

intragenic complementation *Proc. Natl. Acad. Sci. USA* 94: 9063-9068. ([PubMed](#)) ([Full Text in PMC](#))

Amino acid degradation

F. Fusetti, H. Erlandsen, T. Flatmark, and R.C. Stevens. 1998. Structure of tetrameric human phenylalanine hydroxylase and its implications for phenylketonuria *J. Biol. Chem.* 273: 16962-16967. ([PubMed](#))

K. Sugimoto, T. Senda, H. Aoshima, E. Masai, M. Fukuda, and Y. Mitsui. 1999. Crystal structure of an aromatic ring opening dioxygenase LigAB, a protocatechuate 4,5-dioxygenase, under aerobic conditions *Structure Fold Des.* 7: 953-965. ([PubMed](#))

G.P. Titus, H.A. Mueller, J. Burgner, S. Rodriguez De Cordoba, M.A. Penalva, and D.E. Timm. 2000. Crystal structure of human homogentisate dioxygenase *Nat. Struct. Biol.* 7: 542-546. ([PubMed](#))

H. Erlandsen and R.C. Stevens. 1999. The structural basis of phenylketonuria *Mol. Genet. Metab.* 68: 103-125. ([PubMed](#))

Genetic diseases

Striver, C. R., Beaudet, A. L., Sly, W. S., Valle, D., Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S. (Eds.), 1995. *The Metabolic Basis of Inherited Diseases* (7th ed.). McGraw-Hill.

Nyhan, W. L. (Ed.), 1984. *Abnormalities in Amino Acid Metabolism in Clinical Medicine*. Appleton-Century-Crofts.

Historical aspects and the process of discovery

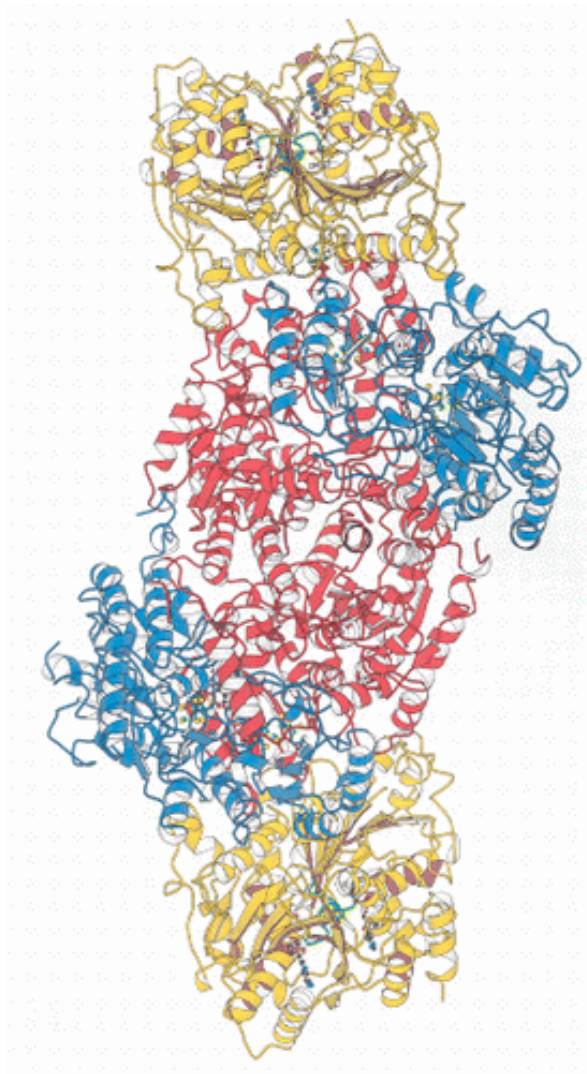
A.J.L. Cooper and A. Meister. 1989. An appreciation of Professor Alexander E. Braunstein: The discovery and scope of enzymatic transamination *Biochimie* 71: 387-404. ([PubMed](#))

Garrod, A. E., 1909. *Inborn Errors in Metabolism*. Oxford University Press (reprinted in 1963 with a supplement by H. Harris).

B. Childs. 1970. Sir Archibald Garrod's conception of chemical individuality: A modern appreciation *N. Engl. J. Med.* 282: 71-78. ([PubMed](#))

F.L. Holmes. 1980. Hans Krebs and the discovery of the ornithine cycle *Fed. Proc.* 39: 216-225. ([PubMed](#))

III. Synthesizing the Molecules of Life



Nitrogenase complex. Nitrogen is an essential component of many biochemical building blocks. The enzyme complex shown here converts nitrogen gas, an abundant but inert compound, into a form that can be used for synthesizing amino acids, nucleotides, and other biochemicals.

24. The Biosynthesis of Amino Acids

The assembly of biological molecules, including proteins and nucleic acids, requires the generation of appropriate starting materials. We have already considered the assembly of carbohydrates in regard to the Calvin cycle and the pentose phosphate pathway ([Chapter 20](#)). The present chapter and the next two examine the assembly of the other important building blocks—namely, amino acids, nucleotides, and lipids.

The pathways for the biosynthesis of these molecules are extremely ancient, going back to the last common ancestor of all living things. Indeed, these pathways probably predate many of the pathways of energy transduction discussed in Part II and may have provided key selective advantages in early evolution. Many of the intermediates present in energy-transduction pathways play a role in biosynthesis as well. These common intermediates allow efficient interplay between energy-transduction (catabolic) and biosynthetic (anabolic) pathways. Thus, cells are able to balance the degradation of compounds for energy mobilization and the synthesis of starting materials for macro-molecular construction.

We begin our consideration of biosynthesis with amino acids. Amino acids are the building blocks of proteins and the nitrogen source for many other important molecules, including nucleotides, neurotransmitters, and prosthetic groups

such as porphyrins.

24.0.1. An Overview of Amino Acid Synthesis

A fundamental problem for biological systems is to obtain nitrogen in an easily usable form. This problem is solved by certain microorganisms capable of reducing the inert $\text{N}\equiv\text{N}$ molecule (nitrogen gas) to two molecules of ammonia in one of the most remarkable reactions in biochemistry. Nitrogen in the form of ammonia is the source of nitrogen for all the amino acids. The carbon backbones come from the glycolytic pathway, the pentose phosphate pathway, or the citric acid cycle.

Anabolism

Biosynthetic processes.

Degradative processes.

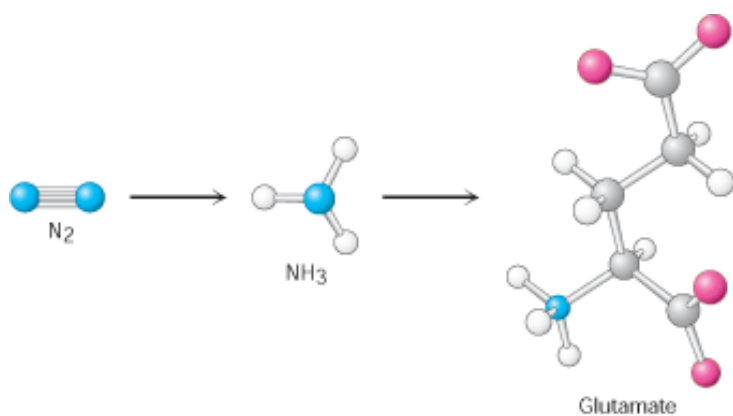
Derived from the Greek *ana*, "up"; *kata*, "down"; *ballein*, "to throw."

In amino acid production, we encounter an important problem in biosynthesis—namely, stereochemical control. Because all amino acids except glycine are chiral, biosynthetic pathways must generate the correct isomer with high fidelity. In each of the 19 pathways for the generation of chiral amino acids, the stereochemistry at the α -carbon atom is established by a transamination reaction that involves pyridoxal phosphate. Almost all the transaminases that catalyze these reactions descend from a common ancestor, illustrating once again that effective solutions to biochemical problems are retained throughout evolution (Figure 24.1).

Biosynthetic pathways are often highly regulated such that building blocks are synthesized only when supplies are low. Very often, a high concentration of the final product of a pathway inhibits the activity of enzymes that function early in the pathway. Often present are allosteric enzymes capable of sensing and responding to concentrations of regulatory species. These enzymes are similar in functional properties to aspartate transcarbamylase and its regulators (Section 10.1). Feedback and allosteric mechanisms ensure that all twenty amino acids are maintained in sufficient amounts for protein synthesis and other processes.



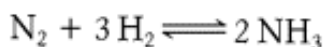
Figure 24.1. Aminotransferase Protein Family. A superposition of the structures of four aminotransferases taking part in amino acid biosynthesis reveals a common fold, indicating that these proteins descended from a common ancestor.



Nitrogen is a key component of amino acids. The atmosphere is rich in nitrogen gas (N_2), a very unreactive molecule. Certain organisms such as bacteria that live in the root nodules of yellow clover (photo at left) can convert nitrogen gas into ammonia. Ammonia can then be used to synthesize first glutamate and then other amino acids. [(Left) Runk/Schoenberger from Grant Heilman]

24.1. Nitrogen Fixation: Microorganisms Use ATP and a Powerful Reductant to Reduce Atmospheric Nitrogen to Ammonia

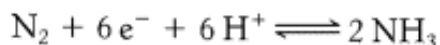
The nitrogen in amino acids, purines, pyrimidines, and other biomolecules ultimately comes from atmospheric nitrogen, N_2 . The biosynthetic process starts with the reduction of N_2 to NH_3 (ammonia), a process called *nitrogen fixation*. Although higher organisms are unable to fix nitrogen, this conversion is carried out by some bacteria and archaea. Symbiotic *Rhizobium* bacteria invade the roots of leguminous plants and form root nodules in which they fix nitrogen, supplying both the bacteria and the plants. The amount of N_2 fixed by *diazotrophic (nitrogen-fixing) microorganisms* has been estimated to be 10^{11} kilograms per year, about 60% of Earth's newly fixed nitrogen. Lightning and ultraviolet radiation fix another 15%; the other 25% is fixed by industrial processes. The industrial process for nitrogen fixation devised by Fritz Haber in 1910 is still being used in fertilizer factories.



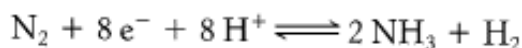
The fixation of N_2 is typically carried out by mixing with H_2 gas over an iron catalyst at about 500°C and a pressure of 300 atmospheres. The extremely strong $\text{N}\equiv\text{N}$ bond, which has a bond energy of $225 \text{ kcal mol}^{-1}$, is highly resistant to chemical attack. Indeed, Lavoisier named nitrogen gas "azote," meaning "without life" because it is so unreactive. Nevertheless, the conversion of nitrogen and hydrogen to form ammonia is thermodynamically favorable; the reaction is difficult kinetically because intermediates along the reaction pathway are unstable.

