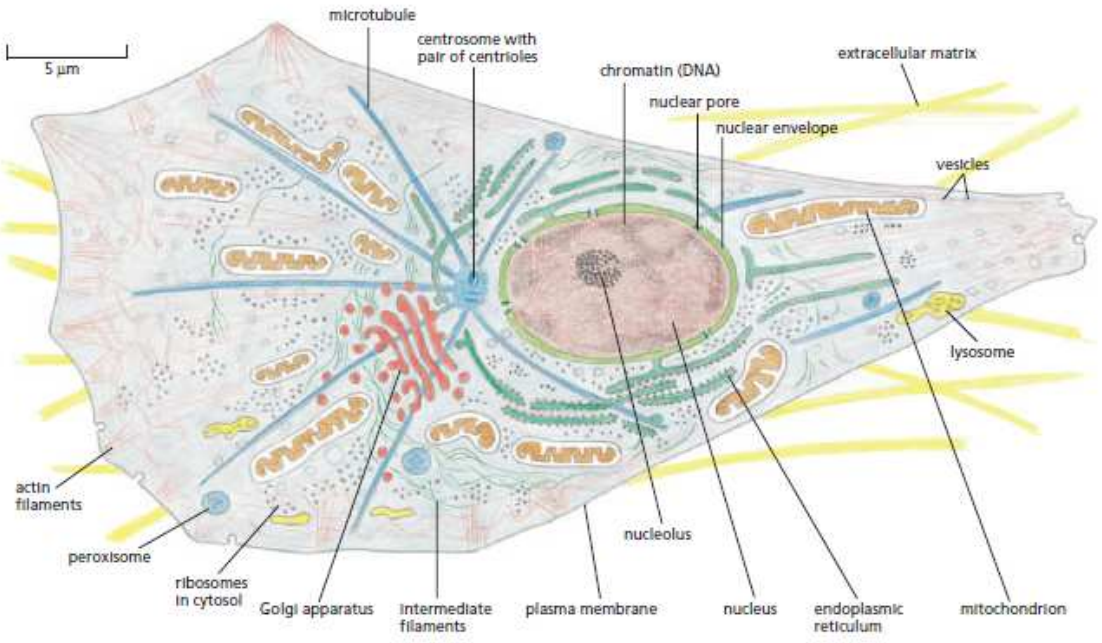


FIGURE 1.20 This bacterium is *Escherichia coli*, a member of the coliform group of bacteria that colonize the intestinal tract of humans. (See Table 1.7.) (Photo, Martin Rotker/Phototake, Inc.; Inset photo, David M. Phillips/The Population Council/Science Source/Photo Researchers, Inc.)


5 µm
microtubule
centrosome with pair of centrioles
chromatin (DNA)
nuclear pore
nuclear envelope
extracellular matrix
vesicles
lysosome
actin filaments
peroxisome
ribosomes in cytosol
Golgi apparatus
intermediate filaments
plasma membrane
nucleolus
nucleus
endoplasmic reticulum
mitochondrion

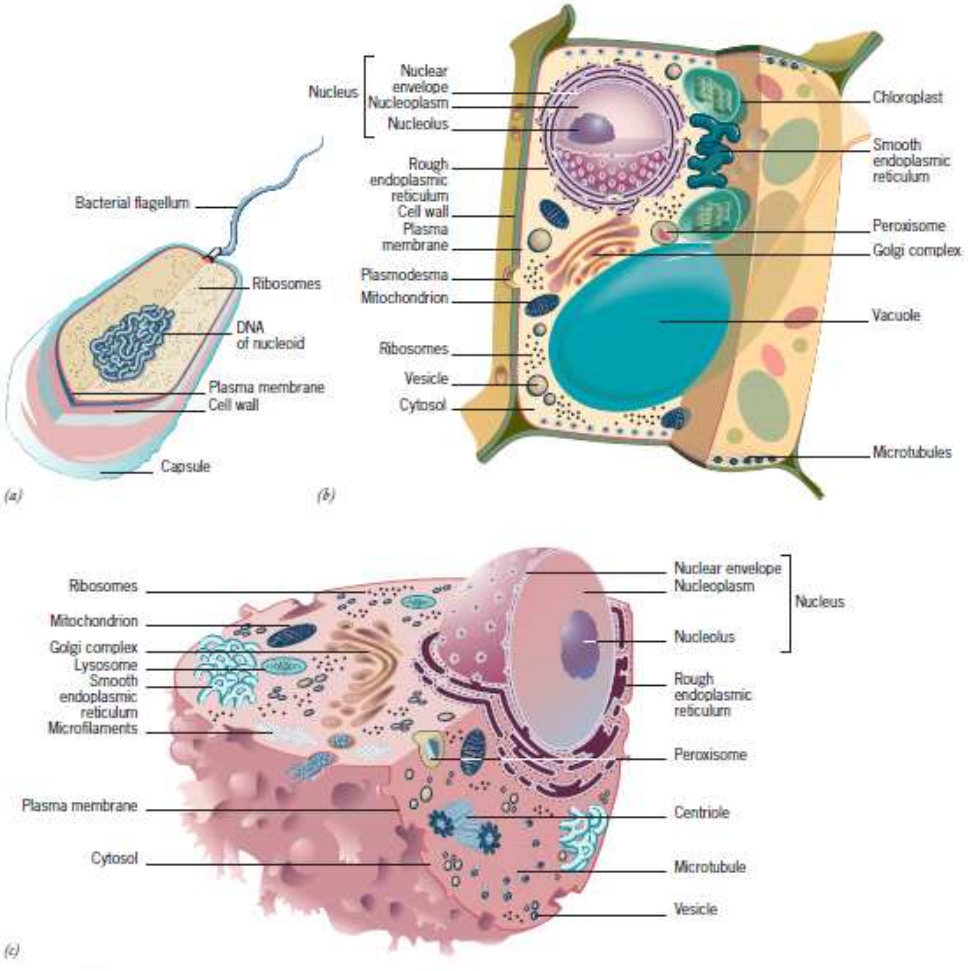


FIGURE 1.8 The structure of cells. Schematic diagrams of a "generalized" bacterial (*a*), plant (*b*), and animal (*c*) cell. Note: Organelles are not drawn to scale.

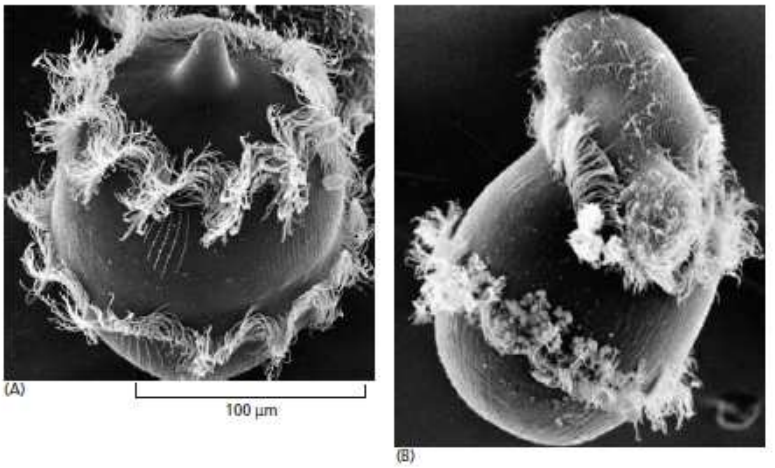


Figure 1–32 A single-celled eucaryote that eats other cells. (A) *Didinium* is a carnivorous protozoan, belonging to the group known as *ciliates*. It has a globular body, about 150 µm in diameter, encircled by two fringes of cilia—sinuous, whiplike appendages that beat continually; its front end is flattened except for a single protrusion, rather like a snout. (B) *Didinium* normally swims around in the water at high speed by means of the synchronous beating of its cilia. When it encounters a suitable prey, usually another type of protozoan, it releases numerous small paralyzing darts from its snout region. Then, the *Didinium* attaches to and devours the other cell by phagocytosis, inverting like a hollow ball to engulf its victim, which is almost as large as itself. (Courtesy of D. Barlow.)

Linnaeus 1735[70] 2 kingdoms	Haeckel 1866[71] 3 kingdoms	Chatton 1925[72] [73] 2 empires	Copeland 1938[64] [74] 4 kingdoms	Whittaker 1969[75] 5 kingdoms	Woese et al. 1977[76] [77] 6 kingdoms	Woese et al. 1990[66] 3 domains	Cavalier-Smith 2004[78] 6 kingdoms
(not treated)	Protista	Prokaryota	Mychota	Monera	Eubacteria	Bacteria	Bacteria
					Archaebacteria	Archaea	
		Eukaryota	Protoctista	Protista	Protista	Eukarya	Protozoa
							Chromista
Vegetabilia	Plantae		Plantae	Plantae	Plantae		Plantae
			Protoctista	Fungi	Fungi		Fungi
Animalia	Animalia		Animalia	Animalia	Animalia		Animalia

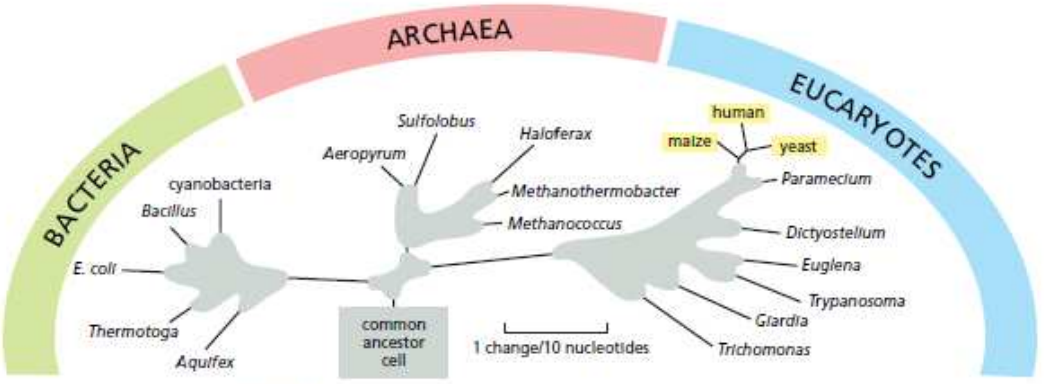


Figure 1–21 The three major divisions (domains) of the living world. Note that traditionally the word *bacteria* has been used to refer to procaryotes in general, but more recently has been redefined to refer to eubacteria specifically. The tree shown here is based on comparisons of the nucleotide sequence of a ribosomal RNA subunit in the different species, and the distances in the diagram represent estimates of the numbers of evolutionary changes that have occurred in this molecule in each lineage (see Figure 1–22). The parts of the tree shrouded in *gray cloud* represent uncertainties about details of the true pattern of species divergence in the course of evolution: comparisons of nucleotide or amino acid sequences of molecules other than rRNA, as well as other arguments, lead to somewhat different trees. There is general agreement, however, as to the early divergence of the three most basic domains—the bacteria, the archaea, and the eucaryotes.

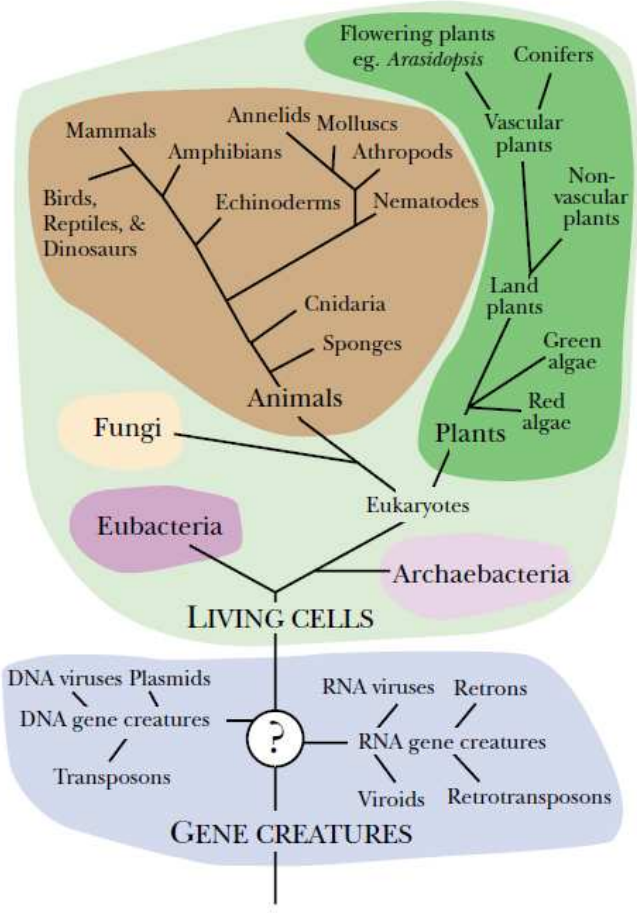
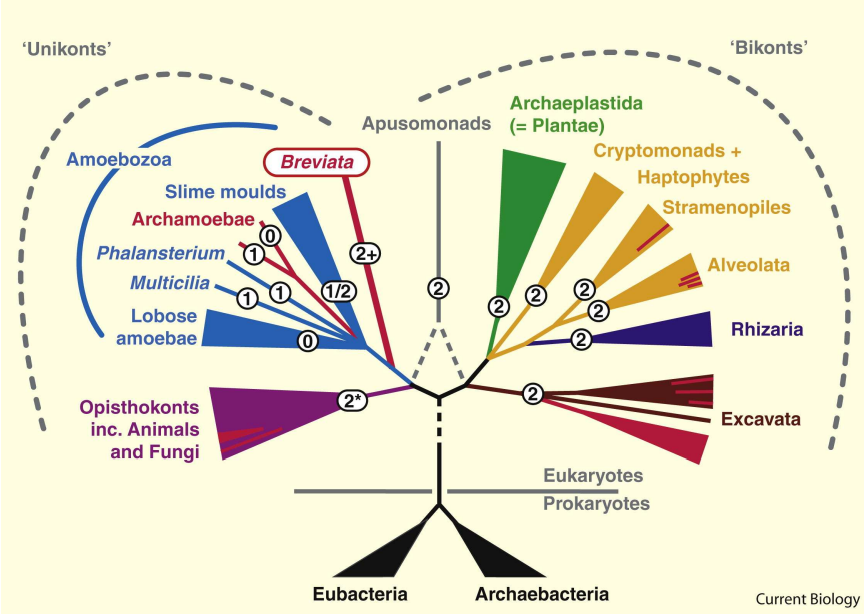


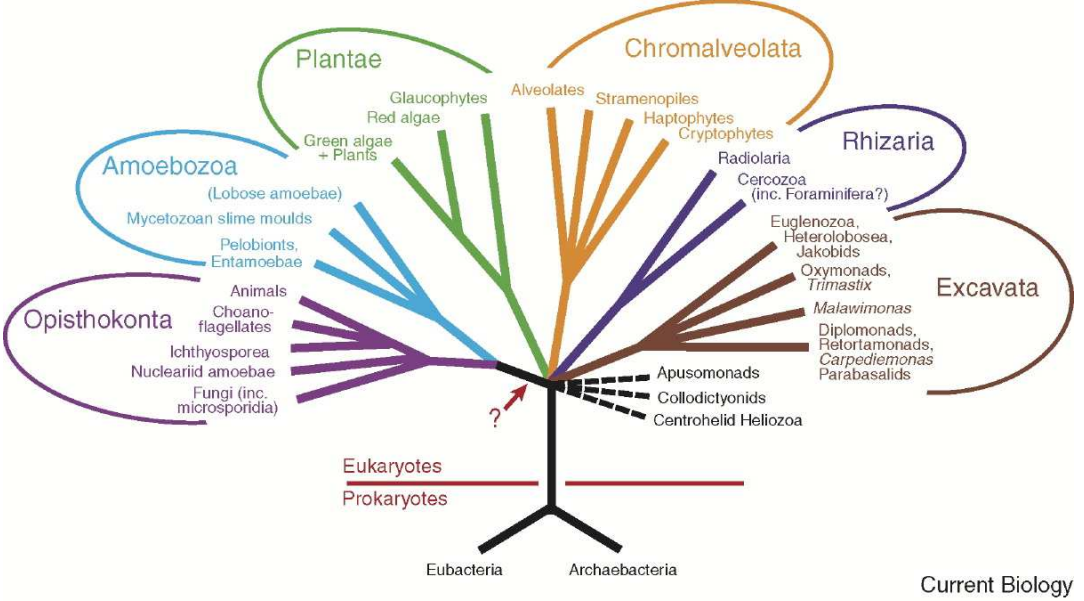
FIGURE 2.31 ***The Molecular Biologist's "Tree of Life"***

This tree of life includes both the traditional living creatures, such as plants and animals, as well as the two genetically distinct types of prokaryotic cell (eubacteria and archaebacteria). At the bottom are shown a variety of gene creatures, whose relationships are still mostly uncertain.

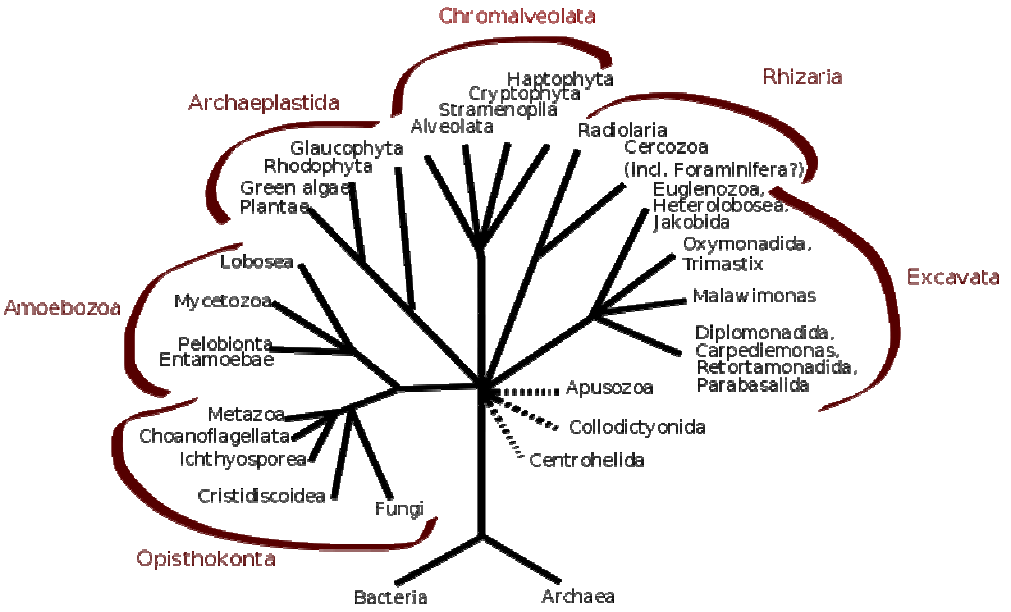


Archaeplastida (or **Plantae** *sensu lato***)**

Opisthokonts (Greek: ὀπίσθιος (*opísthios*) = "rear, posterior" + κοντός (*kontós*) = "pole" i.e. "flagellum") or **"Fungi/Metazoa group"**



Current Biology



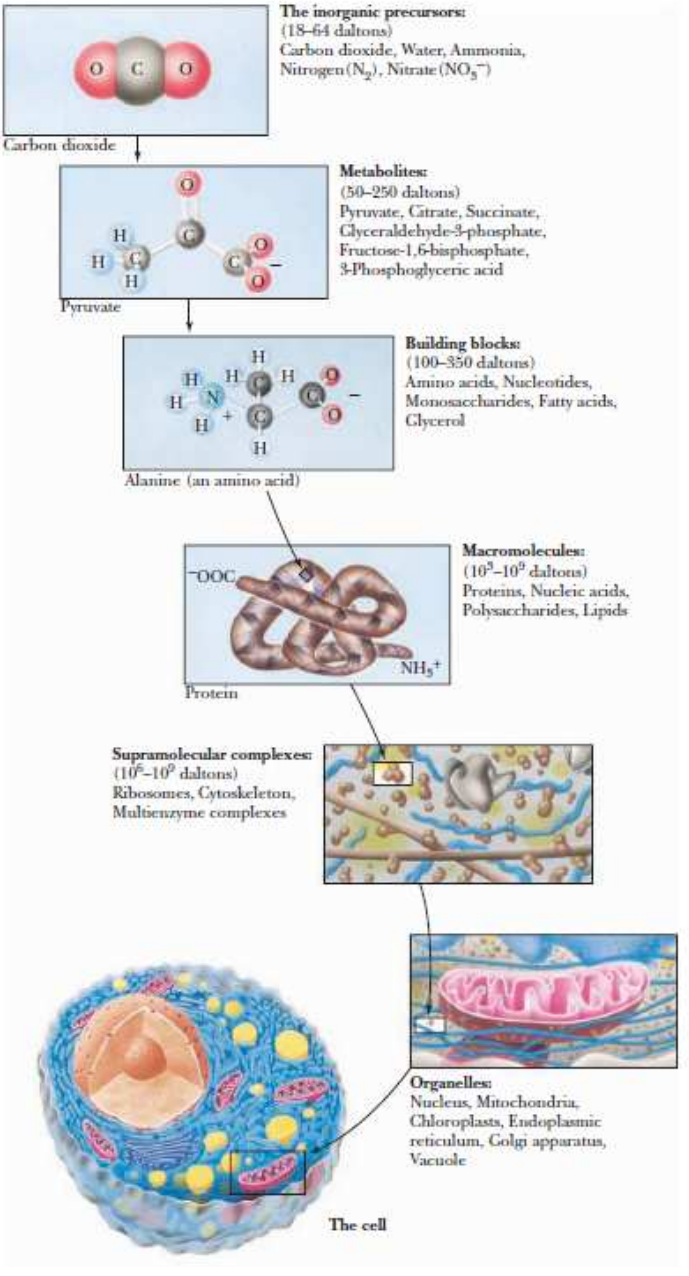


FIGURE 1.8 Molecular organization in the cell is a hierarchy.

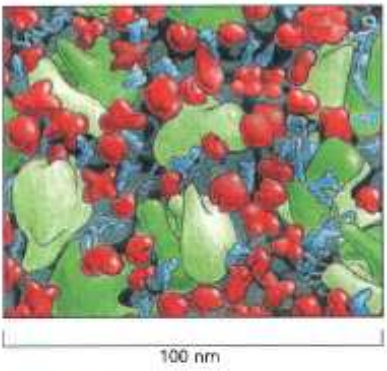


Figure 2–49 The structure of the cytoplasm. The drawing is approximately to scale and emphasizes the crowding in the cytoplasm. Only the macromolecules are shown: RNAs are shown in *blue*, ribosomes in *green*, and proteins in *red*. Enzymes and other macromolecules diffuse relatively slowly in the cytoplasm, in part because they interact with many other macromolecules; small molecules, by contrast, diffuse nearly as rapidly as they do in water. (Adapted from D.S. Goodsell, *Trends Biochem. Sci.* 16:203–206, 1991. With permission from Elsevier.)

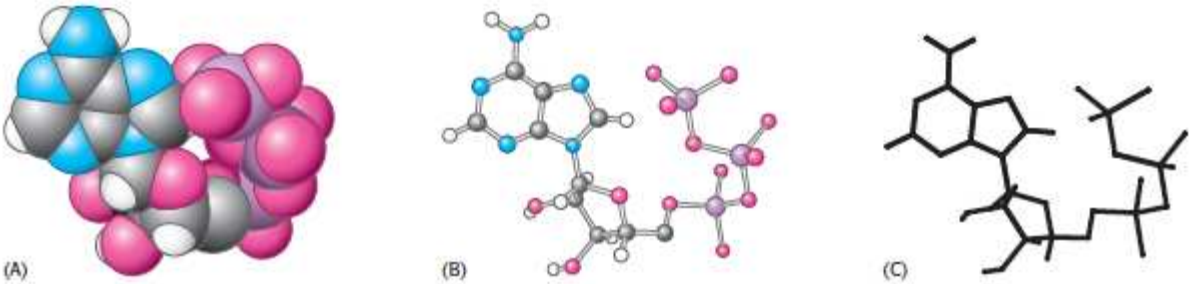


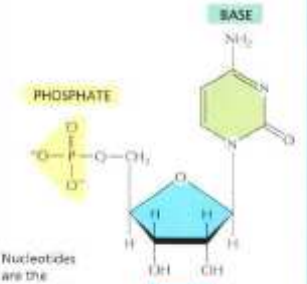
FIGURE 1.16 Molecular representations. Comparison of (A) space-filling, (B) ball-and-stick, and (C) skeletal models of ATP.

Struktura bioloških polimera

Struktura nukleinskih kiselina

NUCLEOTIDES

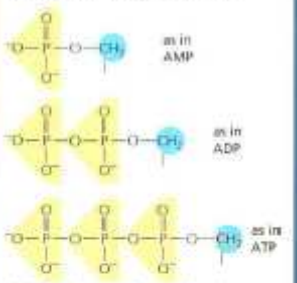
A nucleotide consists of a nitrogen-containing base, a five-carbon sugar, and one or more phosphate groups.



Nucleotides are the subunits of the nucleic acids.

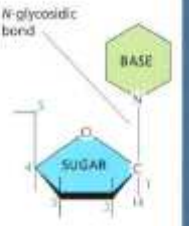
PHOSPHATES

The phosphates are normally joined to the C5 hydroxyl of the ribose or deoxyribose sugar (designated 5'). Mono-, di-, and triphosphates are common.



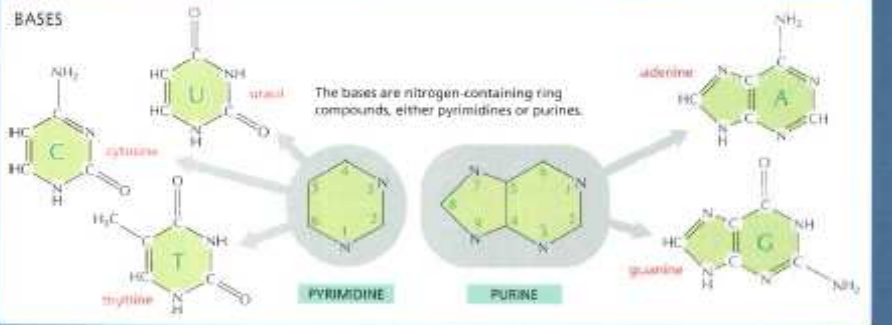
The phosphate makes a nucleotide negatively charged.

BASIC SUGAR LINKAGE



The base is linked to the same carbon (C1) used in sugar-sugar bonds.

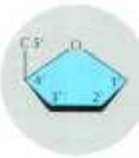
BASES



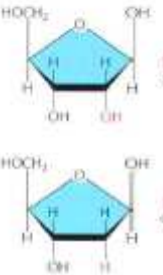
The bases are nitrogen-containing ring compounds, either pyrimidines or purines.

SUGARS

PENTOSE
a five-carbon sugar



two kinds are used



Each numbered carbon on the sugar of a nucleotide is followed by a prime mark; therefore, one speaks of "the 5-prime carbon," etc.

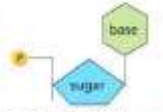
NOMENCLATURE A nucleoside or nucleotide is named according to its nitrogenous base.

BASE	NUCLEOSIDE	ABBR.
adenine	adenosine	A
guanine	guanosine	G
cytosine	cytidine	C
uracil	uridine	U
thymine	thymidine	T

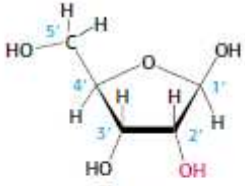
Single letter abbreviations are used variously as shorthand for (1) the base alone, (2) the nucleoside, or (3) the whole nucleotide—the context will usually make clear which of the three entities is meant. When the context is not sufficient, we will add the terms "base", "nucleoside", "nucleotide", or—as in the examples below—use the full 3-letter nucleotide code.

AMP = adenosine monophosphate
dAMP = deoxyadenosine monophosphate
UDP = uridine diphosphate
ATP = adenosine triphosphate

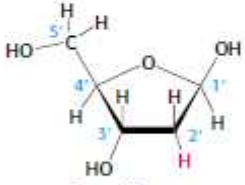
BASE + SUGAR = NUCLEOSIDE



BASE + SUGAR + PHOSPHATE = NUCLEOTIDE



Ribose



Deoxyribose

FIGURE 5.2 Ribose and deoxyribose. Atoms are numbered with primes to distinguish them from atoms in bases (see Figure 5.4).

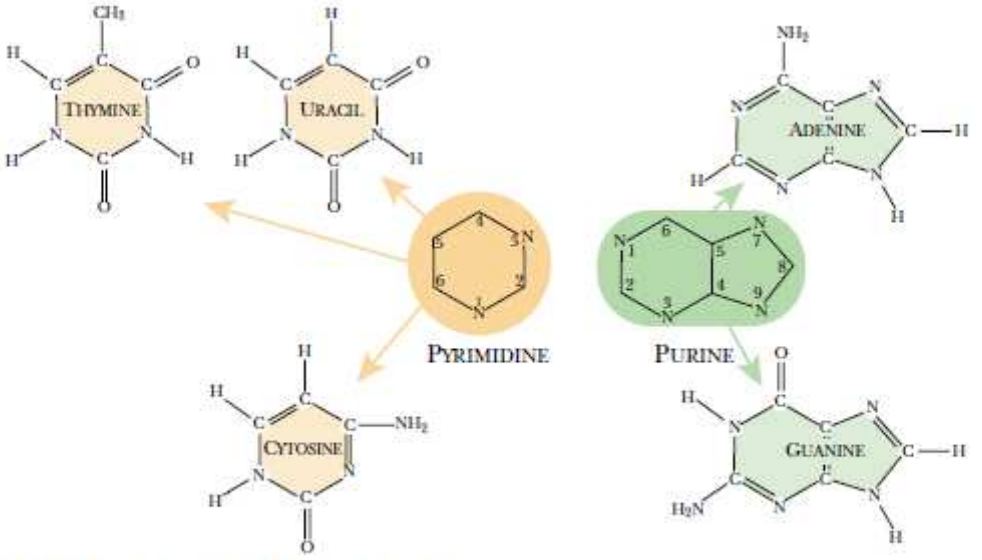
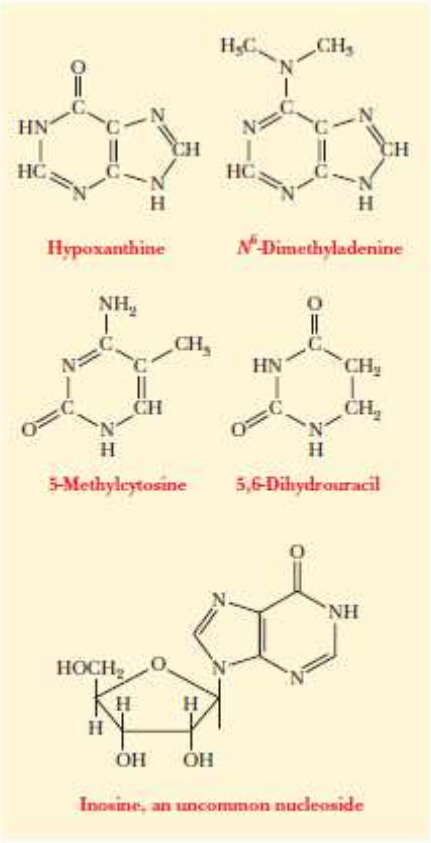


FIGURE 3.05 ***The Bases of the Nucleic Acids***

The four bases of DNA are adenine, guanine, cytosine and thymine. In RNA, uracil replaces thymine. Pyrimidine bases contain one-ring structures, whereas purine bases contain two-ring structures.



■ **FIGURE 9.2 Structures of some of the less common nucleobases.** When hypoxanthine is bonded to a sugar, the corresponding compound is called inosine.

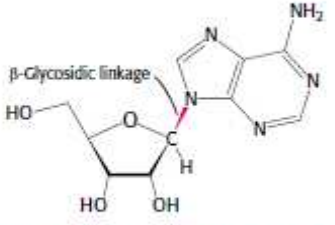


FIGURE 5.5 β-Glycosidic linkage in a nucleoside.

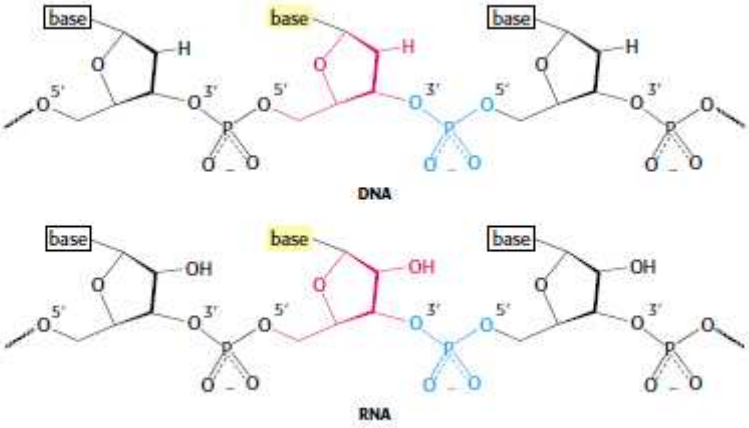
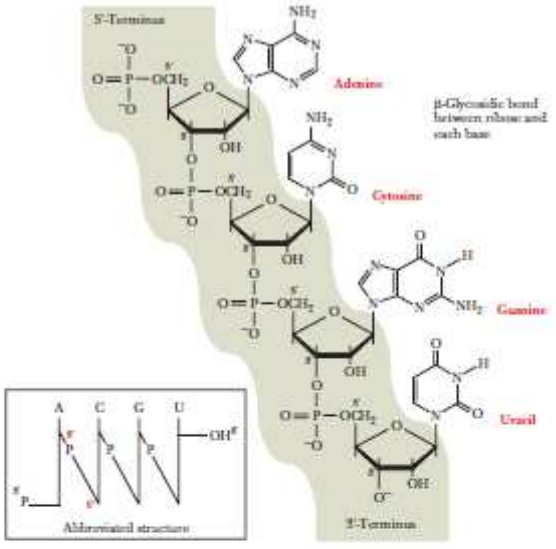
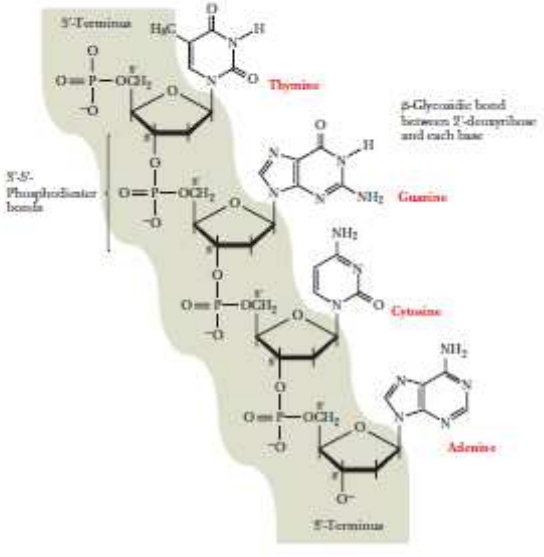


FIGURE 5.3 Backbones of DNA and RNA. The backbones of these nucleic acids are formed by 3'-to-5' phosphodiester linkages. A sugar unit is highlighted in red and a phosphate group in blue.



■ FIGURE 9.5 A fragment of an RNA chain.



■ FIGURE 9.6 A portion of a DNA chain.

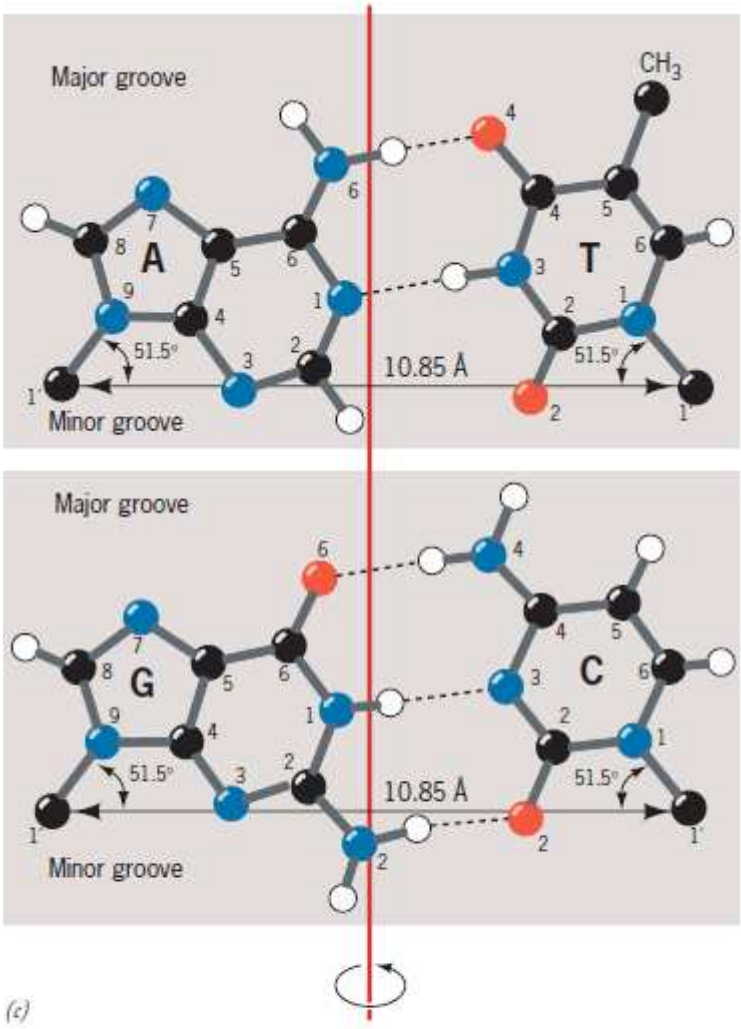


FIGURE 10.10 **The double helix.** (***continued***) (*c*) The Watson-Crick base pairs. The original model showed both A-T and G-C pairs with two hydrogen bonds; the third hydrogen bond in the G-C pair was subsequently identified by Linus Pauling. (*d*) Electron micrograph of DNA being released from the head of a T2 bacteriophage. This linear DNA molecule (note the two free ends) measures 68 μm in length,

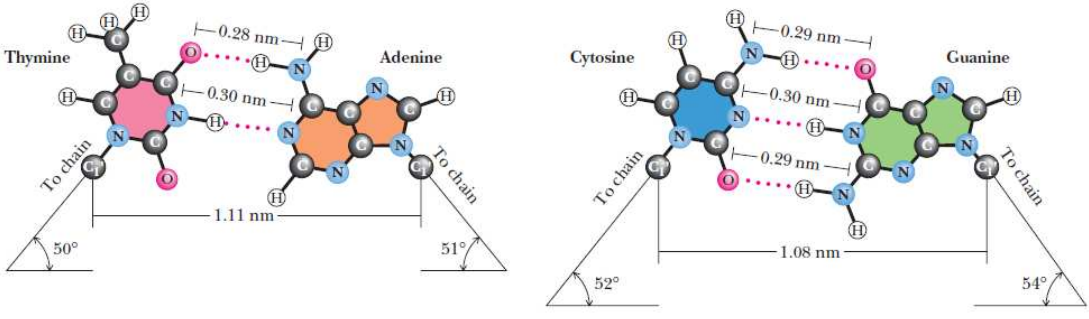


FIGURE 10.17 The Watson–Crick base pairs A:T and G:C.

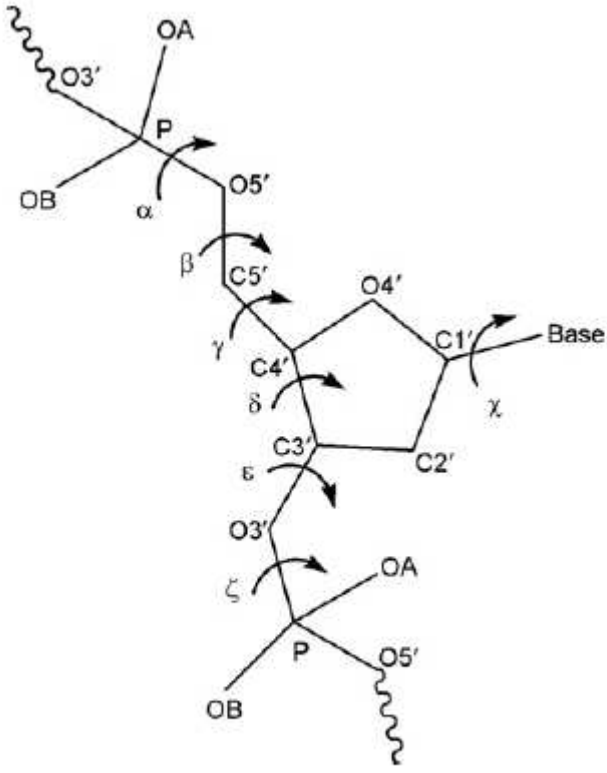


Figure 2.14 The backbone torsion angles in a unit nucleotide. Each rotatable bond is indicated by a curved arrow.