To meet the kinetic challenge, the biological process of nitrogen fixation requires a complex enzyme with multiple redox centers. The *nitrogenase complex*, which carries out this fundamental transformation, consists of two proteins: a *reductase*, which provides electrons with high reducing power, and *nitrogenase*, which uses these electrons to reduce N_2 to NH_3 . The transfer of electrons from the reductase to the nitrogenase component is coupled to the hydrolysis of ATP by the reductase (Figure 24.2). The nitrogenase complex is exquisitely sensitive to inactivation by O_2 . Leguminous plants maintain a very low concentration of free O_2 in their root nodules by binding O_2 to *leghemoglobin*, a homolog of hemoglobin (Section 7.3.1).

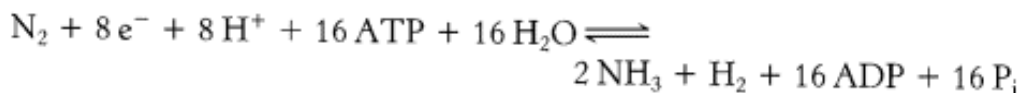
In principle, the reduction of N_2 to NH_3 is a six-electron process.



However, the biological reaction always generates at least 1 mol of H_2 in addition to 2 mol of NH_3 for each mole of $\text{N}\equiv\text{N}$. Hence, an input of two additional electrons is required.




In most nitrogen-fixing microorganisms, *the eight high-potential electrons come from reduced ferredoxin*, generated by photosynthesis or oxidative processes. Two molecules of ATP are hydrolyzed for each electron transferred. Thus, *at least 16 molecules of ATP are hydrolyzed for each molecule of N_2 reduced*.



Again, ATP hydrolysis is not required to make nitrogen reduction favorable thermodynamically. Rather, it is essential to reduce the heights of activation barriers along the reaction pathway, thus making the reaction kinetically feasible.

24.1.1. The Iron-Molybdenum Cofactor of Nitrogenase Binds and Reduces Atmospheric Nitrogen

Both the reductase and the nitrogenase components of the complex are *iron-sulfur proteins*, in which iron is bonded to the sulfur atom of a cysteine residue and to inorganic sulfide. The *reductase* (also called the *iron protein* or the *Fe protein*) is a dimer of identical 30-kd subunits bridged by a 4Fe-4S cluster (Figure 24.3).

 The role of the reductase is to transfer electrons from a suitable donor, such as reduced ferredoxin, to the nitrogenase component. The binding and hydrolysis of ATP triggers a conformational change that moves the reductase closer to the nitrogenase component from whence it is able to transfer its electron to the center of nitrogen reduction. The structure of the ATP-binding region reveals it to be a member of the P-loop NTPase family (Section 9.4.1) that is clearly related to the nucleotide-binding regions found in G proteins and related proteins (Section 15.1.2).

Thus, we see another example of how this domain has been recruited in evolution because of its ability to couple nucleoside triphosphate hydrolysis to conformational changes.

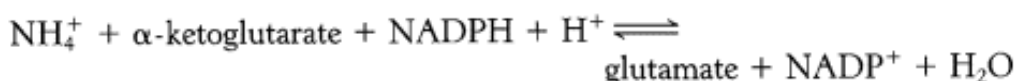
The nitrogenase component is an $\alpha_2 \beta_2$ tetramer (240 kd), in which the α and β subunits are homologous to each other and structurally quite similar (Figure 24.4). Electrons enter at the *P clusters*, which are located at the α - β interface. These clusters are each composed of eight iron atoms and seven sulfide ions. In the reduced form, each cluster takes the form of two 4Fe-3S partial cubes linked by a central sulfide ion. Each cluster is linked to the protein through six cysteinate residues. Electrons flow from the P cluster to the *FeMo cofactor*, a very unusual redox center. Because molybdenum is present in this cluster, the nitrogenase component is also called the *molybdenum-iron protein (MoFe protein)*. The FeMo cofactor consists of two M-3Fe-3S clusters, in which molybdenum occupies the M site in one cluster and iron occupies it in the other. The two clusters are joined by three sulfide ions. The FeMo cofactor is also coordinated to a homocitrate moiety and to the α subunit through one histidine residue and one cysteinate residue. This cofactor is distinct from the molybdenum-containing cofactor found in sulfite oxidase and apparently all other molybdenum-containing enzymes except nitrogenase.

The FeMo cofactor is the site of nitrogen fixation. Note that each of the six central iron atoms is linked to only three atoms, leaving open a binding opportunity for N_2 . It seems likely that N_2 binds in the central cavity of this cofactor (Figure 24.5). The formation of multiple Fe-N interactions in this complex weakens the $N\equiv N$ bond and thereby lowers the activation barrier for reduction.

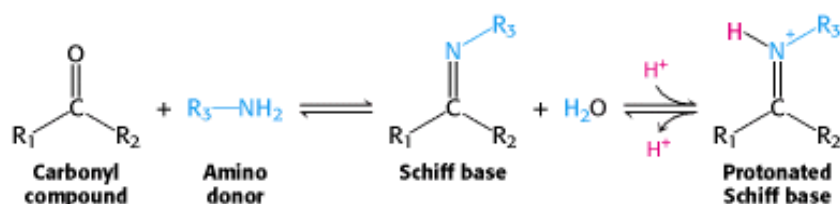
24.1.2. Ammonium Ion Is Assimilated into an Amino Acid Through Glutamate and Glutamine

The next step in the assimilation of nitrogen into biomolecules is the entry of NH_4^+ into amino acids. *Glutamate* and *glutamine* play pivotal roles in this regard. The α -amino group of most amino acids comes from the α -amino group of glutamate by transamination (Section 23.3.1). Glutamine, the other major nitrogen donor, contributes its side-chain nitrogen atom in the biosynthesis of a wide range of important compounds, including the amino acids tryptophan and histidine.

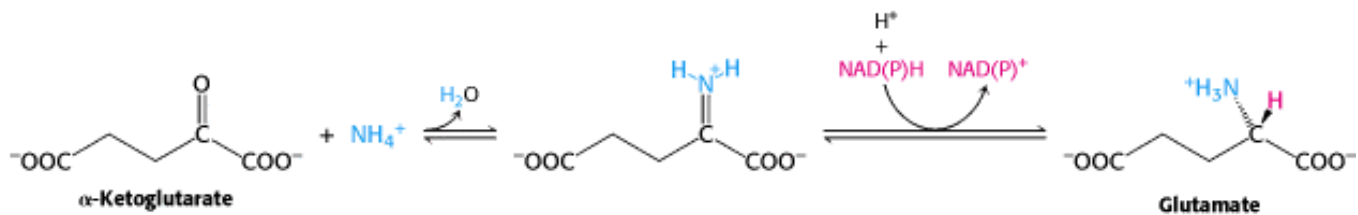
Glutamate is synthesized from NH_4^+ and α -ketoglutarate, a citric acid cycle intermediate, by the action of *glutamate dehydrogenase*. We have already encountered this enzyme in the degradation of amino acids (Section 23.3.1). Recall that NAD^+ is the oxidant in catabolism, whereas $NADPH$ is the reductant in biosyntheses. Glutamate dehydrogenase is unusual in that it does not discriminate between $NADH$ and $NADPH$, at least in some species.



The reaction proceeds in two steps. First, a Schiff base forms between ammonia and α -ketoglutarate. The formation of a Schiff base between an amine and a carbonyl compound is a key reaction that takes place at many stages of amino acid biosynthesis and degradation.

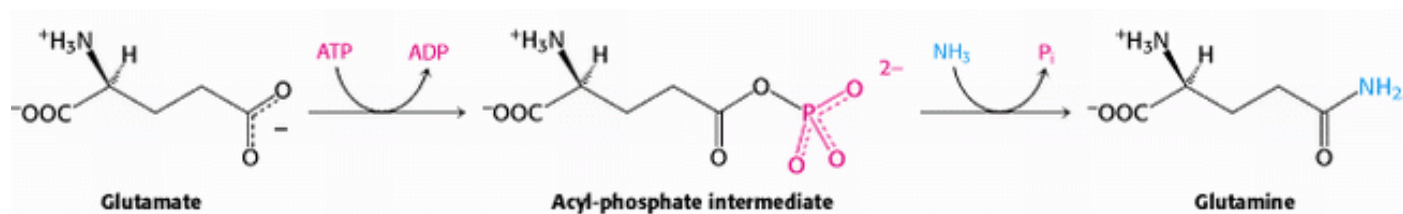


Schiff bases can be easily protonated. With glutamate dehydrogenase, the protonated Schiff base is reduced by the transfer of a hydride ion from $NADPH$ to form glutamate.



This reaction is crucial because it establishes the stereochemistry of the α -carbon atom (*S* absolute configuration) in glutamate. The enzyme binds the α -ketoglutarate substrate in such a way that hydride transferred from NAD(P)H is added to form the *l* isomer of glutamate (Figure 24.6). As we shall see, this stereochemistry is established for other amino acids by transamination reactions that rely on pyridoxal phosphate.

A second ammonium ion is incorporated into glutamate to form glutamine by the action of *glutamine synthetase*. This amidation is driven by the hydrolysis of ATP. ATP participates directly in the reaction by phosphorylating the side chain of glutamate to form an acyl-phosphate intermediate, which then reacts with ammonia to form glutamine.

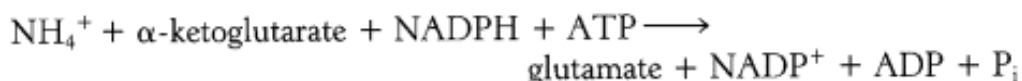


A high-affinity ammonia-binding site is formed only after the formation of the acyl-phosphate intermediate. A specific site for ammonia binding is required to prevent attack by water from hydrolyzing the intermediate and wasting a molecule of ATP. The regulation of glutamine synthetase plays a critical role in controlling nitrogen metabolism (Section 24.3.2).

Glutamate dehydrogenase and glutamine synthetase are present in all organisms. Most prokaryotes also contain an evolutionarily unrelated enzyme, *glutamate synthase*, which catalyzes the reductive amination of α -ketoglutarate with the use of glutamine as the nitrogen donor.



The side-chain amide of glutamine is hydrolyzed to generate ammonia within the enzyme, a recurring theme throughout nitrogen metabolism. When NH_4^+ is limiting, most of the glutamate is made by the sequential action of *glutamine synthetase* and *glutamate synthase*. The sum of these reactions is



Note that this stoichiometry differs from that of the glutamate dehydrogenase reaction in that ATP is hydrolyzed. Why do prokaryotes sometimes use this more expensive pathway? The answer is that the value of K_M of glutamate dehydrogenase for NH_4^+ is high ($\approx 1 \text{ mM}$), and so this enzyme is not saturated when NH_4^+ is limiting. In contrast, glutamine synthetase has very high affinity for NH_4^+ . Thus, ATP hydrolysis is required to capture ammonia when it is scarce.

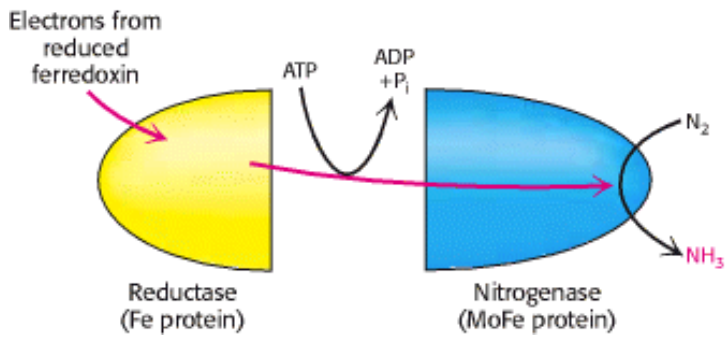


Figure 24.2. Nitrogen Fixation. Electrons flow from ferredoxin to the reductase (iron protein, or Fe protein) to nitrogenase (molybdenum-iron protein, or MoFe protein) to reduce nitrogen to ammonia. ATP hydrolysis within the reductase drives conformational changes necessary for the efficient transfer of electrons.

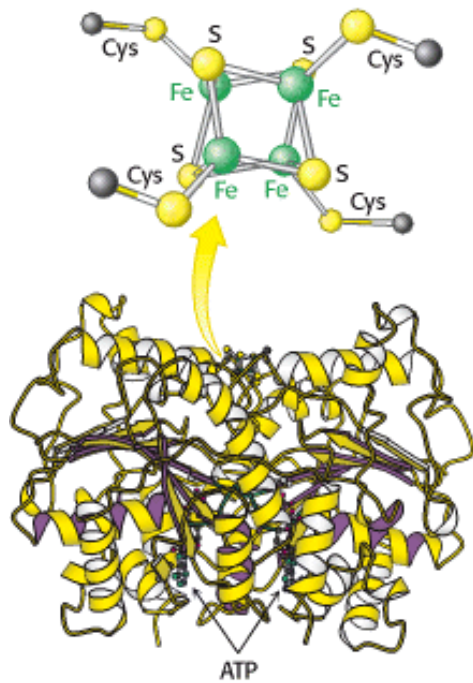


Figure 24.3. Fe Protein. This protein is a dimer composed of two polypeptide chains linked by a 4Fe-4S cluster. Each monomer is a member of the P-loop NTPase family and contains an ATP-binding site.

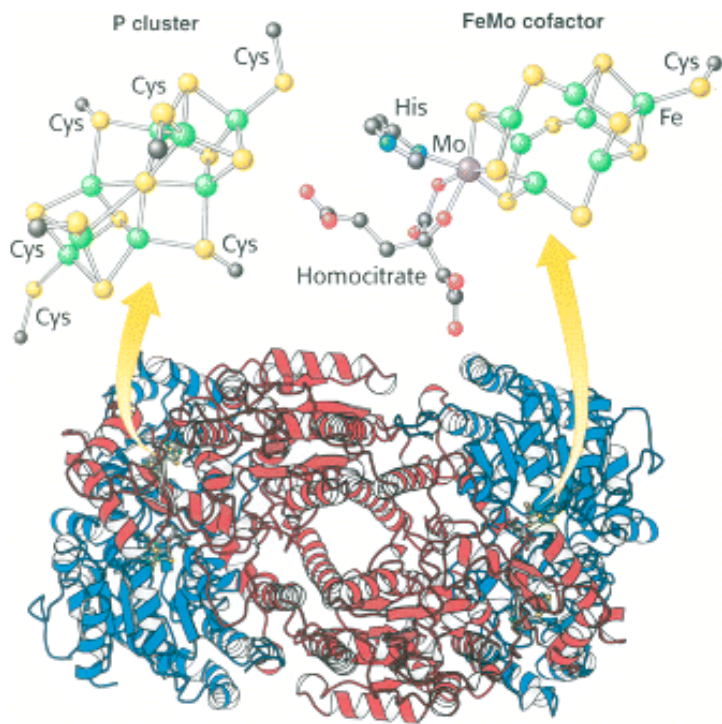


Figure 24.4. MOFe Protein. This protein is a heterotetramer composed of two α subunits (red) and two β subunits (blue). The protein contains two copies each of two types of clusters: P clusters and FeMo cofactors. Each P cluster contains eight iron atoms and seven sulfides linked to the protein by six cysteinate residues. Each FeMo cofactor contains one molybdenum atom, seven iron atoms, nine sulfides, and a homocitrate, and is linked to the protein by one cysteinate residue and one histidine residue.

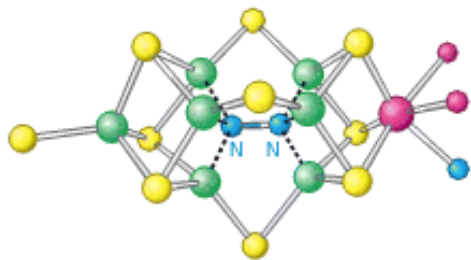


Figure 24.5. Nitrogen-Reduction Site. The FeMo cofactor contains an open center that is the likely site of nitrogen binding and reduction.

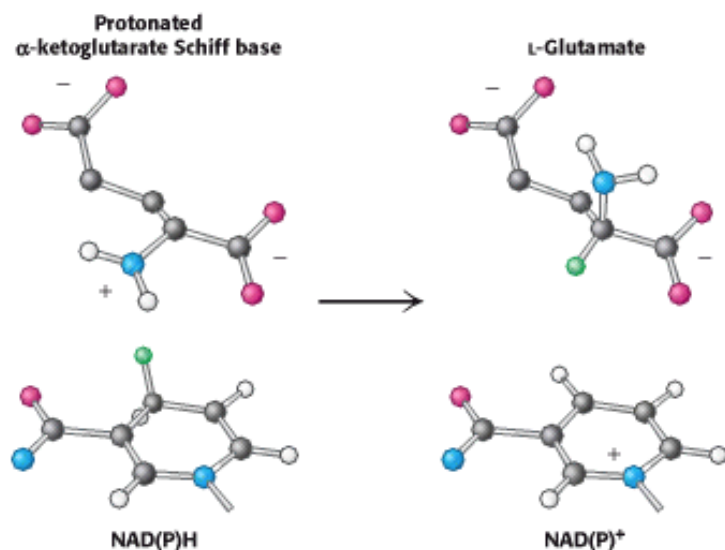


Figure 24.6. Establishment of Chirality. In the active site of glutamate dehydrogenase, hydride transfer from NAD(P)H to a specific face of the achiral protonated Schiff base of α -ketoglutarate establishes the L configuration of glutamate.

24.2. Amino Acids Are Made from Intermediates of the Citric Acid Cycle and Other Major Pathways

Thus far, we have considered the conversion of N_2 into NH_4^+ and the assimilation of NH_4^+ into glutamate and glutamine. We turn now to the biosynthesis of the other amino acids. The pathways for the biosynthesis of amino acids are diverse. However, they have an important common feature: *their carbon skeletons come from intermediates of glycolysis, the pentose phosphate pathway, or the citric acid cycle.* On the basis of these starting materials, amino acids can be grouped into six biosynthetic families (Figure 24.7).

24.2.1. Human Beings Can Synthesize Some Amino Acids but Must Obtain Others from the Diet

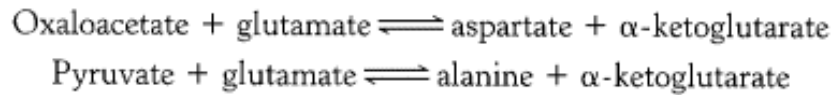
Most microorganisms such as *E. coli* can synthesize the entire basic set of 20 amino acids, whereas human beings cannot make 9 of them. The amino acids that must be supplied in the diet are called *essential amino acids*, whereas the others are termed *nonessential amino acids* (Table 24.1). These designations refer to the needs of an organism under a particular set of conditions. For example, enough arginine is synthesized by the urea cycle to meet the needs of an adult but perhaps not those of a growing child. A deficiency of even one amino acid results in a *negative nitrogen balance*. In this state, more protein is degraded than is synthesized, and so more nitrogen is excreted than is ingested.

The nonessential amino acids are synthesized by quite simple reactions, whereas the pathways for the formation of the essential amino acids are quite complex. For example, the nonessential amino acids *alanine* and *aspartate* are synthesized in a single step from pyruvate and oxaloacetate, respectively. In contrast, the pathways for the essential amino acids require from 5 to 16 steps (Figure 24.8). The sole exception to this pattern is arginine, inasmuch as the synthesis of this nonessential amino acid de novo requires 10 steps. Typically, though, it is made in only 3 steps from ornithine as part of the urea cycle. Tyrosine, classified as a nonessential amino acid because it can be synthesized in 1 step from phenylalanine, requires 10 steps to be synthesized from scratch and is essential if phenylalanine is not abundant. We begin with the biosynthesis of nonessential amino acids.

24.2.2. A Common Step Determines the Chirality of All Amino Acids

Three α -ketoacids— α -ketoglutarate, oxaloacetate, and pyruvate—can be converted into amino acids in one step

through the addition of an amino group. We have seen that α -ketoglutarate can be converted into glutamate by reductive amination (Section 24.1.2). The amino group from glutamate can be transferred to other α -ketoacids by transamination reactions. Thus, aspartate and alanine can be made from the addition of an amino group to oxaloacetate and pyruvate, respectively.



These reactions are carried out by *pyridoxal phosphate-dependent transaminases*. Transamination reactions participate in the synthesis of most amino acids. We shall review the transaminase mechanism (Section 23.3.1) as it applies to amino acid biosynthesis (see Figure 23.10).

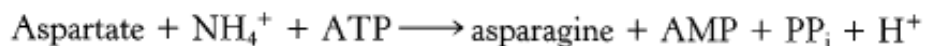
The reaction pathway begins with pyridoxal phosphate in a Schiff-base linkage with lysine at the transaminase active site, forming an internal aldimine (Figure 24.9). An external aldimine forms between PLP and the amino-group donor, glutamate, which displaces the lysine residue. The transfer of the amino group from glutamate to pyridoxal phosphate forms pyridoxamine phosphate, the actual amino donor. When the amino group has been incorporated into pyridoxamine, the reaction pathway proceeds in reverse, and the amino group is transferred to an α -ketoacid to form an amino acid.

Aspartate aminotransferase is the prototype of a large family of PLP-dependent enzymes. Comparisons of amino acid sequences as well as several three-dimensional structures reveal that almost all transaminases having roles in amino acid biosynthesis are related to aspartate aminotransferase by divergent evolution. An examination of the aligned amino acid sequences reveals that two residues are completely conserved. These residues are the lysine residue that forms the Schiff base with the pyridoxal phosphate cofactor (lysine 258 in aspartate aminotransferase) and an arginine residue that interacts with the α -carboxylate group of the ketoacid (see Figure 23.11).