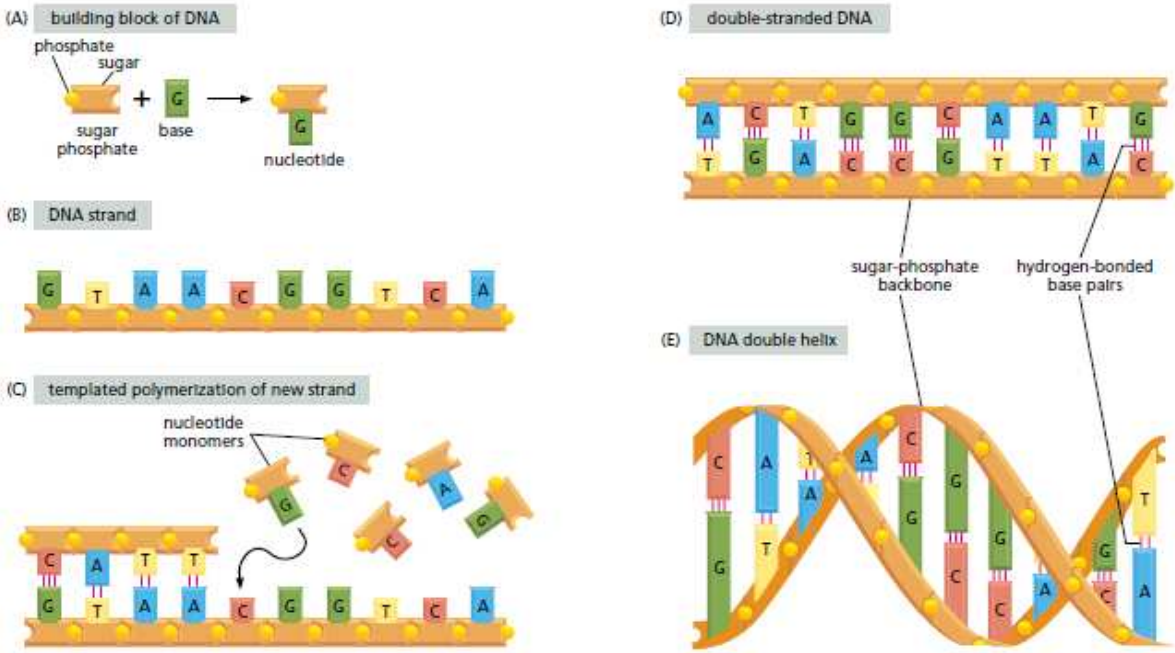
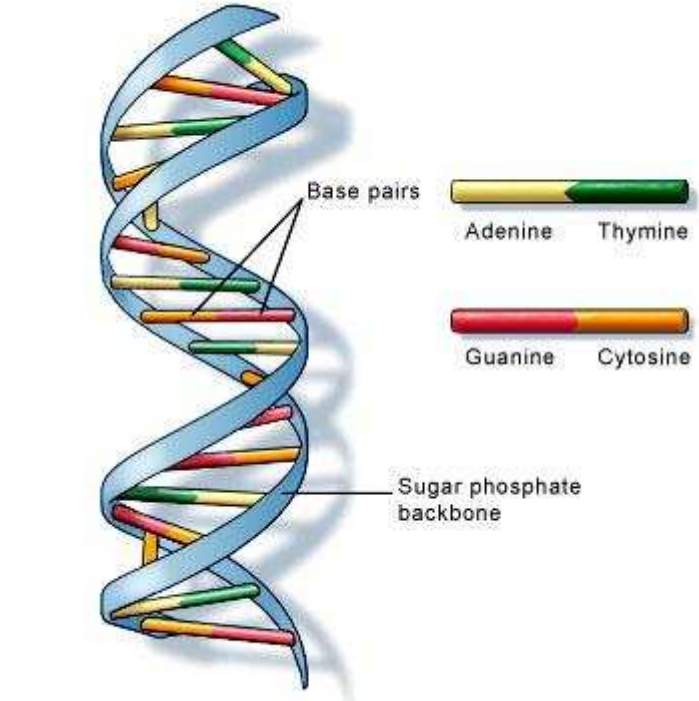
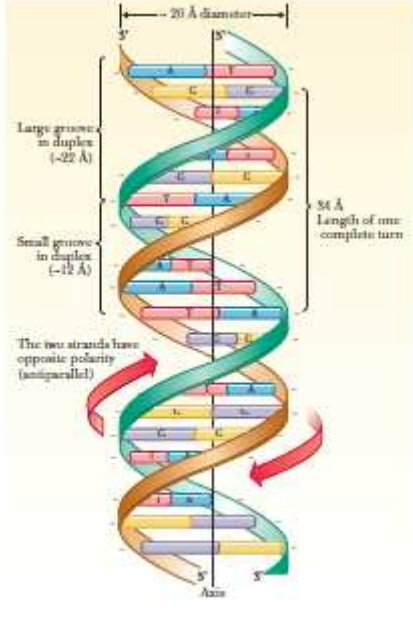


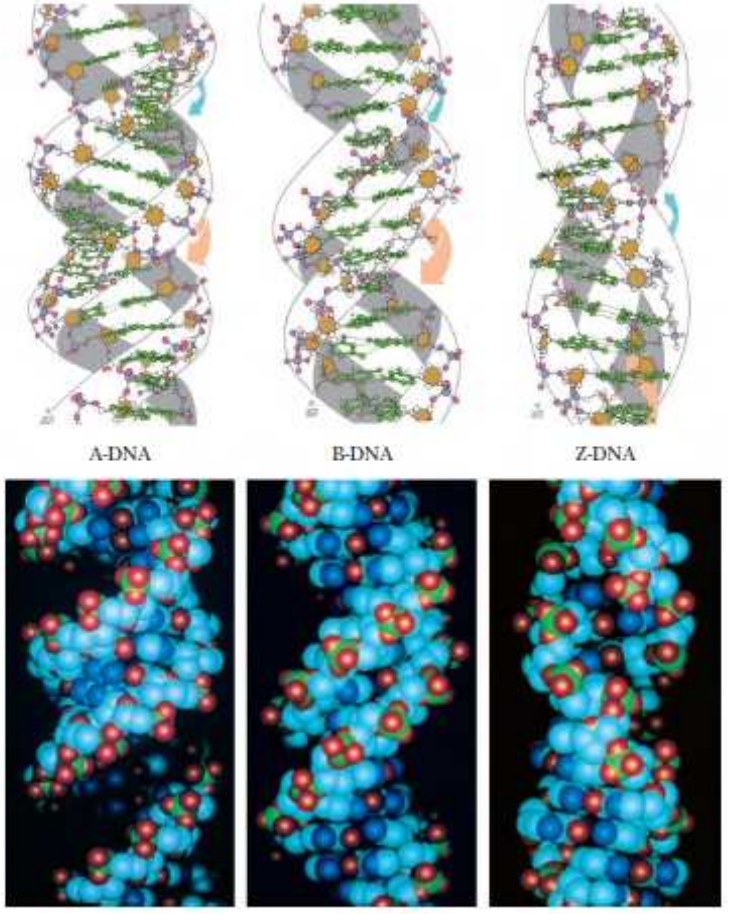
Figure 1–2 DNA and its building blocks. (A) DNA is made from simple subunits, called nucleotides, each consisting of a sugar-phosphate molecule with a nitrogen-containing sidegroup, or base, attached to it. The bases are of four types (adenine, guanine, cytosine, and thymine), corresponding to four distinct nucleotides, labeled A, G, C, and T. (B) A single strand of DNA consists of nucleotides joined together by sugar-phosphate linkages. Note that the individual sugar-phosphate units are asymmetric, giving the backbone of the strand a definite directionality, or polarity. This directionality guides the molecular processes by which the information in DNA is interpreted and copied in cells: the information is always "read" in a consistent order, just as written English text is read from left to right. (C) Through templated polymerization, the sequence of nucleotides in an existing DNA strand controls the sequence in which nucleotides are joined together in a new DNA strand; T in one strand pairs with A in the other, and G in one strand with C in the other. The new strand has a nucleotide sequence complementary to that of the old strand, and a backbone with opposite directionality: corresponding to the GTAA... of the original strand, it has ...TTAC. (D) A normal DNA molecule consists of two such complementary strands. The nucleotides within each strand are linked by strong (covalent) chemical bonds; the complementary nucleotides on opposite strands are held together more weakly, by hydrogen bonds. (E) The two strands twist around each other to form a double helix—a robust structure that can accommodate any sequence of nucleotides without altering its basic structure.



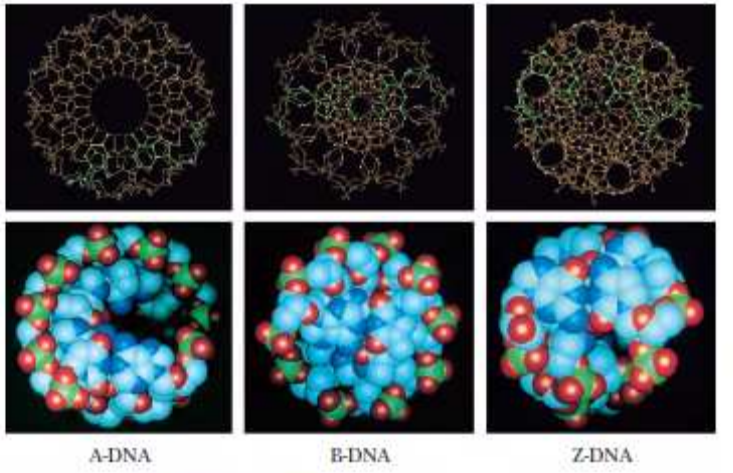
U.S. National Library of Medicine



■ **FIGURE 9.7 The double helix.** A complete turn of the helix spans ten base pairs, covering a distance of 34 Å (3.4 nm). The individual base pairs are spaced 3.4 Å (0.34 nm) apart. The places where the strands cross hide base pairs that extend perpendicular to the viewer. The inside diameter is 11 Å (1.1 nm), and the outside diameter is 20 Å (2.0 nm). Within the cylindrical outline of the double helix are two grooves, a small one and a large one. Both are large enough to accommodate polypeptide chains. The minus signs alongside the strands represent the many negatively charged phosphate groups along the entire length of each strand.

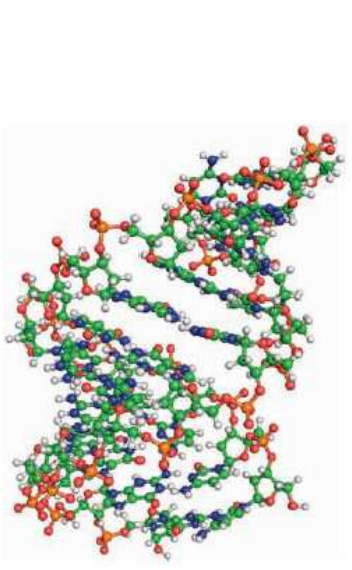


■ **FIGURE 9.9** Comparison of the A, B, and Z forms of DNA. (a) Side views.

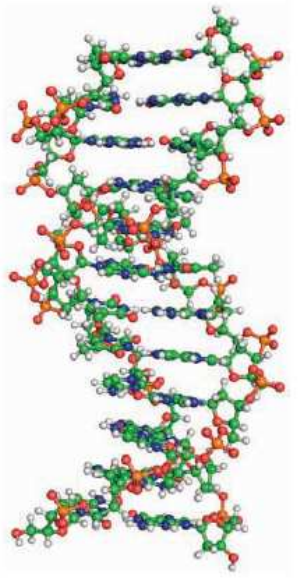


■ **FIGURE 9.9–cont'd** (b) Top views. Both parts include computer-generated space-filling models (bottom). The top half of each part shows corresponding ball-and-stick drawings. In the A form, the base pairs have a marked propeller-twist with respect to the helix axis. In the B form, the base pairs lie in a plane that is close to perpendicular to the helix axis. Z-DNA is a left-handed helix and in this respect differs from A-DNA and B-DNA, both of which are right-handed helices. *(Robert Sinsheimer, Fox Chase Cancer Research Center. Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)*

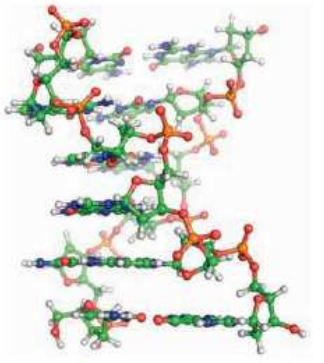
■ FIGURE 9.10 Right- and left-handed helices are related to each other in the same way as right and left hands.



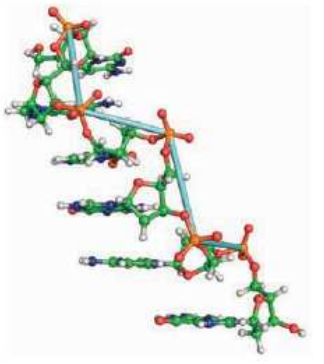
A-DNA

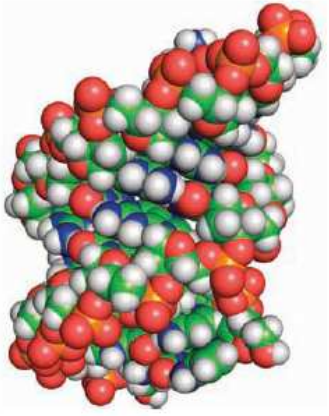


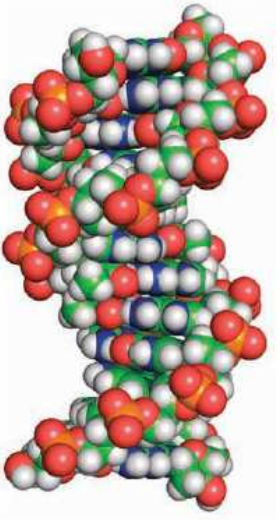
B-DNA



Z-DNA







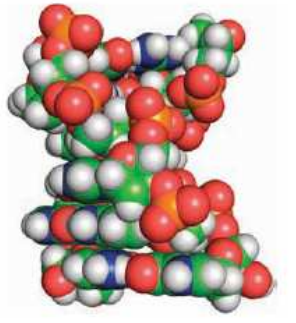


FIGURE 11.9 Comparison of the A-, B-, and Z-forms of the DNA double helix. The A- and B-structures show 12 bp of DNA; the Z-structures, 6 bp. The middle Z-structure shows just one strand of a Z-DNA double helix to illustrate better the left-handed zigzag path of the polynucleotide backbones in Z-DNA. (The light blue line was added to show the imaginary zigzag path.) A-DNA: pdb id = 2D47, B-DNA: pdb id = 355D, Z-DNA: pdb id = 1DCG.

TABLE 11.1 Comparison of the Structural Properties of A-, B-, and Z-DNA

	Double Helix Type		
	A	B	Z
Overall proportions	Short and broad	Longer and thinner	Elongated and slim
Rise per base pair	2.3 Å	3.32 Å ± 0.19 Å	3.8 Å
Helix packing diameter	25.5 Å	23.7 Å	18.4 Å
Helix rotation sense	Right-handed	Right-handed	Left-handed
Base pairs per helix repeat	1	1	2
Base pairs per turn of helix	~11	~10	12
Mean rotation per base pair	33.6°	35.9° ± 4.2°	−60°/2
Pitch per turn of helix	24.6 Å	33.2 Å	45.6 Å
Base-pair tilt from the perpendicular	+19°	−1.2° ± 4.1°	−9°
Base-pair mean propeller twist	+18°	+16° ± 7°	~0°
Helix axis location	Major groove	Through base pairs	Minor groove
Major groove proportions	Extremely narrow but very deep	Wide and with intermediate depth	Flattened out on he surface
Minor groove proportions	Very broad but shallow	Narrow and with intermediate depth	Extremely narrow b deep
Glycosyl bond conformation	anti	anti	anti at C, syn at G

Adapted from Dickerson, R.L., et al., 1983. Helix geometry and hydration in A-DNA, B-DNA, and Z-DNA. *Cold Spring Harbor Symposium on Quantitative Biology* **47**:13–24.

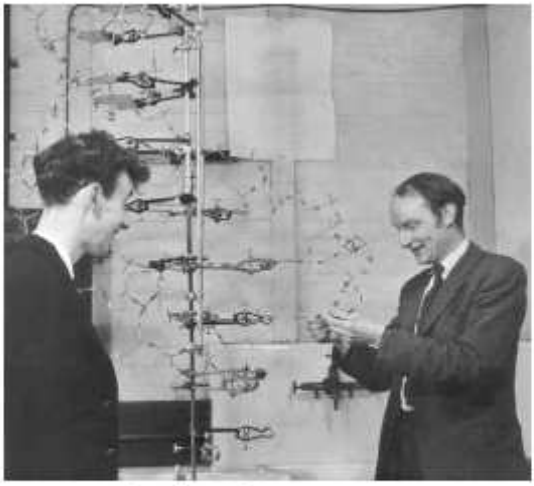


FIGURE 4.01 ***Watson and Crick in the 1950s***

James Watson (b.1928) at left and Francis Crick (b.1916), with their model of part of a DNA molecule in 1953. Courtesy of: A. Barrington Brown, Science Photo Library.

http://upload.wikimedia.org/wikipedia/commons/8/81/ADN_animation.gif

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1906).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

- **A** = adenine
- **C** = cytosine
- **G** = guanine
- **T** = thymine
- **R** = G A (purine)
- **Y** = T C (pyrimidine)

- **K** = G T (keto)
- **M** = A C (amino)
- **S** = G C (strong bonds)
- **W** = A T (weak bonds)
- **B** = G T C (all but A)
- **D** = G A T (all but C)
- **H** = A C T (all but G)
- **V** = G C A (all but T)
- **N** = A G C T (any)

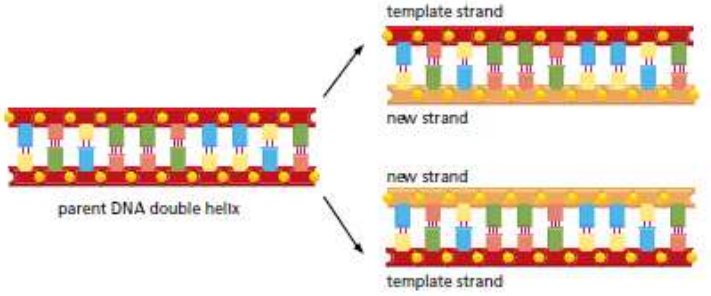
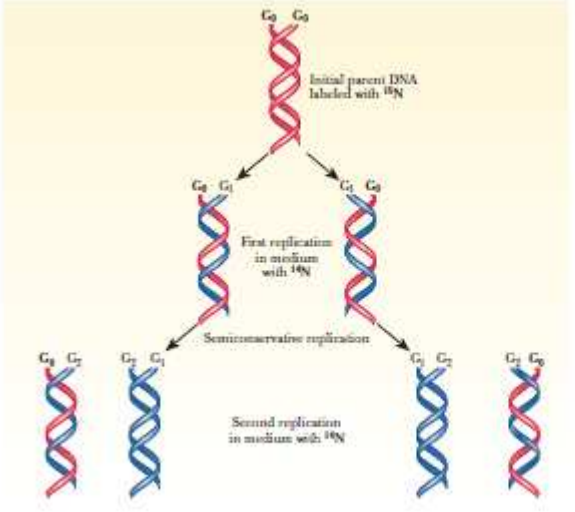
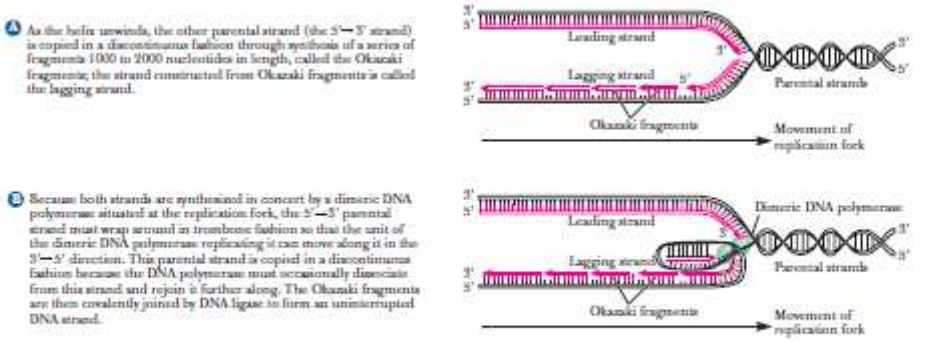


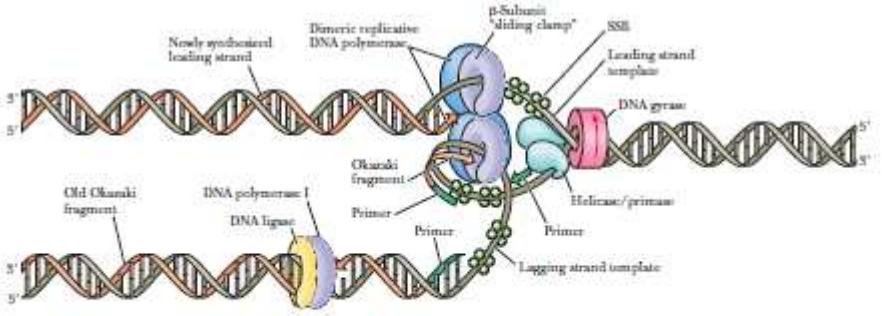
Figure 1–3 The copying of genetic information by DNA replication. In this process, the two strands of a DNA double helix are pulled apart, and each serves as a template for synthesis of a new complementary strand.



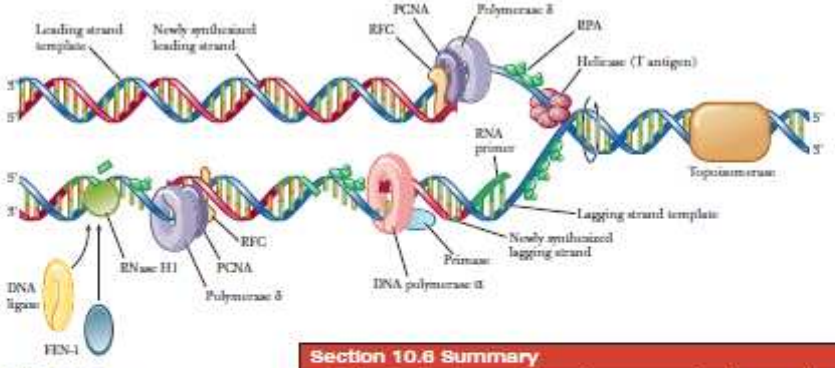
■ **FIGURE 10.2 The labeling pattern of ^{15}N strands in semiconservative replication.** (G_0 indicates original strands; G_1 indicates new strands after the first generation; G_2 indicates new strands after the second generation.)



■ **FIGURE 10.8 The semidiscontinuous model for DNA replication.** Newly synthesized DNA is shown in red. Because DNA polymerases only polymerize nucleotides 5′ → 3′, both strands must be synthesized in the 5′ → 3′ direction. Thus, the copy of the parental 5′ → 3′ strand is synthesized continuously; this newly made strand is designated the leading strand.



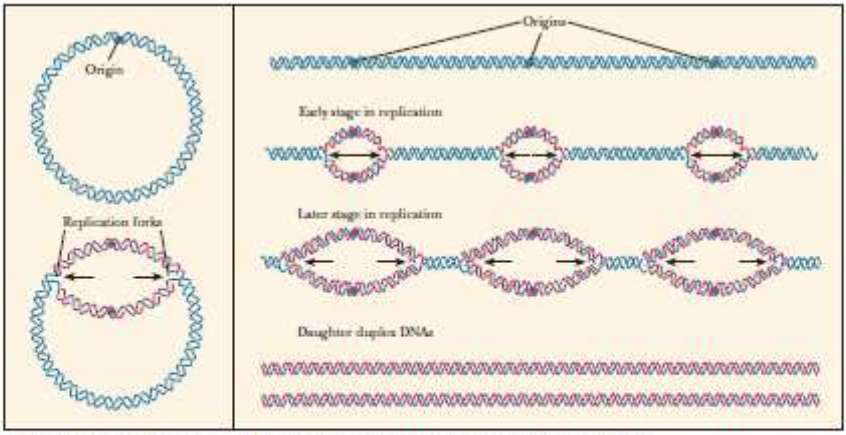
■ **FIGURE 10.9 General features of a replication fork.** The DNA duplex is unwound by the action of DNA gyrase and helicase, and the single strands are coated with SSB (ssDNA-binding protein). Primase periodically primes synthesis on the lagging strand. Each half of the dimeric replicative polymerase is a holoenzyme bound to its template strand by a β-subunit sliding clamp. DNA polymerase I and DNA ligase act downstream on the lagging strand to remove RNA primers, replace them with DNA, and ligate the Okazaki fragments.



■ **FIGURE 10.20 The basics of the eukaryotic replication fork.** The primase activity is associated with DNA polymerase α. After a few nucleotides are incorporated, DNA polymerase δ, with its associated proteins called PCNA and RFC, bind and do the majority of the synthesis. The enzymes FEN-1 and RNase H1 degrade the RNA primers in eukaryotic replication. *(From Cellular and Molecular Biology by Karp, Figure 13-22. Used by permission of John Wiley & Sons, Inc.)*

Section 10.6 Summary

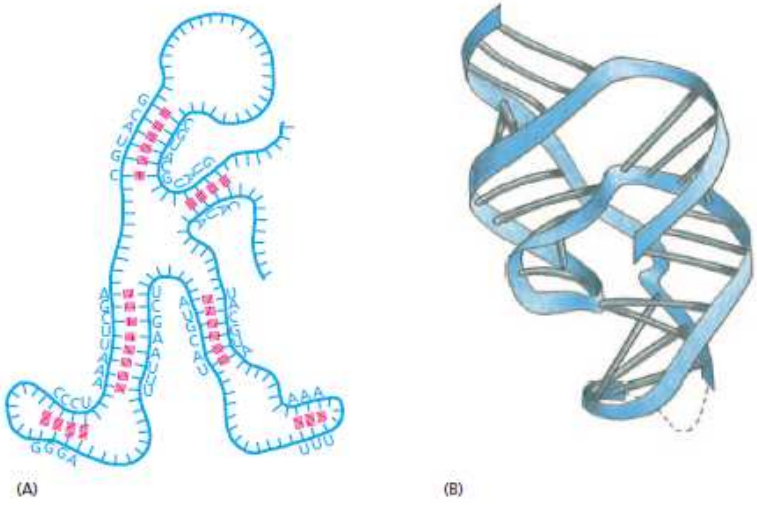
- Replication in eukaryotes follows the same general outline as replication in prokaryotes, with the most important difference being the presence of histone proteins complexed to eukaryotic DNA.
- Different proteins are used, and the system is more complex than it is in prokaryotes. Replication is controlled so that it occurs only once during a cell-division cycle, during the S phase.
- Five different DNA polymerases are present in eukaryotes: α, β, γ, δ, and ε. Polymerase δ is the principal synthesizer of DNA and is the equivalent of Pol III in prokaryotes.



(A) Replication of the chromosome of *E. coli*, a typical prokaryote. There is one origin of replication, and there are two replication forks.

(B) Replication of a eukaryotic chromosome. There are several origins of replication, and there are two replication forks for each origin. The "bubbles" that arise from each origin eventually coalesce.

■ **FIGURE 10.4 Bidirectional replication.** Bidirectional replication of DNA is shown for prokaryotes (one origin of replication) and eukaryotes (several origins). Bidirectional replication refers to overall synthesis (compare this with Figure 10.6).



(A) (B)

Figure 1–6 The conformation of an RNA molecule. (A) Nucleotide pairing between different regions of the same RNA polymer chain causes the molecule to adopt a distinctive shape. (B) The three-dimensional structure of an actual RNA molecule, from hepatitis delta virus, that catalyzes RNA strand cleavage. The *blue* ribbon represents the sugar-phosphate backbone; the bars represent base pairs. (B, based on A.R. Ferré D'Amaré, K. Zhou and J.A. Doudna, *Nature* 395:567–574, 1998. With permission from Macmillan Publishers Ltd.)

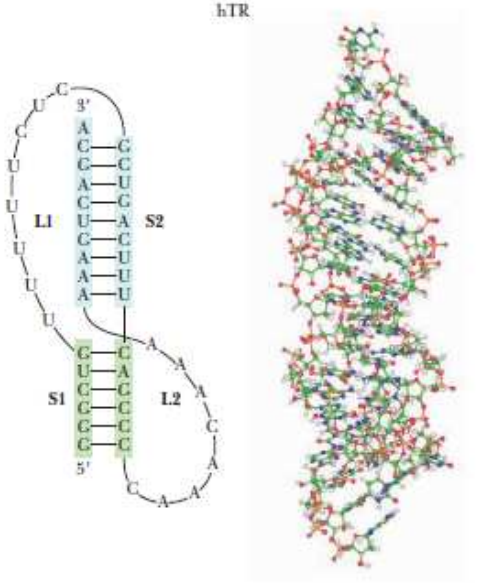


FIGURE 11.32 RNA pseudoknots are formed when a single-stranded region of RNA folds to base-pair with a hairpin loop. Loops L1 and L2, as shown on the sequence representation of human telomerase RNA (hTR) on the left, form a pseudoknot. The three-dimensional structure of an hTR pseudoknot is shown on the right (pdb id = 1YMO). (Adapted from Figure 2 in Staple, D. W., and Butcher, S. E., 2005. Pseudoknots: RNA structures with diverse functions. *PLoS Biology* **3**:e213.)

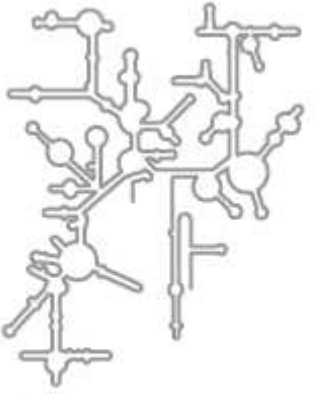


FIGURE 10.21 Ribosomal RNA has a complex secondary structure due to many intrastrand hydrogen bonds. The gray line in this figure traces a polynucleotide chain consisting of more than 1000 nucleotides. Aligned regions represent H-bonded complementary base sequences.

(a) *E. coli* (a eubacterium)

(b) *H. volcanii* (an archaebacterium)

(c) *S. cerevisiae* (yeast, a lower eukaryote

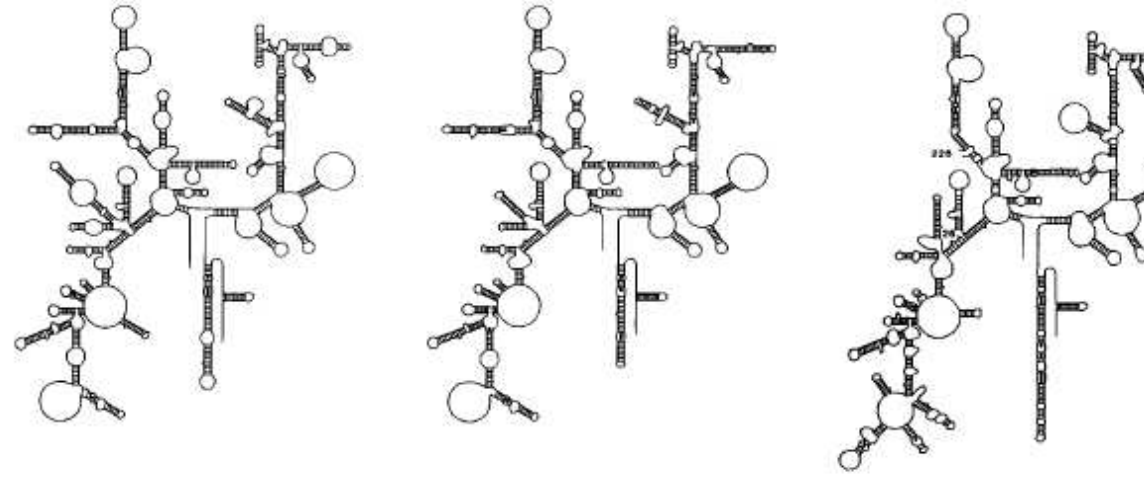


FIGURE 11.36 Comparison of secondary structures of 16S-like rRNAs from **(a)** a bacterium (*E.coli*), **(b)** chaeon (*H.volcanii*), and **(c)** a eukaryote (*S.cerevisiae*, a yeast).

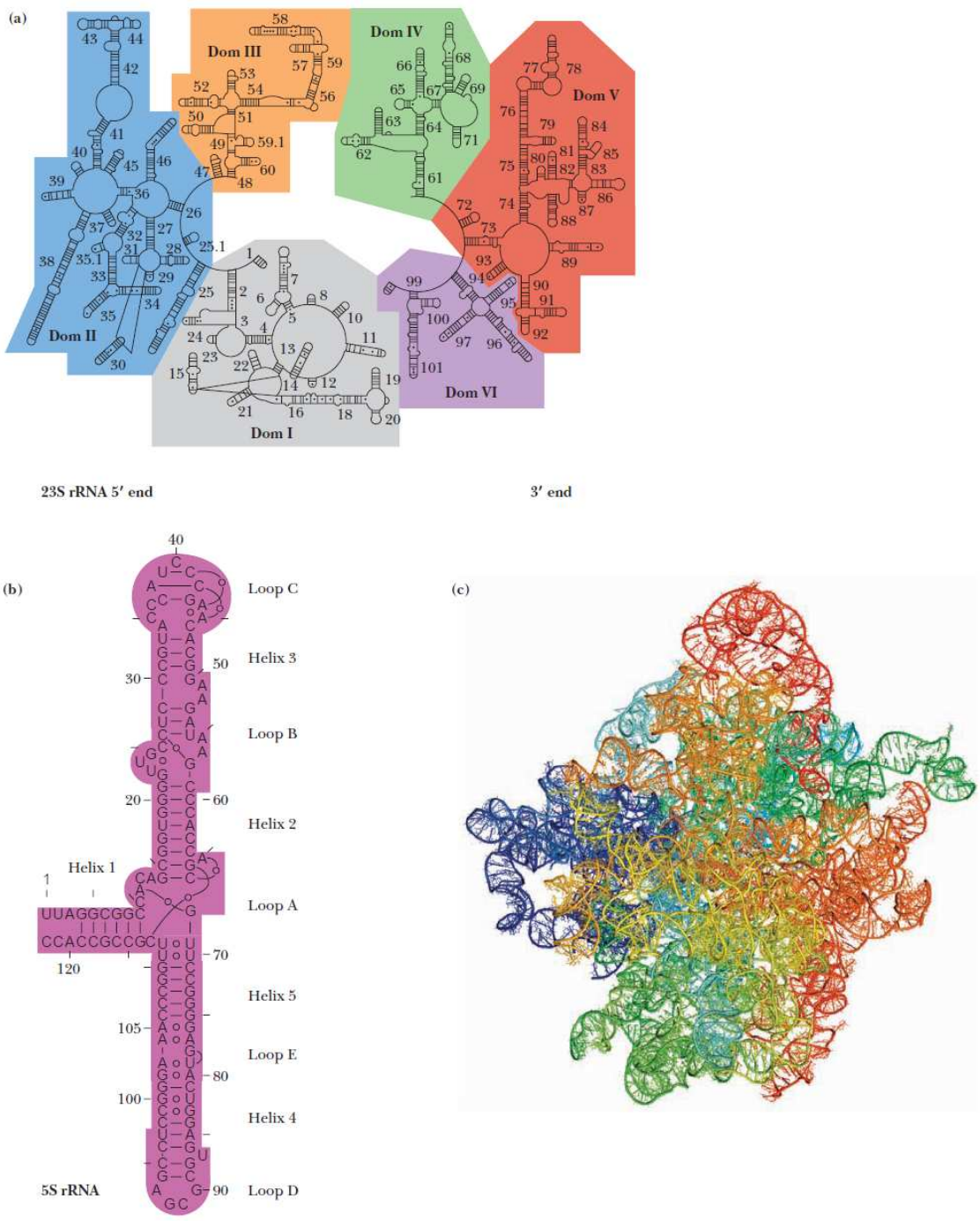


FIGURE 11.37 The secondary and tertiary structures of rRNAs in the 50S ribosomal subunit from the archaeon *Haloarcula marismortui* (pdb id = 1FFk). **(a)** Secondary structure of the 23S rRNA, with various domains color-coded. **(b)** Secondary structure of 5S rRNA. **(c)** Tertiary structure of the 5S and 23S rRNAs within the 50S ribosomal subunit. The 5S rRNA (red) lies atop the 23S rRNA. (Adapted from Figure 4 in Ban, N., et al., 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289:**905–920.)

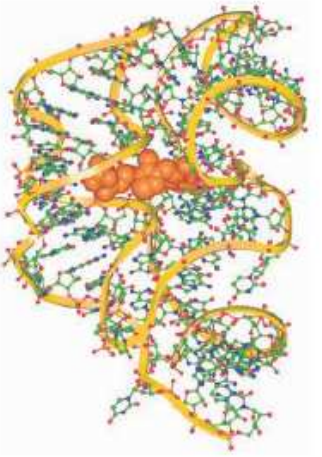


FIGURE 11.38 Structure of the thiamine pyrophosphate (TPP) riboswitch, a conserved region within the mRNA that encodes enzymes for synthesis of this coenzyme (pdb id = 2CKY). TTP, a pyrimidine-containing compound, is shown in orange. (From Figure 1b in Thore, S., Leibundgdut, M., and Ban, N., 2006. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. Science **312**:1208–1211.)

Riboswitches, a naturally occurring class of aptamers, are conserved regions of mRNAs that reversibly bind specific metabolites and coenzymes and usually act as regulators of gene expression. Riboswitches are usually buried within the 5′- or 3′-untranslated regions of the mRNAs whose expression they regulate. Binding of the metabolite to the riboswitch typically blocks expression of the mRNA. Figure 11.38 shows the structure of the thiamine pyrophosphate riboswitch.

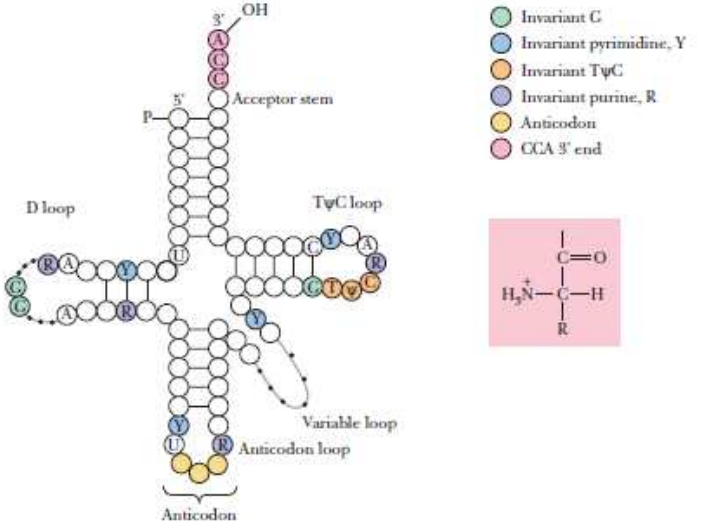


FIGURE 11.33 A general diagram for the structure of tRNA. The positions of invariant bases as well as bases that seldom vary are shown in color. R = purine; Y = pyrimidine. Dotted lines denote sites in the D loop and variable loop regions where varying numbers of nucleotides are found in different tRNAs. Inset: An aminoacyl group can add to the 3′-OH to create an aminoacyl-tRNA.

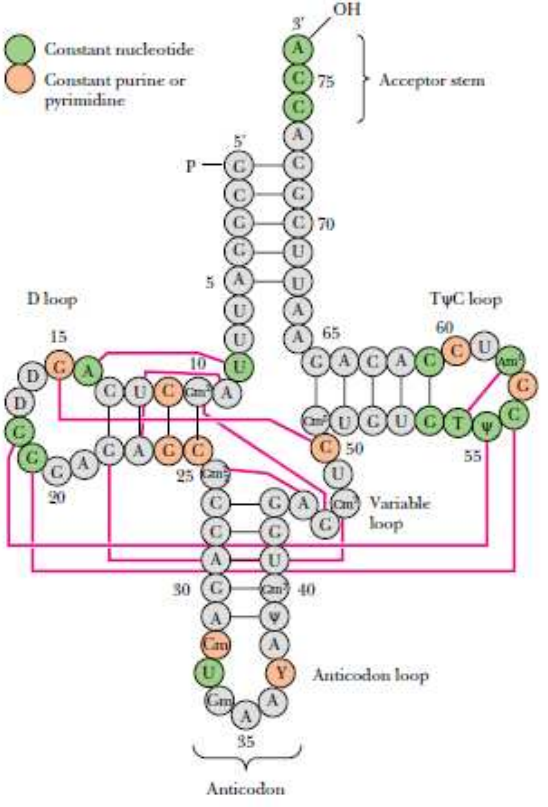
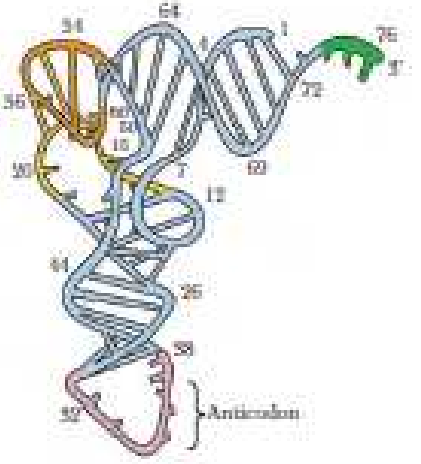
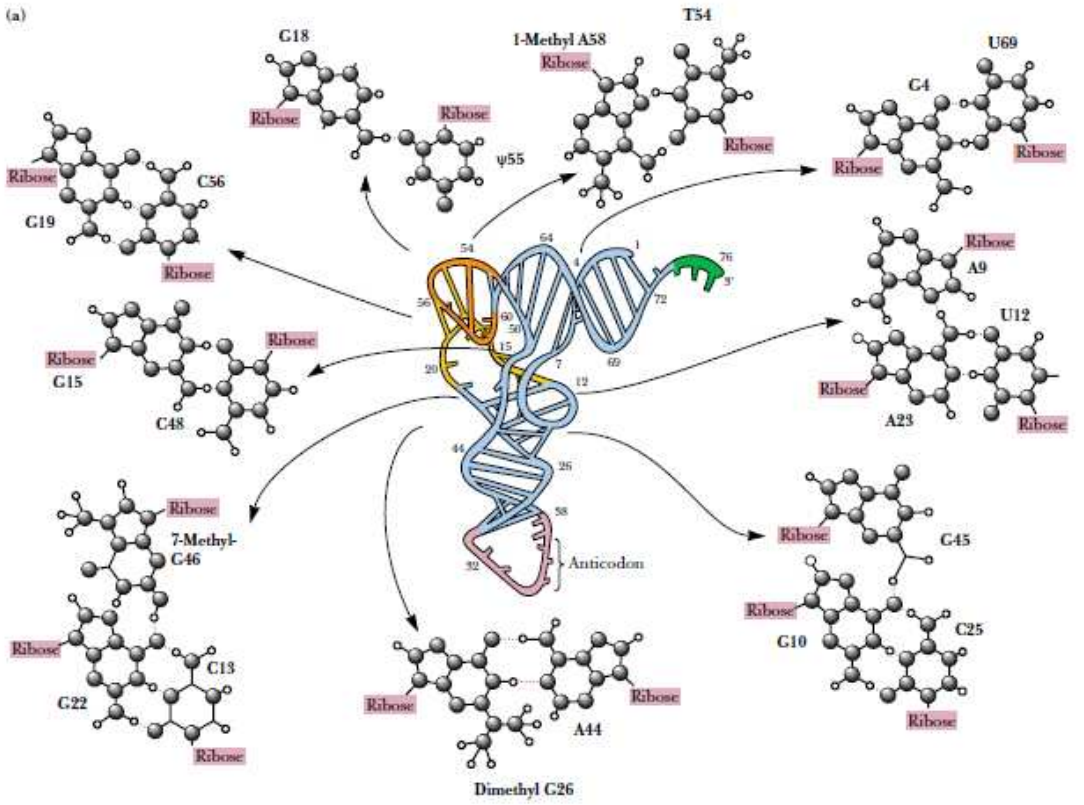


FIGURE 11.34 Tertiary interactions in yeast phenylalanine tRNA. The molecule is presented in the conventional cloverleaf secondary structure generated by intrastrand hydrogen bonding. Solid lines connect bases that are hydrogen bonded when this cloverleaf pattern is folded into the characteristic tRNA tertiary structure (see also Figure 11.35).