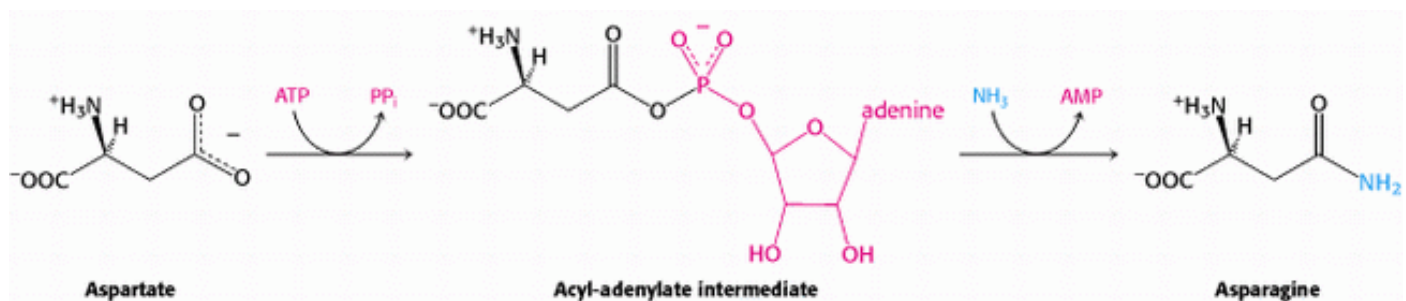
An essential step in the transamination reaction is the protonation of the quinonoid intermediate to form the external aldimine. *The chirality of the amino acid formed is determined by the direction from which this proton is added to the quinonoid form.* (Figure 24.10). This protonation step determines the l configuration of the amino acids produced. The interaction between the conserved arginine residue and the α -carboxylate group helps orient the substrate so that, when the lysine residue transfers a proton to the face of the quinonoid intermediate, it generates an aldimine with an l configuration at the C _{α} center.

24.2.3. An Adenylyated Intermediate Is Required to Form Asparagine from Aspartate

The formation of asparagine from aspartate is chemically analogous to the formation of glutamine from glutamate. Both transformations are amidation reactions and both are driven by the hydrolysis of ATP. The actual reactions are different, however. In bacteria, the reaction for the asparagine synthesis is



Thus, the products of ATP hydrolysis are AMP and PP_i rather than ADP and P_i. Aspartate is activated by adenylation rather than by phosphorylation.

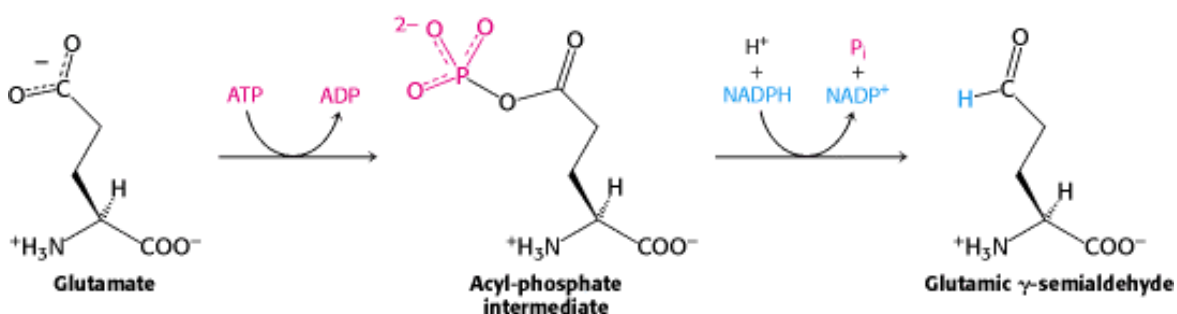


We have encountered this mode of activation in fatty acid degradation and will see it again in lipid and protein synthesis.

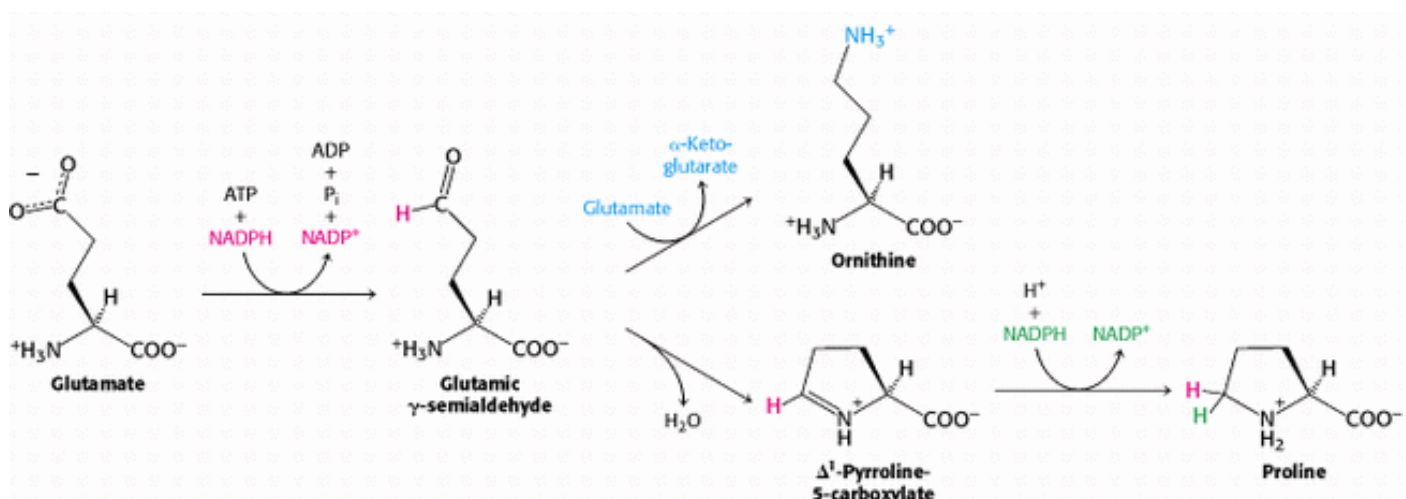
In mammals, the nitrogen donor for asparagine is glutamine rather than ammonia as in bacteria. Ammonia is generated by hydrolysis of the side chain of glutamine and directly transferred to activated aspartate, bound in the active site. An advantage is that the cell is not directly exposed to NH_4^+ , which is toxic at high levels to human beings and other mammals. *The use of glutamine hydrolysis as a mechanism for generating ammonia for use within the same enzyme is a motif common throughout biosynthetic pathways.*

24.2.4. Glutamate Is the Precursor of Glutamine, Proline, and Arginine

The synthesis of glutamate by the reductive amination of α -ketoglutarate has already been discussed, as has the conversion of glutamate into glutamine (Section 24.1.2). Glutamate is the precursor of two other nonessential amino acids: *proline* and *arginine*. First, the γ -carboxyl group of glutamate reacts with ATP to form an acyl phosphate. This mixed anhydride is then reduced by NADPH to an aldehyde.

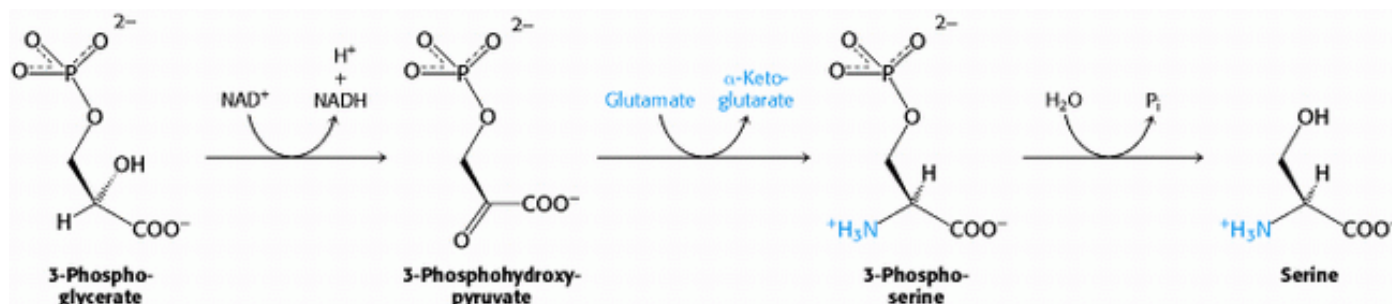


Glutamic γ -semialdehyde cyclizes with a loss of H_2O in a nonenzymatic process to give Δ^1 -pyrroline-5-carboxylate, which is reduced by NADPH to proline. Alternatively, the semialdehyde can be transaminated to ornithine, which is converted in several steps into arginine (Section 23.4.1).

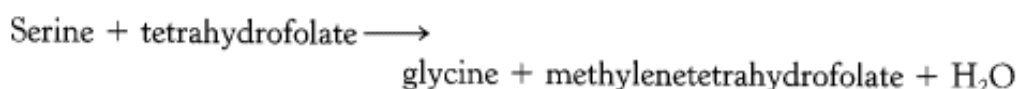


24.2.5. Serine, Cysteine, and Glycine Are Formed from 3-Phosphoglycerate

Serine is synthesized from 3-phosphoglycerate, an intermediate in glycolysis. The first step is an oxidation to 3-phosphohydroxypyruvate. This α -ketoacid is transaminated to 3-phosphoserine, which is then hydrolyzed to serine.



Serine is the precursor of *glycine* and *cysteine*. In the formation of glycine, the side-chain methylene group of serine is transferred to *tetrahydrofolate*, a carrier of one-carbon units that will be discussed shortly.



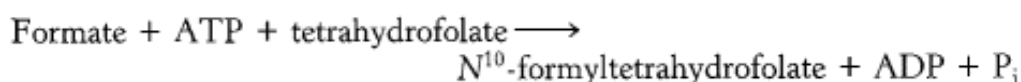
This interconversion is catalyzed by *serine transhydroxymethylase*, another PLP enzyme that is homologous to aspartate aminotransferase (Figure 24.11). The bond between the α - and β -carbon atoms of serine is labilized by the formation of a Schiff base between serine and PLP (Section 23.3.3). The side-chain methylene group of serine is then transferred to tetrahydrofolate. The conversion of serine into cysteine requires the substitution of a sulfur atom derived from methionine for the side-chain oxygen atom (Section 24.2.8).

24.2.6. Tetrahydrofolate Carries Activated One-Carbon Units at Several Oxidation Levels

Tetrahydrofolate (also called *tetrahydropteroylglutamate*), a highly versatile carrier of activated one-carbon units, consists of three groups: a substituted pteridine, *p*-aminobenzoate, and a chain of one or more glutamate residues (Figure 24.12). Mammals can synthesize the pteridine ring, but they are unable to conjugate it to the other two units. They obtain tetrahydrofolate from their diets or from microorganisms in their intestinal tracts.

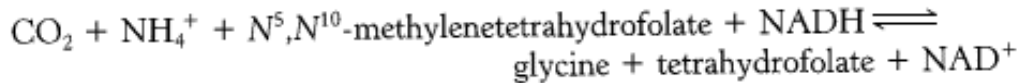
The one-carbon group carried by tetrahydrofolate is bonded to its N-5 or N-10 nitrogen atom (denoted as N^5 and N^{10}) or to both. This unit can exist in three oxidation states (Table 24.2). The most-reduced form carries a *methyl* group, whereas the intermediate form carries a *methylene* group. More-oxidized forms carry a *formyl*, *formimino*, or *methenyl* group. The fully oxidized one-carbon unit, CO_2 , is carried by biotin rather than by tetrahydrofolate.