■ **FIGURE 9.24 The three-dimensional structure of yeast phenylalanine tRNA as deduced from X-ray diffraction studies of its crystals.** The tertiary folding is illustrated, and the ribose–phosphate backbone is presented as a continuous ribbon; H bonds are indicated by crossbars. Unpaired bases are shown as short, unconnected rods. The anticodon loop is at the bottom and the —CCA 3'—OH acceptor end is at the top right.



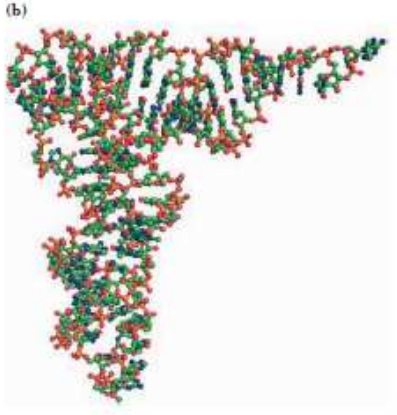
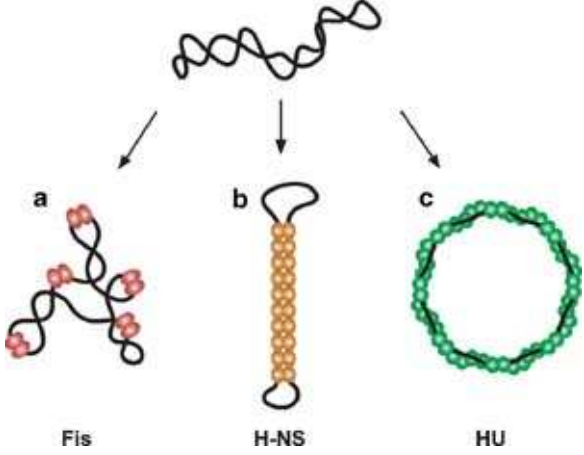
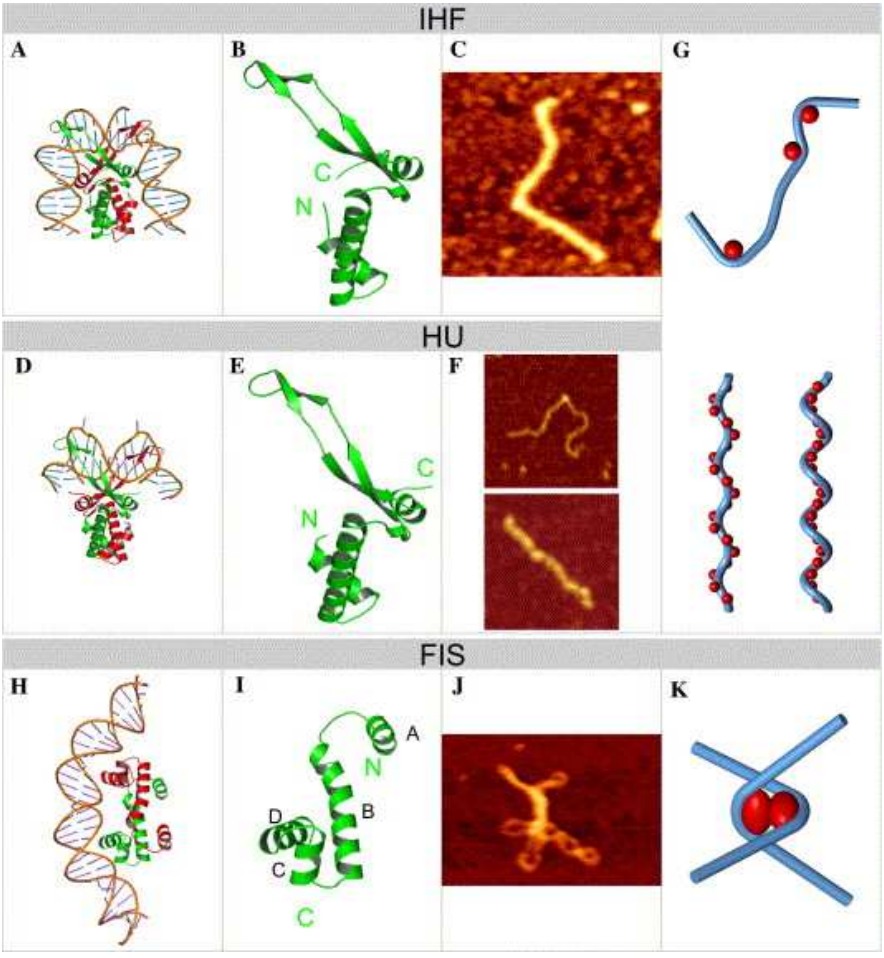
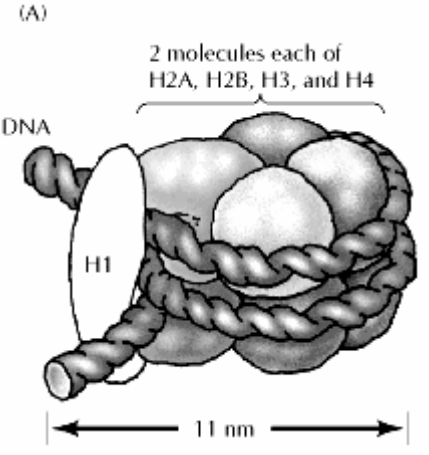


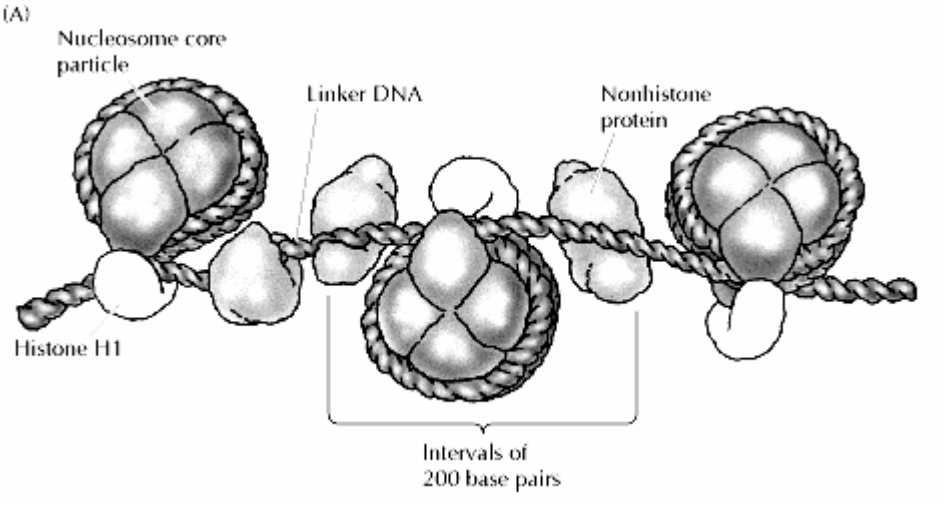
FIGURE 11.35 **(a)** The three-dimensional structure of yeast phenylalanine tRNA. The tertiary folding is illustrated in the center of the diagram with the ribose–phosphate backbone presented as a continuous ribbon; H bonds are indicated by crossbars. Unpaired bases are shown as short, unconnected rods. The anticodon loop is at the bottom and the -CCA 3′-OH acceptor end is at the top right. **(b)** A space-filling model of the molecule (pdb id = 6TNA).

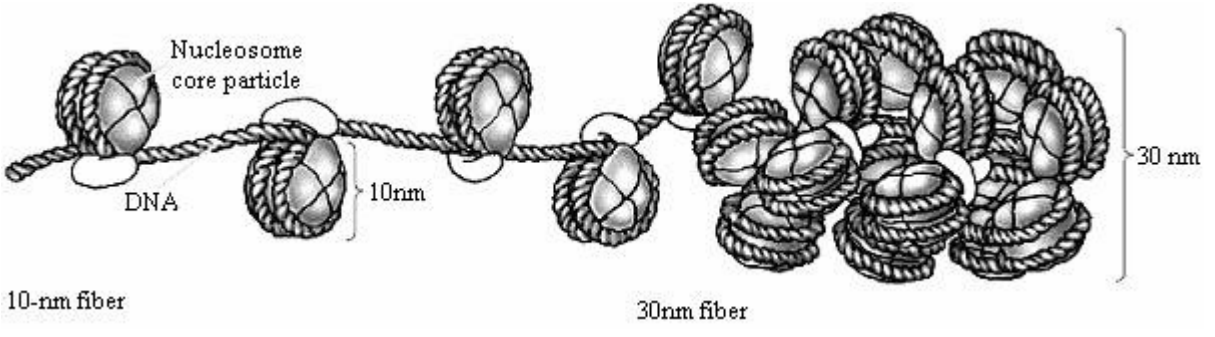


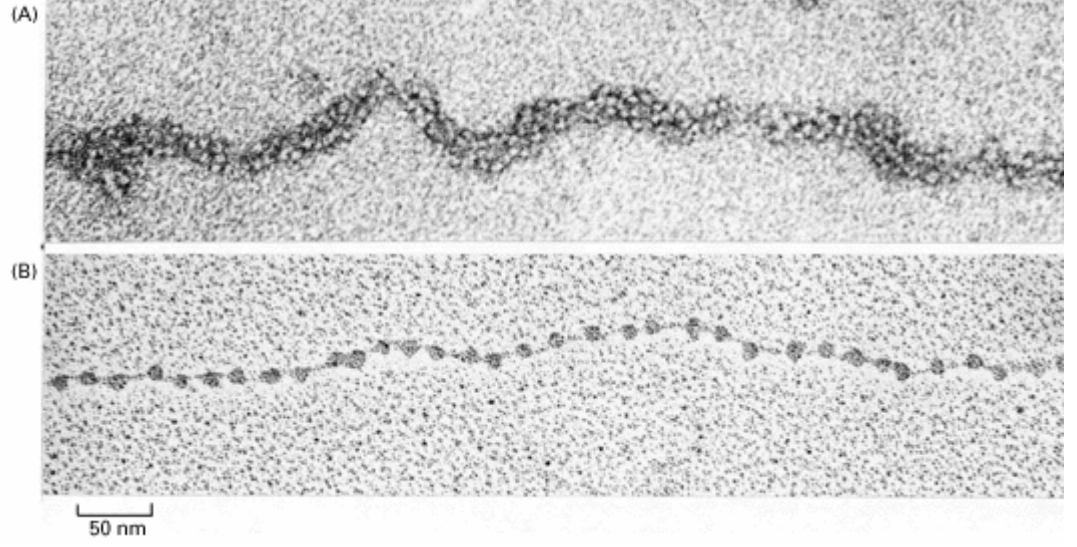
Fis, H-NS i HU nukleoid asocirani protein

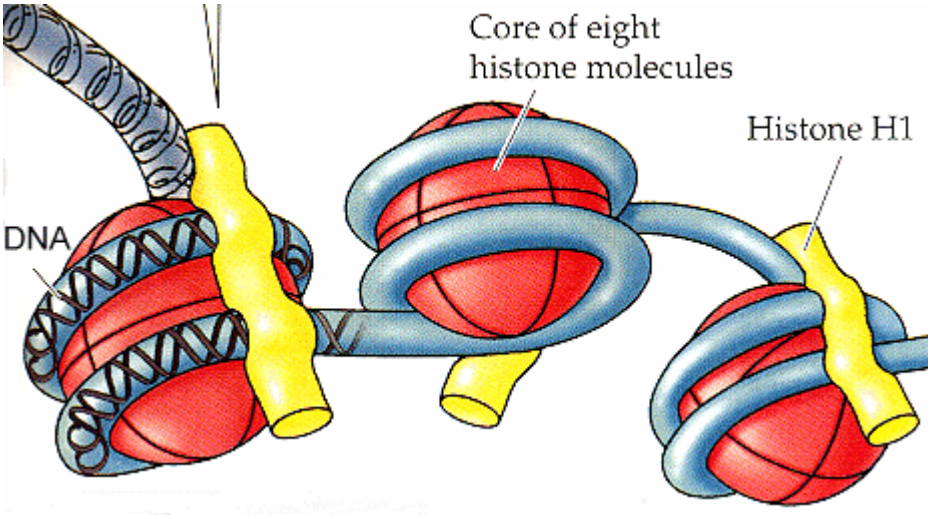


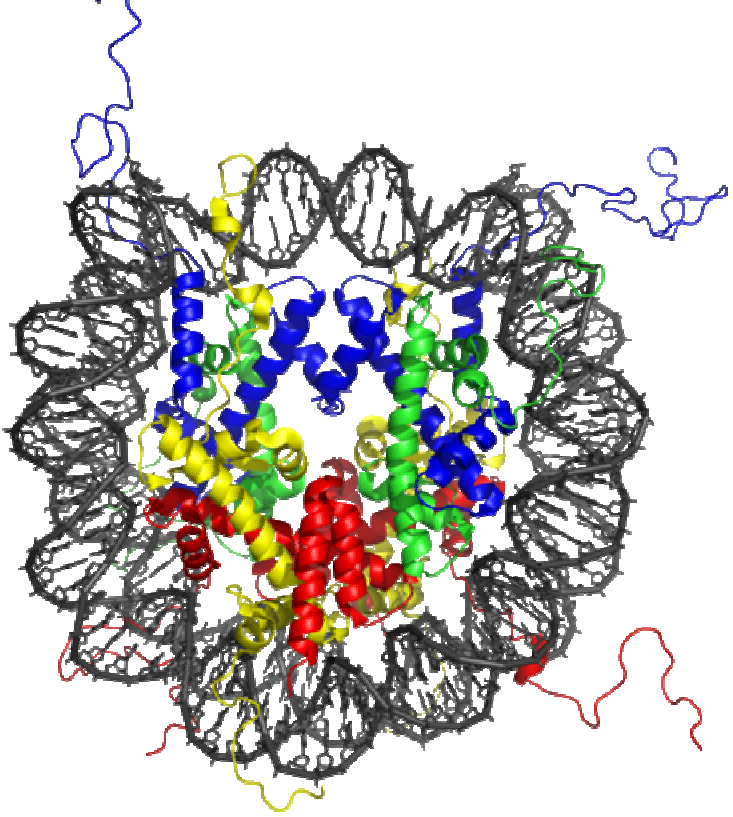

(A)
2 molecules each of
H2A, H2B, H3, and H4
DNA
H1
11 nm

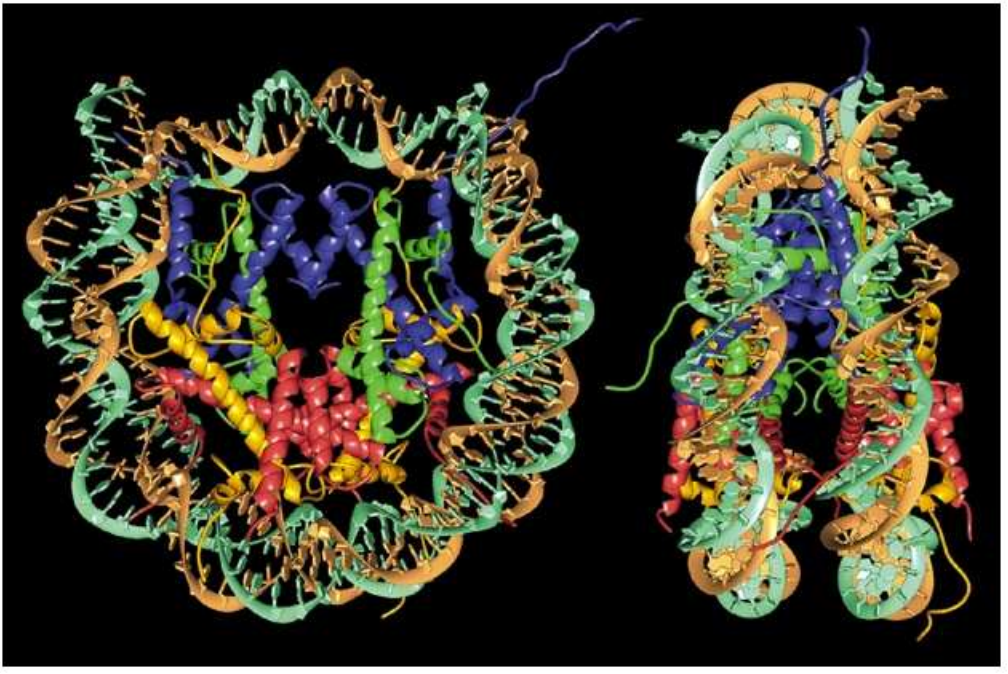

(A)
Nucleosome core particle
Linker DNA
Nonhistone protein
Histone H1
Intervals of
200 base pairs

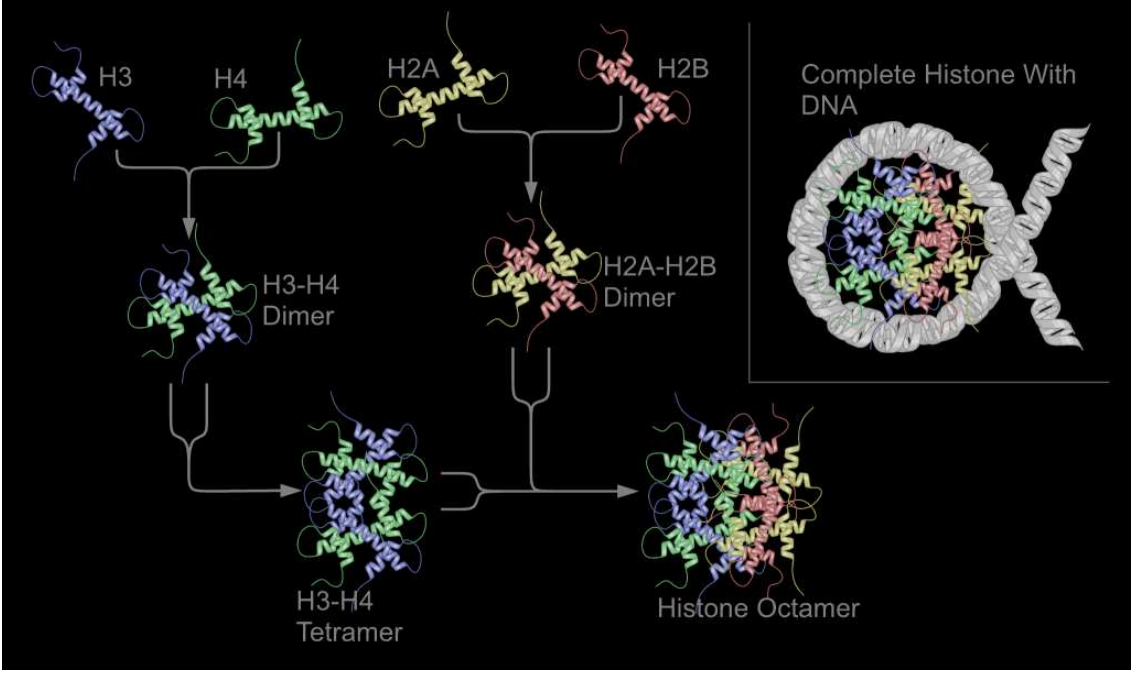

Nucleosome
core particle
DNA
10nm
30 nm
10-nm fiber
30nm fiber

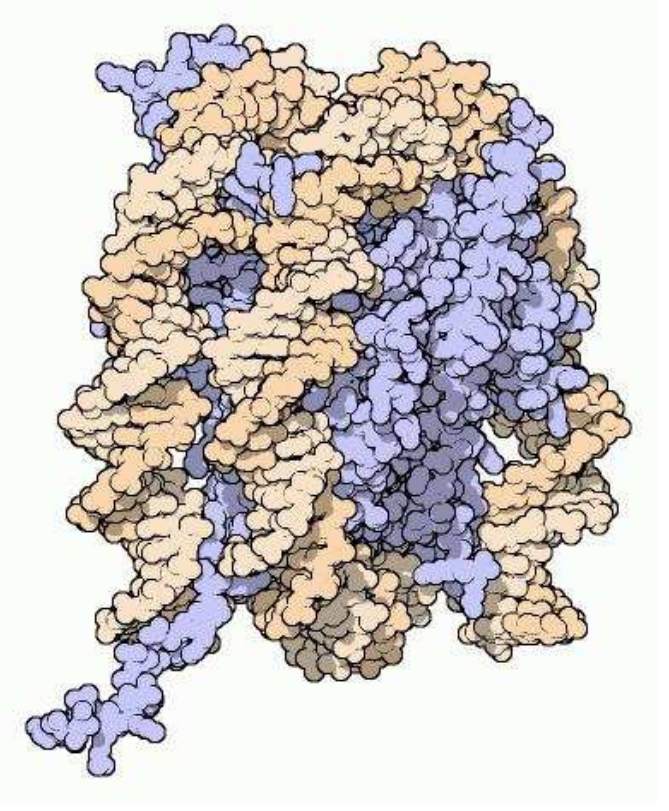

(A)
(B)
50 nm


Core of eight
histone molecules
Histone H1
DNA





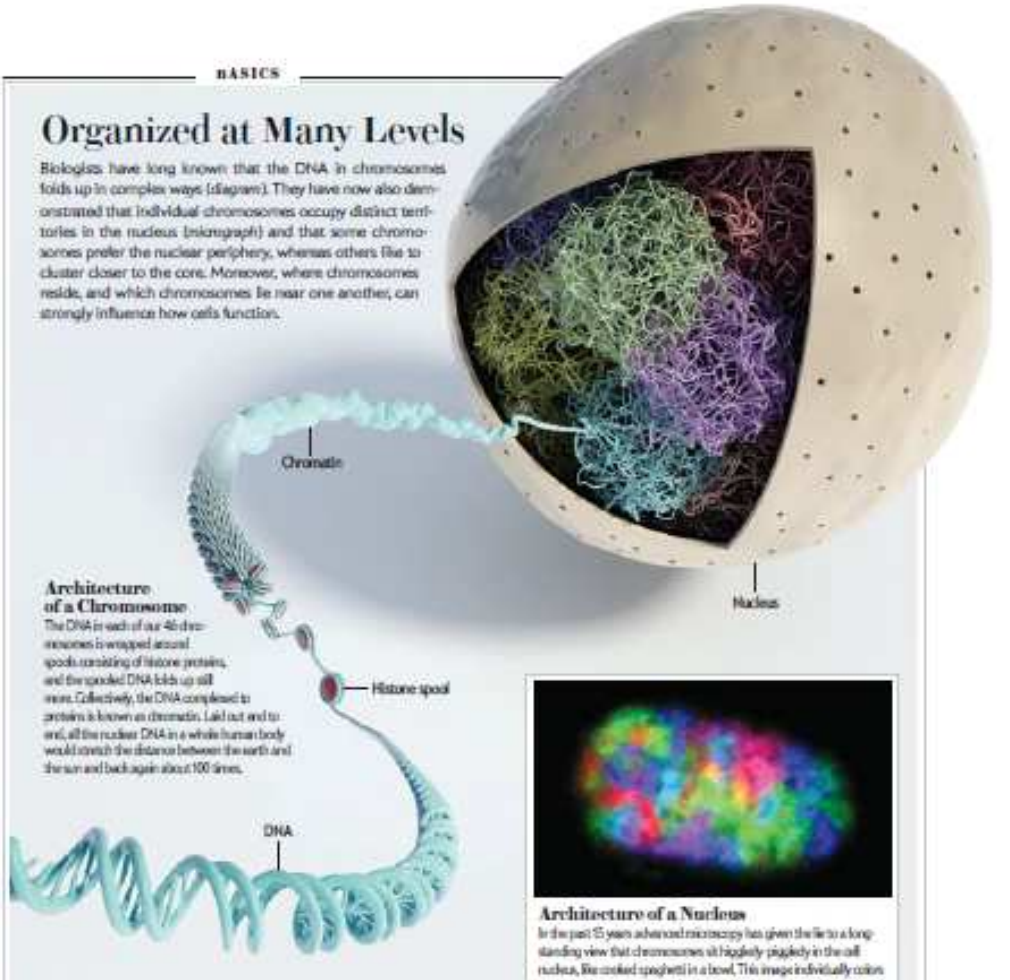

H3
H4
H2A
H2B
Complete Histone With DNA
H3-H4 Dimer
H2A-H2B Dimer
H3-H4 Tetramer
Histone Octamer



Hromozomi

Organized at Many Levels

Biologists have long known that the DNA in chromosomes folds up in complex ways (diagram). They have now also demonstrated that individual chromosomes occupy distinct territories in the nucleus (micrograph) and that some chromosomes prefer the nuclear periphery, whereas others like to cluster closer to the core. Moreover, where chromosomes reside, and which chromosomes lie near one another, can strongly influence how cells function.



Architecture of a Chromosome
The DNA in each of our 46 chromosomes is wrapped around spools consisting of histone proteins, and the spooled DNA folds up still more. Collectively, the DNA complexed to protein is known as chromatin. Laid out end to end, all the nuclear DNA in a whole human body would stretch the distance between the earth and the sun and back again about 100 times.

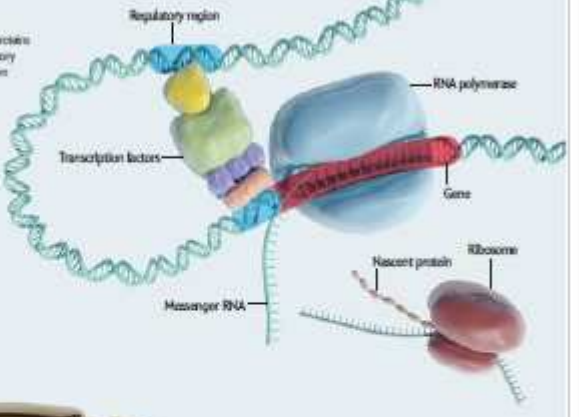
Architecture of a Nucleus
In the past 15 years advanced microscopy has given the lie to a long-standing view that chromosomes sit higgledy-piggledy in the cell nucleus, like cooked spaghetti in a bowl. This image individually colors the chromosomes in the nucleus of a human fibroblast.

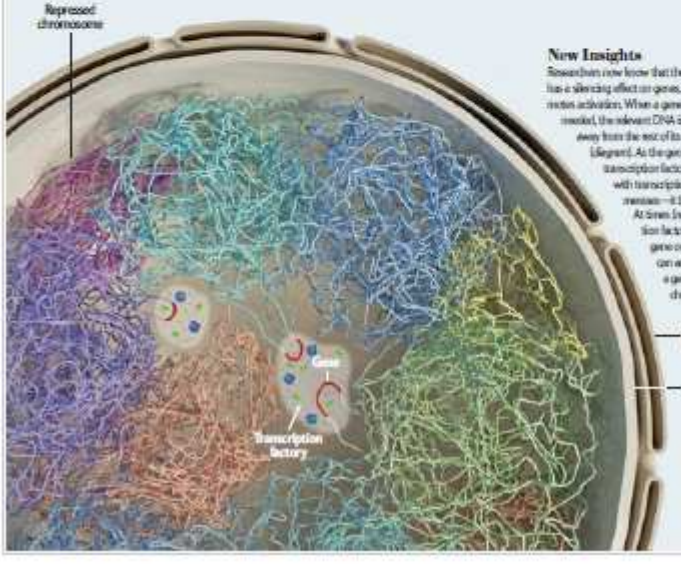
Fresh Clues to Gene Activation

For many years investigators have had a good understanding of the molecular machines that switch on genes (top image), the parts of chromosomes that encode the proteins and RNA molecules produced in cells. But now, thanks to new tools, they also have insight into a higher level of control: that exerted by the architecture of the nucleus (bottom).

Basics of Gene Activation

A gene gets switched on, or read out, after proteins called transcription factors collect on regulatory regions of the gene, enabling enzymes known as RNA polymerases to transcribe the gene's DNA code letters, or nucleotides, into mobile RNA copies. In the case of protein-coding genes, the RNA molecules, known as messenger RNAs, migrate to the cytoplasm, where structures called ribosomes translate them into the specified proteins.





New Insights

Researchers now know that the nuclear periphery has a silencing effect on genes, and the center promotes activation. When a gene that is quiet is needed, the relevant DNA is thought to loop away from the rest of its chromosome (diagram). As the gene finds itself in a transcription factory—a zone buzzing with transcription factors and polymerases—it becomes fully active. At times (not shown), transcription factors attached to a gene on one chromosome can actually help activate a gene on a nearby chromosome.

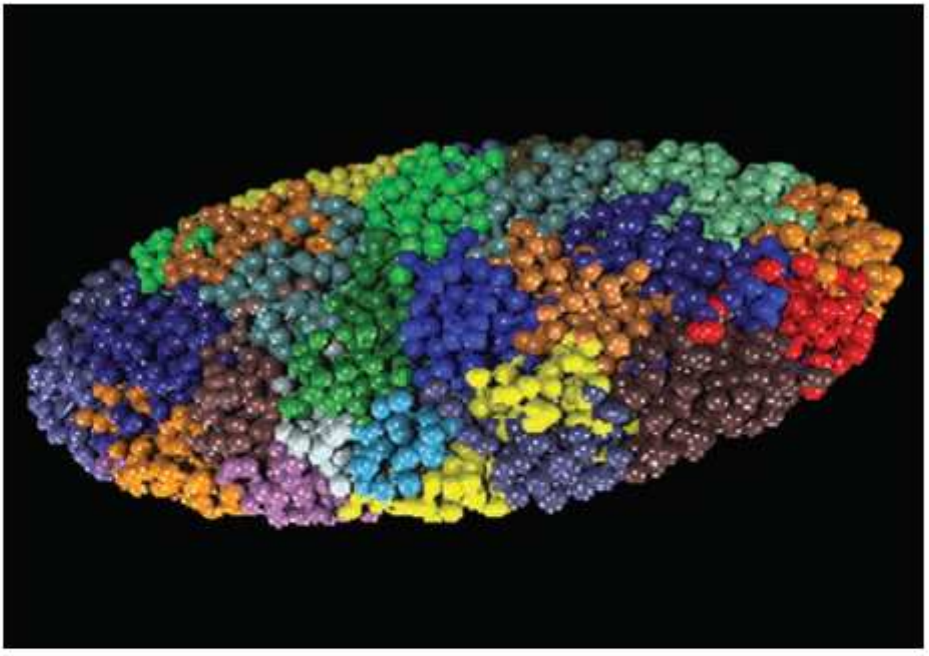
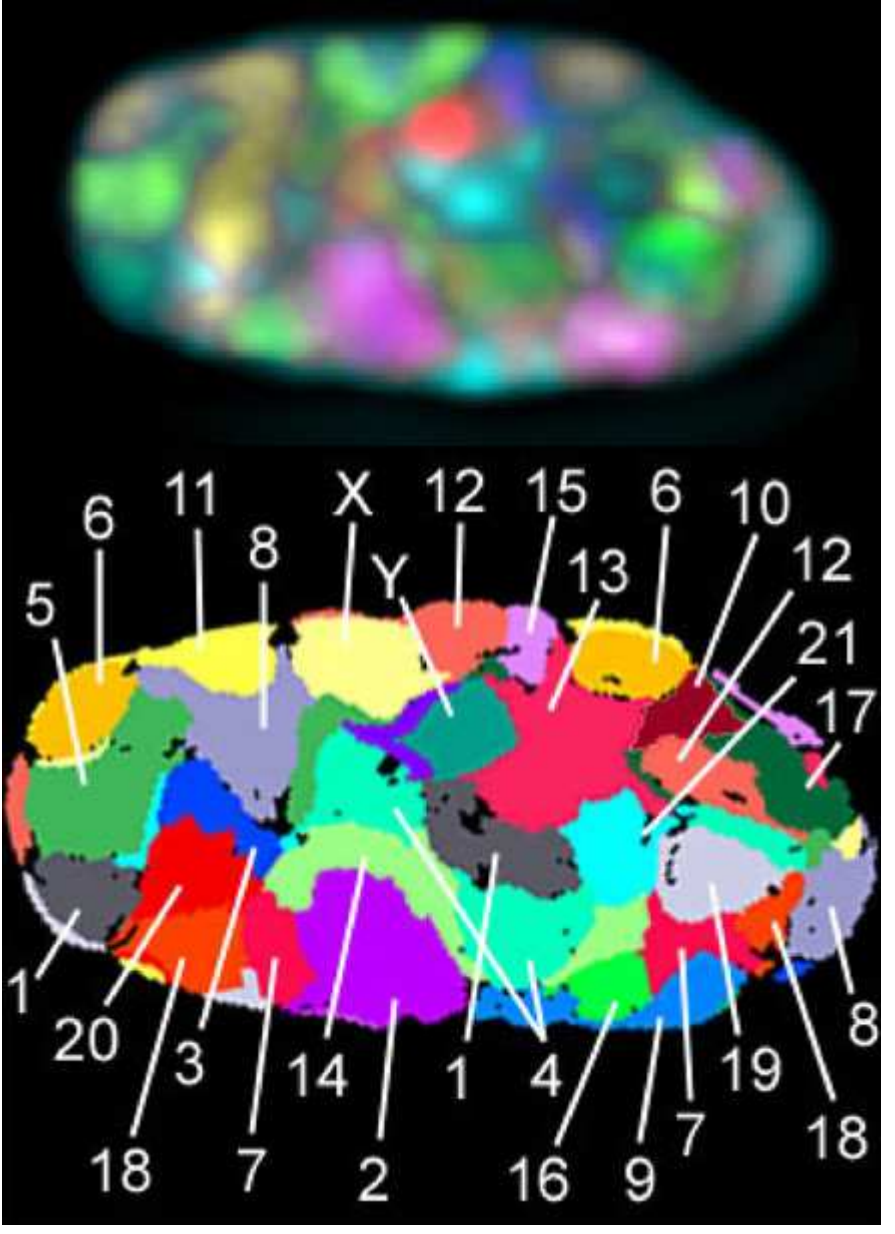


FIGURE 12.23 Three-dimensional map of all of the chromosomes present in a human fibroblast nucleus. This computer-generated image is based on fluorescence in situ hybridization analysis similar to that described for Figure 12.18*b* that allows each human chromosome to be distinguished from others and represented by an identifiable color. Each chromosome is found to occupy a distinct territory within the nucleus. (FROM ANDREAS BOLZER ET AL, PLOS BIOL. 3:E157, 2005, COURTESY OF THOMAS CREMER.)


11
X
12
15
6
10
6
8
Y
13
12
5
21
17
1
20
3
14
1
4
19
8
18
7
2
16
9
7
18

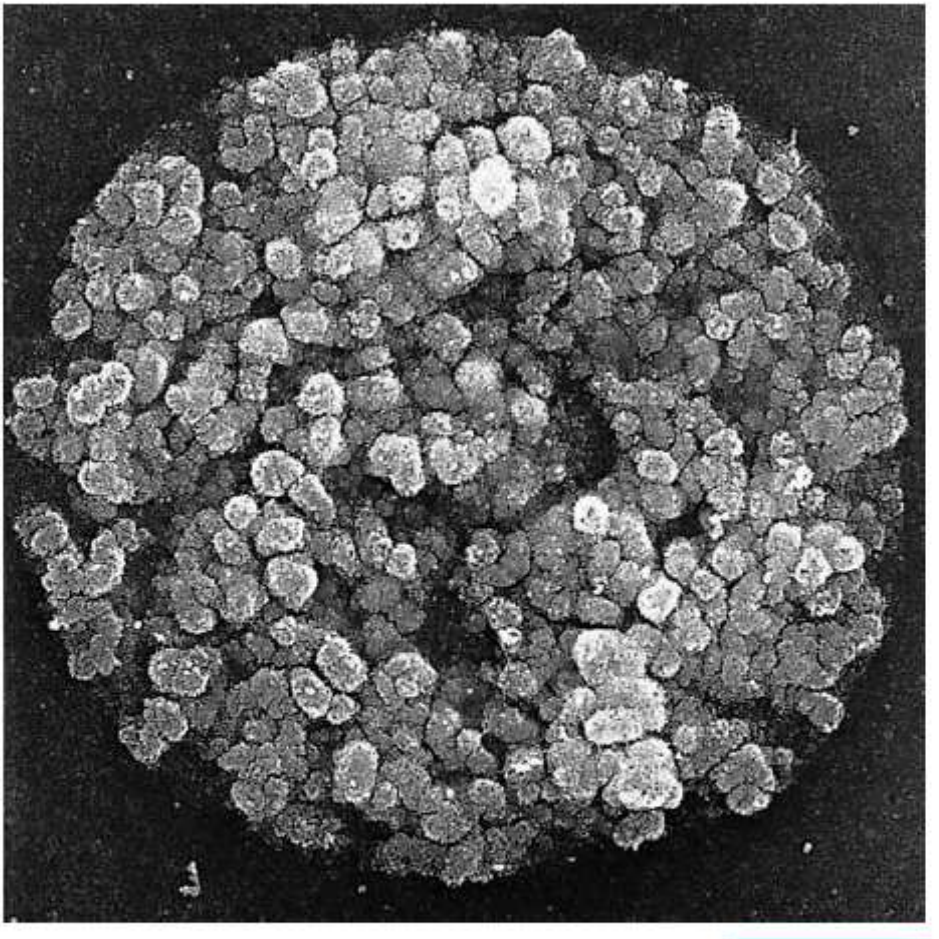
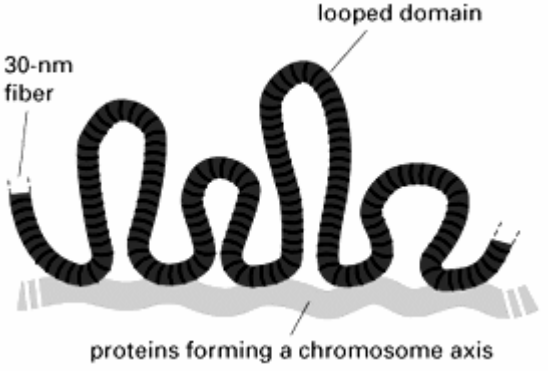
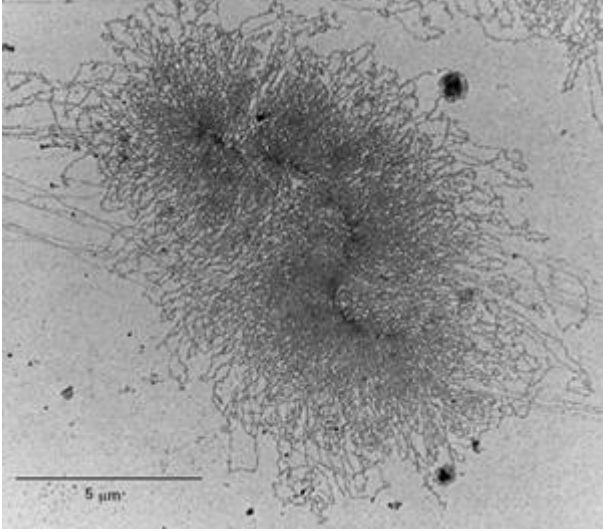
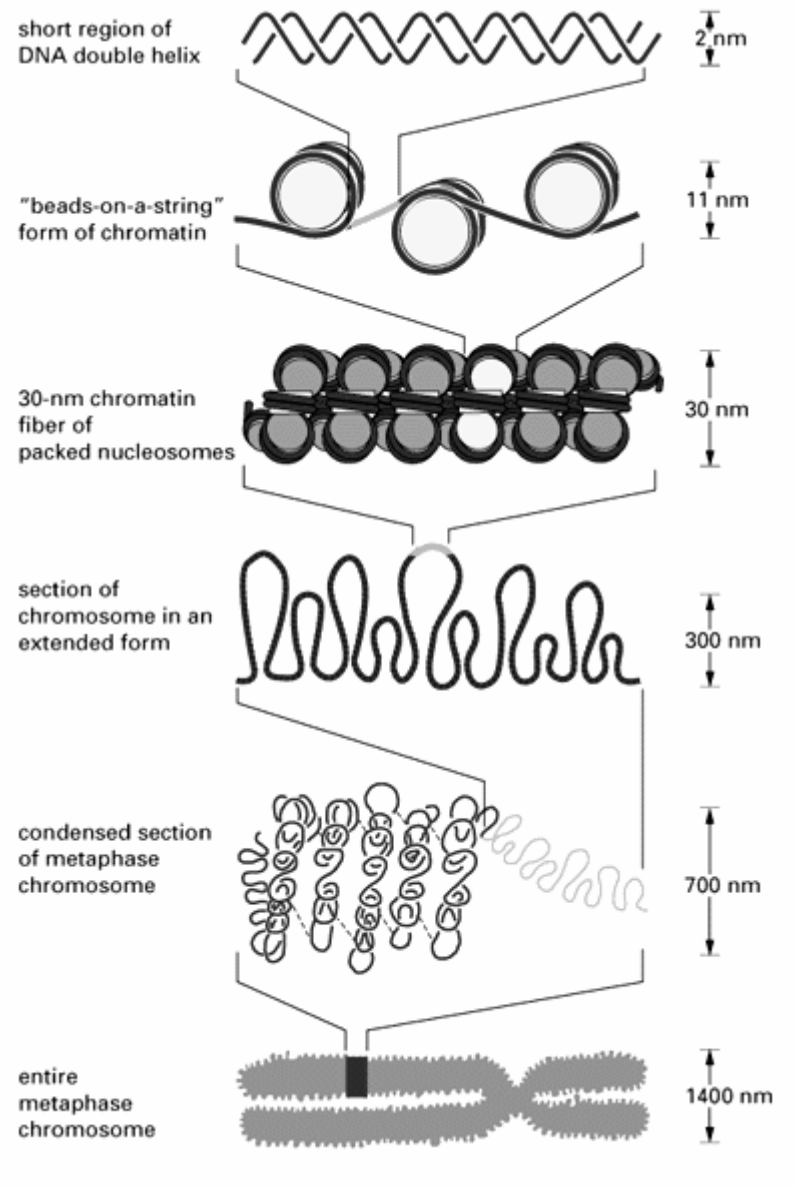
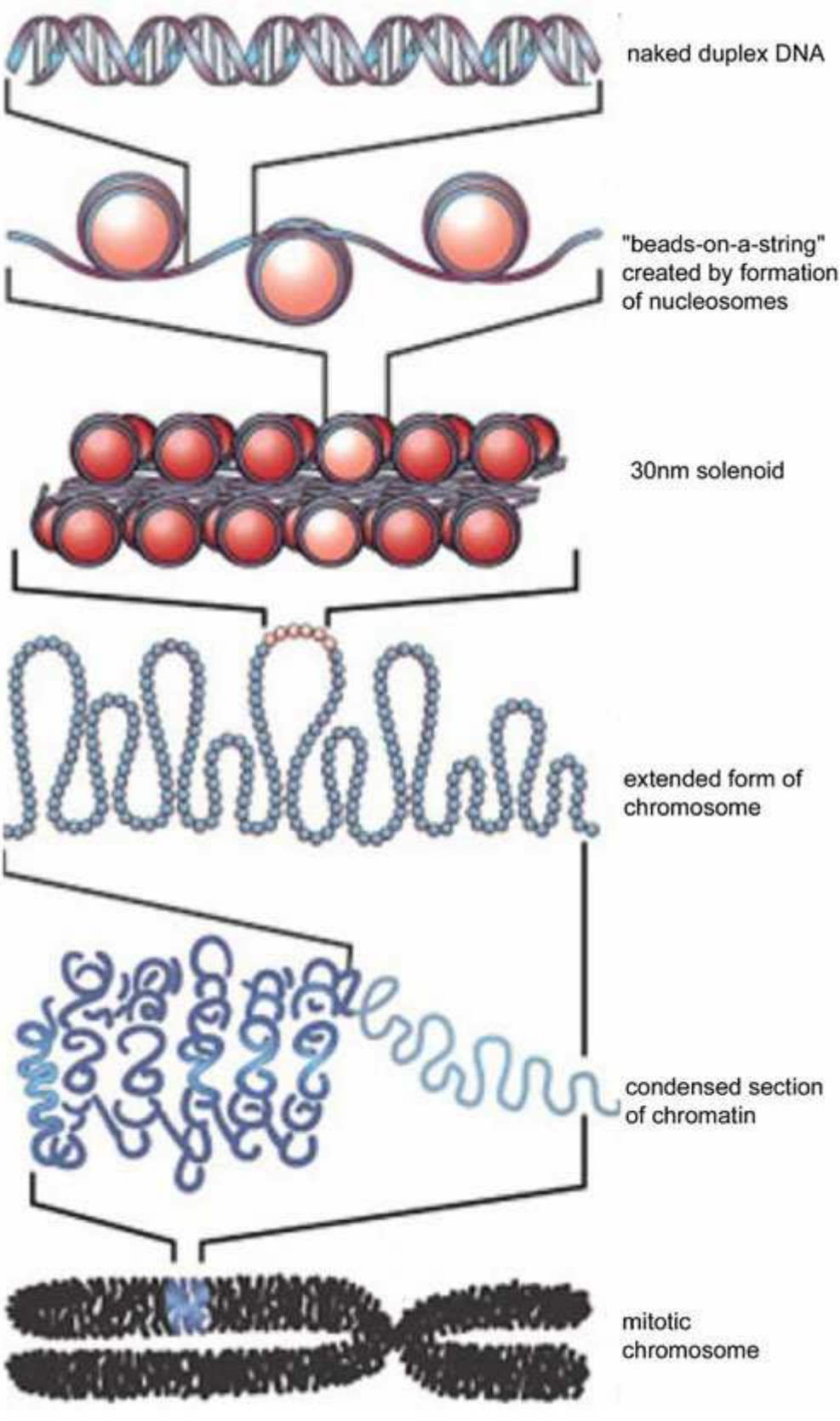


FIGURE 14.12 The chromosomes of this early prophase nucleus have begun the process of compaction that converts them into short, rod-like mitotic chromosomes that separate at a later stage in mitosis. (FROM A. T. SUMNER, CHROMOSOMA 100:411, 1991.)