The one-carbon units carried by tetrahydrofolate are interconvertible (Figure 24.13). N^5, N^{10} -Methylenetetrahydrofolate can be reduced to N^5 -methyltetrahydrofolate or oxidized to N^5, N^{10} -methenyltetrahydrofolate. N^5, N^{10} -Methenyltetrahydrofolate can be converted into N^5 -formiminetetrahydrofolate or N^{10} -formyltetrahydrofolate, both of which are at the same oxidation level. N^{10} -Formyltetrahydrofolate can also be synthesized from tetrahydrofolate, formate, and ATP.



N^5 -Formyltetrahydrofolate can be reversibly isomerized to N^{10} -formyltetrahydrofolate or it can be converted into N^5 , N^{10} -methylenetetrahydrofolate.

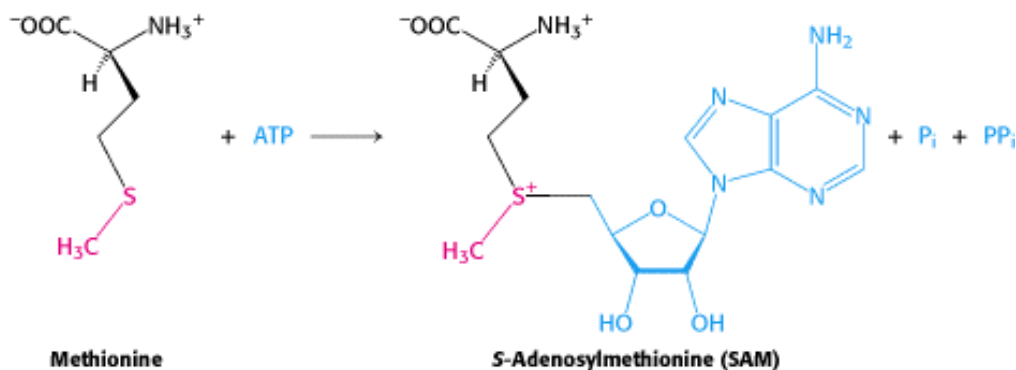
These tetrahydrofolate derivatives serve as donors of one-carbon units in a variety of biosyntheses. Methionine is regenerated from homocysteine by transfer of the methyl group of N^5 -methyltetrahydrofolate, as will be discussed shortly. We shall see in [Chapter 25](#) that some of the carbon atoms of *purines* are acquired from derivatives of N^{10} -formyltetrahydrofolate. The methyl group of *thymine*, a pyrimidine, comes from N^5, N^{10} -methylenetetrahydrofolate. This tetrahydrofolate derivative can also donate a one-carbon unit in an alternative synthesis of *glycine* that starts with CO_2 and NH_4^+ , a reaction catalyzed by *glycine synthase* (called the *glycine cleavage enzyme* when it operates in the reverse direction).



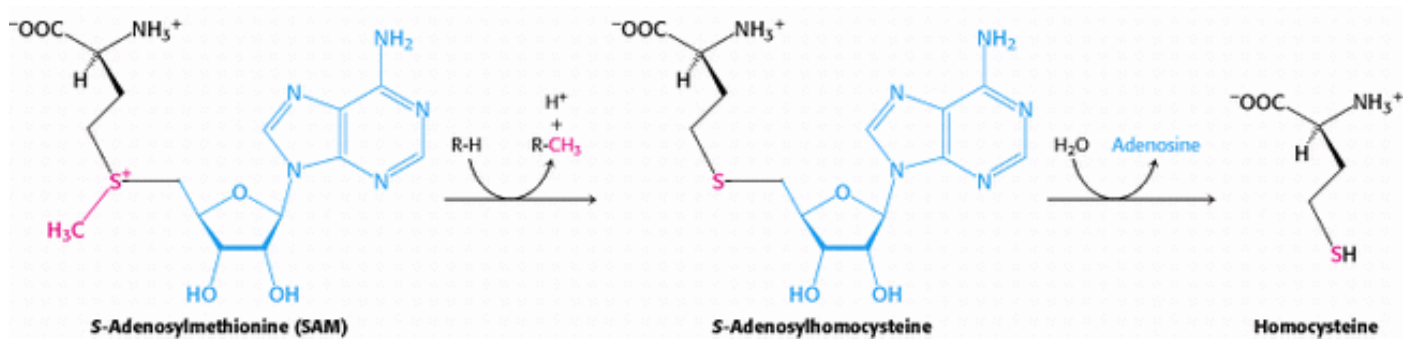
Thus, one-carbon units at each of the three oxidation levels are utilized in biosyntheses. Furthermore, *tetrahydrofolate* serves as an acceptor of one-carbon units in degradative reactions. The major source of one-carbon units is the facile conversion of serine into glycine, which yields N^5, N^{10} -methylenetetrahydrofolate. Serine can be derived from 3-phosphoglycerate, and so *this pathway enables one-carbon units to be formed de novo from carbohydrates*.

24.2.7. S-Adenosylmethionine Is the Major Donor of Methyl Groups

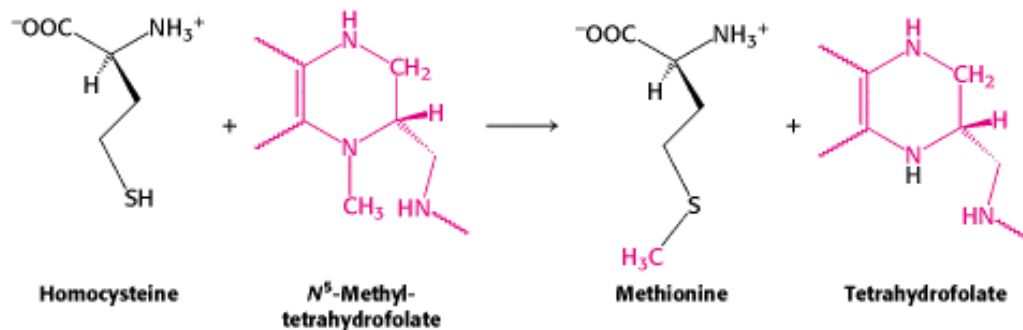
Tetrahydrofolate can carry a methyl group on its N-5 atom, but its transfer potential is not sufficiently high for most biosynthetic methylations. Rather, the activated methyl donor is usually *S-adenosylmethionine (SAM)*, which is synthesized by the transfer of an adenosyl group from ATP to the sulfur atom of methionine.



The methyl group of the methionine unit is activated by the positive charge on the adjacent sulfur atom, which makes the molecule much more reactive than N^5 -methyltetrahydrofolate. The synthesis of *S-adenosylmethionine* is unusual in that the triphosphate group of ATP is split into pyrophosphate and orthophosphate; the pyrophosphate is subsequently hydrolyzed to two molecules of P_i . *S-Adenosylhomocysteine* is formed when the methyl group of *S-adenosylmethionine* is transferred to an acceptor. *S-Adenosylhomocysteine* is then hydrolyzed to *homocysteine* and adenosine.



Methionine can be regenerated by the transfer of a methyl group to homocysteine from *N*⁵-methyltetrahydrofolate, a reaction catalyzed by *methionine synthase* (also known as *homocysteine methyltransferase*).



The coenzyme that mediates this transfer of a methyl group is *methylcobalamin*, derived from vitamin B₁₂. In fact, this reaction and the rearrangement of 1-methylmalonyl CoA to succinyl CoA (Section 23.5.4), catalyzed by a homologous enzyme, are the only two B₁₂-dependent reactions known to take place in mammals. Another enzyme that converts homocysteine into methionine without vitamin B₁₂ also is present in many organisms.

These reactions constitute the *activated methyl cycle* (Figure 24.14). Methyl groups enter the cycle in the conversion of homocysteine into methionine and are then made highly reactive by the addition of adenosyl groups, which make the sulfur atoms positively charged and the methyl groups much more electrophilic. The high transfer potential of the *S*-methyl group enables it to be transferred to a wide variety of acceptors.

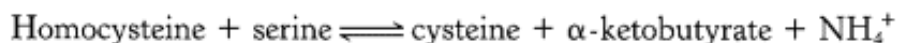
Among the acceptors modified by *S*-adenosylmethionine are specific bases in DNA. The methylation of DNA protects bacterial DNA from cleavage by restriction enzymes (Section 9.3). The base to be methylated is flipped out of the DNA double helix into the active site where it can accept a methyl group from *S*-adenosylmethionine (Figure 24.15). A recurring *S*-adenosylmethionine-binding domain is present in many SAM-dependent methylases.

S-Adenosylmethionine is also the precursor of *ethylene*, a gaseous plant hormone that induces the ripening of fruit. *S*-Adenosylmethionine is cyclized to a cyclopropane derivative that is then oxidized to form ethylene. The Greek philosopher Theophrastus recognized more than 2000 years ago that sycamore figs do not ripen unless they are scraped with an iron claw. The reason is now known: *wounding triggers ethylene production, which in turn induces ripening*. Much effort is being made to understand this biosynthetic pathway because ethylene is a culprit in the spoilage of fruit.

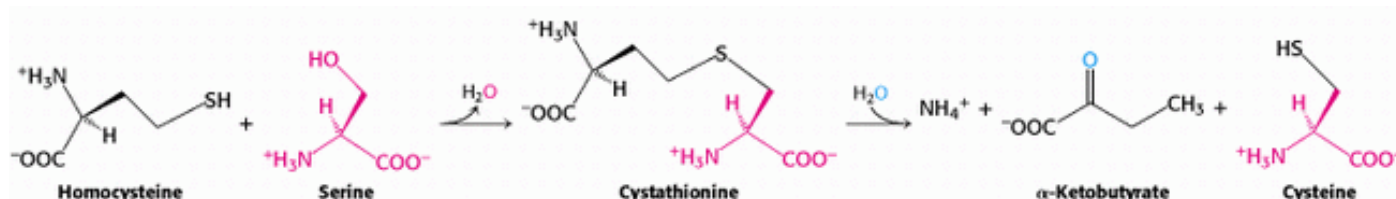


24.2.8. Cysteine Is Synthesized from Serine and Homocysteine


In addition to being a precursor of methionine in the activated methyl cycle, homocysteine is an intermediate in the synthesis of cysteine. Serine and homocysteine condense to form *cystathionine*. This reaction is catalyzed by *cystathionine* β -synthase. Cystathionine is then deaminated and cleaved to cysteine and α -ketobutyrate by *cystathioninase*. Both of these enzymes utilize PLP and are homologous to aspartate aminotransferase. The net reaction is



Note that the sulfur atom of cysteine is derived from homocysteine, whereas the carbon skeleton comes from serine.



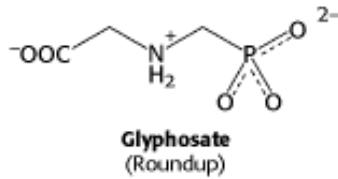
24.2.9. High Homocysteine Levels Are Associated with Vascular Disease

 People with elevated serum levels of homocysteine or the disulfide-linked dimer homocystine have an unusually high risk for coronary heart disease and arteriosclerosis. The most common genetic cause of high homocysteine levels is a mutation within the gene encoding cystathionine β -synthase. The molecular basis of homocysteine's action has not been clearly identified, although it appears to damage cells lining blood vessels and to increase the growth of vascular smooth muscle. The amino acid raises oxidative stress as well. Vitamin treatments are effective in reducing homocysteine levels in some people. Treatment with vitamins maximizes the activity of the two major metabolic pathways processing homocysteine. Pyridoxal phosphate, a vitamin B₆ derivative, is necessary for the activity of cystathionine β -synthase, which converts homocysteine into cystathionine; tetrahydrofolate, and vitamin B₁₂, supports the methylation of homocysteine to methionine.

24.2.10. Shikimate and Chorismate Are Intermediates in the Biosynthesis of Aromatic Amino Acids

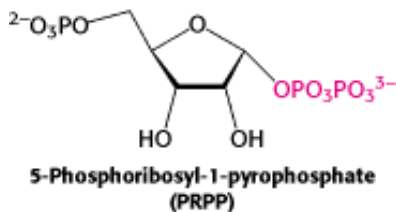
We turn now to the biosynthesis of essential amino acids. These amino acids are synthesized by plants and microorganisms, and those in the human diet are ultimately derived primarily from plants. The essential amino acids are formed by much more complex routes than are the nonessential amino acids. The pathways for the synthesis of aromatic amino acids in bacteria have been selected for discussion here because they are well understood and exemplify recurring mechanistic motifs.

Phenylalanine, tyrosine, and tryptophan are synthesized by a common pathway in *E. coli* (Figure 24.16). The initial step is the condensation of phosphoenolpyruvate (a glycolytic intermediate) with erythrose 4-phosphate (a pentose phosphate pathway intermediate). The resulting seven-carbon open-chain sugar is oxidized, loses its phosphoryl group, and cyclizes to 3-dehydroquinate. Dehydration then yields 3-dehydroshikimate, which is reduced by NADPH to shikimate. Phosphorylation of shikimate by ATP gives shikimate 3-phosphate, which condenses with a second molecule of phosphoenolpyruvate. This 5-enolpyruvyl intermediate loses its phosphoryl group, yielding chorismate, the common precursor of all three aromatic amino acids. The importance of this pathway is revealed by the effectiveness of glyphosate (Roundup), a broad-spectrum herbicide. This compound inhibits the enzyme that produces 5-enolpyruvylshikimate 3-phosphate and, hence, blocks aromatic amino acid biosynthesis in plants. Because animals lack this enzyme, the herbicide is fairly nontoxic.



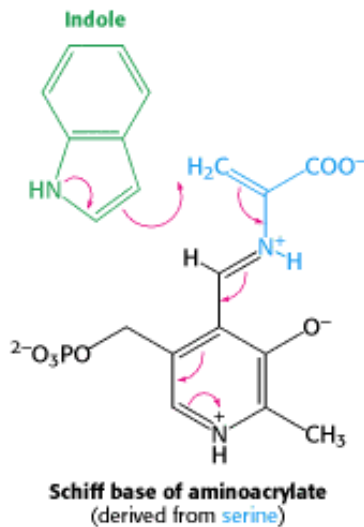
The pathway bifurcates at chorismate. Let us first follow the *prephenate branch* (Figure 24.17). A mutase converts chorismate into prephenate, the immediate precursor of the aromatic ring of phenylalanine and tyrosine. This fascinating conversion is a rare example of an electrocyclic reaction in biochemistry, mechanistically similar to the well-known Diels-Alder reaction from organic chemistry. Dehydration and decarboxylation yield *phenylpyruvate*. Alternatively, prephenate can be oxidatively decarboxylated to *p-hydroxyphenylpyruvate*. These α -ketoacids are then transaminated to form *phenylalanine* and *tyrosine*.

The branch starting with *anthranilate* leads to the synthesis of tryptophan (Figure 24.18). Chorismate acquires an amino group derived from the hydrolysis of the side chain of glutamine and releases pyruvate to form anthranilate. Then anthranilate condenses with *5-phosphoribosyl-1-pyrophosphate (PRPP)*, an activated form of ribose phosphate. PRPP is also an important intermediate in the synthesis of histidine, purine nucleotides, and pyrimidine nucleotides (Sections 25.1.4 and 25.2.2). The C-1 atom of ribose 5-phosphate becomes bonded to the nitrogen atom of anthranilate in a reaction that is driven by the release and hydrolysis of pyrophosphate. The ribose moiety of phosphoribosylanthranilate undergoes rearrangement to yield 1-(*o*-carboxyphenylamino)-1-deoxyribulose 5-phosphate. This intermediate is dehydrated and then decarboxylated to indole-3-glycerol phosphate, which is cleaved to indole. Then indole reacts with serine to form tryptophan. In these final steps, which are catalyzed by tryptophan synthetase, the side chain of indole-3-glycerol phosphate is removed as glyceraldehyde 3-phosphate and replaced by the carbon skeleton of serine.



24.2.11. Tryptophan Synthetase Illustrates Substrate Channeling in Enzymatic Catalysis

Tryptophan synthetase of *E. coli*, an $\alpha_2 \beta_2$ tetramer, can be dissociated into two α subunits and a β_2 subunit (Figure 24.19). The α subunit catalyzes the formation of indole from indole-3-glycerol phosphate, whereas each β subunit has a PLP-containing active site that catalyzes the condensation of indole and serine to form tryptophan. The overall three-dimensional structure of this enzyme is distinct from that of aspartate aminotransferase and the other PLP enzymes already discussed. Serine forms a Schiff base with this PLP, which is then dehydrated to give the *Schiff base of aminoacrylate*. This reactive intermediate is attacked by indole to give tryptophan.



The synthesis of tryptophan poses a challenge. Indole, a hydrophobic molecule, readily traverses membranes and would be lost from the cell if it were allowed to diffuse away from the enzyme. This problem is solved in an ingenious way. A 25-Å-long channel connects the active site of the α subunit with that of the adjacent β subunit in the $\alpha_2\beta_2$ tetramer (Figure 24.20). Thus, indole can diffuse from one active site to the other without being released into bulk solvent. Indeed, the results of isotopic-labeling experiments showed that indole formed by the α subunit does not leave the enzyme when serine is present. Furthermore, the two partial reactions are coordinated. Indole is not formed by the α subunit until the highly reactive aminoacrylate is ready and waiting in the β subunit. We see here a clear-cut example of *substrate channeling* in catalysis by a multienzyme complex. Channeling substantially increases the catalytic rate. Furthermore, a deleterious side reaction—in this case, the potential loss of an intermediate—is prevented. We shall encounter other examples of substrate channeling in [Chapter 25](#).

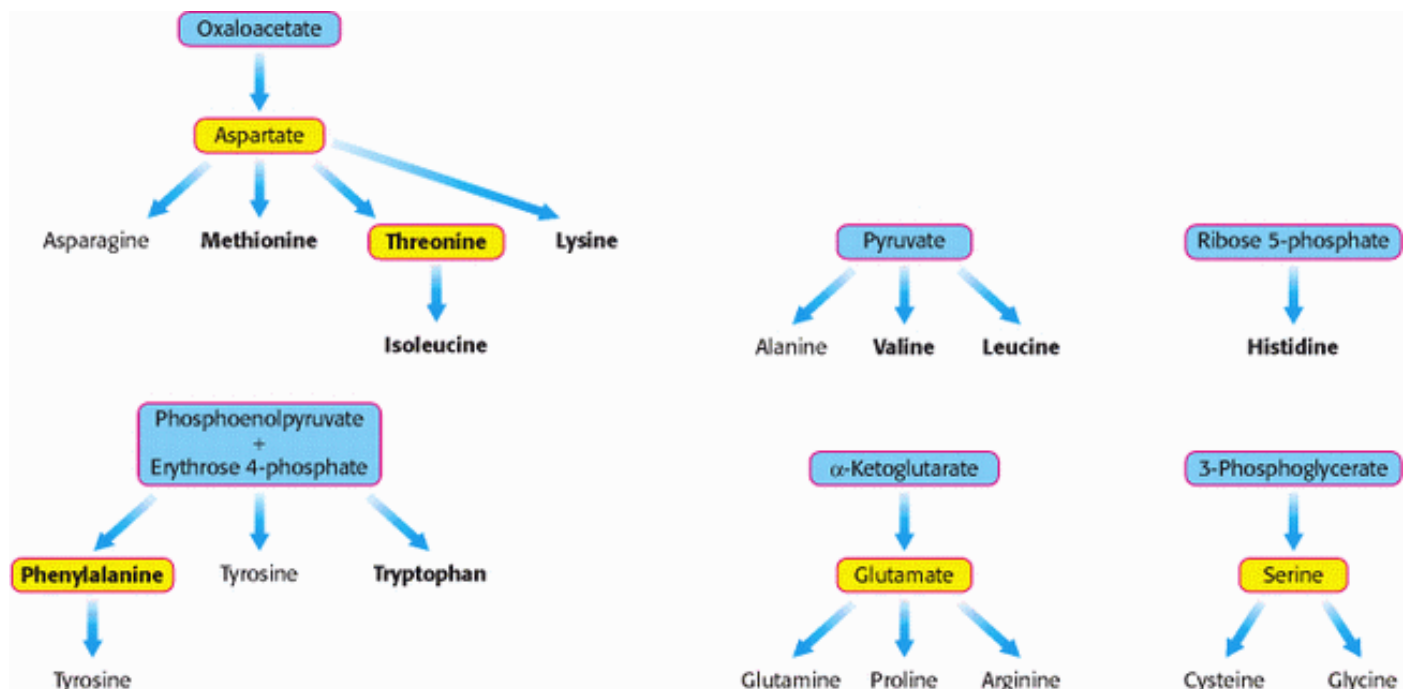


Figure 24.7. Biosynthetic Families of Amino Acids in Bacteria and Plants. Major metabolic precursors are shaded blue. Amino acids that give rise to other amino acids are shaded yellow. Essential amino acids are in boldface type.