5 µm


short region of
DNA double helix
2 nm
"beads-on-a-string"
form of chromatin
11 nm
30-nm chromatin
fiber of
packed nucleosomes
30 nm
section of
chromosome in an
extended form
300 nm
condensed section
of metaphase
chromosome
700 nm
entire
metaphase
chromosome
1400 nm


naked duplex DNA
"beads-on-a-string" created by formation of nucleosomes
30nm solenoid
extended form of chromosome
condensed section of chromatin
mitotic chromosome

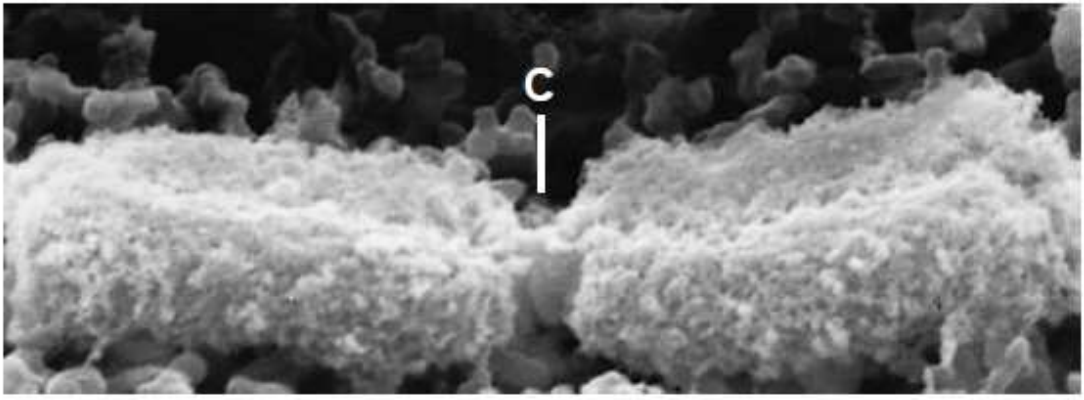
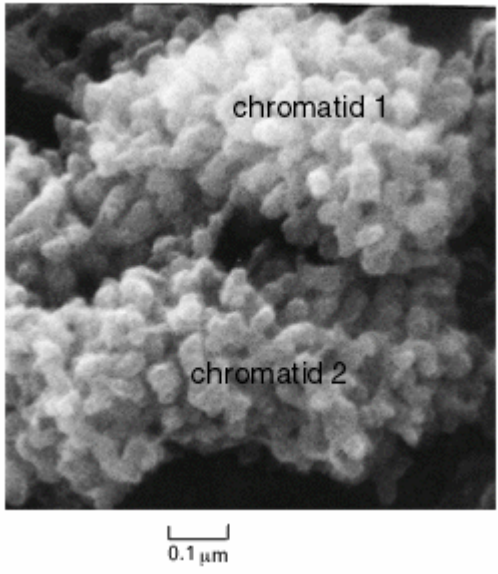
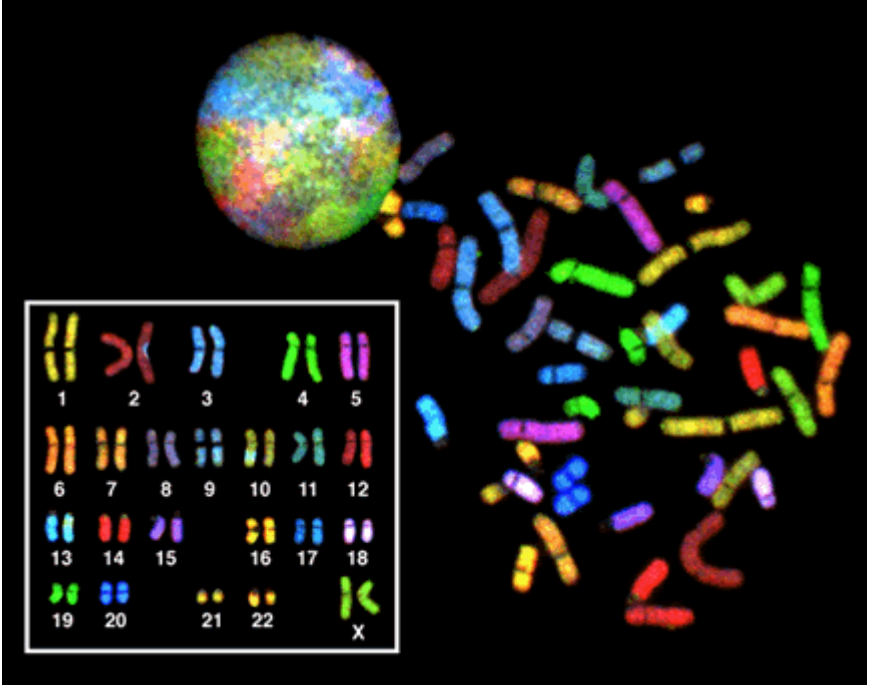
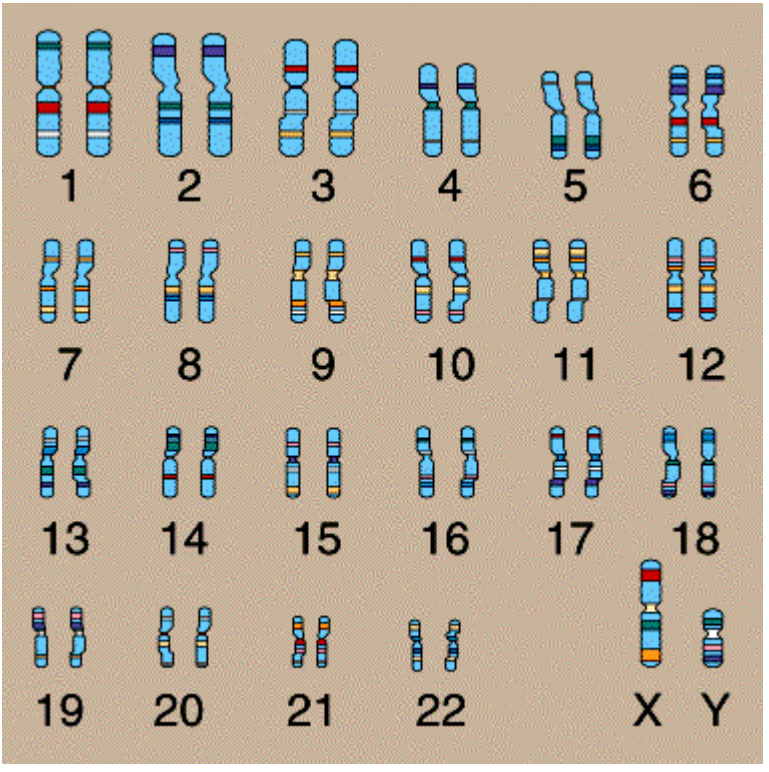


FIGURE 12.22 Each mitotic chromosome has a centromere whose site is marked by a distinct indentation. Scanning electron micrograph of a mitotic chromosome. The centromere (C) contains highly repeated DNA sequences (satellite DNA) and a protein-containing structure called the kinetochore that serves as a site for the attachment of spindle microtubules during mitosis and meiosis (discussed in Chapter 14). (FROM JEROME B. RATTNER, BIOESS. 13:51, 1991.)




1 2 3 4 5
6 7 8 9 10 11 12
13 14 15 16 17 18
19 20 21 22 X


1 2 3 4 5 6
7 8 9 10 11 12
13 14 15 16 17 18
19 20 21 22 X Y

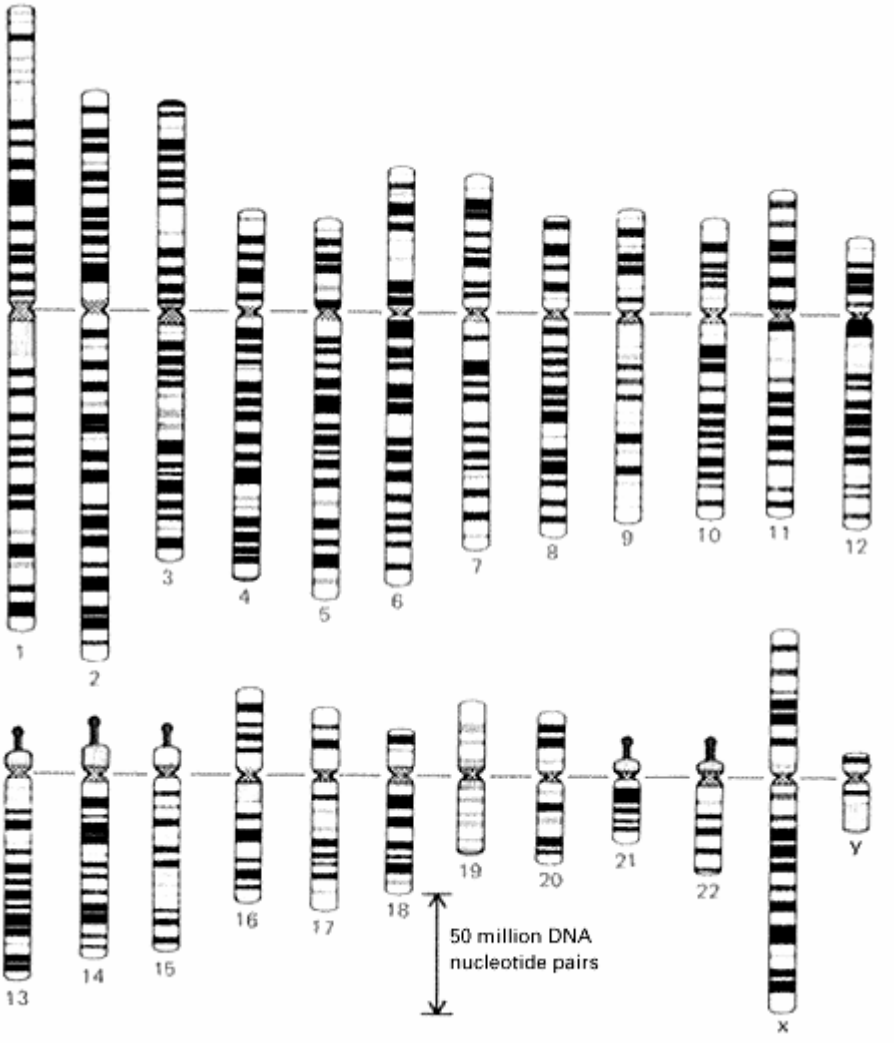
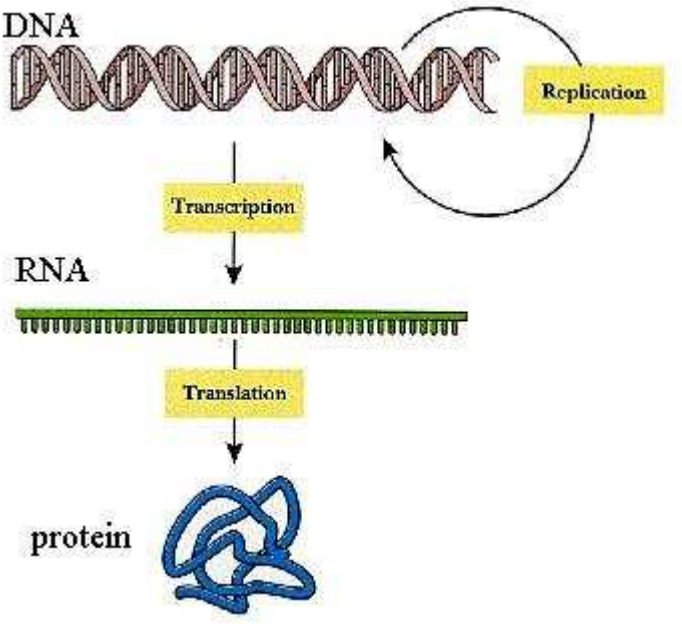
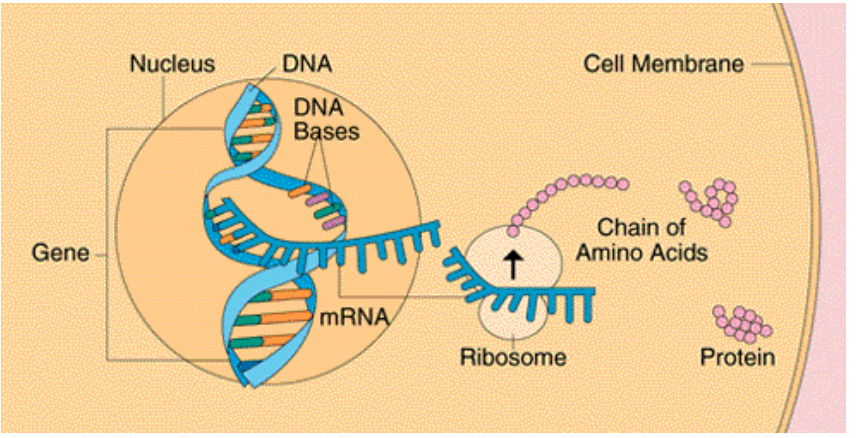

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
x
y
50 million DNA
nucleotide pairs

FIGURE 3.16 ***Banding Patterns of Human Chromosomes***

Representation of the banding patterns seen in metaphase chromosomes during meiosis. The bands are originally visualized by dyes. The relative distances between these bands are the same for an individual chromosome, so this is a useful way of identifying a particular chromosome. Courtesy of Dept. of Clinical Cytogenetics, Addenbrookes Hospital, Cambridge, UK, Science Photo Library.

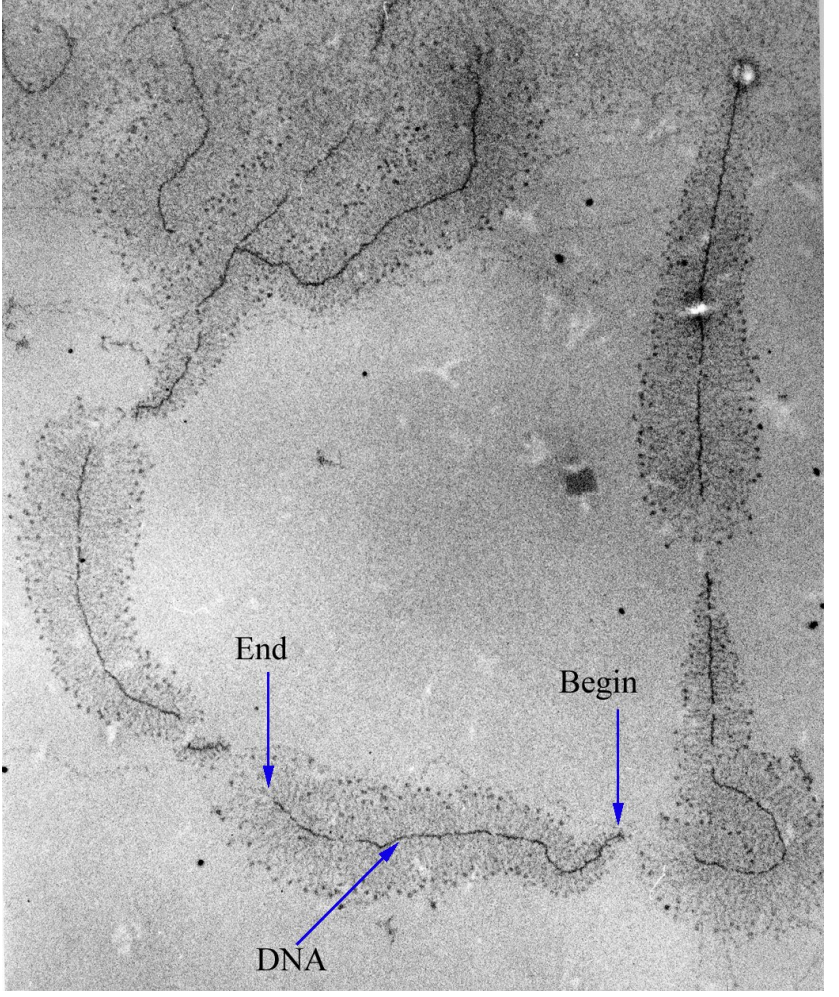
Centralna dogma molekulske biologije



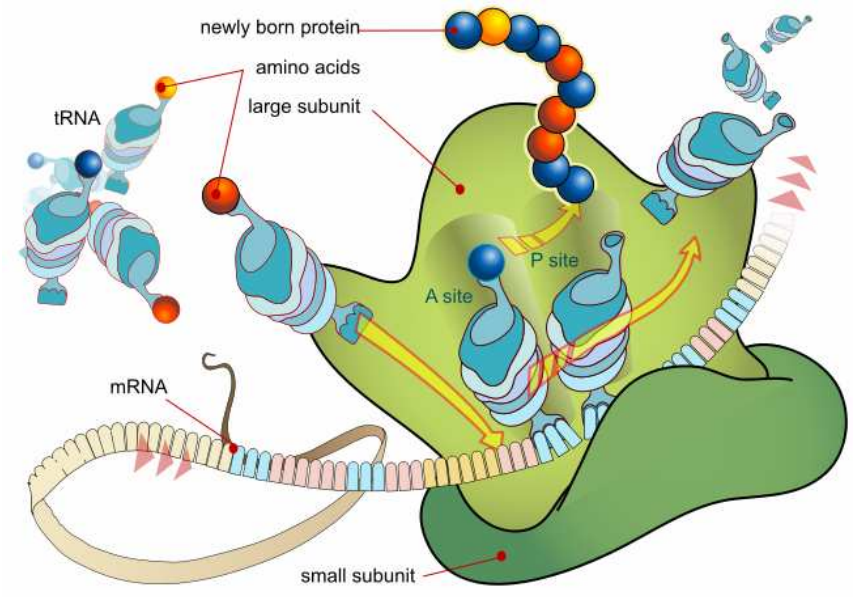


DNA replikacija

Transkripcija



Biosinteza proteina translacija



Geni

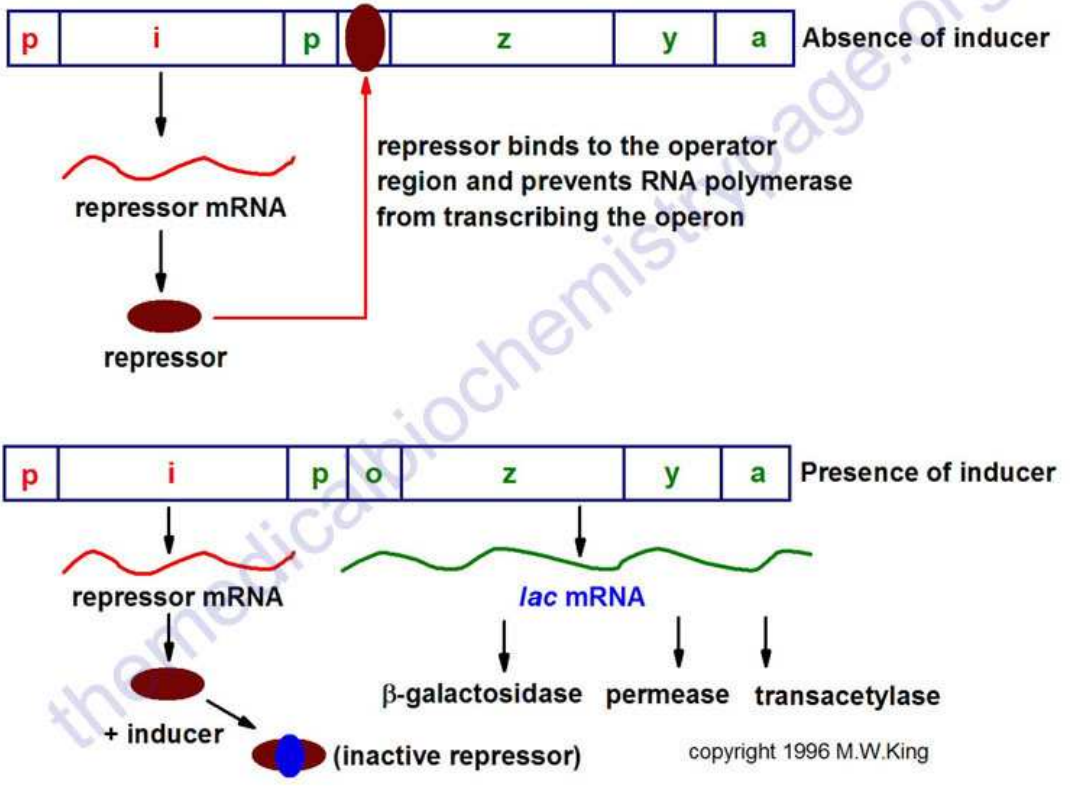
Protein nekodirajući

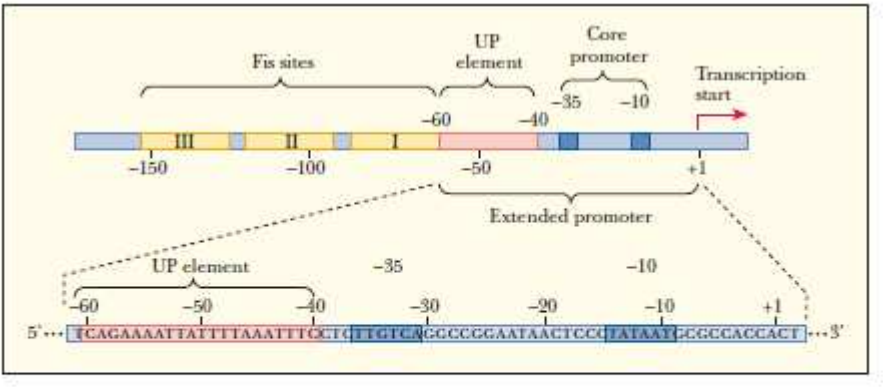
TABLE 3.02 Major Classes of Non-Translated RNA

Name	Function
Ribosomal RNA	comprises major portion of ribosome and is involved in synthesis of polypeptide chains
Transfer RNA	carries amino acids to ribosome and recognizes codons on mRNA
Small nuclear RNA	involved in the processing of messenger RNA molecules in the nucleus of eukaryotic cells (also called snRNA, or "snurps")
Guide RNA	involved in processing of RNA or DNA in some organisms
Regulatory RNA	functions in the regulation of gene expression by binding to proteins or DNA or to other RNA molecules
Antisense RNA	functions in regulating gene expression by base pairing to mRNA
Recognition RNA	part of a few enzymes (e.g., telomerase); enables them to recognize certain short DNA sequences
Ribozymes	enzymatically active RNA molecules

Protein kodirajući

The *lac* Operon





■ **FIGURE 11.9 Elements of a bacterial promoter.** The core promoter includes the −10 and −35 regions. The extended promoter includes the UP element. Upstream of the UP element, there may be enhancers, such as the Fis sites seen in the promoters for genes that code for ribosomal RNA in *E. coli*. The protein Fis is a transcription factor. *(Adapted by permission from* Molecular Biology, *by R. F. Weaver, McGraw-Hill, 1999.)*

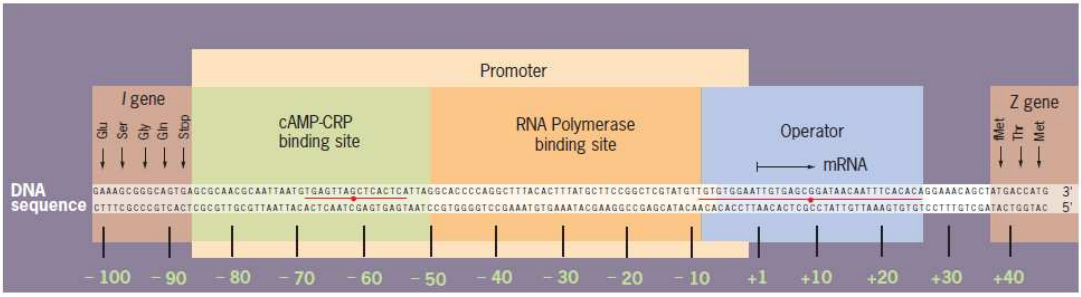
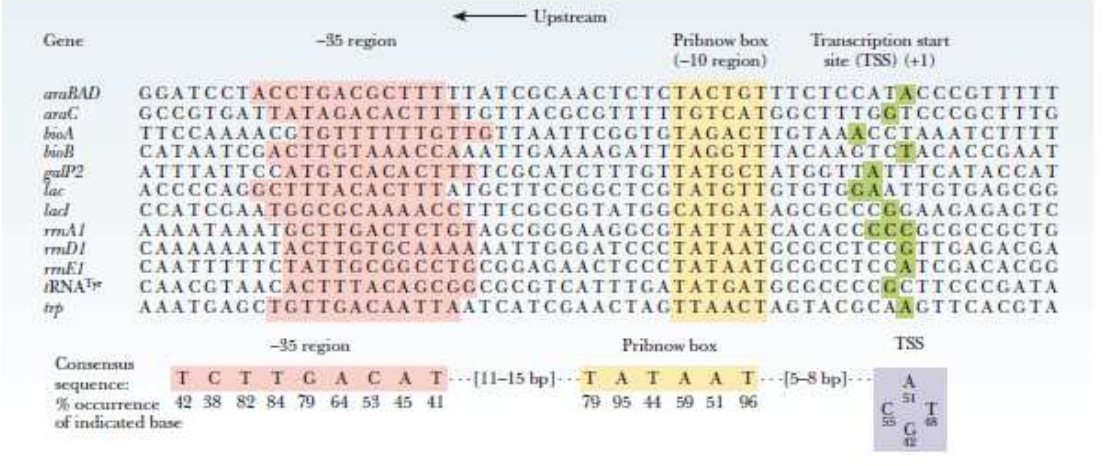
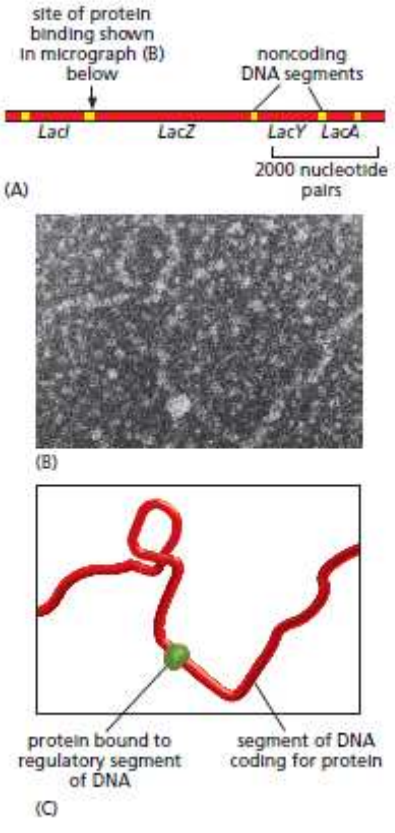


FIGURE 12.30 The nucleotide sequence of binding sites in the control region of the *lac* operon. The promoter region contains the binding site for the CRP protein as well as the RNA polymerase. The site of initiation of transcription is denoted as +1, which is approximately 40 nucleotides upstream from the site at which translation is initiated. Regions of sequence symmetry in the CRP site and operator are indicated by the red horizontal line. (REPRINTED WITH PERMISSION FROM R. C. DICKSON ET AL., SCIENCE 187:32, 1975; © COPYRIGHT 1975, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE.)



■ **FIGURE 11.2 Sequences of representative promoters from *E. coli*.** By convention, these are given as the sequence that would be found on the coding strand going from left to right as the 5' to 3' direction. The numbers below the consensus sequences indicate the percentage of the time that a certain position is occupied by the indicated nucleotide.


site of protein binding shown in micrograph (B) below
noncoding DNA segments
LacI
LacZ
LacY
LacA
2000 nucleotide pairs
(A)
(B)
protein bound to regulatory segment of DNA
segment of DNA coding for protein
(C)

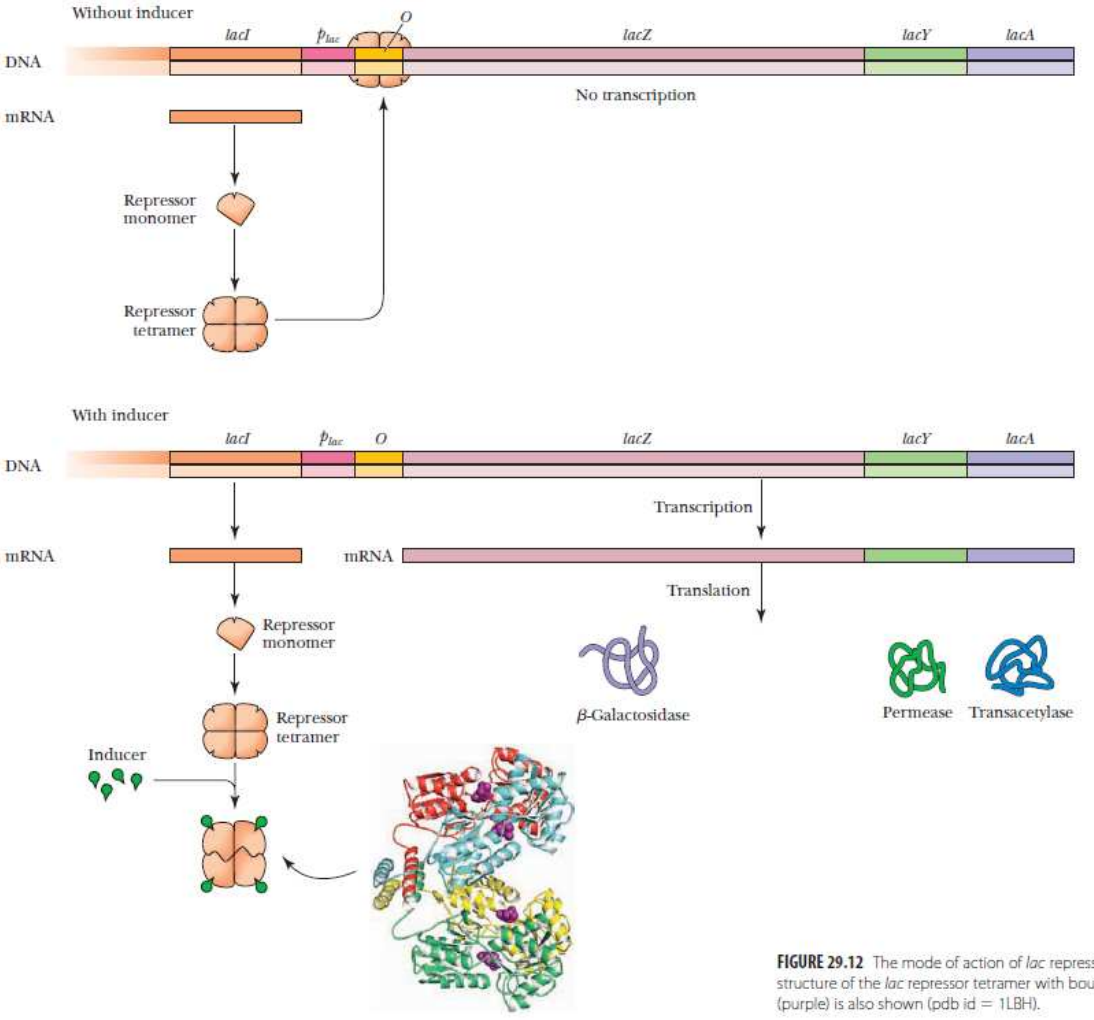
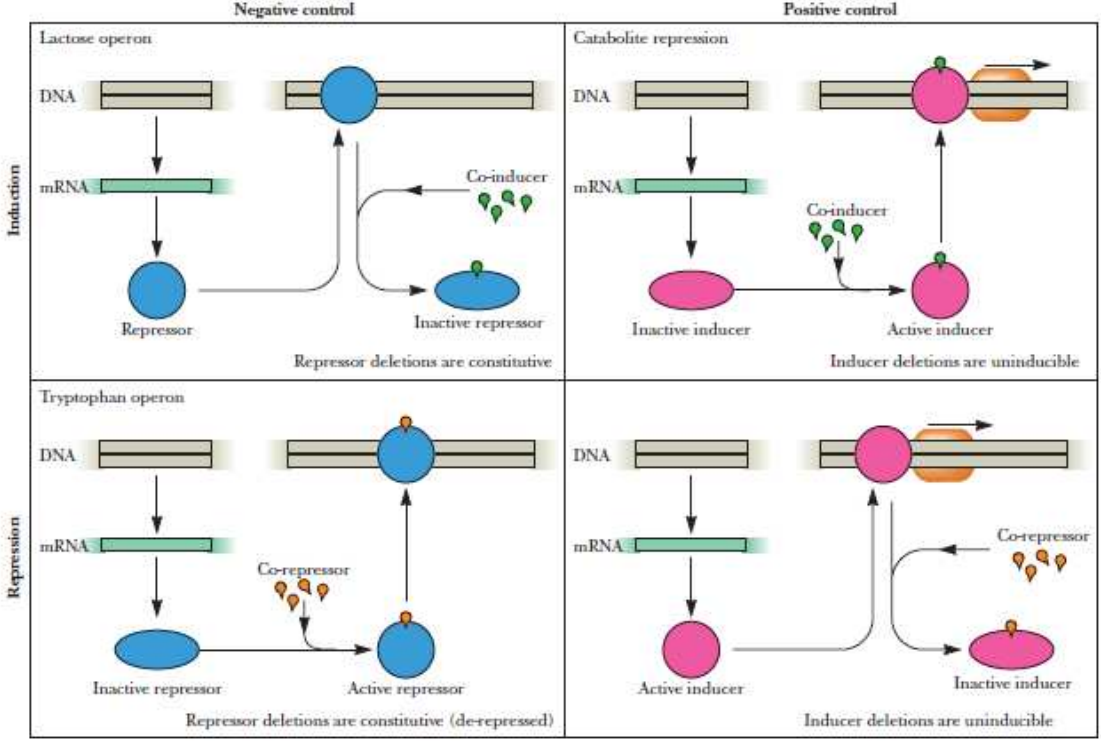


FIGURE 29.12 The mode of action of *lac* repressor. The structure of the *lac* repressor tetramer with bound IPTG (purple) is also shown (pdb id = 1LBH).



■ **FIGURE 11.13** **Basic control mechanisms seen in the control of genes.** They may be inducible or repressible, and they may be positively or negatively controlled.

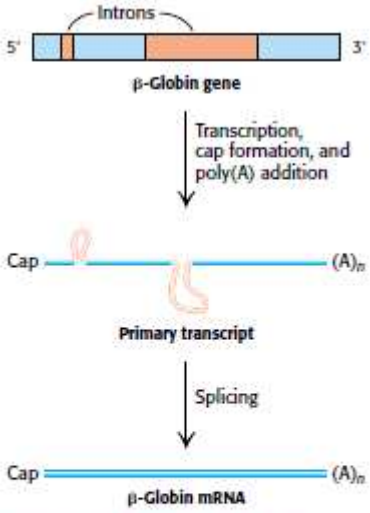
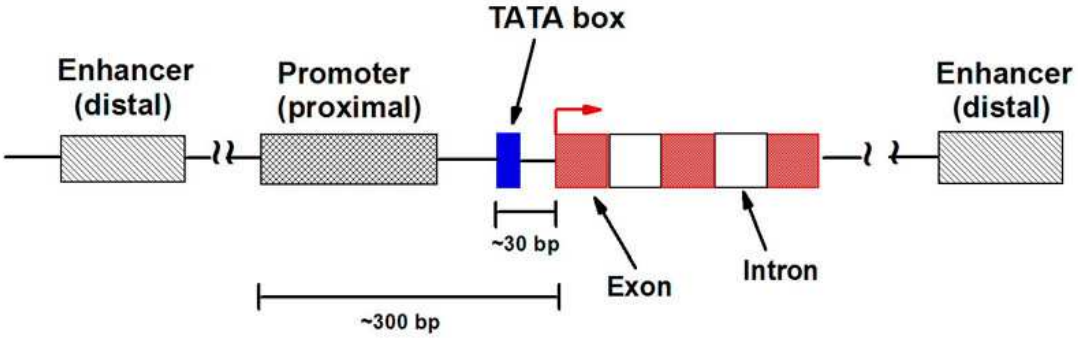
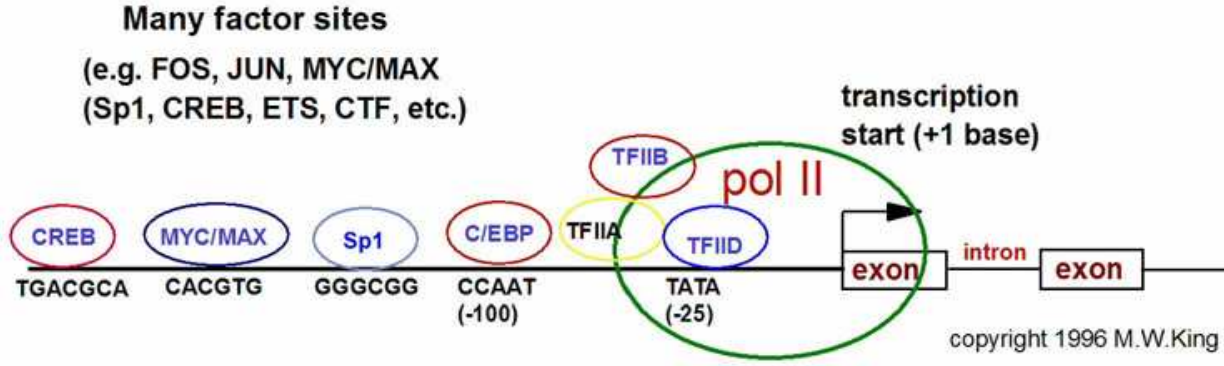


FIGURE 5.34 Transcription and processing of the β-globin gene. The gene is transcribed to yield the primary transcript, which is modified by cap and poly(A) addition. The intervening sequences in the primary RNA transcript are removed to form the mRNA.





[8]Although the term *exon* is commonly used to refer to the protein-coding regions of an interrupted or split gene, a more precise definition would specify exons as sequences that are represented in mature RNA molecules. This definition encompasses not only protein-coding genes but also the genes for various RNAs (such as tRNAs or rRNAs) from which intervening sequences must be excised in order to generate the mature gene product.