Table 24.1. Basic set of 20 amino acids

Nonessential	Essential
Alanine	Histidine
Arginine	Isoleucine
Asparagine	Leucine
Aspartate	Lysine
Cysteine	Methionine
Glutamate	Phenylalanine
Glutamine	Threonine
Glycine	Tryptophan
Proline	Valine
Serine	
Tyrosine	

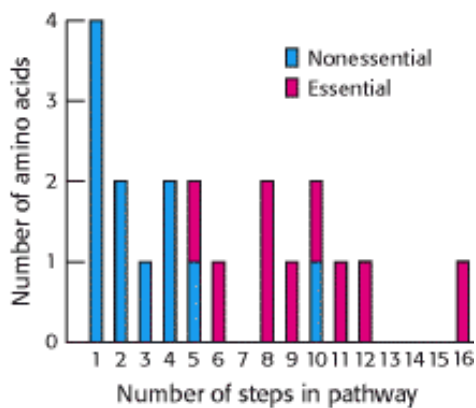


Figure 24.8. Essential and Nonessential Amino Acids. Some amino acids are nonessential to human beings because they can be biosynthesized in a small number of steps. Those amino acids requiring a large number of steps for their synthesis are essential in the diet because some of the enzymes for these steps have been lost in the course of evolution.

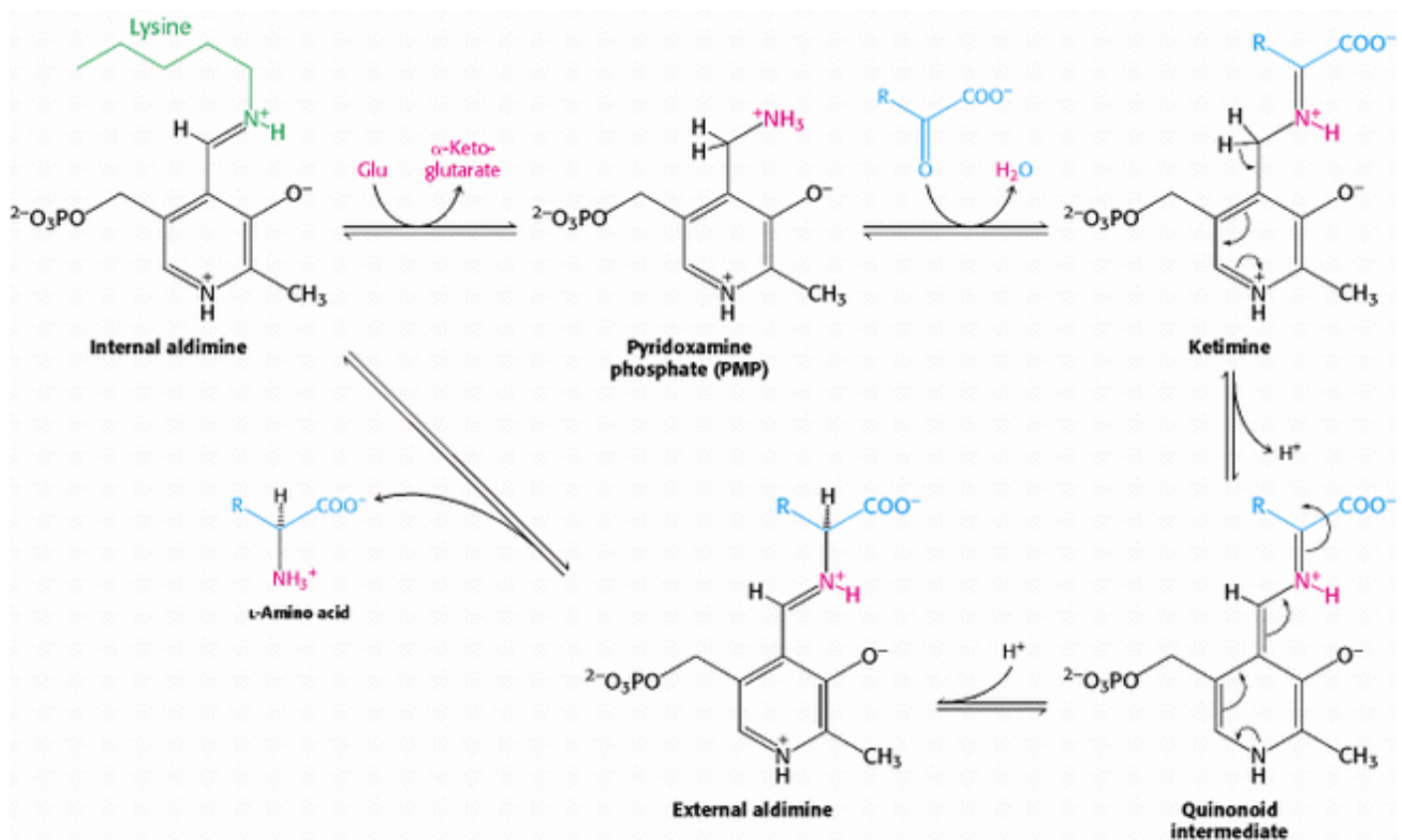


Figure 24.9. Amino Acid Biosynthesis by Transamination. Within a transaminase, the internal aldimine is converted into pyridoxamine phosphate (PMP) by reaction with glutamate. PMP then reacts with an α -ketoacid to generate a ketimine. This intermediate is converted into a quinonoid intermediate, which in turn yields an external aldimine. The aldimine is cleaved to release the newly formed amino acid to complete the cycle.

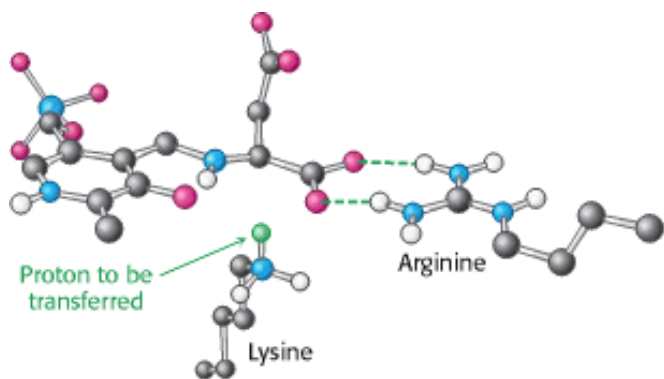


Figure 24.10. Stereochemistry of Proton Addition. In a transaminase active site, the addition of a proton from the lysine residue to the bottom face of the quinonoid intermediate determines the L configuration of the amino acid product. The conserved arginine residue interacts with the α -carboxylate group and helps establish the appropriate geometry of the quinonoid intermediate.

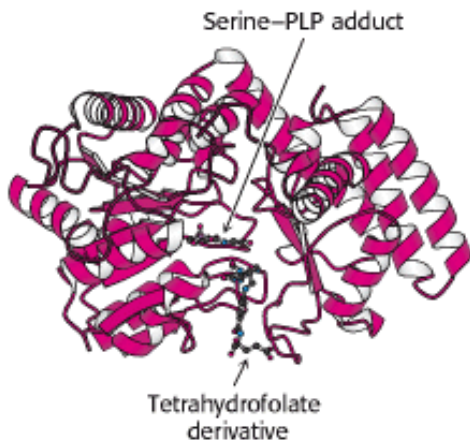


Figure 24.11. Structure of Serine Hydroxymethyltransferase. This enzyme transfers a one-carbon unit from the side chain of serine to tetrahydrofolate. One subunit of the dimeric enzyme is shown.

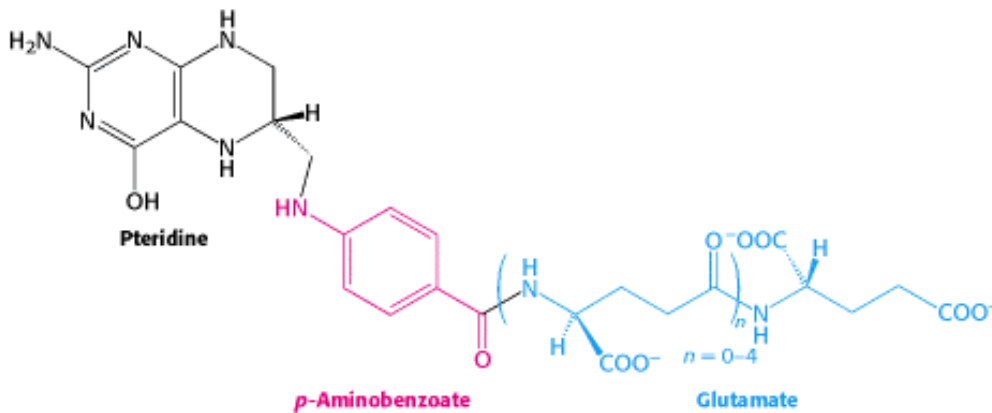


Figure 24.12. Tetrahydrofolate. This cofactor includes three components: a pteridine ring, *p*-aminobenzoate, and one or more glutamate residues.

Table 24.2. One-carbon groups carried by tetrahydrofolate

Oxidation state	Group
Most reduced (= methanol)	-CH ₃ Methyl
Intermediate (= formaldehyde)	-CH ₂ - Methylene
Most oxidized (= formic acid)	-CHO Formyl
	-CHNH Formimino
	-CH= Methenyl

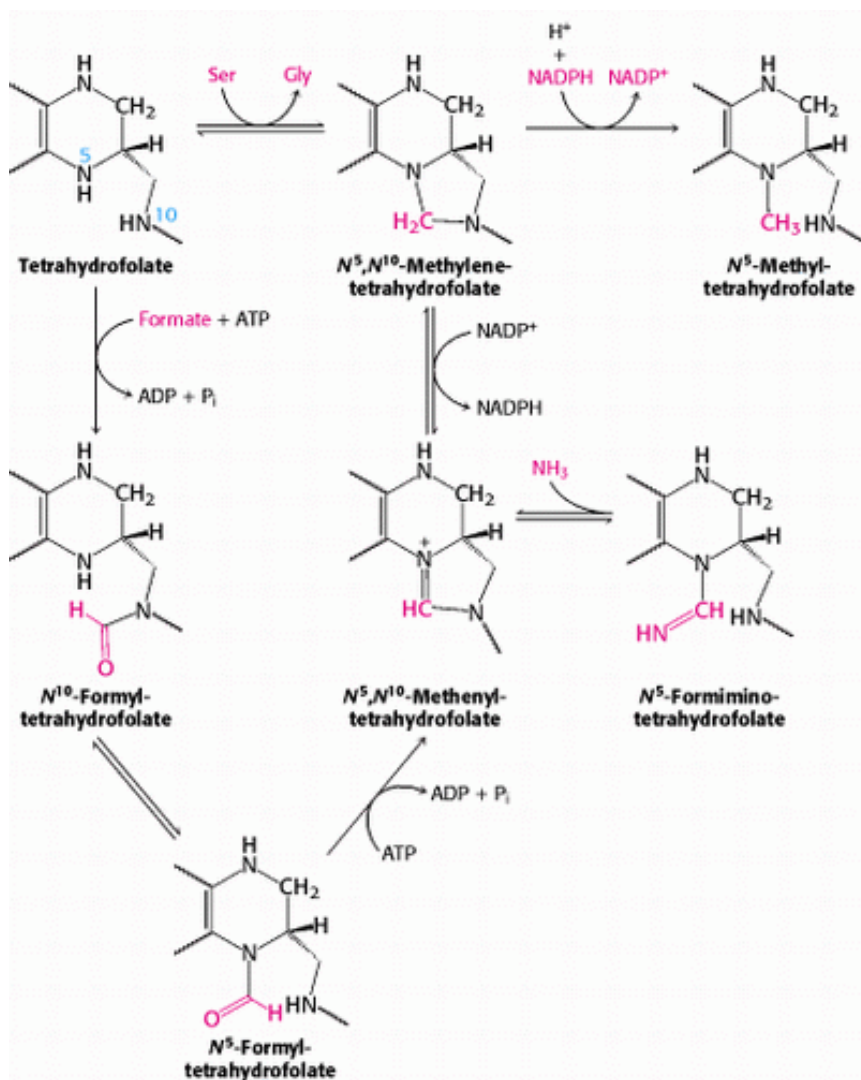


Figure 24.13. Conversions of One-Carbon Units Attached to Tetrahydrofolate.