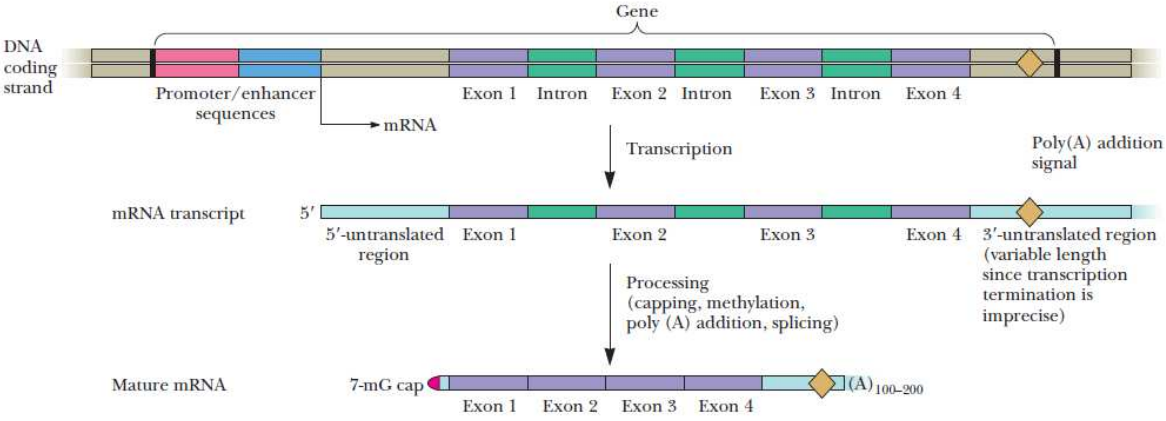


FIGURE 29.36 The organization of split eukaryotic genes.

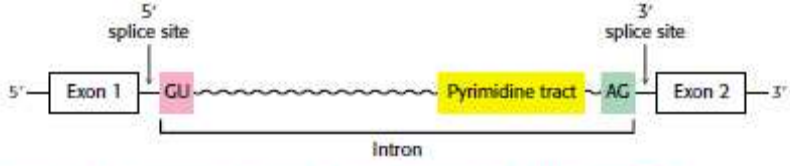
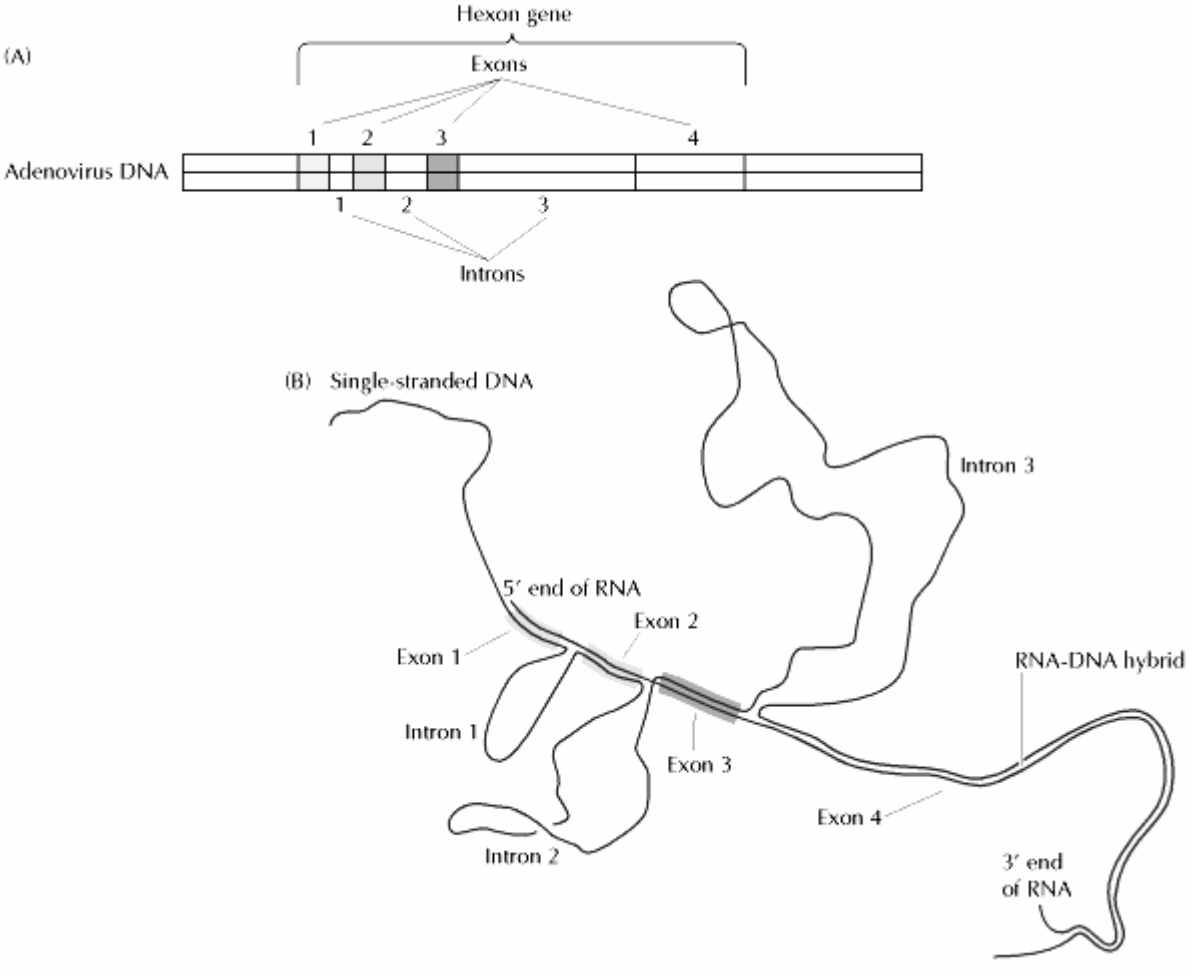


FIGURE 5.35 Consensus sequence for the splicing of mRNA precursors.


(A)
Hexon gene
Exons
1
2
3
4
Adenovirus DNA
1
2
3
Introns
(B) Single-stranded DNA
Intron 3
5′ end of RNA
Exon 2
Exon 1
RNA-DNA hybrid
Intron 1
Exon 3
Exon 4
Intron 2
3′ end
of RNA

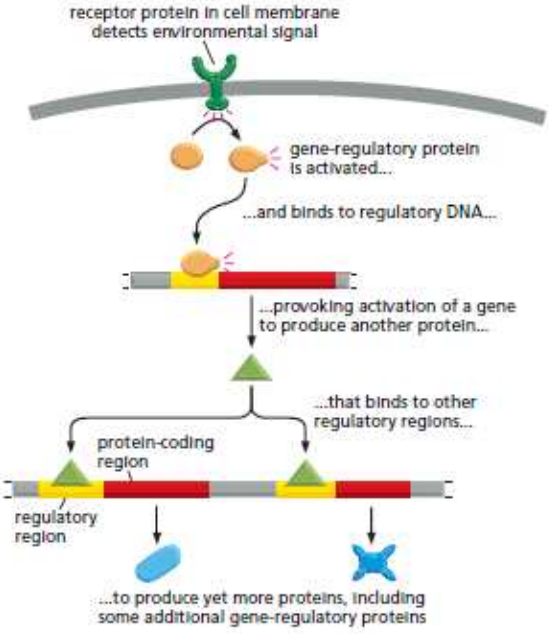


Figure 1–39 Controlling gene readout by environmental signals. Regulatory DNA allows gene expression to be controlled by regulatory proteins, which are in turn the products of other genes. This diagram shows how a cell's gene expression is adjusted according to a signal from the cell's environment. The initial effect of the signal is to activate a regulatory protein already present in the cell; the signal may, for example, trigger the attachment of a phosphate group to the regulatory protein, altering its chemical properties.

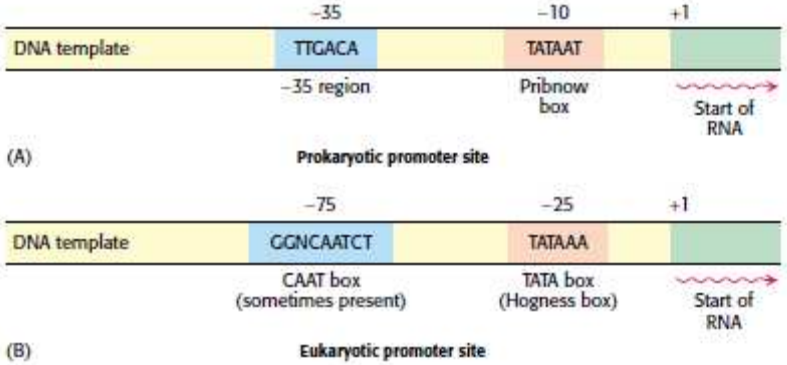


FIGURE 5.27 Promoter sites for transcription. Promoter sites are required for the initiation of transcription in both (A) prokaryotes and (B) eukaryotes. Consensus sequences are shown. The first nucleotide to be transcribed is numbered +1. The adjacent nucleotide on the 5′ side is numbered −1. The sequences shown are those of the coding strand of DNA.

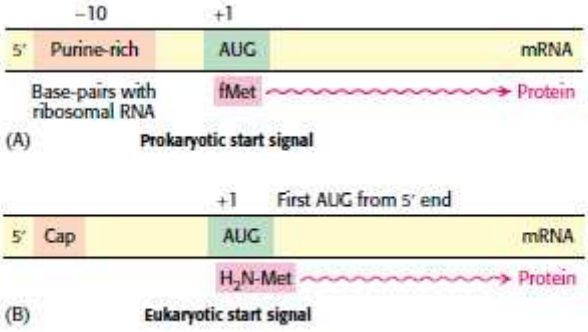
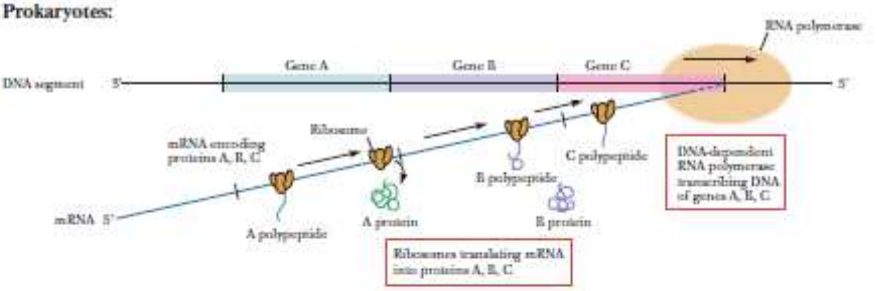
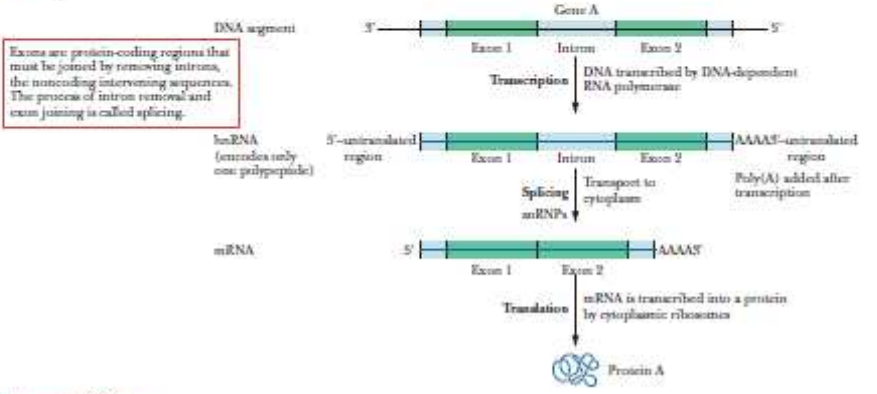


FIGURE 5.32 Initiation of protein synthesis. Start signals are required for the initiation of protein synthesis in (A) prokaryotes and (B) eukaryotes.

Prokaryotes:

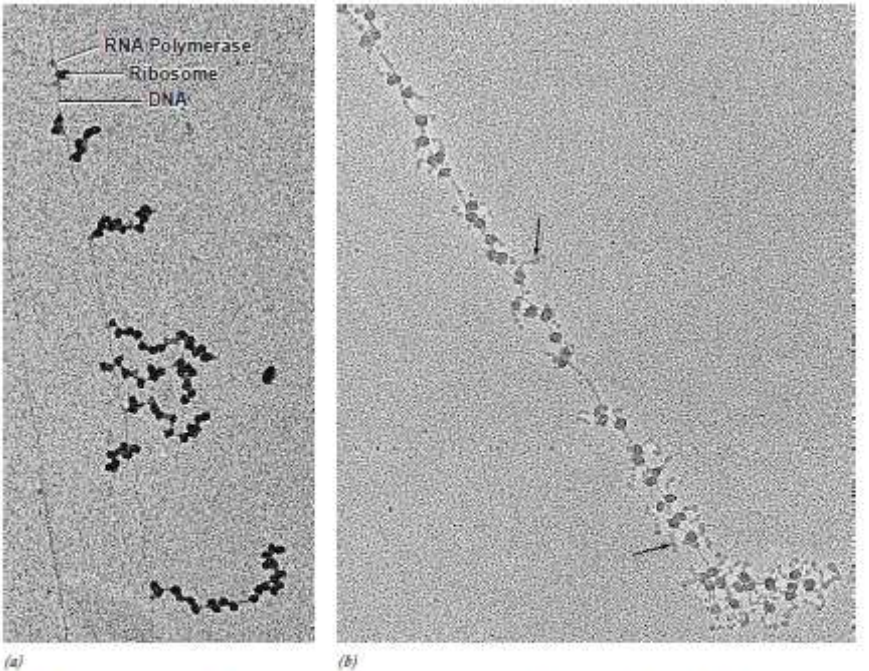


Eukaryotes:



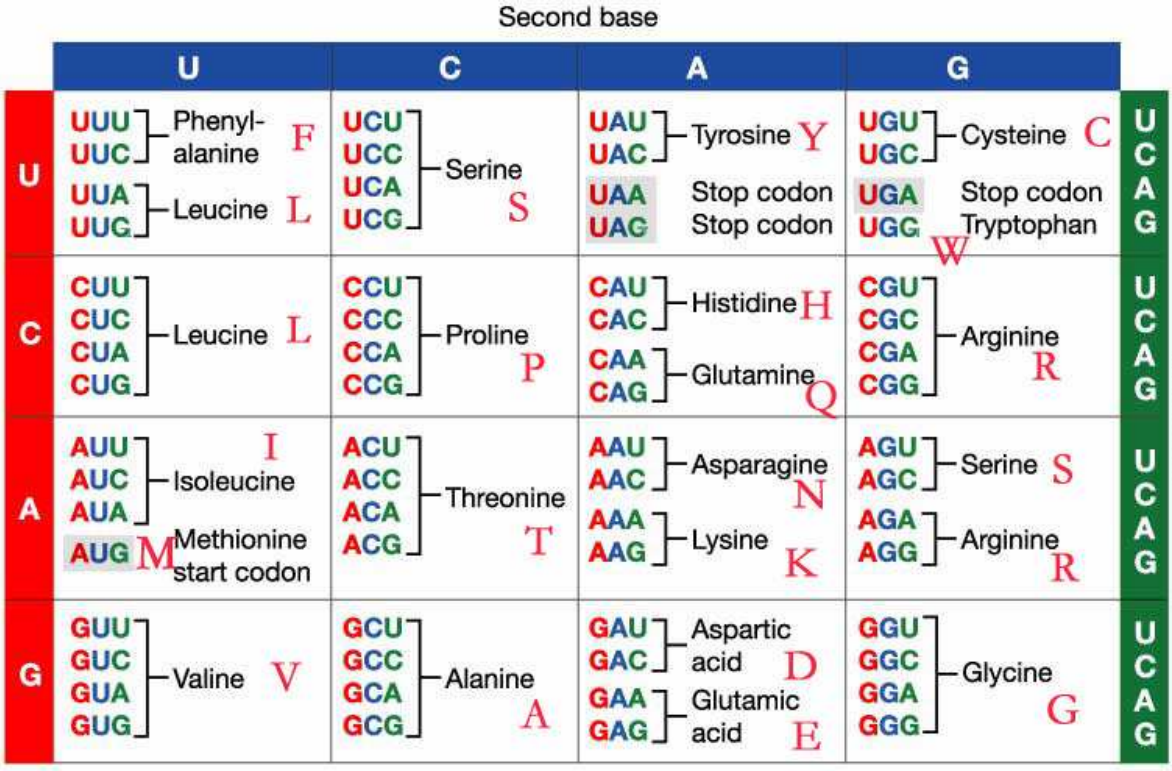
BiochemistryNow™

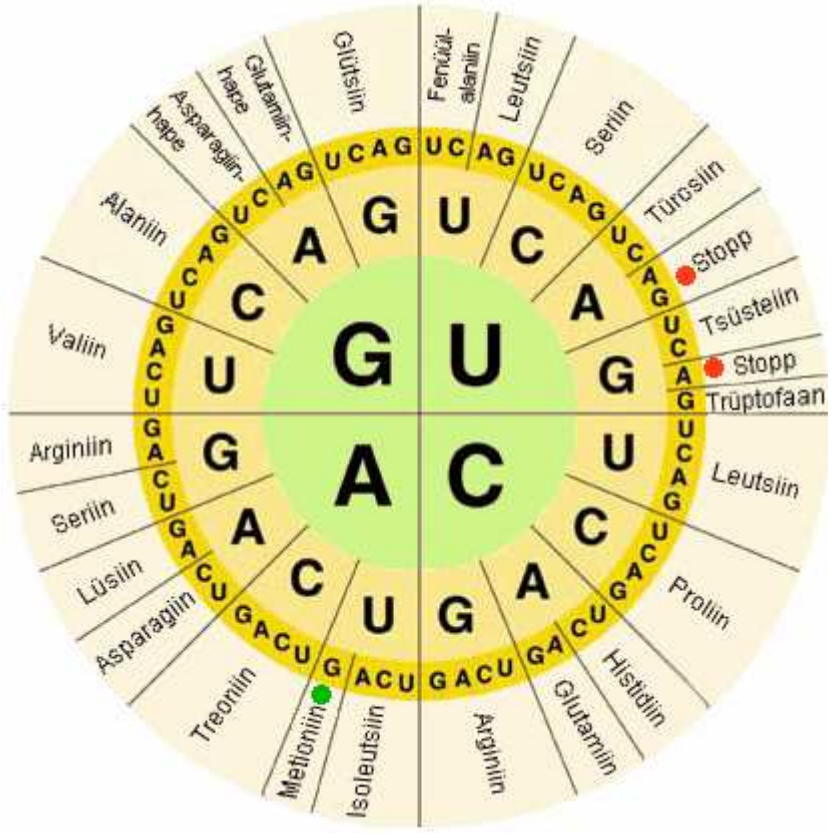
■ **ACTIVE FIGURE 9.21** **The role of mRNA in transcription.** The properties of mRNA molecules in prokaryotic versus eukaryotic cells during transcription and translation. **Sign in at www.thomsonedu.com/login to explore an interactive version of this figure.**

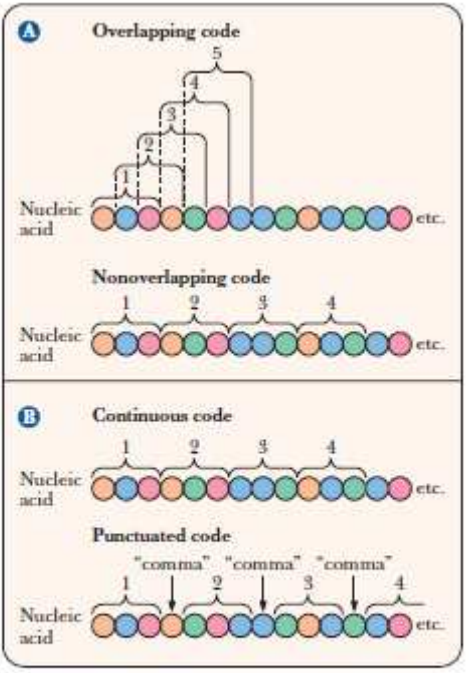


(a) (b)

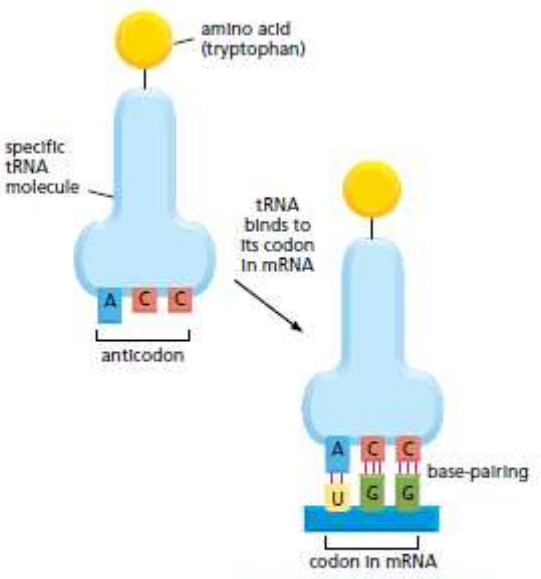
FIGURE 11.51 Visualizing transcription and translation. (*a*) Electron micrograph of portions of an *E. coli* chromosome engaged in transcription. The DNA is seen as faint lines running the length of the photo, whereas the nascent mRNA chains are seen to be attached at one of their ends, presumably by an RNA polymerase molecule. The particles associated with the nascent RNAs are ribosomes in the act of translation; in bacteria, transcription and translation occur simultaneously. The RNA molecules increase in length as the distance from the initiation site increases. (*b*) Electron micrograph of a polyribosome isolated from cells of the silk gland of the silkworm, which produces large quantities of the silk protein fibroin. This protein is large enough to be visible in the micrograph (arrows point to nascent polypeptide chains). (A: REPRINTED WITH PERMISSION FROM OSCAR L. MILLER, JR., BARBARA A. HAMKALO, AND C. A. THOMAS, SCIENCE 169:392, 1970; © COPYRIGHT 1970, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE; B: COURTESY OF STEVEN L. MCKNIGHT AND OSCAR L. MILLER, JR.)


Second base
U
C
A
G
First base
U
C
A
G
UUU UUC Phenyl-alanine F
UUA UUG Leucine L
UCU UCC UCA UCG Serine S
UAU UAC Tyrosine Y
UAA Stop codon
UAG Stop codon
UGU UGC Cysteine C
UGA Stop codon
UGG Tryptophan W
CUU CUC CUA CUG Leucine L
CCU CCC CCA CCG Proline P
CAU CAC Histidine H
CAA CAG Glutamine Q
CGU CGC CGA CGG Arginine R
AUU AUC AUA Isoleucine I
AUG Methionine start codon M
ACU ACC ACA ACG Threonine T
AAU AAC Asparagine N
AAA AAG Lysine K
AGU AGC Serine S
AGA AGG Arginine R
GUU GUC GUA GUG Valine V
GCU GCC GCA GCG Alanine A
GAU GAC Aspartic acid D
GAA GAG Glutamic acid E
GGU GGC GGA GGG Glycine G
Third base
U C A G

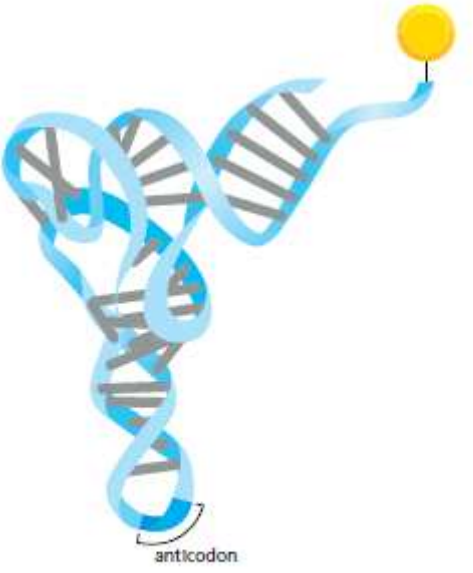

G
U
A
C
Glütsiin
Fenüülalaniin
Leutsiin
Seriin
Türosiin
Stopp
Tsüsteiin
Stopp
Trüptofaan
Leutsiin
Prolüin
Histidiin
Glutamiin
Arginiin
Isoleutsiin
Metioniin
Treoniin
Asparagiin
Lüsiin
Seriin
Arginiin
Valiin
Alaniin
Asparagiinhape
Glutamiinhape



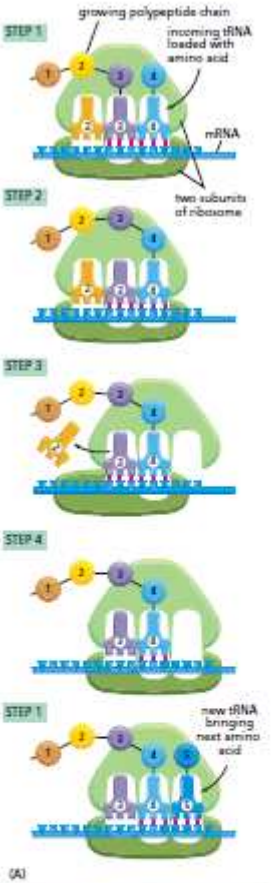
■ **FIGURE 12.2 Theoretically possible genetic codes.** (a) An overlapping versus a nonoverlapping code. (b) A continuous versus a punctuated code.

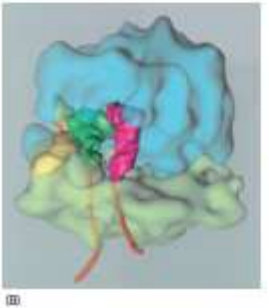

amino acid
(tryptophan)
specific
tRNA
molecule
A
C
C
anticodon
tRNA
binds to
its codon
in mRNA
A
C
C
U
G
G
base-pairing
codon in mRNA
NET RESULT: AMINO ACID IS
SELECTED BY ITS CODON

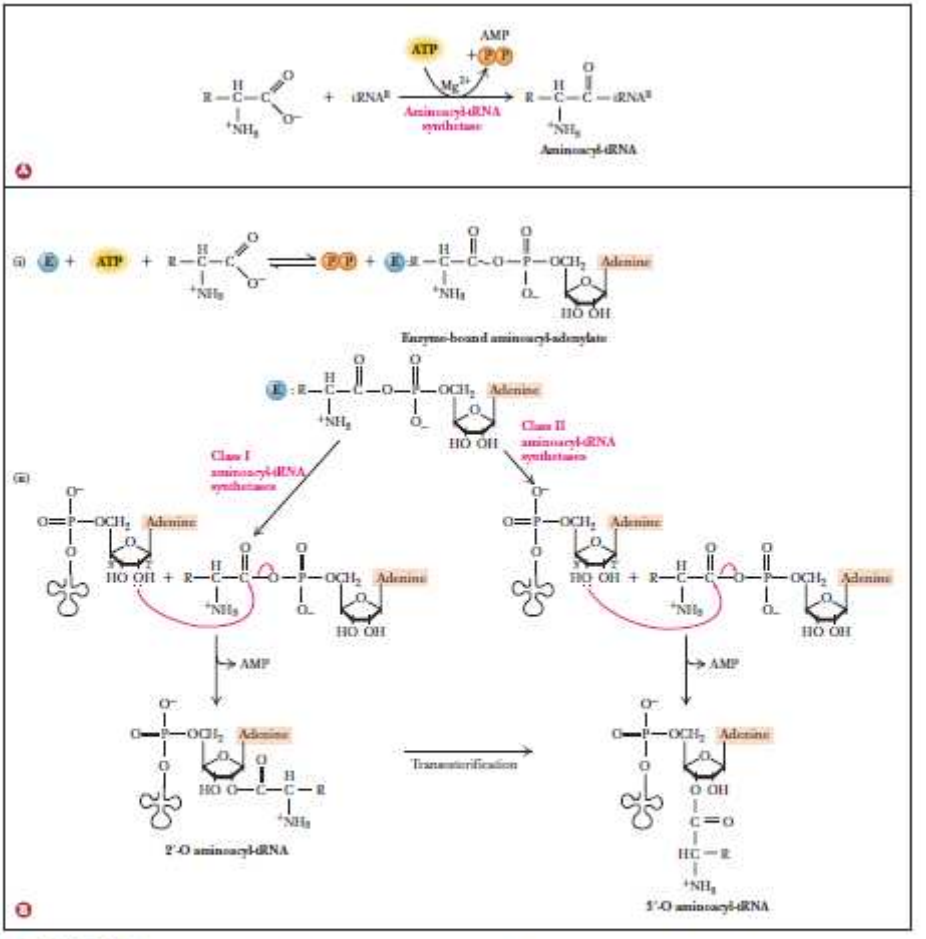
(A)


anticodon

(B)


growing polypeptide chain
STEP 1
incoming tRNA loaded with amino acid
mRNA
two subunits of ribosome
STEP 2
STEP 3
STEP 4
STEP 1
new tRNA bringing next amino acid
(A)


(B)



■ **FIGURE 12.6 The aminoacyl-tRNA synthetase reaction.** (a) The overall reaction. Ever-present pyrophosphatases in cells quickly hydrolyze the PP_i produced in the aminoacyl-tRNA synthetase reaction, rendering aminoacyl-tRNA synthesis thermodynamically favorable and essentially irreversible. (b) The overall reaction commonly proceeds in two steps: (i) formation of an aminoacyl-adenylate and (ii) transfer of the activated amino acid moiety of the mixed anhydride to either the 2′-OH (class I aminoacyl-tRNA synthetases) or 3′-OH (class II aminoacyl-tRNA synthetases) of the ribose on the terminal adenylic acid at the 3′-OH terminus common to all tRNAs. Those aminoacyl-tRNAs formed as 2′-OH esters undergo a transesterification that moves the aminoacyl group to the 3′-OH of tRNA. Only the 3′-esters are substrates for protein synthesis.

Comparison of different genome sizes

Organism type	Organism	Genome size (base pairs)	Genome size (in human-readable format)	mass - in pg	Note
Virus	Bacteriophage MS2	3,569	3.5kb	0.000002	First sequenced RNA-genome[8]
Virus	SV40	5,224	5.2kb		[9]
Virus	Phage Φ-X174	5,386	5.4kb		First sequenced DNA-genome[10]
Virus	HIV	9,749	9.7kb		[11]
Virus	Phage λ	48,502	48kb		
Virus	Mimivirus	1,181,404	1.2Mb		Largest known viral genome
Bacterium	*Haemophilus influenzae*	1,830,000	1.8Mb		First genome of a living organism sequenced, July 1995[12]
Bacterium	*Carsonella ruddii*	159,662	160kb		Smallest non-viral genome.[13]
Bacterium	*Buchnera aphidicola*	600,000	600kb		
Bacterium	*Wigglesworthia glossinidia*	700,000	700Kb		
Bacterium	*Escherichia coli*	4,600,000	4.6Mb		[14]
Bacterium	*Solibacter usitatus* (strain Ellin 6076)	9,970,000	10Mb		Largest known Bacterial genome
Amoeboid	*Polychaos dubium* (*"Amoeba" dubia*)	670,000,000,000	670Gb	737	Largest known genome.[15] (Disputed [16])
Plant	*Arabidopsis thaliana*	157,000,000	157Mb		First plant genome sequenced, December 2000.[17]
Plant	*Genlisea margaretae*	63,400,000	63Mb		Smallest recorded flowering plant genome, 2006.[17]
Plant	*Fritillaria assyrica*	130,000,000,000	130Gb		
Plant	*Populus trichocarpa*	480,000,000	480Mb		First tree genome sequenced, September 2006
Plant	*Paris japonica* (Japanese-native, pale-petal)	150,000,000,000	150Gb	152.23	Largest plant genome known
Moss	*Physcomitrella patens*	480,000,000	480Mb		First genome of a bryophyte sequenced, January 2008.[18]
Yeast	*Saccharomyces cerevisiae*	12,100,000	12.1Mb		First eukaryotic genome sequenced, 1996[19]
Fungus	*Aspergillus nidulans*	30,000,000	30Mb		
Nematode	*Caenorhabditis elegans*	100,300,000	100Mb		First multicellular animal genome sequenced, December 1998[20]
Nematode	*Pratylenchus coffeae*	20,000,000	20Mb		Smallest animal genome known[21]

Insect	*Drosophila melanogaster* (fruit fly)	130,000,000	130Mb		[22]
Insect	*Bombyx mori* (silk moth)	530,000,000	530Mb		
Insect	*Apis mellifera* (honey bee)	236,000,000	236Mb		
Insect	*Solenopsis invicta* (fire ant)	480,000,000	480Mb		[23]
Fish	*Tetraodon nigroviridis* (type of puffer fish)	385,000,000	390Mb		Smallest vertebrate genome known
Mammal	*Homo sapiens*	3,200,000,000	3.2Gb	3	
Fish	*Protopterus aethiopicus* (marbled lungfish)	130,000,000,000	130Gb	143	Largest vertebrate genome known

TABLE 1.6 How Many Genes Does It Take To Make An Organism?

Organism	Number of Cells in Adult*	Number of Genes
Mycobacterium genitalium Pathogenic bacterium	1	523
Methanococcus jannaschii Archaeal methanogen	1	1,800
Escherichia coli K12 Intestinal bacterium	1	4,400
Saccharomyces cereviseae Baker's yeast (eukaryote)	1	6,000
Caenorhabditis elegans Nematode worm	959	19,000
Drosophila melanogaster Fruit fly	10^4	13,500
Arabidopsis thaliana Flowering plant	10^7	27,000
Fugu rubripes Pufferfish	10^{12}	38,000 (est.)
Homo sapiens Human	10^{14}	20,500 (est.)

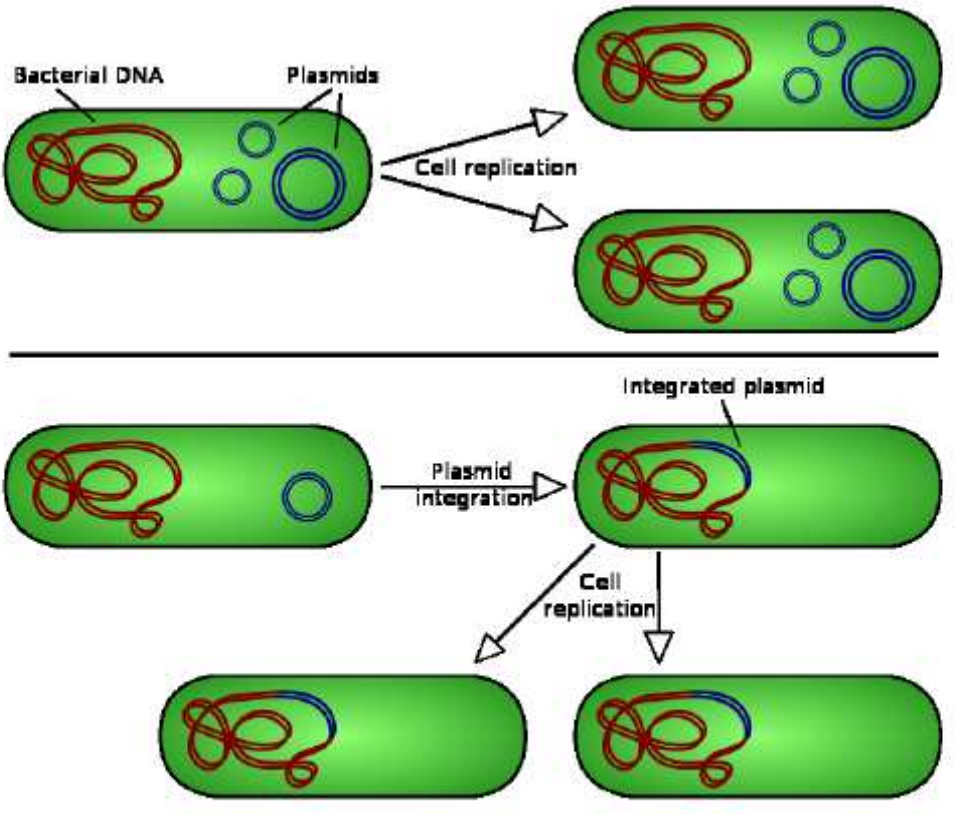
The first four of the nine organisms in the table are single-celled microbes; the last six are eukaryotes; the last five are multicellular, four of which are animals; the final two are vertebrates. Although pufferfish and humans have roughly the same number of genes, the pufferfish genome, at 0.365 billion nucleotide pairs, is only one-eighth the size of the human genome.

*Numbers for *Arabidopsis thaliana*, the pufferfish, and human are "order-of-magnitude" rough estimates.

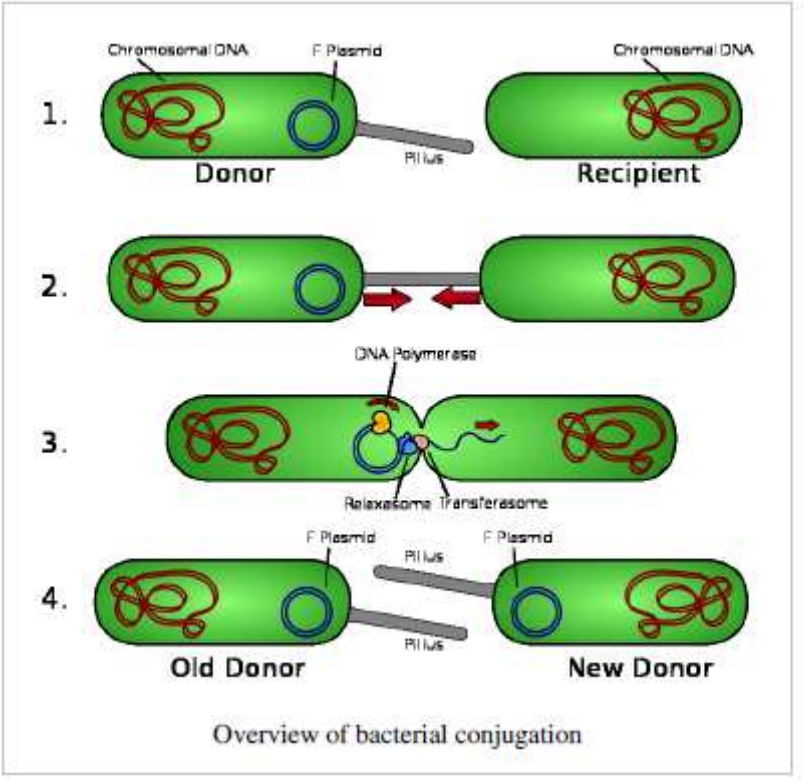
TABLE 4.03 Components of the Eukaryotic Genome

(Numbers of copies given is for the human genome.)

Unique sequences	
Protein encoding genes—comprising upstream regulatory region, exons and introns	
Genes encoding non-translated RNA (snRNA, snoRNA, 7SL RNA, telomerase RNA, Xist RNA, a variety of small regulatory RNAs)	
Non-repetitive intragenic non-coding DNA	
Interspersed Repetitive DNA	
Pseudogenes	
Short Interspersed Elements (SINEs)	
Alu element (300 bp)	~1,000,000 copies
MIR families (average ~130 bp) (mammalian-wide interspersed repeat)	~400,000 copies
Long Interspersed Elements (LINEs)	
LINE-1 family (average ~800 bp)	~200,000–500,000 copies
LINE-2 family (average ~250 bp)	~270,000 copies
Retrovirus like elements (500–1300 bp)	~250,000 copies
DNA transposons (variable; average ~250 bp)	~200,000 copies
Tandem Repetitive DNA	
Ribosomal RNA genes	5 clusters of about 50 tandem repeats on 5 different chromosomes
Transfer RNA genes	multiple copies plus several pseudogenes
Telomere sequences	several kb of a 6 bp tandem repeat
Mini-satellites (= VNTRs)	blocks of 0.1 to 20 kbp of short tandem repeats (5–50 bp), most located close to telomeres
Centromere sequence (α-satellite DNA)	171 bp repeat, binds centromere proteins
Satellite DNA	blocks of 100 kbp or longer of tandem repeats of 20 to 200 bp, most located close to centromeres
Mega-satellite DNA	blocks of 100 kbp or longer of tandem repeats of 1 to 5 kbp, various locations



There are two types of plasmid integration into a host bacteria: Non-integrating plasmids replicate as with the top instance; whereas episomes, the lower example, integrate into the host chromosome.



Overview of bacterial conjugation

Konjugacija transfer plazmida izmedju bakterija

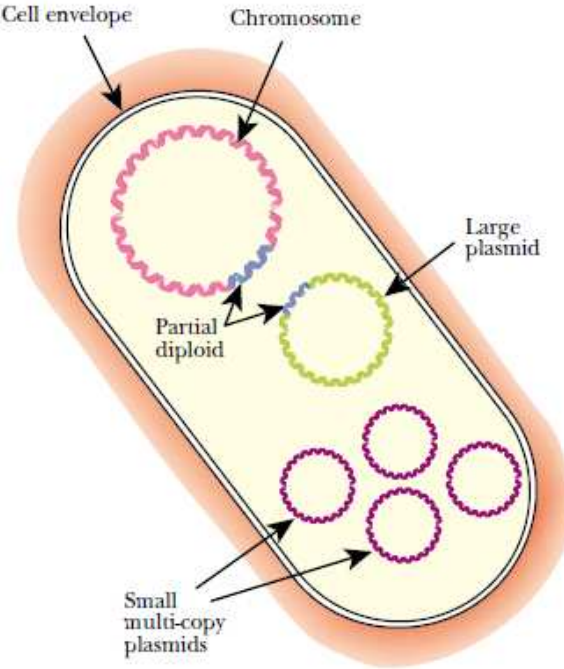
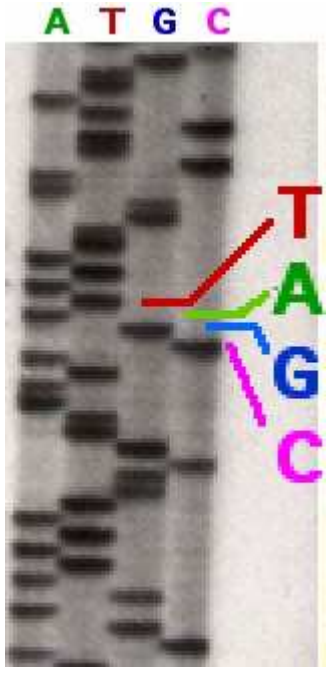
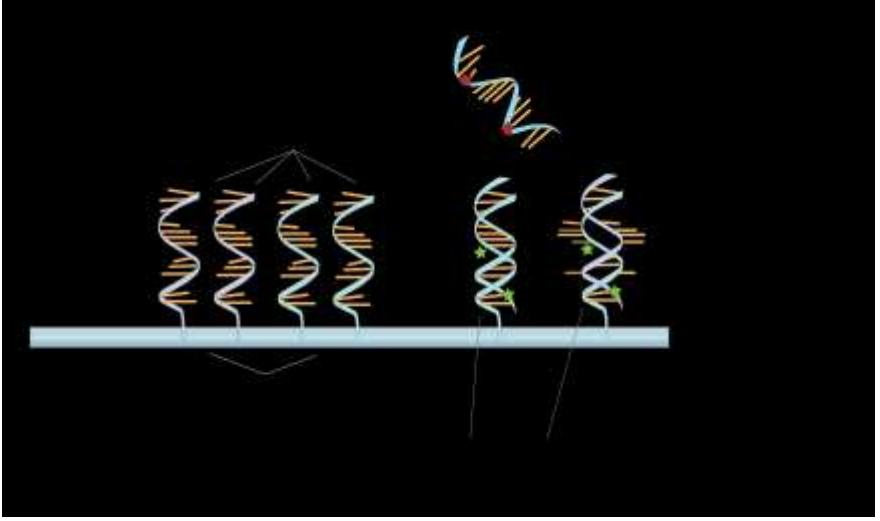


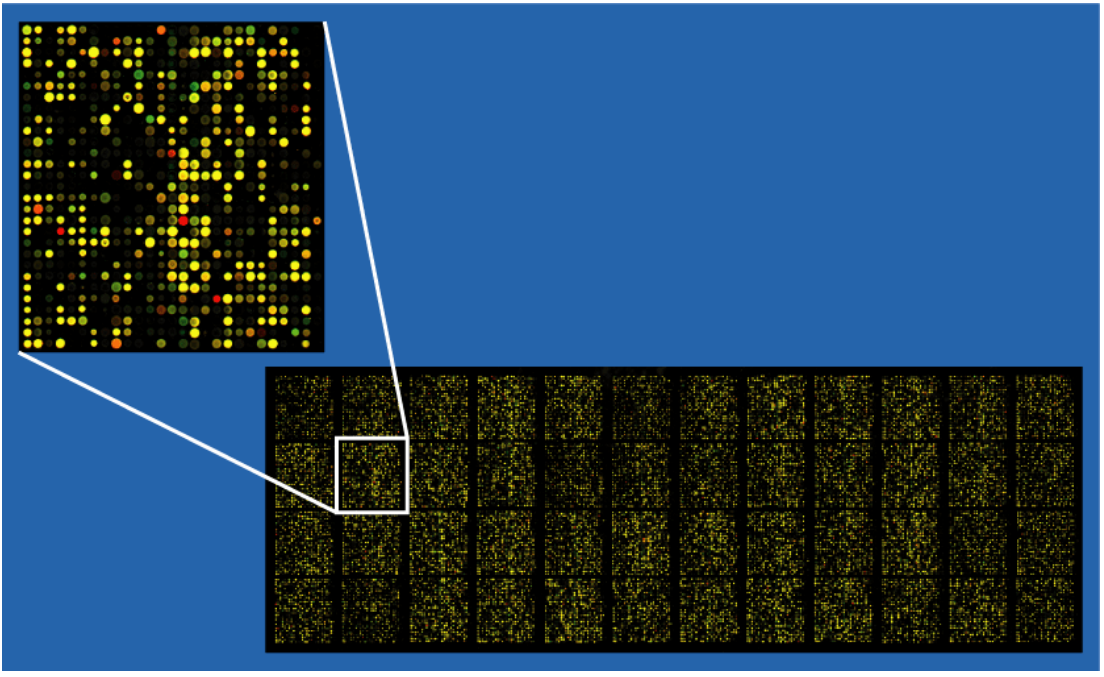
FIGURE 1.26 ***Plasmid within an* E. coli *Cell Shows Partial Diploidy***

The bacterial chromosome is indicated in red; the large single-copy plasmid is indicated in green and the small multiple-copy plasmid in purple. Note that a segment of the bacterial chromosome, colored blue, has been duplicated and is carried by the larger plasmid, making the cell a partial diploid.

Transpozon


A T G C
T
A
G
C





standard **Amino Acids**

A L-Alanine ALA

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Nonpolar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

C L-Cysteine CYS

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Polar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

D L-Aspartic Acid ASP

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Polar
Acidity/ Basicity	Acidic
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

E L-Glutamic Acid GLU

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	[illegible]
Acidity/ Basicity	[illegible]
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

F L-Phenylalanine PHE

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Nonpolar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

G Glycine GLY

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Nonpolar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

H L-Histidine HIS

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Polar
Acidity/ Basicity	Base (Weak)
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

I L-Isoleucine ILE

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	[illegible]
Acidity/ Basicity	[illegible]
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

K L-Lysine LYS

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Polar
Acidity/ Basicity	Basic
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

L L-Leucine LEU

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Nonpolar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

M L-Methionine MET

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Nonpolar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

N L-Asparagine ASN

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	[illegible]
Acidity/ Basicity	[illegible]
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

P L-Proline PRO

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Nonpolar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

Q L-Glutamine GLN

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Polar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

R L-Arginine ARG

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Polar
Acidity/ Basicity	Basic (Strong)
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

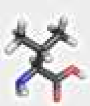
S L-Serine SER

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	[illegible]
Acidity/ Basicity	[illegible]
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

T L-Threonine THR

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Polar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

V L-Valine VAL



Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Nonpolar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

W L-Tryptophan TRP

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	Nonpolar
Acidity/ Basicity	Neutral
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

Y L-Tyrosine TYR

Chemical Formula	[illegible]
Molecular Weight	[illegible]
Polarity	[illegible]
Acidity/ Basicity	[illegible]
Hydrophobicity	[illegible]
Hydropathy Index	[illegible]
Isoelectric Point	[illegible]
pKa (α-COOH)	[illegible]
pKa (α-NH3)	[illegible]
CAS Number	[illegible]
PubChem CID	[illegible]
Canonical SMILES	[illegible]

B L-Asparagine or L-Aspartic Acid ASX

J L-Leucine or L-Isoleucine XLE

X Unspecified or Unknown XAA

Z L-Glutamine or L-Glutamic Acid GLX

Essential amino acids are marked with yellow spots. Polarity refers to Side Chain Polarity. Acidity/Basicity refers to Side Chain Acidity/Basicity.

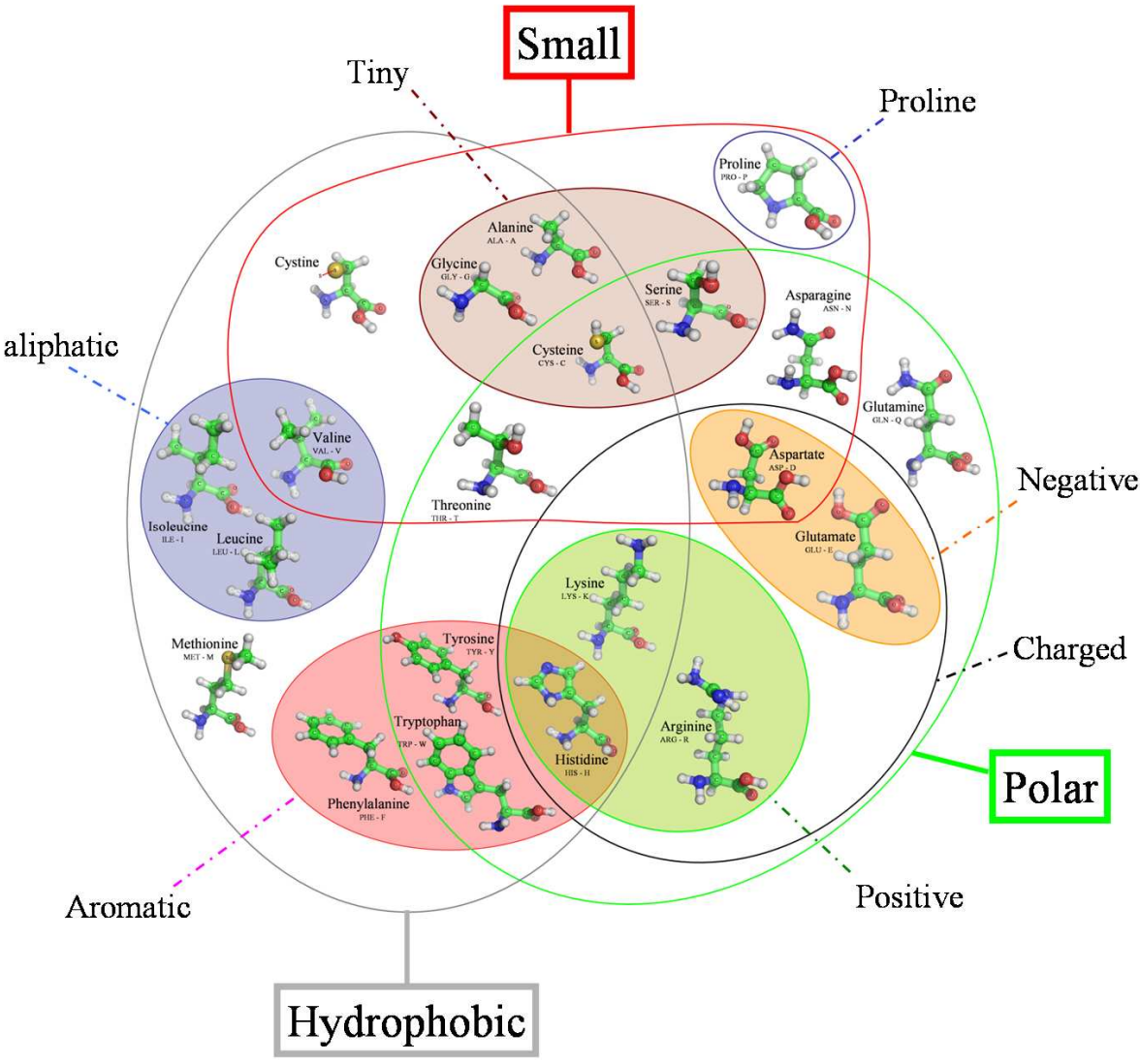

Small
Tiny
Proline
Proline
PRO - P
Alanine
ALA - A
Glycine
GLY - G
Cystine
Serine
SER - S
Asparagine
ASN - N
Cysteine
CYS - C
aliphatic
Glutamine
GLN - Q
Valine
VAL - V
Threonine
THR - T
Negative
Aspartate
ASP - D
Isoleucine
ILE - I
Leucine
LEU - L
Glutamate
GLU - E
Lysine
LYS - K
Charged
Methionine
MET - M
Tyrosine
TYR - Y
Tryptophan
TRP - W
Arginine
ARG - R
Histidine
HIS - H
Polar
Phenylalanine
PHE - F
Aromatic
Positive
Hydrophobic

Table 4. Values of the 12 Previously Published Scales of Amino Acid Hydrophobicity Coefficients[a]

amino acid	ZIMM	FAUC	BULL	CHOTHIA	JANIN	GUY	HEIJ	HW	KD	MEEK	MEEKR	GUO
Ala	0.83	0.31	−200	0.38	1.70	0.10	−12.04	−0.5	1.8	0.5	1.1	2.2
Arg	0.83	−1.01	−120	0.01	01.0	1.90	39.23	3.0	−4.5	0.8	−0.4	0.9
Asn	0.09	−0.60	80	0.12	0.40	0.48	4.25	0.2	−3.5	0.8	−4.2	−0.8
Asp	0.64	−0.77	−200	0.15	0.40	0.78	23.22	3.0	−3.5	−8.2	−1.6	−2.6
Cys	1.48	1.54	−450	0.50	4.60	−1.42	3.95	−1.0	2.5	−6.8	7.1	2.6
Gln	0.00	−0.22	160	0.07	0.30	0.95	2.16	0.2	−3.5	−4.8	−2.9	−0.2
Glu	0.65	−0.64	−300	0.18	0.30	0.83	16.81	3.0	−3.5	−16.9	0.7	−0.2
Gly	0.10	0.00	0	0.36	1.80	0.33	−7.85	0.0	−0.4	0.0	−0.2	0.0
His	1.10	0.13	−120	0.17	0.80	−0.50	6.28	−0.5	−3.2	−3.5	−0.7	2.2
Ile	2.52	1.80	2260	0.60	3.10	−1.13	−18.32	−1.8	4.5	13.9	8.5	8.3
Leu	3.07	1.70	2460	0.45	2.40	−1.18	−17.79	−1.8	3.8	8.8	11.0	9.0
Lys	1.60	−0.99	−350	0.03	0.05	1.40	9.71	3.0	−3.9	0.1	−1.9	0.0
Met	1.40	1.23	1470	0.40	1.90	−1.59	−8.86	−1.3	1.9	4.8	5.4	6.0
Phe	2.75	1.79	2330	0.50	2.20	−2.12	−21.98	−2.5	2.8	13.2	13.0	9.0
Pro	2.70	0.72	−980	0.18	0.60	0.73	5.82	0.0	−1.6	6.1	4.4	2.2
Ser	0.14	−0.04	−300	0.22	0.80	0.52	−1.54	0.3	−0.8	1.2	−3.2	2.0
Thr	0.54	0.26	−520	0.23	0.70	0.07	−4.15	−0.4	−0.7	2.7	−1.7	0.3
Trp	0.31	2.25	2010	0.27	1.60	−0.51	−16.19	−3.4	−0.9	14.9	17.0	9.5
Tyr	2.97	0.96	2240	0.15	1.50	−0.21	−1.51	−2.3	−1.3	6.1	7.4	4.6
Val	1.79	1.22	1560	0.54	2.90	−1.27	−16.22	−1.5	4.2	2.7	5.9	5.7

[a] For reference codes, see Table 3.

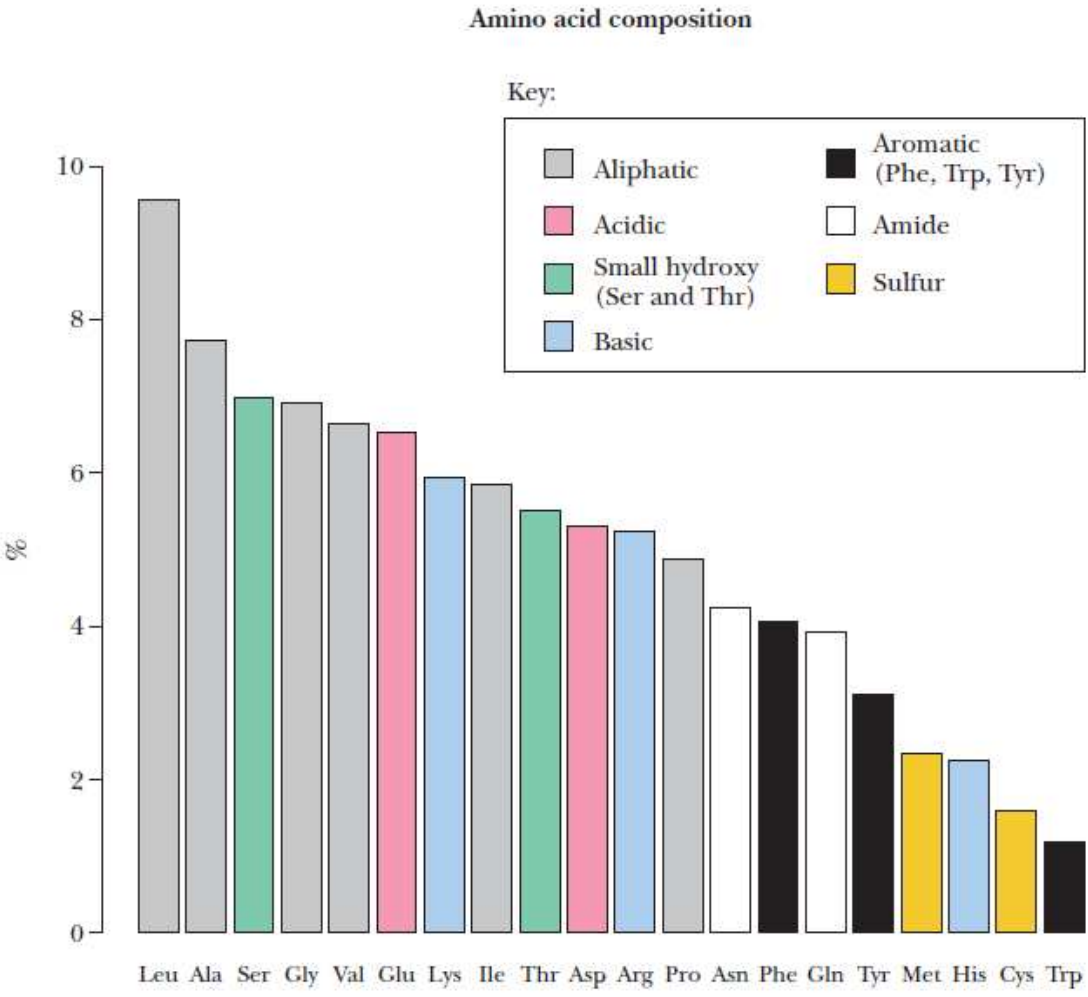
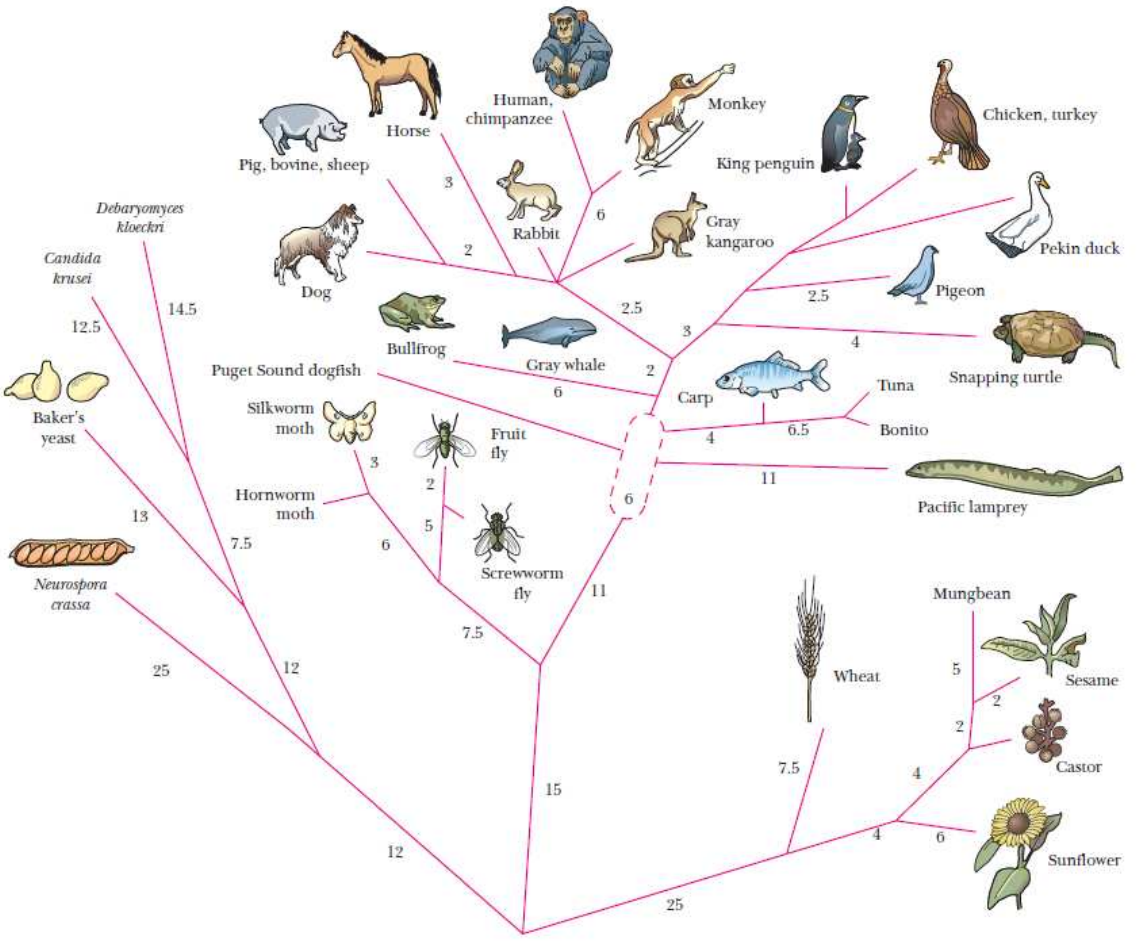


FIGURE 5.16 Amino acid composition: frequencies of the various amino acids in proteins for all the proteins in the SWISS-PROT protein knowedgebase. These data are derived from the amino acid composition of more than 100,000 different proteins (representing more than 40,000,000 amino acid residues). The range is from leucine at 9.55% to tryptophan at 1.18% of all residues.



	1									10										20										
Ancestral cytochrome *c*	–Pro–	Ala–	Gly–	Asp–	?	–Lys–	Lys	–Gly–	Ala–	Lys–	Ile	–Phe–	Lys	–Thr–	?	–Cys–	Ala	–Gln–	Cys	–His–	Thr–	Val–	Glu–	?	–	Gly–	Gly–	?	–	
Human cytochrome *c*		Gly–	Asp–	Val–	Glu–	Lys	–Gly–	Lys	–Lys–	Ile	–Phe–	Ile	–Met–	Lys	–Cys–	Ser	–Gln–	Cys	–His–	Thr–	Val–	Glu–	Lys	–	Gly–	Gly–	Lys–			

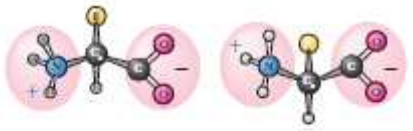
	30									40										50										
	–His–	Lys	–Val–	Gly–	Pro	–Asn–	Leu–	His–	Gly–	Leu–	Phe–	Gly	–Arg–	Lys–	?	–Gly–	Gln–	Ala–	?	–	Gly–	Tyr	–Ser–	Tyr	–Thr–	Asp–			
	–His–	Lys–	Thr–	Gly–	Pro	–Asn–	Leu–	His–	Gly–	Leu–	Phe–	Gly	–Arg–	Lys–	Thr–	Gly–	Gln–	Ala–	Pro	–	Gly–	Tyr	–Ser–	Tyr	–Thr–	Ala–				

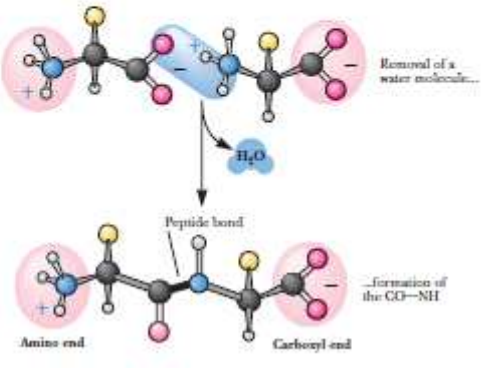
	60									70																				
	–Ala–	Asn–	Lys–	Asn–	Lys	–Gly–	?	–	?	–Trp–	?	–Glu–	Asn–	Thr–	Leu–	Phe–	Glu–	Tyr–	Leu	–Glu–	Asn–	Pro	–Lys–	Lys	–Tyr–	Ile	–			
	–Ala–	Asn–	Lys–	Asn–	Lys	–Gly–	Ile	–	Ile	–Trp–	Gly	–Glu–	Asp–	Thr–	Leu–	Met–	Gln–	Tyr–	Leu	–Glu–	Asn–	Pro	–Lys–	Lys	–Tyr–	Pro–				

	80									90										100										
	–Pro–	Gly–	Thr–	Lys–	Met–	?	–Phe–	?	–	Gly–	Leu–	Lys	–Lys–	?	–	?	–Asp–	Arg–	Ala–	Asp	–Leu–	Ile–	Ala–	Tyr	–Leu–	Lys–	?	–		
	–Pro–	Gly–	Thr–	Lys–	Met–	Ile	–Phe–	Val	–Gly–	Ile	–	Lys	–Lys–	Lys	–Glu–	Glu	–Arg–	Ala–	Asp	–Leu–	Ile–	Ala–	Tyr	–Leu–	Lys–	Lys–				

–Ala–Thr–Ala
–Ala–Thr–Asn–Glu

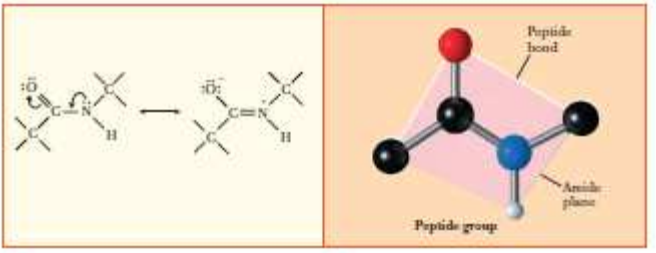
FIGURE 5.20 This phylogenetic tree depicts the evolutionary relationships among organisms as determined by the similarity of their cytochrome *c* amino acid sequences. The numbers along the branches give the amino acid changes between a species and a hypothetical progenitor. Note that extant species are located only at the tips of branches. Below, the sequence of human cytochrome *c* is compared with an inferred ancestral sequence represented by the base of the tree. Uncertainties are denoted by question marks. (Adapted from Creighton, T. E., 1983. *Proteins: Structure and Molecular Properties.* San Francisco: W. H. Freeman.)





Biochemistry Now™

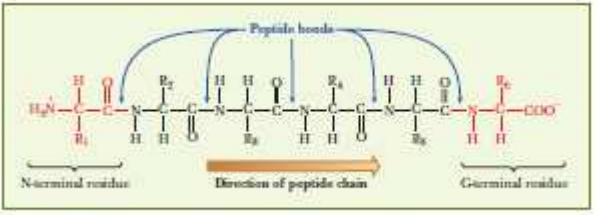
■ **ANIMATED FIGURE 3.8** **Formation of the peptide bond.** *(Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)* **Sign in at www.thomsonedu.com/login to see an animated version of this figure.**



Ⓐ Resonance structures of the peptide group.

Ⓑ The planar peptide group.

■ **FIGURE 3.10** **The resonance structures of the peptide bond lead to a planar group.** *(Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)*



■ **FIGURE 3.9** **A small peptide showing the direction of the peptide chain (N-terminal to C-terminal).**

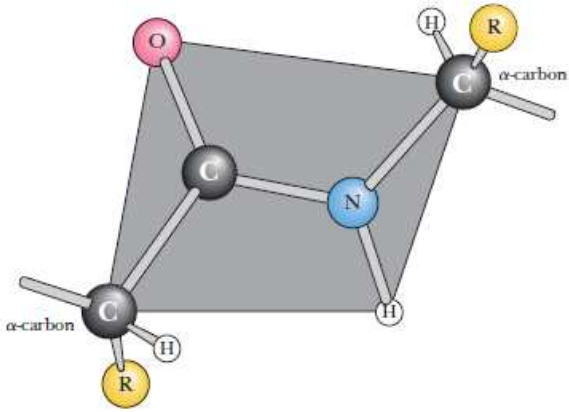


FIGURE 4.17 The coplanar relationship of the atoms in the amide group is highlighted as an imaginary shaded plane lying between two successive α-carbon atoms in the peptide backbone. (Illustration: Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

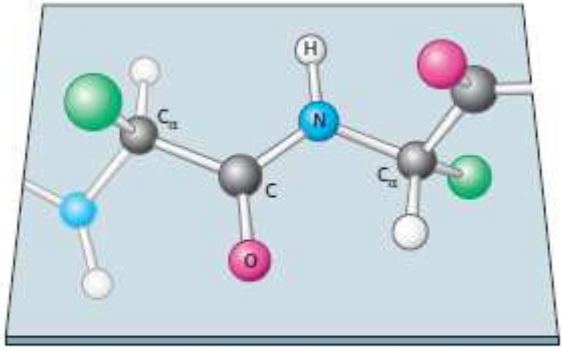
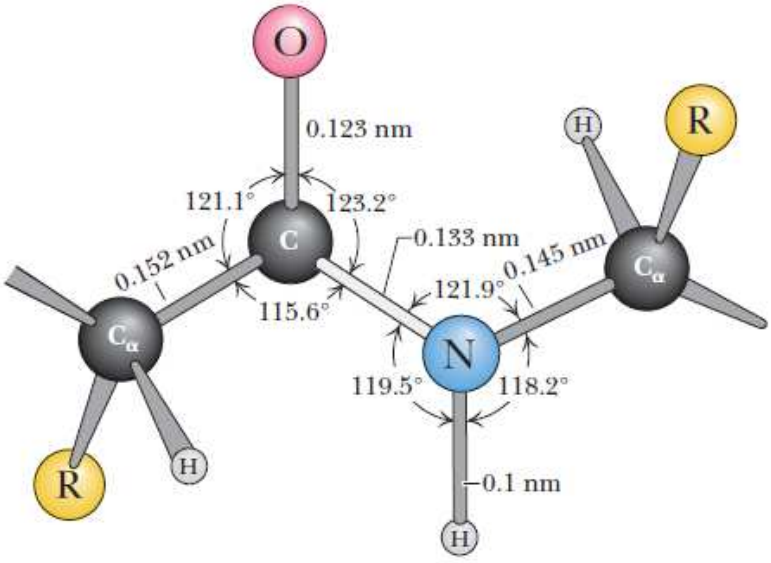


FIGURE 3.23 Peptide bonds are planar. In a pair of linked amino acids, six atoms (C_α, C, O, N, H, and C_α) lie in a plane. Side chains are shown as green balls.



CENGAGENOW™ **ANIMATED FIGURE 4.15** The peptide bond is shown in its usual *trans* conformation of carbonyl O and amide H. The C_α atoms are the α-carbons of two adjacent amino acids joined in peptide linkage. The dimensions and angles are the average values observed by crystallographic analysis of amino acids and small peptides. The peptide bond is the light-colored bond between C and N. (Adapted from Ramachandran, G. N., et al., 1974. The mean geometry of the peptide unit from crystal structure data. *Biochimica et Biophysica Acta* **359:**298–302.) **See this figure animated at www.cengage.com/login**

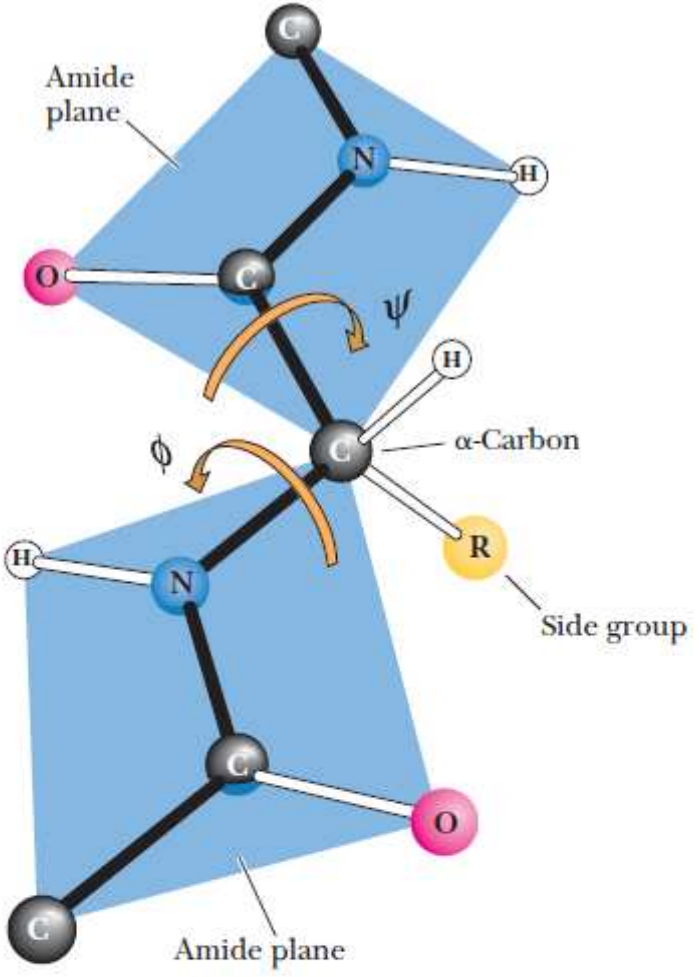


FIGURE 6.2 The amide or peptide bond planes are joined by the tetrahedral bonds of the α-carbon. The rotation parameters are ϕ and ψ. The conformation shown corresponds to $\phi = 180°$ and $\psi = 180°$. Note that positive values of ϕ and ψ correspond to clockwise rotation as viewed from C_α. Starting from 0°, a rotation of 180° in the clockwise direction (+180°) is equivalent to a rotation of 180° in the counterclockwise direction (−180°). (Illustration: Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

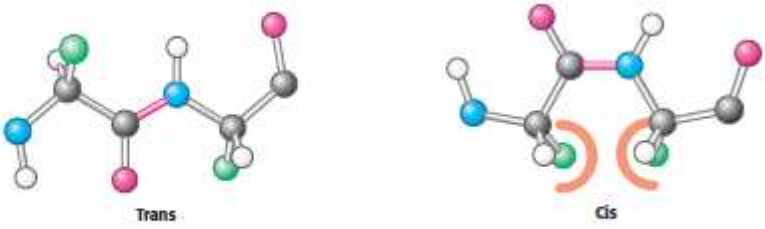


FIGURE 3.25 Trans and cis peptide bonds. The trans form is strongly favored because of steric clashes that occur in the cis form.

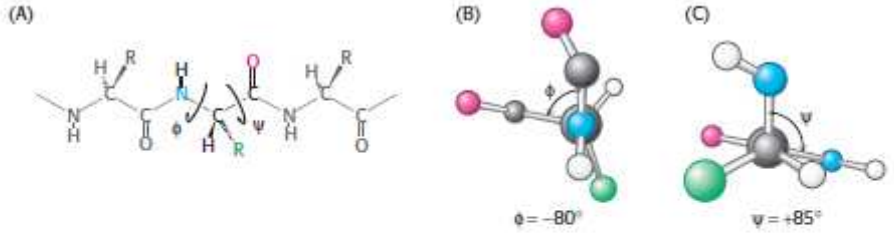
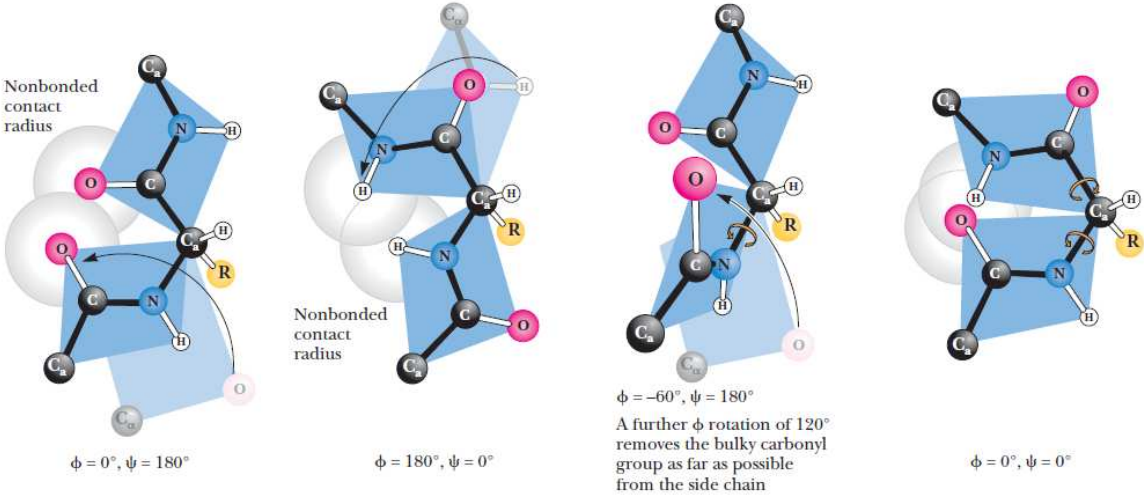


FIGURE 3.27 Rotation about bonds in a polypeptide. The structure of each amino acid in a polypeptide can be adjusted by rotation about two single bonds. (A) Phi (ϕ) is the angle of rotation about the bond between the nitrogen and the α-carbon atoms, whereas psi (ψ) is the angle of rotation about the bond between the α-carbon and the carbonyl carbon atoms. (B) A view down the bond between the nitrogen and the α-carbon atoms, showing how ϕ is measured. (C) A view down the bond between the α-carbon and the carbonyl carbon atoms, showing how ψ is measured.

permission.)



CENGAGENOW™ **ACTIVE FIGURE 6.3** Many of the possible conformations about an α-carbon between two peptide planes are forbidden because of steric crowding. Several noteworthy examples are shown here.

Note: The formal IUPAC-IUB Commission on Biochemical Nomenclature convention for the definition of the torsion angles ϕ and ψ in a polypeptide chain (*Biochemistry* **9:**3471–3479, 1970) is different from that used here, where the C_α atom serves as the point of reference for both rotations, but the result is the same. (Illustration: Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.) **Test yourself on the concepts in this figure at www.cengage.com/login.**

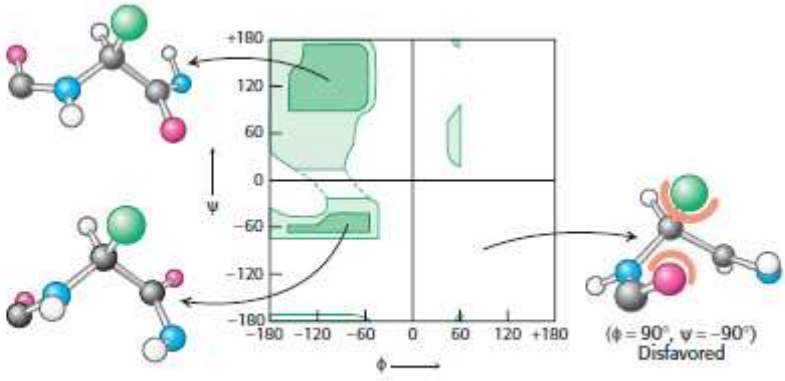


FIGURE 3.28 A Ramachandran diagram showing the values of ϕ and ψ. Not all ϕ and ψ values are possible without collisions between atoms. The most favorable regions are shown in dark green; borderline regions are shown in light green. The structure on the right is disfavored because of steric clashes.

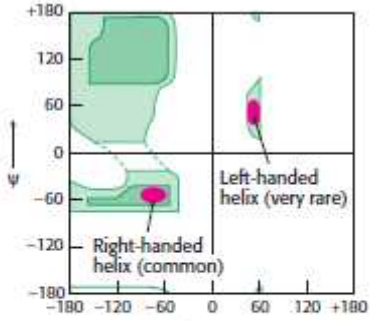
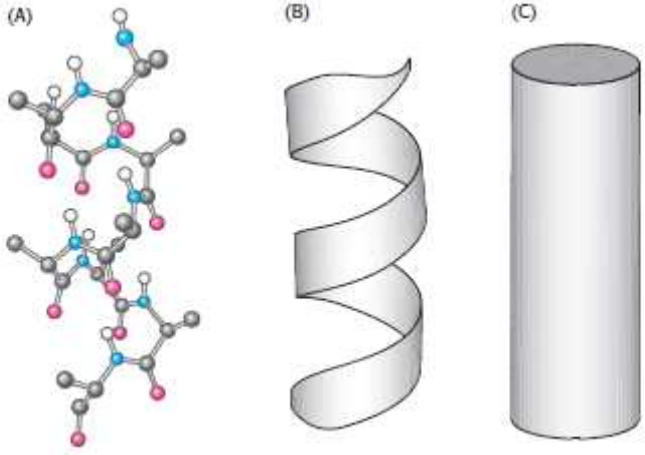


FIGURE 3.31 Ramachandran diagram for helices. Both right- and left-handed helices lie in regions of allowed conformations in the Ramachandran diagram. However, essentially all α helices in proteins are right-handed.



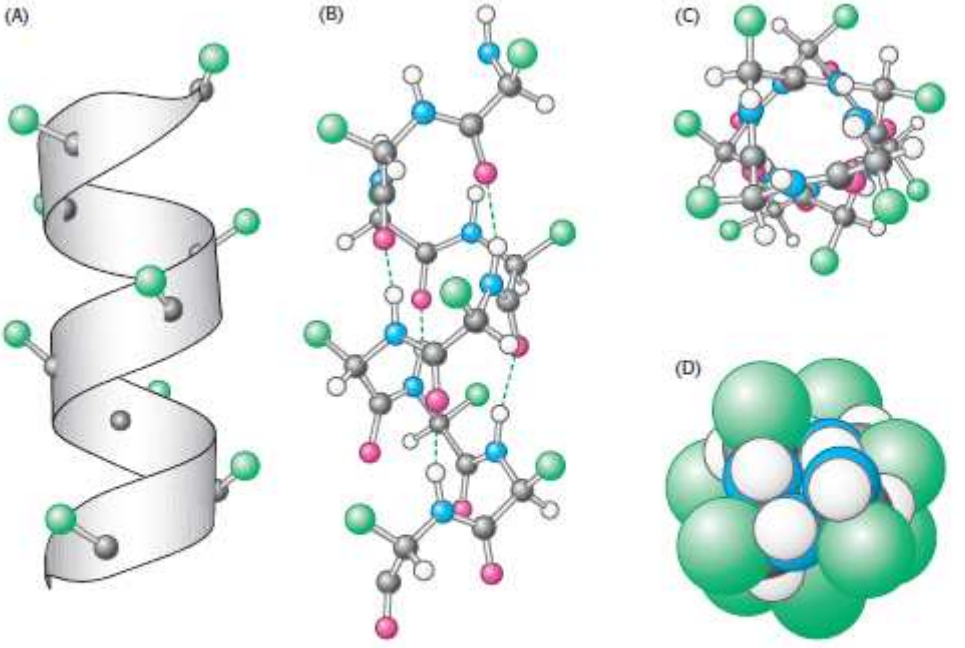
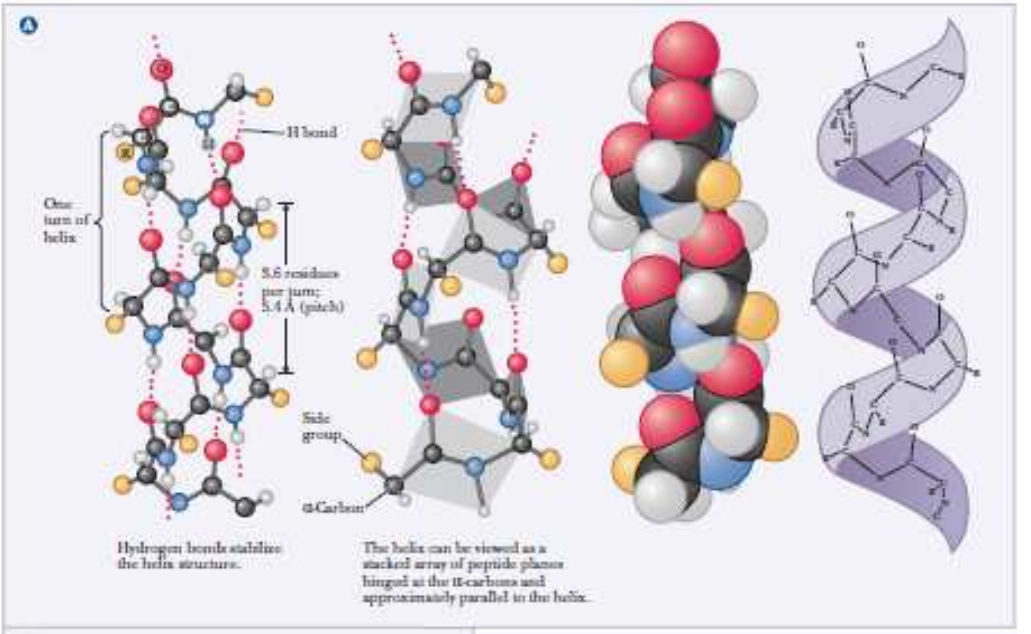
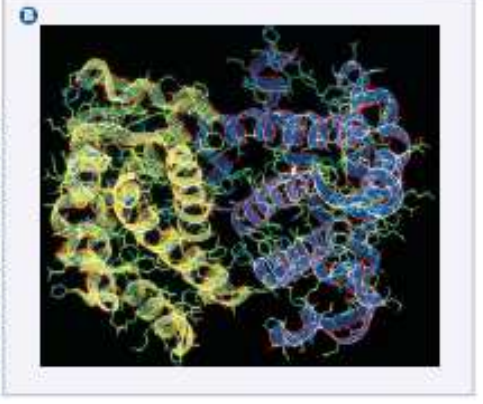
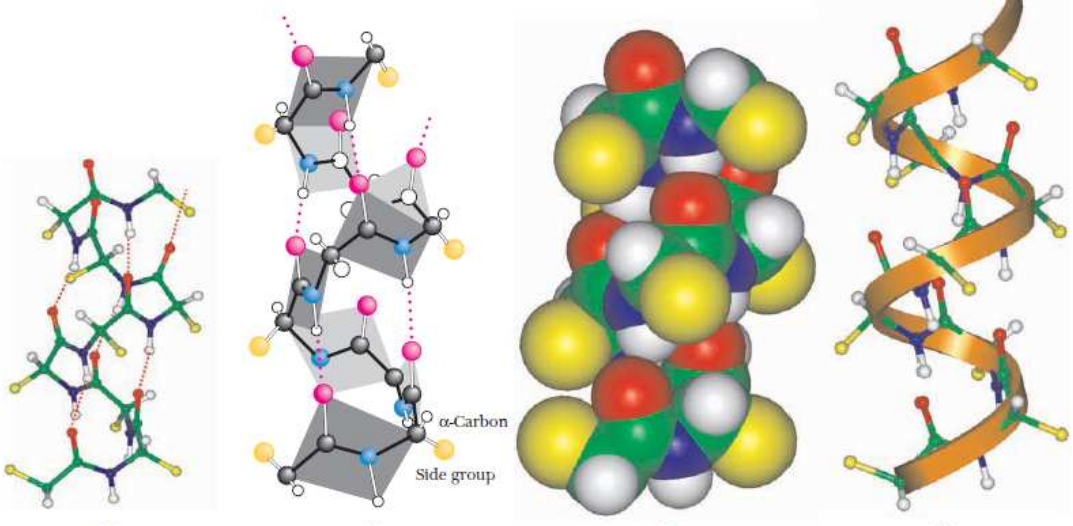


FIGURE 3.29 Structure of the α helix. (A) A ribbon depiction with the α-carbon atoms and side chains (green) shown. (B) A side view of a ball-and-stick version depicts the hydrogen bonds (dashed lines) between NH and CO groups. (C) An end view shows the coiled backbone as the inside of the helix and the side chains (green) projecting outward. (D) A space-filling view of part C shows the tightly packed interior core of the helix.





■ **FIGURE 4.2 The α-helix.** (a) From left to right, ball-and-stick model of the α-helix, showing terminology; ball-and-stick model with planar peptide groups shaded; computer-generated space-filling model of the α-helix; outline of the α-helix. (b) Model of the protein hemoglobin, showing the helical regions. *(Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)*



(a)
Hydrogen bonds stabilize the helix structure.

(b)
The helix can be viewed as a stacked array of peptide planes hinged at the α-carbons and approximately parallel to the helix.

(c)

(d)

FIGURE 6.6 Four different graphic representations of the α-helix. **(a)** A stick representation with H bonds as dotted lines, as originally conceptualized in Pauling's 1960 *The Nature of the Chemical Bond*. **(b)** Showing the arrangement of peptide planes in the helix. (Illustration: Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.) **(c)** A space-filling computer graphic presentation. **(d)** A "ribbon structure" with an inlaid stick figure, showing how the ribbon indicates the path of the polypeptide backbone.

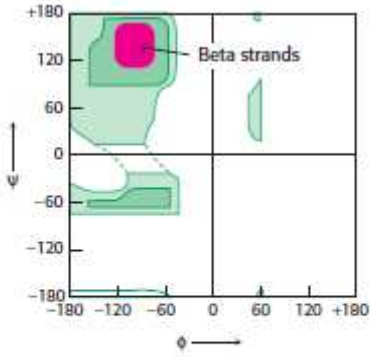


FIGURE 3.35 Ramachandran diagram for β strands. The red area shows the sterically allowed conformations of extended, β-strand-like structures.

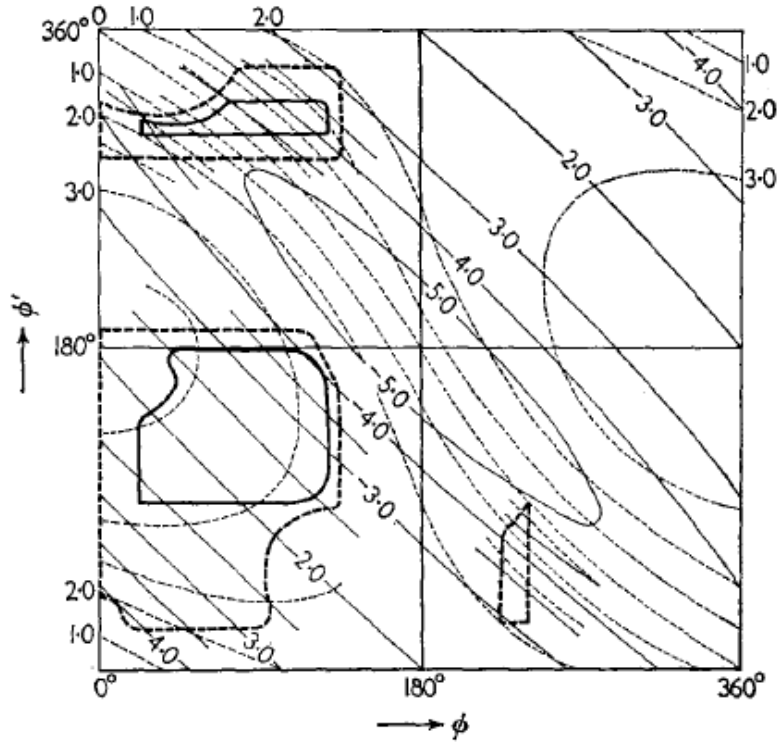


FIG. 2. The fully allowed (———) and outer limit (— — —) regions of (ϕ, ϕ') for angl
N—αC—C′ = 110° along with the configurations of various known di-, tri- and polypeptide an
protein structures.

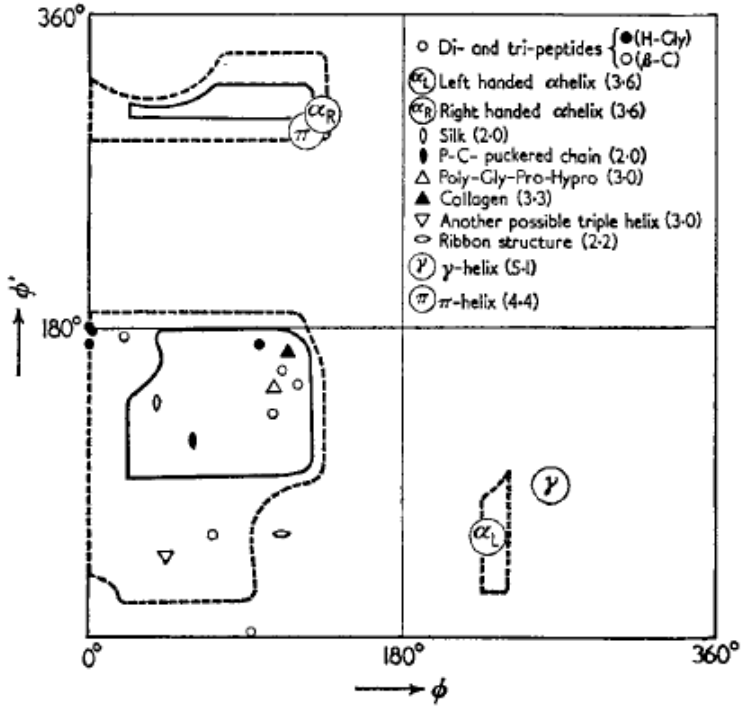


FIG. 3. Contours of constant n (———) and constant h (– – – – – –) corresponding to the angle N—αC—C′ = 110°. The boundaries of the fully allowed and outer limit regions are also shown.

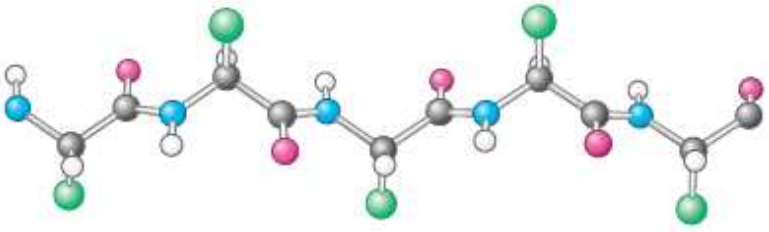


FIGURE 3.36 Structure of a β strand. The side chains (green) are alternately above and below the plane of the strand.

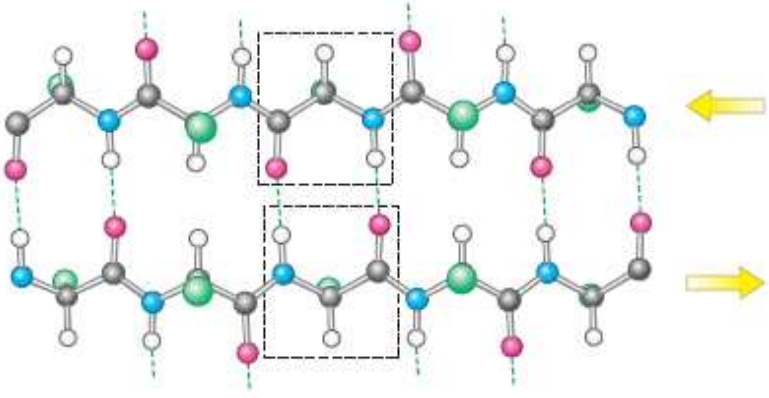


FIGURE 3.37 An antiparallel β sheet. Adjacent β strands run in opposite directions. Hydrogen bonds between NH and CO groups connect each amino acid to a single amino acid on an adjacent strand, stabilizing the structure.

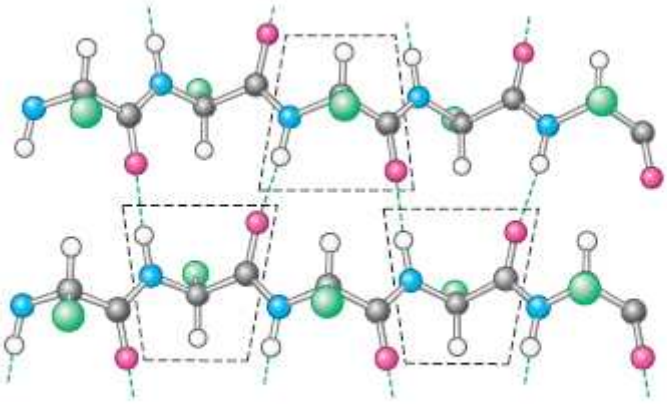


FIGURE 3.38 A parallel β sheet. Adjacent β strands run in the same direction. Hydrogen bonds connect each amino acid on one strand with two different amino acids on the adjacent strand.

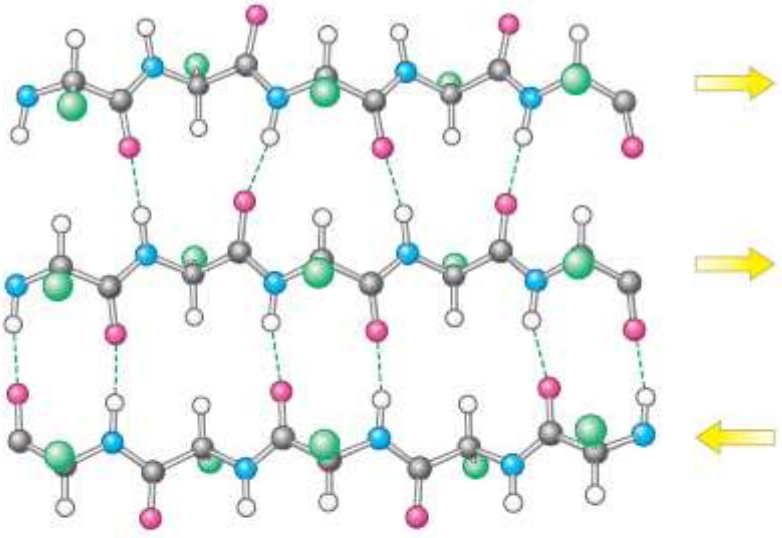


FIGURE 3.39 Structure of a mixed β sheet.

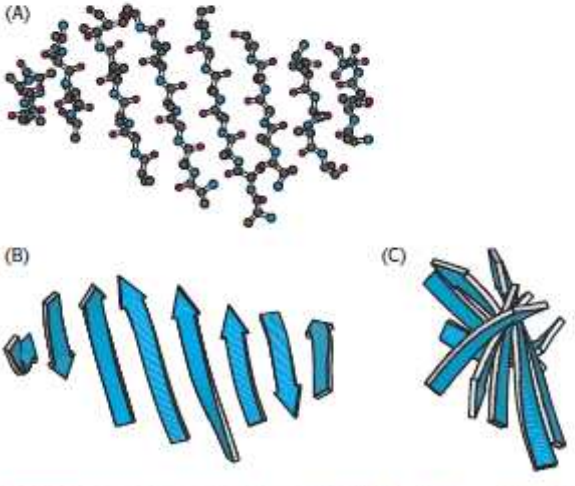


FIGURE 3.40 A twisted β sheet. (A) A ball-and-stick model. (B) A schematic model. (C) The schematic view rotated by 90 degrees to illustrate the twist more clearly.

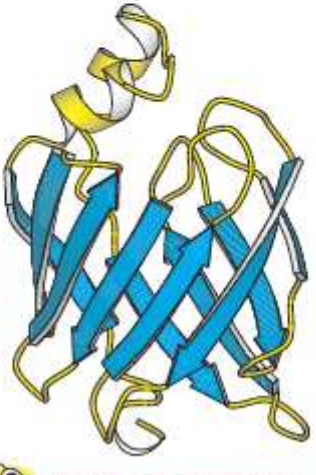


FIGURE 3.41 A protein rich in β sheets. The structure of a fatty acid-binding protein.

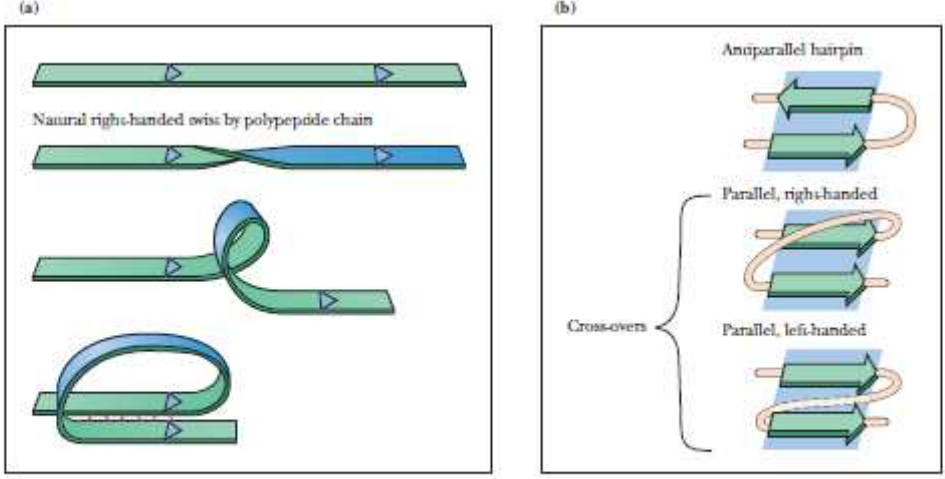


FIGURE 6.37 **(a)** The natural right-handed twist exhibited by polypeptide chains, and **(b)** the types of connections between β-strands.

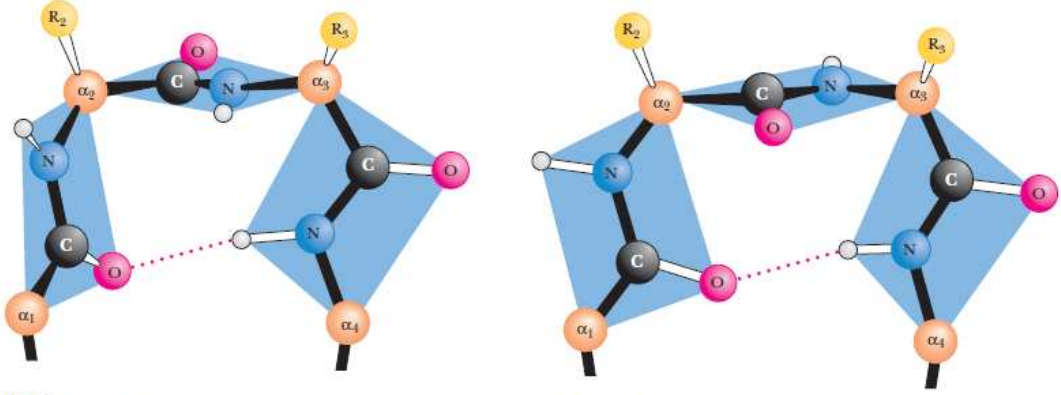


FIGURE 6.12 The structures of two kinds of β-turns (also called tight turns or β-bends). Four residues are required to form a β-turn. Left: Type I; right: Type II. (Illustration: Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

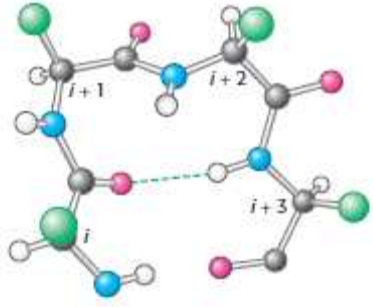
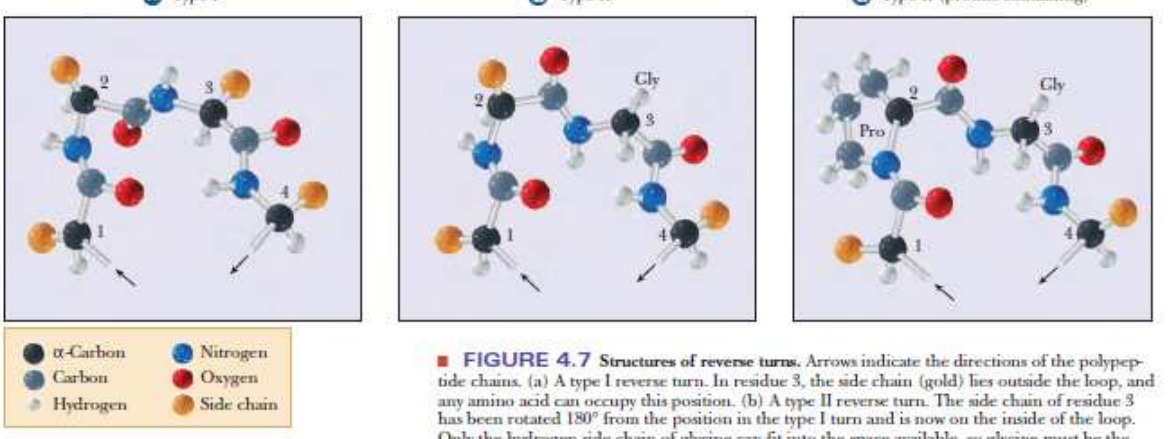
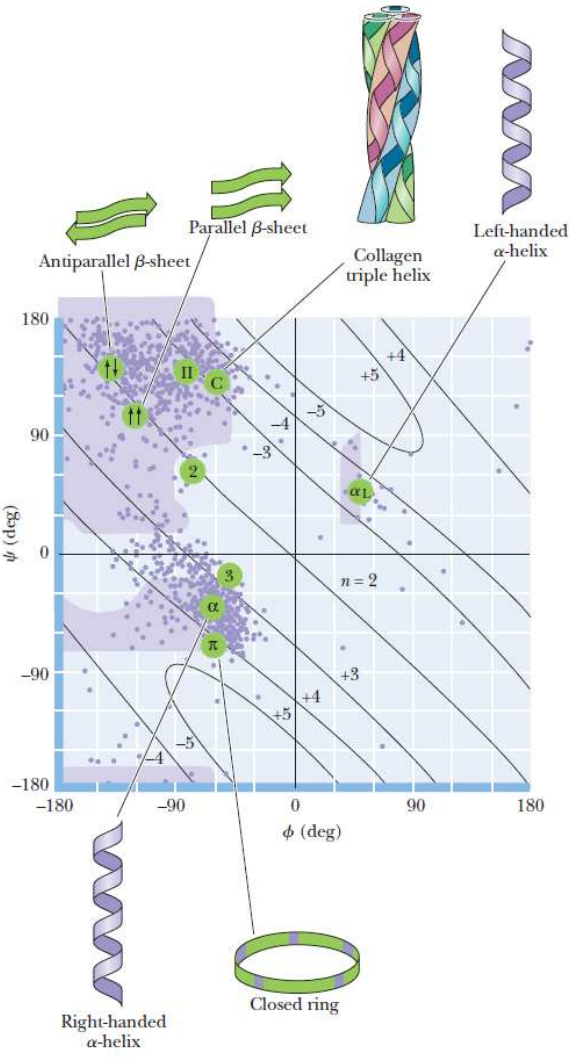


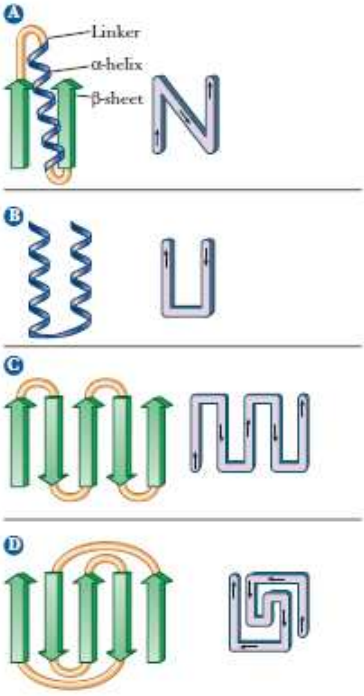
FIGURE 3.42 Structure of a reverse turn. The CO group of residue *i* of the polypeptide chain is hydrogen bonded to the NH group of residue *i* + 3 to stabilize the turn.



■ **FIGURE 4.7 Structures of reverse turns.** Arrows indicate the directions of the polypeptide chains. (a) A type I reverse turn. In residue 3, the side chain (gold) lies outside the loop, and any amino acid can occupy this position. (b) A type II reverse turn. The side chain of residue 3 has been rotated 180° from the position in the type I turn and is now on the inside of the loop. Only the hydrogen side chain of glycine can fit into the space available, so glycine must be the third residue in a type II reverse turn. (c) The five-membered ring of proline has the correct geometry for a reverse turn; this residue normally occurs as the second residue of a reverse turn. The turn shown here is type II, with glycine as the third residue.

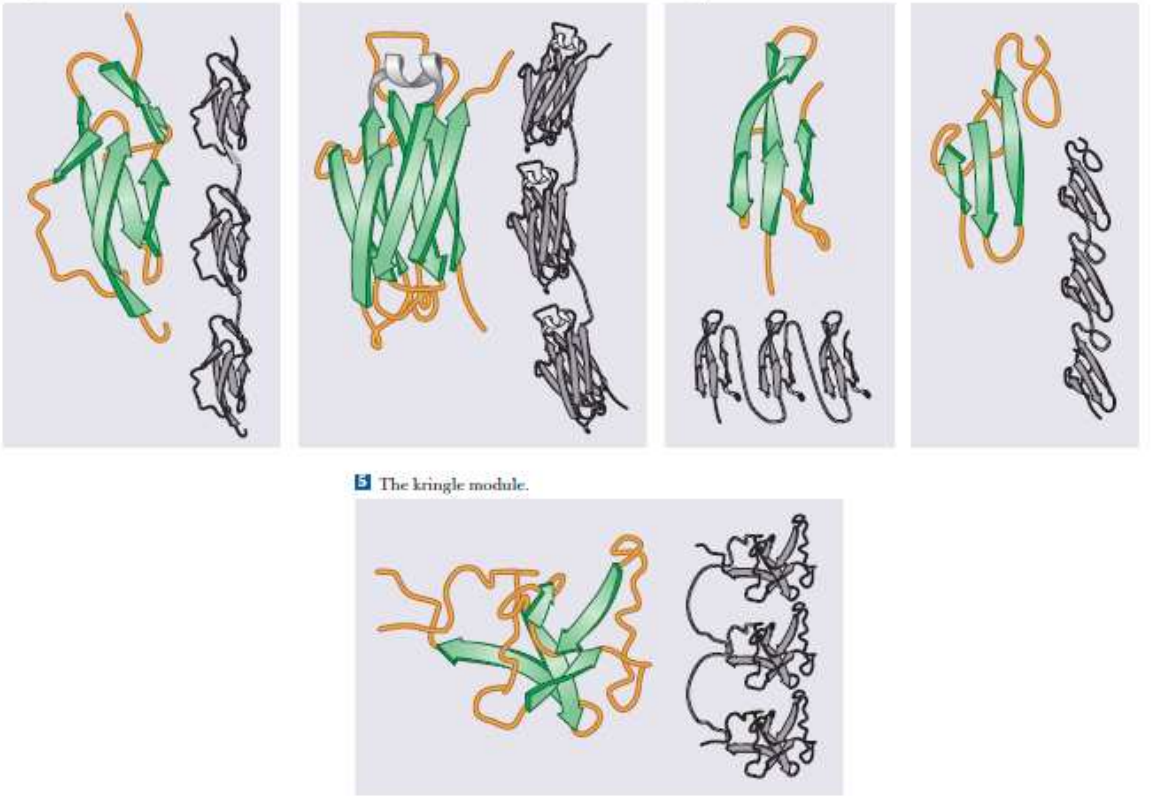


CENGAGENOW™ **ACTIVE FIGURE 6.4** A Ramachandran diagram showing the sterically reasonable values of the angles ϕ and ψ. The shaded regions indicate particularly favorable values of these angles. Dots in purple indicate actual angles measured for 1000 residues (excluding glycine, for which a wider range of angles is permitted) in eight proteins. The lines running across the diagram (numbered +5 through 2 and −5 through −3) signify the number of amino acid residues per turn of the helix; "+" means right-handed helices; "−" means left-handed helices. (After Richardson, J. S., 1981. The anatomy and taxonomy of protein structure. *Advances in Protein Chemistry* **34:**167–339.) **Test yourself on the concepts in this figure at www.cengage.com/login.**

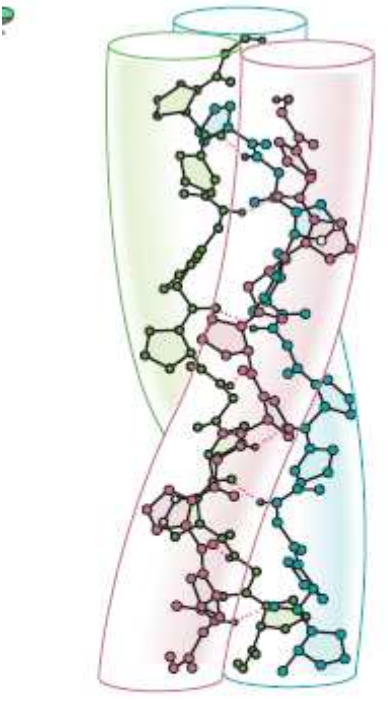


National Archaeological Museum, Athens/The Bridgeman Art Library International Ltd., London

■ **FIGURE 4.8 Schematic diagrams of supersecondary structures.** Arrows indicate the directions of the polypeptide chains. (a) A $\beta\alpha\beta$ unit, (b) an $\alpha\alpha$ unit, (c) a β-meander, and (d) the Greek key. (e) The Greek key motif in protein structure resembles the geometric patterns on this ancient Greek vase, giving rise to the name.

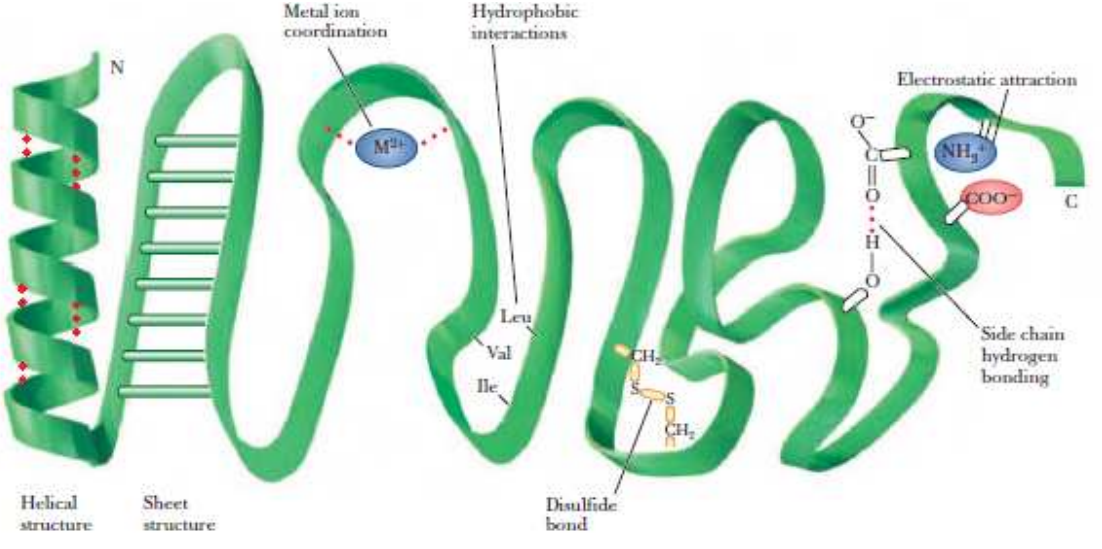


■ **FIGURE 4.9 Motifs and modules.** Motifs are repeated supersecondary structures, sometimes called modules. All of these have a particular secondary structure that is repeated in the protein. *(Reprinted from "Protein Modules," Trends in Biochemical Sciences, Vol. 16, pp. 13–17, Copyright © 1991, with permission from Elsevier.)*



Biochemistry Now™

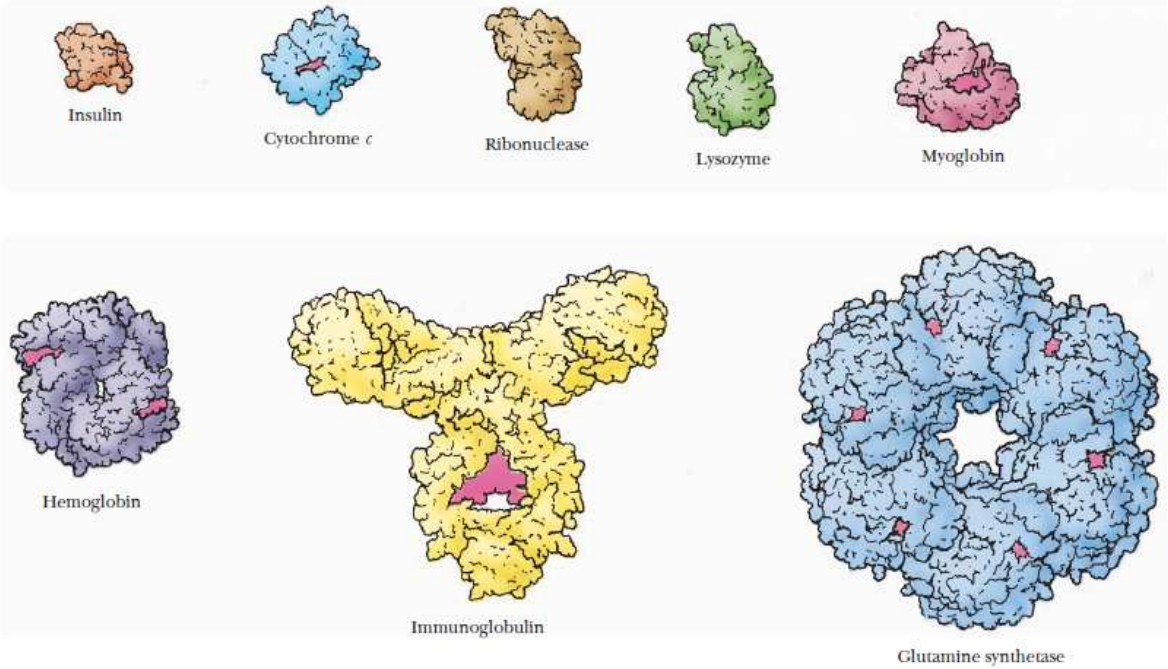
■ **ACTIVE FIGURE 4.11 A triple helix.** Poly (Gly—Pro—Pro) is a collagen-like right-handed triple helix composed of three left-handed helical chains. *(Adapted from M. H. Miller and H. A. Scheraga, 1976, Calculation of the structures of collagen models. Role of interchain interactions in determining the triple-helical coiled-coil conformations. I. Poly(glycyl-prolyl-prolyl).* Journal of Polymer Science Symposium *54:171–200.*)



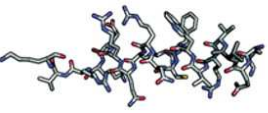
■ **FIGURE 4.13 Forces that stabilize the tertiary structure of proteins.** Note that the helical structure and sheet structure are two kinds of backbone hydrogen bonding. Although backbone hydrogen bonding is part of secondary structure, the conformation of the backbone constrains the possible arrangement of the side chains.

TABLE 4.2 Size of Protein Molecules*

Protein	M_r	Number of Residues per Chain	Subunit Organization
Insulin (bovine)	5,733	21 (A)	$\alpha\beta$
		30 (B)	
Cytochrome *c* (equine)	12,500	104	α_1
Ribonuclease A (bovine pancreas)	12,640	124	α_1
Lysozyme (egg white)	13,930	129	α_1
Myoglobin (horse)	16,980	153	α_1
Chymotrypsin (bovine pancreas)	22,600	13 (α)	$\alpha\beta\gamma$
		132 (β)	
		97 (γ)	
Hemoglobin (human)	64,500	141 (α)	$\alpha_2\beta_2$
		146 (β)	
Serum albumin (human)	68,500	550	α_1
Hexokinase (yeast)	96,000	200	α_4
γ-Globulin (horse)	149,900	214 (α)	$\alpha_2\beta_2$
		446 (β)	
Glutamate dehydrogenase (liver)	332,694	500	α_6
Myosin (rabbit)	470,000	2,000 (heavy, *h*)	$h_2\alpha_1\alpha'_2\beta_2$
		190 (α)	
		149 (α')	
		160 (β)	
Ribulose bisphosphate carboxylase (spinach)	560,000	475 (α)	$\alpha_8\beta_8$
		123 (β)	
Glutamine synthetase *(E. coli)*	600,000	468	α_{12}

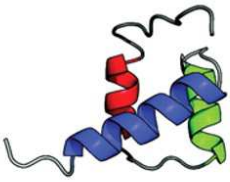


*Illustrations of selected proteins listed in Table 4.2 are drawn to constant scale.
Adapted from Goodsell, D. S., and Olson, A. J., 1993. Soluble proteins: Size, shape and function. *Trends in Biochemical Sciences* **18**:65–68.

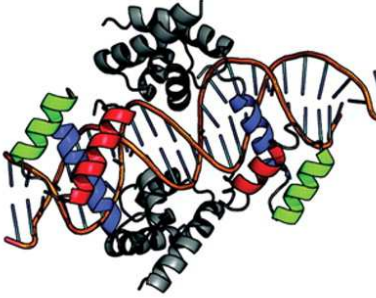


Secondary structure

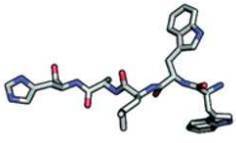
Motif

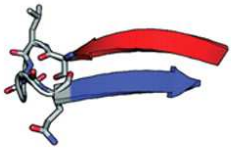


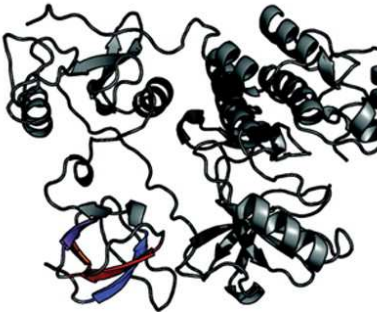
Domain



Quaternary/tertiary structure







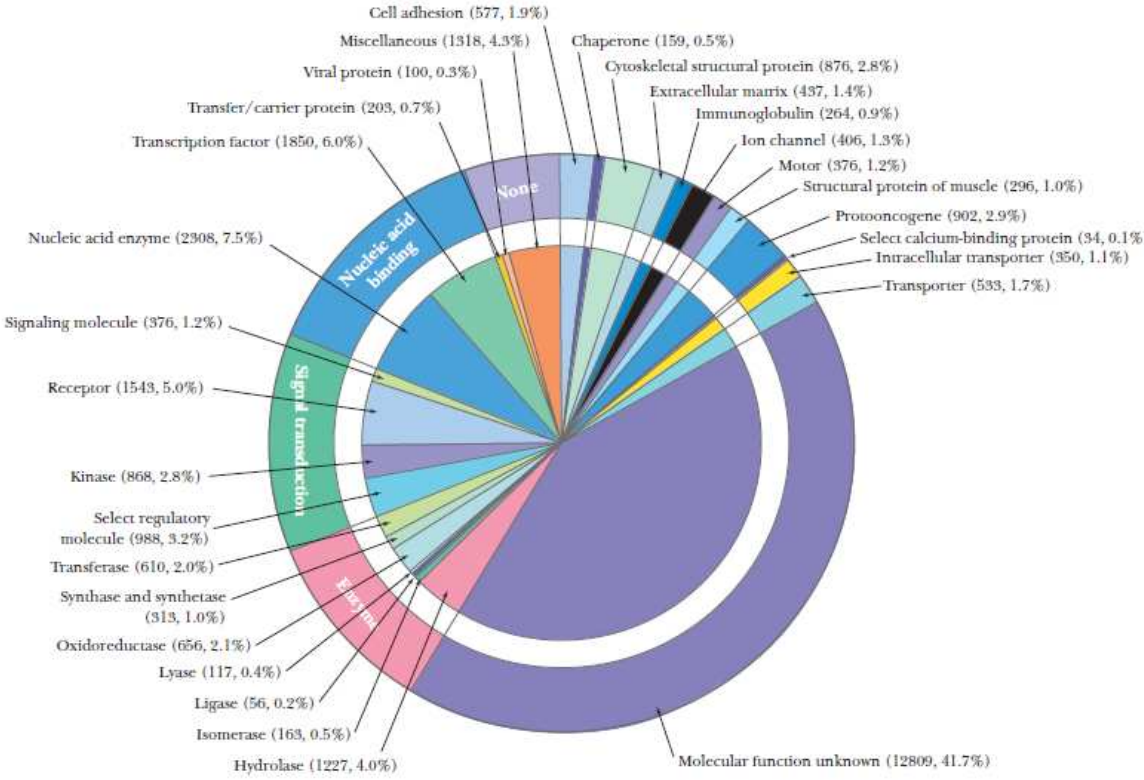
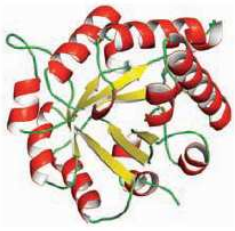
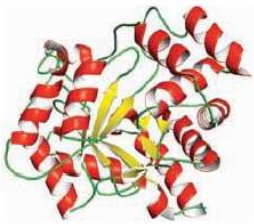


FIGURE 5.26 Proteins of the human genome grouped according to their molecular function. The numbers and percentages within each functional category are enclosed in parentheses. Note that the function of more than 40% of the proteins encoded by the human genome remains unknown. Considering those of known function, enzymes (including kinases and nucleic acid enzymes) account for about 20% of the total number of proteins; nucleic acid–binding proteins of various kinds, about 14%, among which almost half are gene-regulatory proteins (transcription factors). Transport proteins collectively constitute about 5% of the total; and structural proteins, another 5%. (Adapted from Figure 15 in Venter, J. C., et al., 2001. The sequence of the human genome. *Science* **291:**1304–1351.)

(a) Same domain type, different functions:



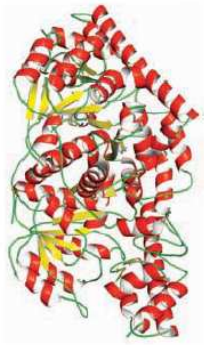
Triose phosphate isomerase



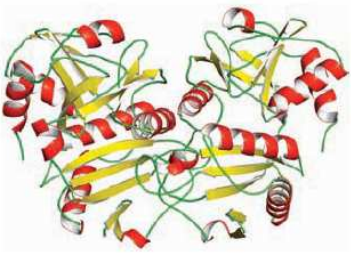
Aldose reductase

Phosphotriesterase

(b) Same function, different structures:



Aspartate aminotransferase



D-amino acid aminotransferase

FIGURE 6.28 **(a)** Some proteins share similar structural features but carry out quite different functions (triose phosphate isomerase, pdb id = 8TIM; aldose reductase, pdb id = 1ADS; phosphotriesterase, pdb id = 1DPM). **(b)** Proteins with quite different structures can carry out similar functions (yeast aspartate aminotransferase, pdb id = 1YAA); D-amino acid aminotransferase, pdb id = 3DAA).

Protein moonlighting (or **gene sharing**) is a phenomenon by which a protein can perform more than one function.

Examples of moonlighting proteins[9]

Kingdom	Protein	Organism	Function	
			primary	moonlighting
Animal				
	Aconitase	*H. sapiens*	TCA cycle enzyme	Iron homeostasis
	ATF2	*H. sapiens*	Transcription factor	DNA damage respons
	Crystallins	Various	Lens structural protein	Various enzyme
	Cytochrome c	Various	Energy metabolism	Apoptosis
	DLD	*H. sapiens*	Energy metabolism	Protease
	ERK2	*H. sapiens*	MAP kinase	Transcriptional repressor
	ESCRT-II complex	*D. melanogaster*	Endosomal protein sorting	Biocoid mRNA localization
	STAT3	*M. musculus*	Transcription factor	Electron transport chain
Plant				
	Hexokinase	*A. thaliana*	Glucose metabolism	Glucose signaling
	Presenilin	*P. patens*	γ-secretase	Cystoskeletal function
Fungus				
	Aconitase	*S. cerevisiae*	TCA cycle enzyme	mtDNA stability
	Aldolase	*S. cerevisiae*	Glycolytic enzyme	V-ATPase assembly
	Arg5,6	*S. cerevisiae*	Arginine biosynthesis	Transcriptional control
	Enolase	*S. cerevisiae*	Glycolytic enzyme	• Homotypic vacuole fusion • Mitochondrial tRNA import
	Galactokinase	*K. lactis*	Galactose catabolism enyzme	Induction galactose genes
	Hal3	*S. cerevisiae*	Halotolerance determinant	Coenzyme A biosynthesis
	HSP60	*S. cerevisiae*	Mitochondrial chaperone	Stabilization active DNA ori's
	Phosphofructokinase	*P. pastoris*	Glycolytic enzyme	Autophagy peroxisomes
	Pyruvate carboxylase	*H. polymorpha*	Anaplerotic enzyme	Assembly of alcohol oxidase
	Vhs3	*S. cerevisiae*	Halotolerance determinant	Coenzyme A biosynthesis
Prokaryotes				
	Aconitase	*M. tuberculosis*	TCA cycle enzyme	Iron-responsive protein
	CYP170A1	*S. coelicolor*	Albaflavenone synthase	Terpene synthase
	Enolase	*S. pneumoniae*	Glycolytic enzyme	Plasminogen binding
	GroEL	*E. aerogenes*	Chaperone	Insect toxin
	Glutamate racemase (MurI)	*E. coli*	cell wall biosynthesis	gyrase inhibition
	Thioredoxin	*E. coli*	Anti-oxidant	T7 DNA polymerase subunit
Protist				
	Aldolase	*P. vivax*	Glycolytic enzyme	Host-cell invasion

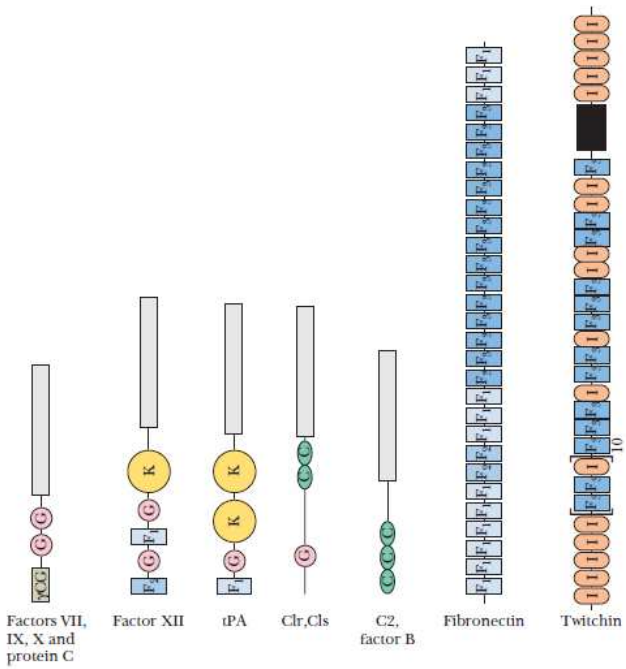
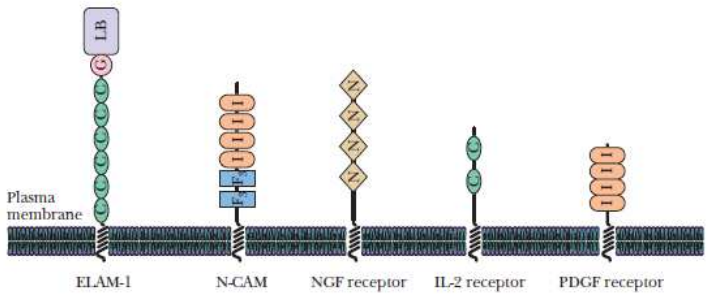


FIGURE 6.26 A sampling of proteins that consist of mosaics of individual protein modules. The modules shown include γCG, a module containing γ-carboxyglutamate residues; G, an epidermal growth factor–like module; K, the "kringle" domain, named for a Danish pastry; C, which is found in complement proteins; F1, F2, and F3, first found in fibronectin; I, the immunoglobulin superfamily domain; N, found in some growth factor receptors; E, a module homologous to the calcium-binding E–F hand domain; and LB, a lectin module found in some cell surface proteins. (Adapted from Baron, M., Norman, D., and Campbell, I. 1991. Protein modules. *Trends in Biochemical Sciences* **16:**13–17.)



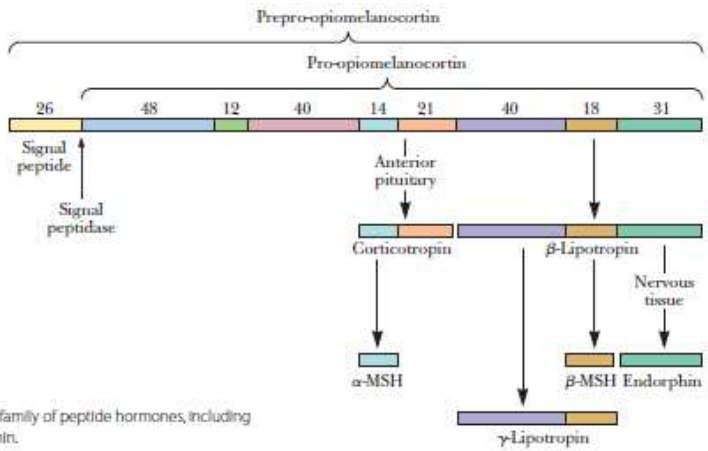
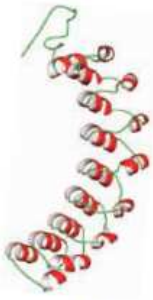


FIGURE 32.3 The conversion of prepro-opiomelanocortin to a family of peptide hormones, including corticotropin, β- and γ-lipotropin, α- and β-MSH, and endorphin.

All α proteins:

Human growth hormone (pdb id = 1HGU)



Leucine-rich repeat variant (pdb id = 1LRV)

Peridinin-chlorophyll protein (a "solenoid"—pdb id = 1PPR)

Endoglucanase A (an α-helical barrel—pdb id = 1CEM)

Cat allergen (pdb id = 1PUO)

All β proteins:

Mannose-specific agglutinin (a prism—(pdb id = 1JPC)

Rieske iron protein (a 3-layer β-sandwich—(pdb id = 1RIE)

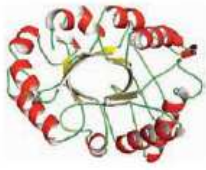
Hemopexin C-terminal domain (a 4-bladed propellor—pdb id = 1HXN)

Lectin from *R. solanacearum* (a 6-bladed propellor—pdb id = 1BT9)

Pleckstrin domain of protein kinase B/AKT (pdb id = 1UNQ)

α/β proteins:

Human bactericidal permeability-increasing protein (pdb id = 1BP1)

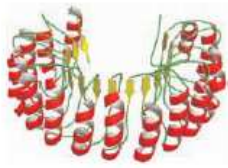


Hevamine (a "TIM barrel" —pdb id = 2HVM)

Hepatocyte growth factor (N-terminal domain —pdb id = 2HGF)

Prokaryotic ribosomal protein L9 (pdb id = 1DIV)

MurA (an α–β prism —pdb id = 1EYN)



Porcine ribonuclease inhibitor (a "horseshoe"—pdb id = 2BNH)

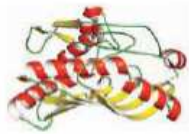
$\alpha+\beta$ proteins:

Equine leucocyte elastase inhibitor (pdb id = 1HLE)

RuvA protein (pdb id = 1CUK)

Ribonuclease H (pdb id = 1RNH)

L-Arginine: glycine amidinotransferase (a metabolic enzyme—pdb id = 4JDW)



Thymidylate synthase (pdb id = 3TMS)

FIGURE 6.39 Four major classes of protein structure (as defined in the SCOP database). **(a) All α proteins,** where α-helices dominate the structure; **(b) All β proteins,** in which β-sheets are the primary feature; **(c) α/β proteins,** where α-helices and β-sheets are mixed within a domain; **(d) $\alpha+\beta$ proteins,** in which α-helical and β-sheet domains are separated to at least some extent.

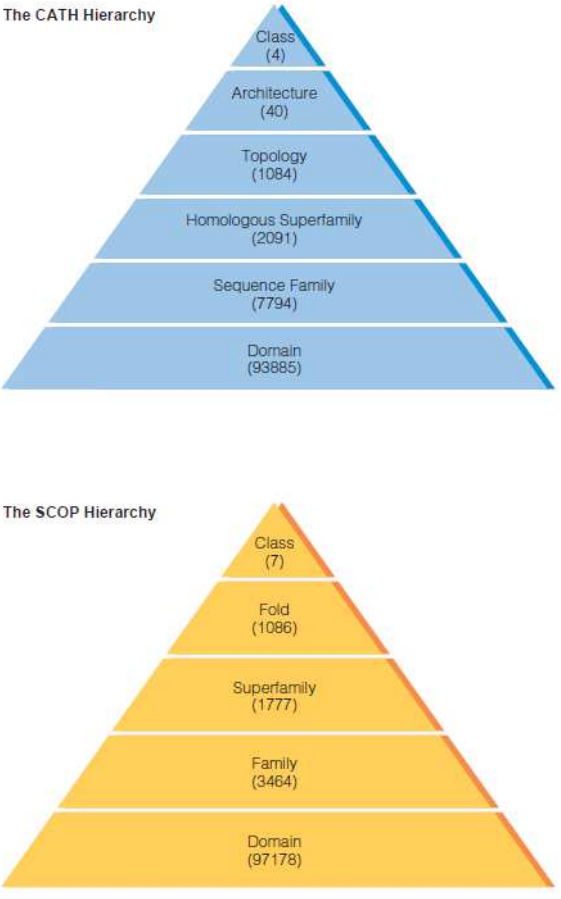


FIGURE 6.27 SCOP and CATH are hierarchical classification systems for the known proteins. Proteins are classified in SCOP by a manual process, whereas CATH combines manual and automated procedures. Numbers indicate the population of each category.

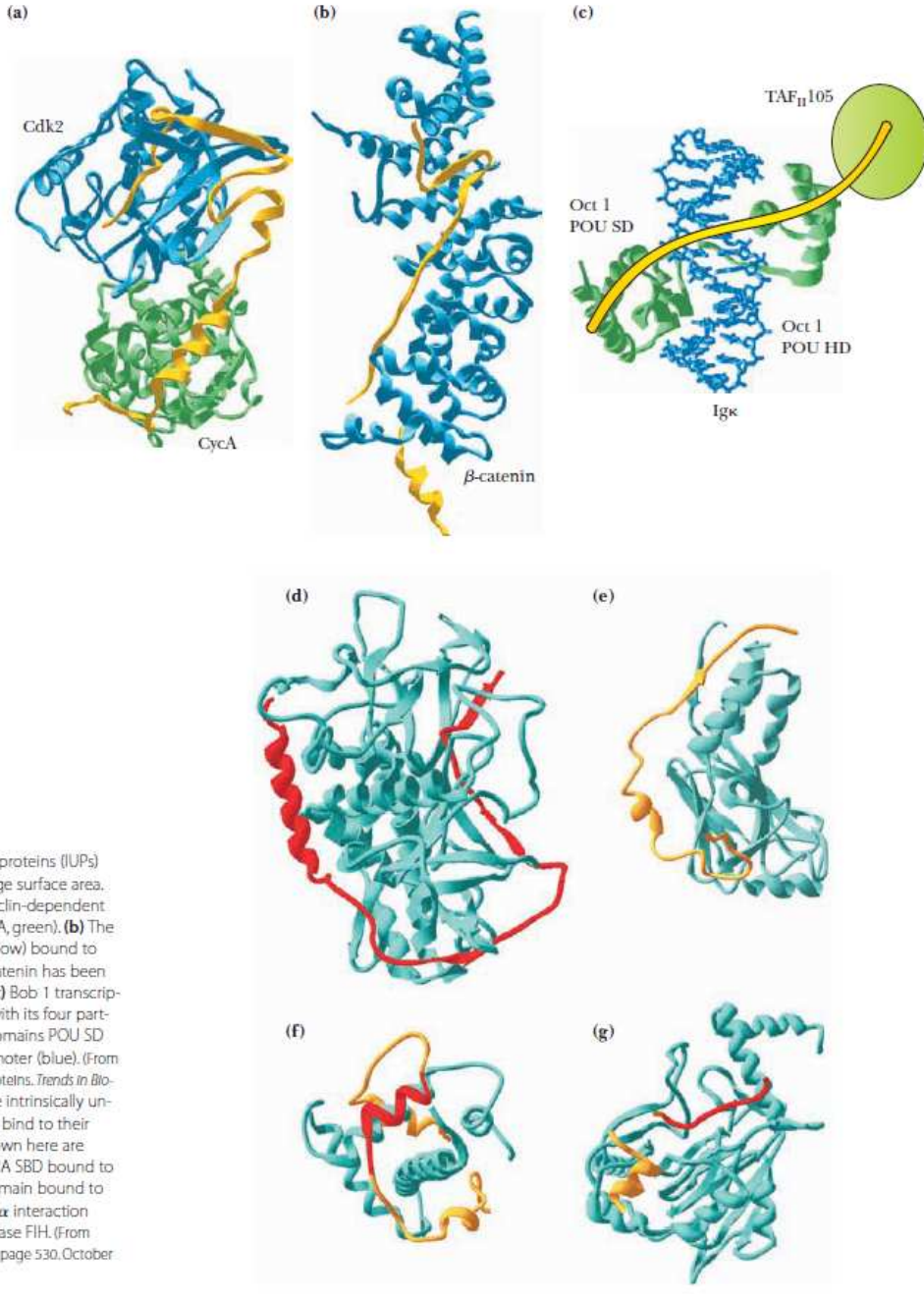


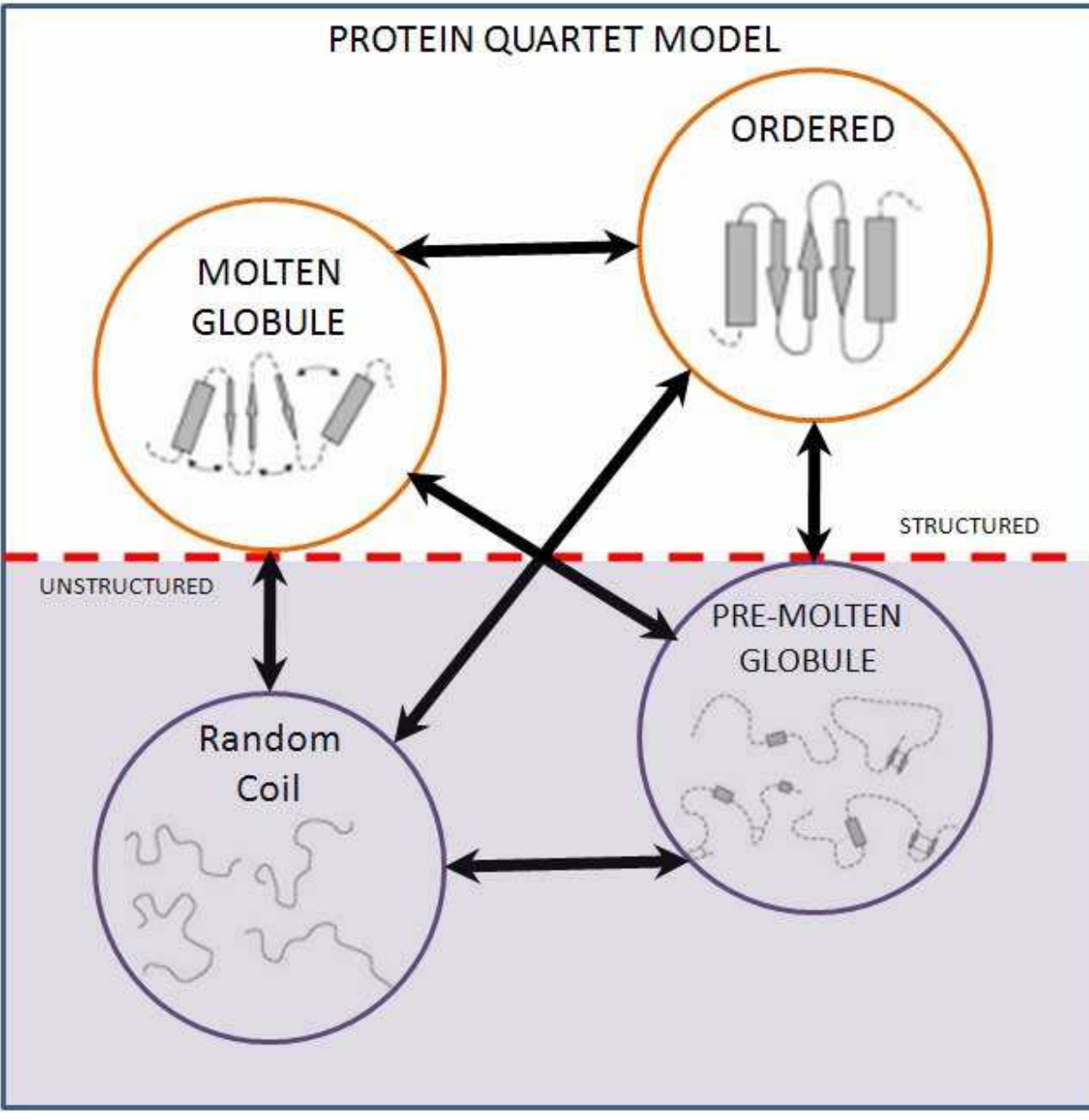
FIGURE 6.40 Intrinsically unstructured proteins (IUPs) contact their target proteins over a large surface area. **(a)** $p27^{Kip1}$ (yellow) complexed with cyclin-dependent kinase 2 (Cdk2, blue) and cyclin A (CycA, green). **(b)** The transactivator domain CBD of Tcf3 (yellow) bound to β-catenin (blue). **Note:** Part of the β-catenin has been removed for a clear view of the CBD. **(c)** Bob 1 transcriptional coactivator (yellow) in contact with its four partners: $TAF_{II}105$ (green oval), the Oct 1 domains POU SD and POU HD (green), and the Igκ promoter (blue). (From Tompa, P., 2002. Intrinsically unstructured proteins. *Trends in Biochemical Sciences* **27:**527–533.) **(d-g)** Some intrinsically unstructured proteins (in red and yellow) bind to their targets by wrapping around them. Shown here are **(d)** SNAP-25 bound to BoNT/A, **(e)** SARA SBD bound to Smad 2 MH2, **(f)** HIF-1α interaction domain bound to the TAZ1 domain of CBP, and **(g)** HIF-1α interaction domain bound to asparagine hydroxylase FIH. (From *Trends in Biochemical Sciences*, Vol. 27, No. 10, page 530. October 2002.)

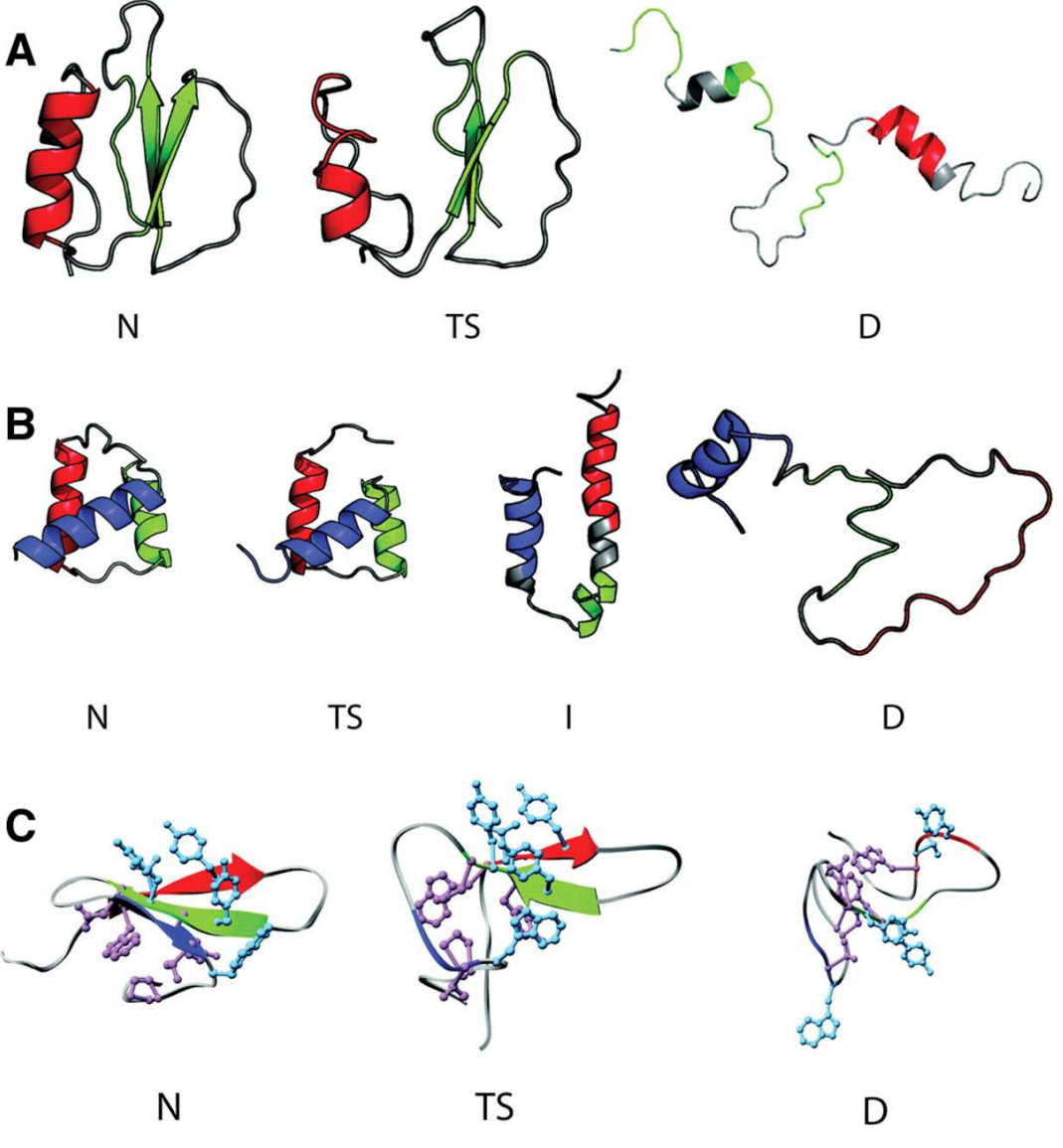
TABLE 3.3 Relative frequencies of amino acid residues in secondary structures

Amino acid	α helix	β sheet	Turn
Ala	**1.29**	0.90	0.78
Cys	**1.11**	0.74	0.80
Leu	**1.30**	1.02	0.59
Met	**1.47**	0.97	0.39
Glu	**1.44**	0.75	1.00
Gln	**1.27**	0.80	0.97
His	**1.22**	1.08	0.69
Lys	**1.23**	0.77	0.96
Val	0.91	**1.49**	0.47
Ile	0.97	**1.45**	0.51
Phe	1.07	**1.32**	0.58
Tyr	0.72	**1.25**	1.05
Trp	0.99	**1.14**	0.75
Thr	0.82	**1.21**	1.03
Gly	0.56	0.92	**1.64**
Ser	0.82	0.95	**1.33**
Asp	1.04	0.72	**1.41**
Asn	0.90	0.76	**1.28**
Pro	0.52	0.64	**1.91**
Arg	0.96	0.99	0.88

Note: The amino acids are grouped according to their preference for α helices (top group), β sheets (second group), or turns (third group). Arginine shows no significant preference for any of the structures.
After T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2d ed. (W. H. Freeman and Company, 1992), p. 256.

http://upload.wikimedia.org/wikipedia/en/0/00/Protein_Dynamics_Cytochrome_C_2NEW_small.gif


PROTEIN QUARTET MODEL
ORDERED
MOLTEN GLOBULE
STRUCTURED
UNSTRUCTURED
PRE-MOLTEN GLOBULE
Random Coil


A
N
TS
D
B
N
TS
I
D
C
N
TS
D

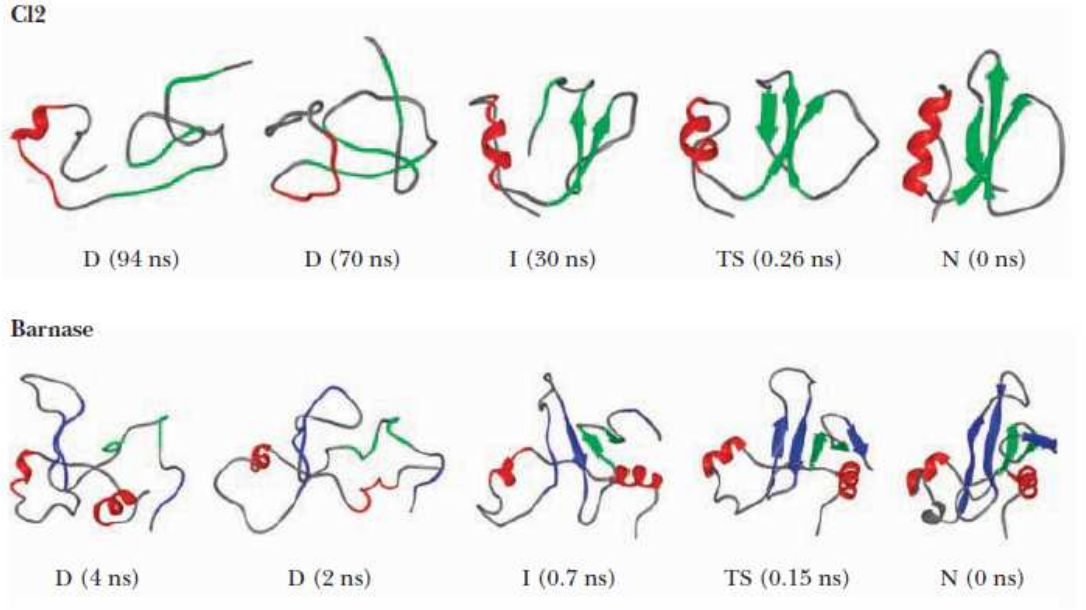


FIGURE 6.33 Computer simulations of folding and unfolding of proteins can reveal possible folding pathways. Molecular dynamics simulations of the unfolding of small proteins such as chymotrypsin inhibitor 2 (CI2) and barnase are presented here on a reversed time scale, to show how folding may occur. D = denatured, I = intermediate, TS = transition state, N = native. (Adapted from Daggett, V., and Fersht, A. R., 2003. Is there a unifying mechanism for protein folding? *Trends in Biochemical Sciences* **28:**18-25. Figures provided by Alan Fersht and Valerie Daggett.)

FIGURE 6.34 A model for the steps involved in the folding of globular proteins. The funnel represents a free energy surface or energy landscape for the folding process. The protein folding process is highly cooperative. Rapid and reversible formation of local secondary structures is followed by a slower phase in which establishment of partially folded intermediates leads to the final tertiary structure. Substantial exclusion of water occurs very early in the folding process.

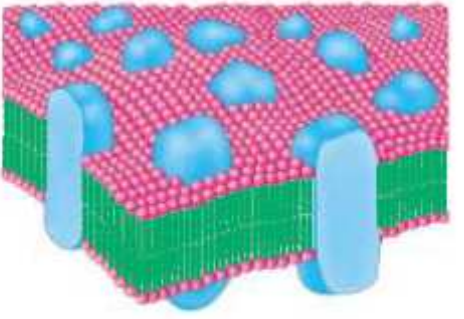


FIGURE 12.30 Fluid mosaic model. (After S. J. Singer and G. L. Nicolson, *Science* 175(1972):723.)

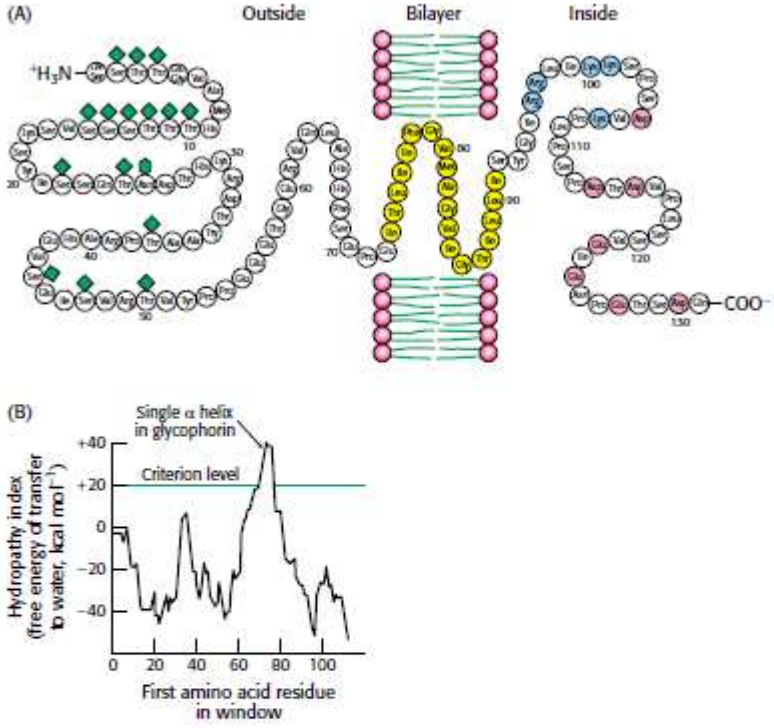


FIGURE 12.27 Locating the membrane-spanning helix of glycophorin. (A) Amino acid sequence and transmembrane disposition of glycophorin A from the red-cell membrane. Fifteen *O*-linked carbohydrate units are shown as diamond shapes, and an *N*-linked unit is shown as a lozenge shape. The hydrophobic residues (yellow) buried in the bilayer form a transmembrane α helix. The carboxyl-terminal part of the molecule, located on the cytosolic side of the membrane, is rich in negatively charged (red) and positively charged (blue) residues. (B) Hydropathy plot for glycophorin. The free energy for transfering a helix of 20 residues from the membrane to water is plotted as a function of the position of the first residue of the helix in the sequence of the protein. Peaks of greater than +20 kcal mol^{-1} in hydropathy plots are indicative of potential transmembrane helices. [(A) Courtesy of Dr. Vincent Marchesi; (B) after D. M. Engelman, T. A. Steitz, and A. Goldman, Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu. Rev. Biophys. Biophys. Chem.* 15(1986):343.]

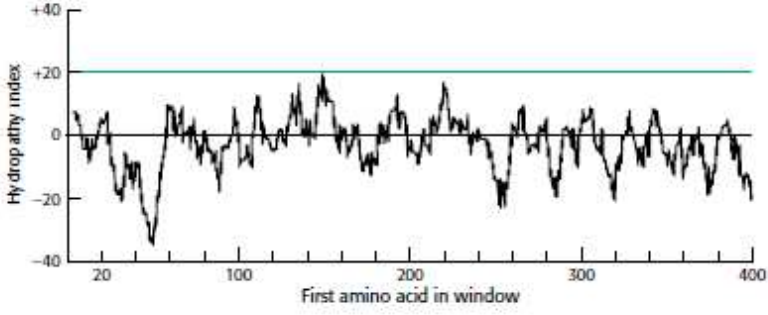


FIGURE 12.28 Hydropathy plot for porin. No strong peaks are observed for this intrinsic membrane protein because it is constructed from membrane-spanning β strands rather than α helices.

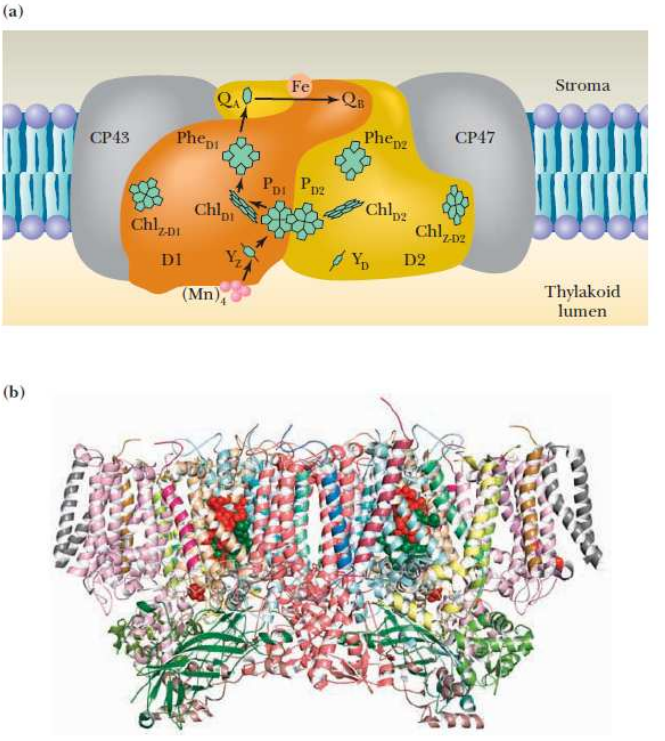


FIGURE 21.16 Molecular architecture of the *Synechococcus elongatus* PSII dimer. **(a)** The arrow shows the path of electron transfer from P680* to Chl_{D1} to $Pheo_{D1}$ to Q_A on D2 and then, via the Fe atom, to Q_B on D1. The Tyr^{161} residue of D1, symbolized by Y_z, is situated between P680 and the $(Mn)_4$ cluster. **(b)** Structure of *S. elongatus* PSII (pdb id = 1S5L). Chlorophylls of the reaction center and electron transfer path are shown in green; pheophytins, in blue. The OECs are shown in brick red.
(Adapted from Barber, J., 2003. Photosystem II: The engine of life. *Quarterly Review of Biophysics* **36:**71–89.)

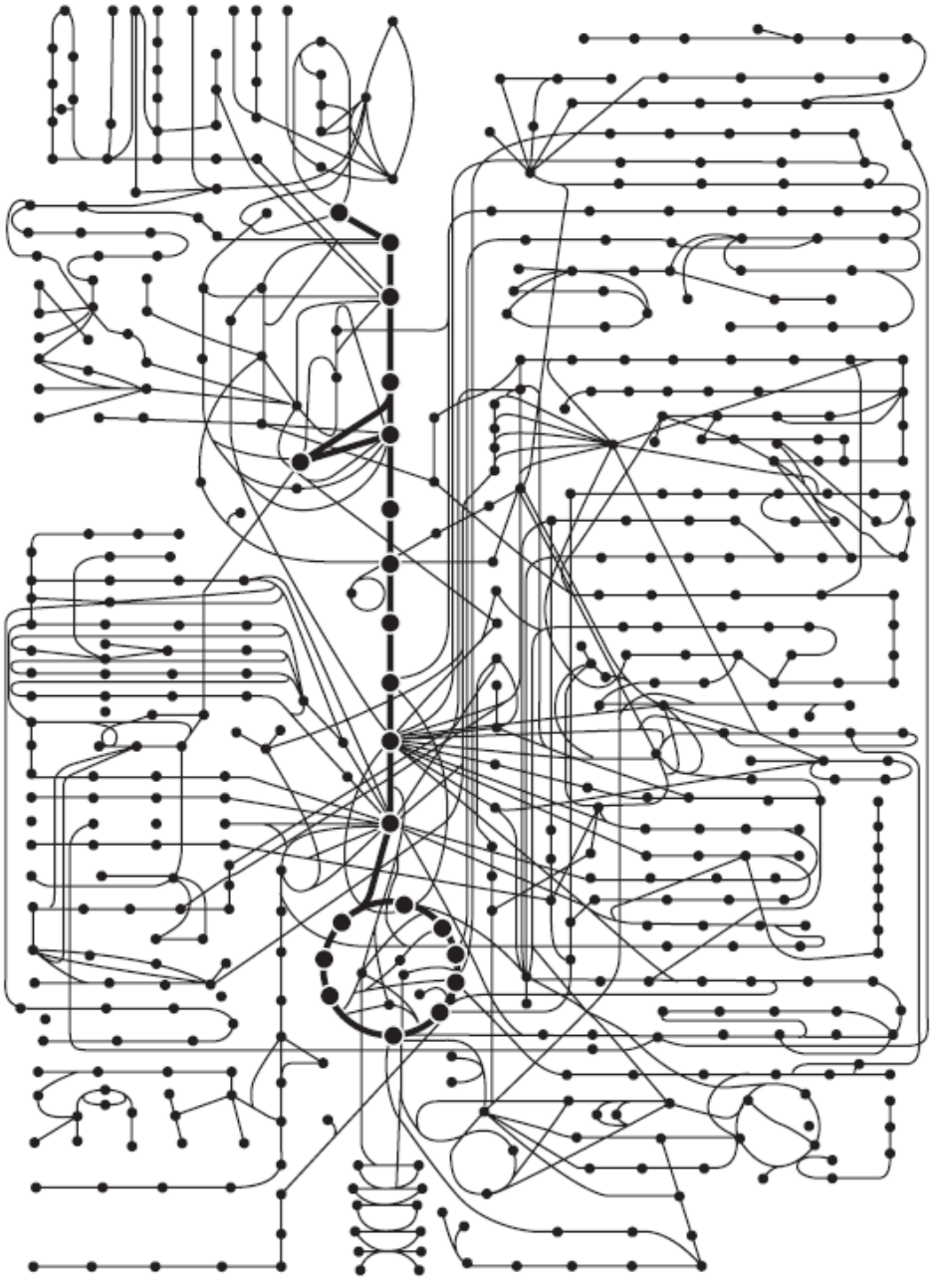
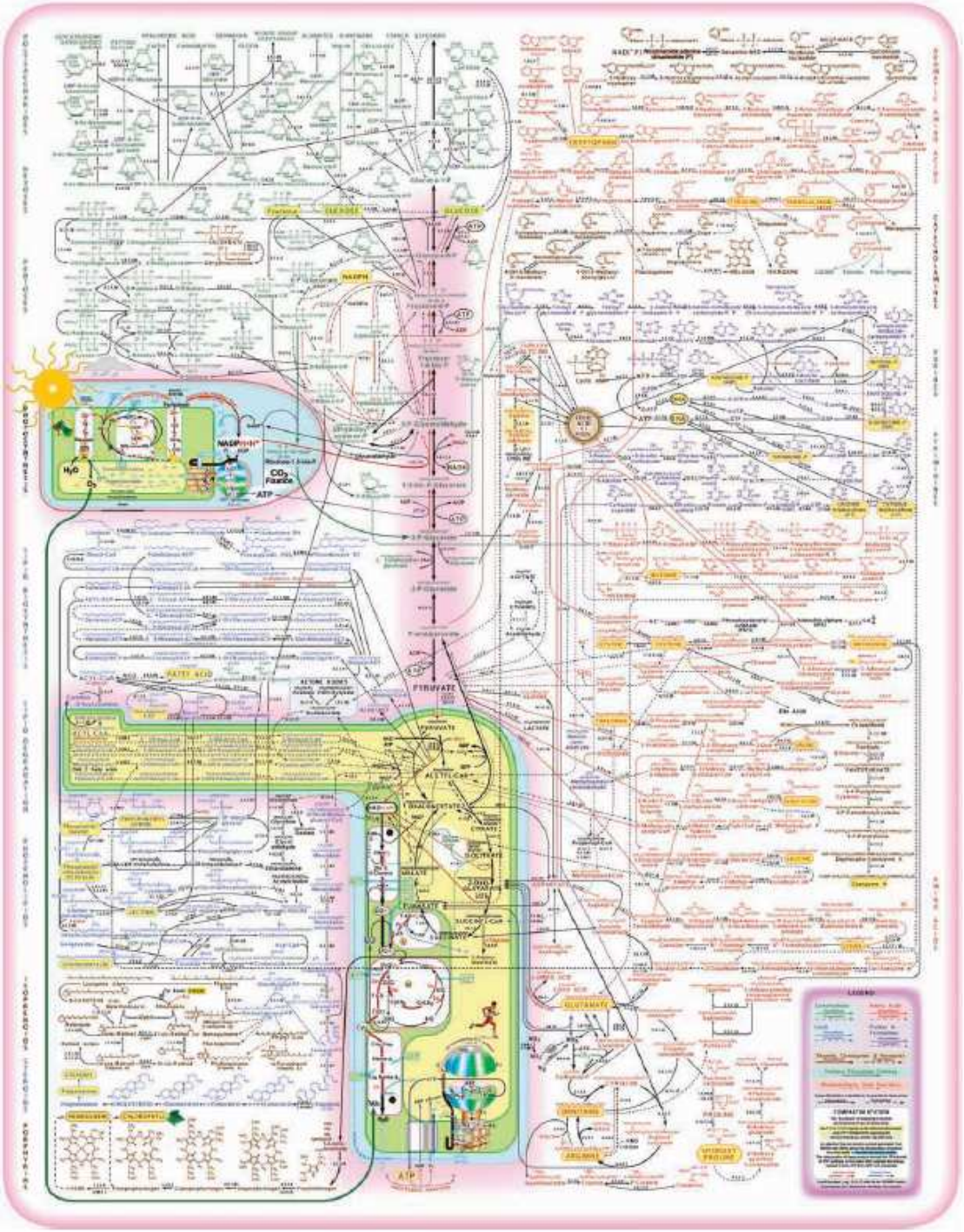


FIGURE 17.3 The metabolic map as a set of dots and lines. The heavy dots and lines trace the central energy-releasing pathways known as glycolysis and the citric acid cycle. (Adapted from Alberts, B., et al., 1989. *Molecular Biology of the Cell*, 2nd ed. New York: Garland Publishing Co.)



CENGAGENOW™ **ANIMATED FIGURE 17.2** A metabolic map, indicating the reactions of intermediary metabolism and the enzymes that catalyze them. More than 500 different chemical intermediates, or metabolites, and a greater number of enzymes are represented here. (Source: From Donald Nicholson, Map #22, © International Union of Biochemistry and Molecular Biology) **See this figure animated at www.cengage.com/login.**