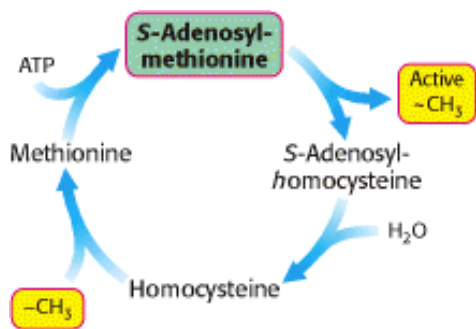


Figure 24.14. Activated Methyl Cycle. The methyl group of methionine is activated by the formation of S-adenosylmethionine.

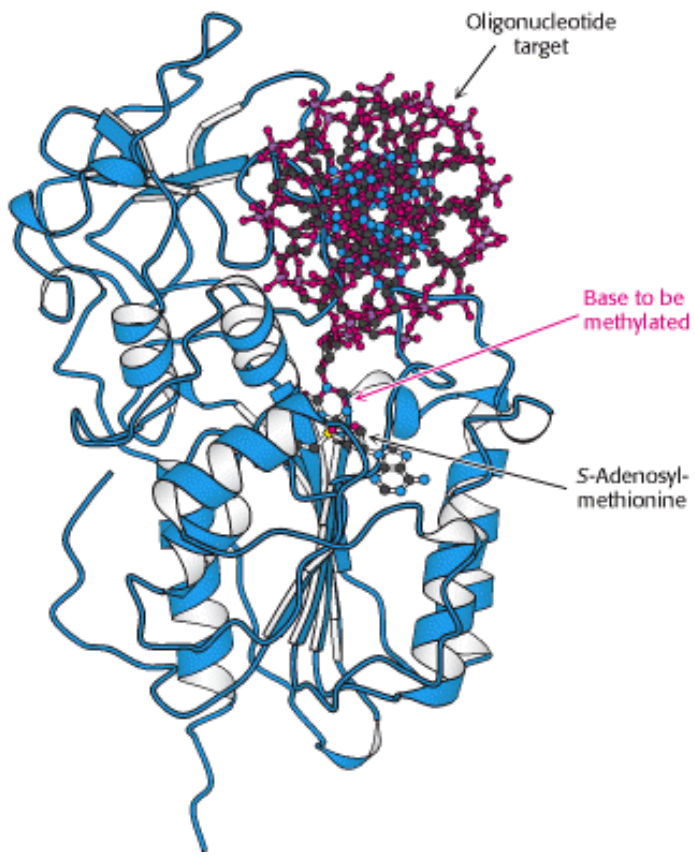



Figure 24.15. DNA Methylation. The structure of a DNA methylase bound to an oligonucleotide target shows that the  base to be methylated is flipped out of the DNA helix into the active site of a SAM-dependent methylase.

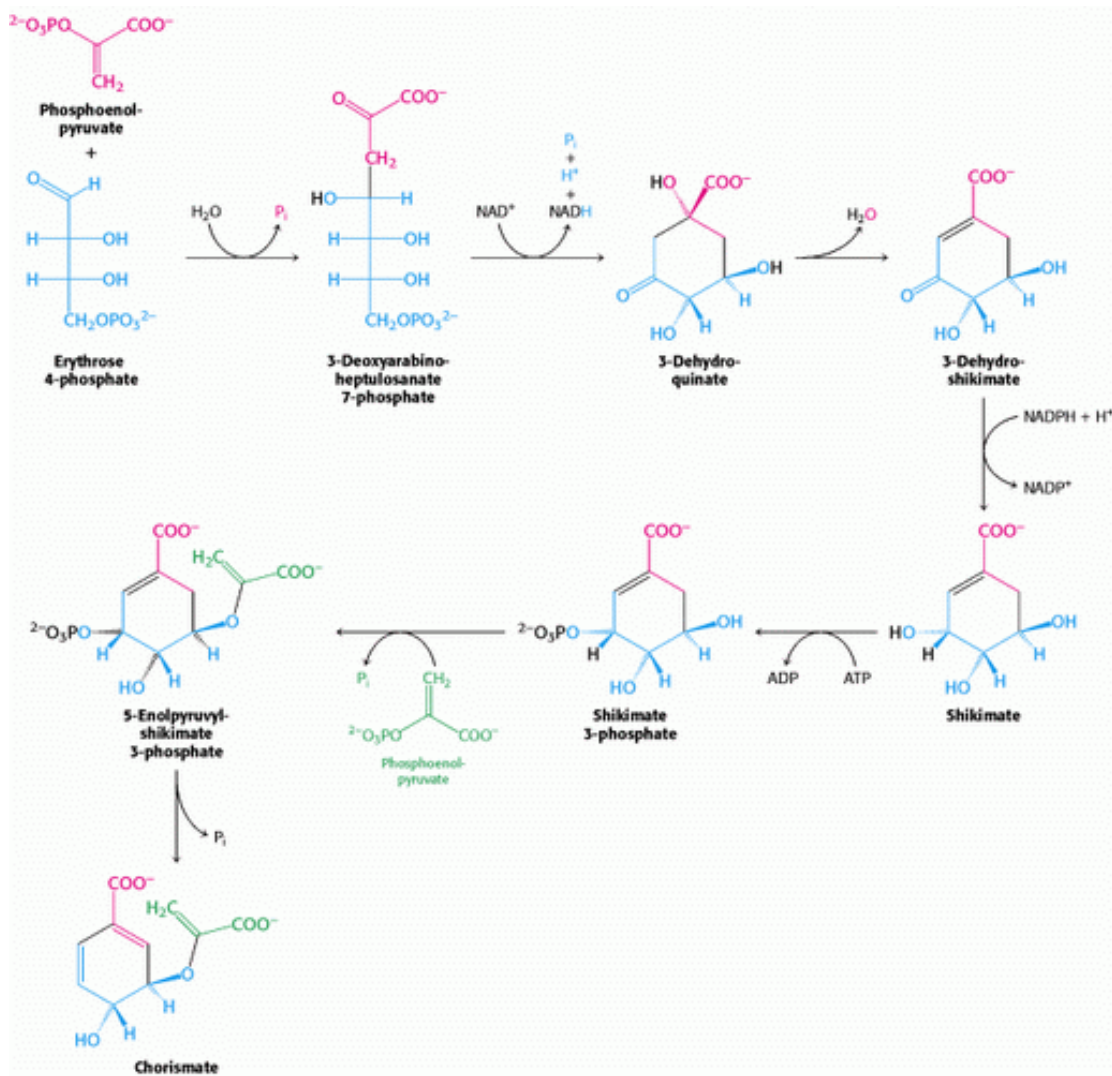


Figure 24.16. Pathway to Chorismate. Chorismate is an intermediate in the biosynthesis of phenylalanine, tyrosine, and tryptophan.

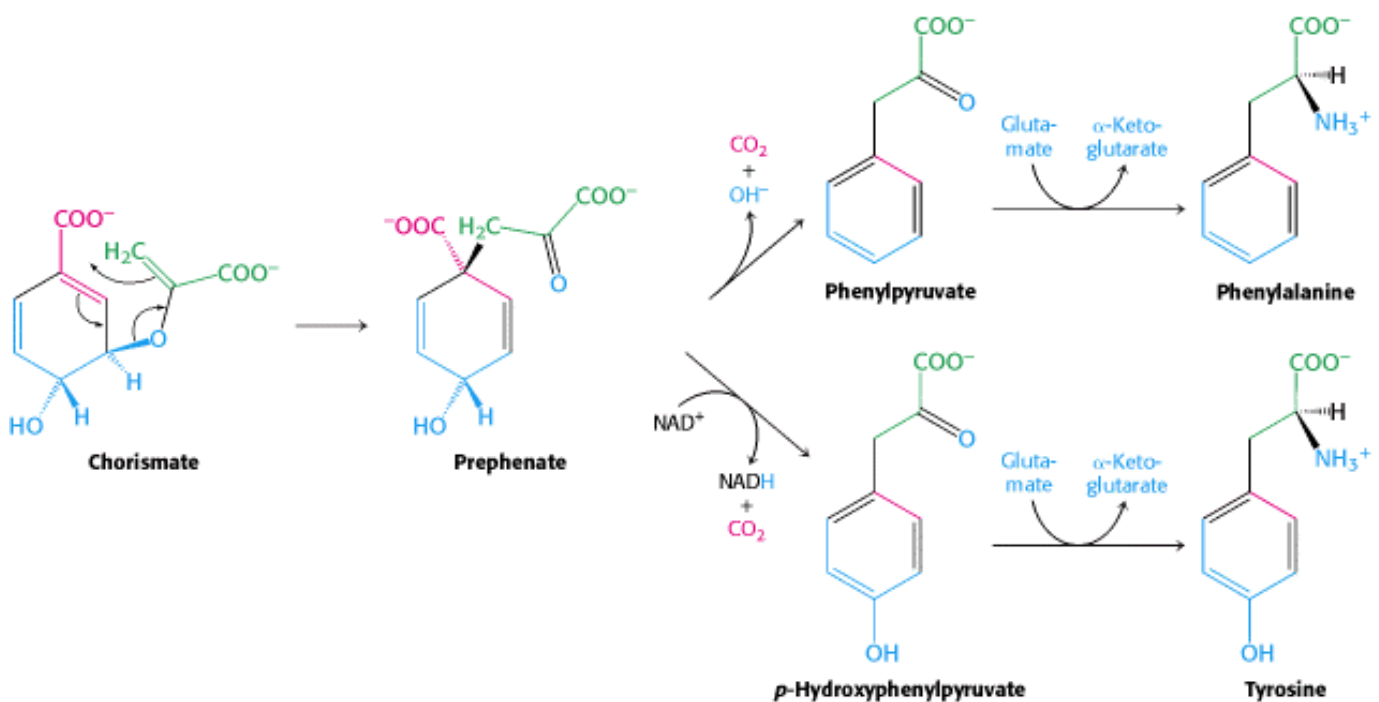


Figure 24.17. Synthesis of Phenylalanine and Tyrosine. Chorismate can be converted into prephenate, which is subsequently converted into phenylalanine and tyrosine.

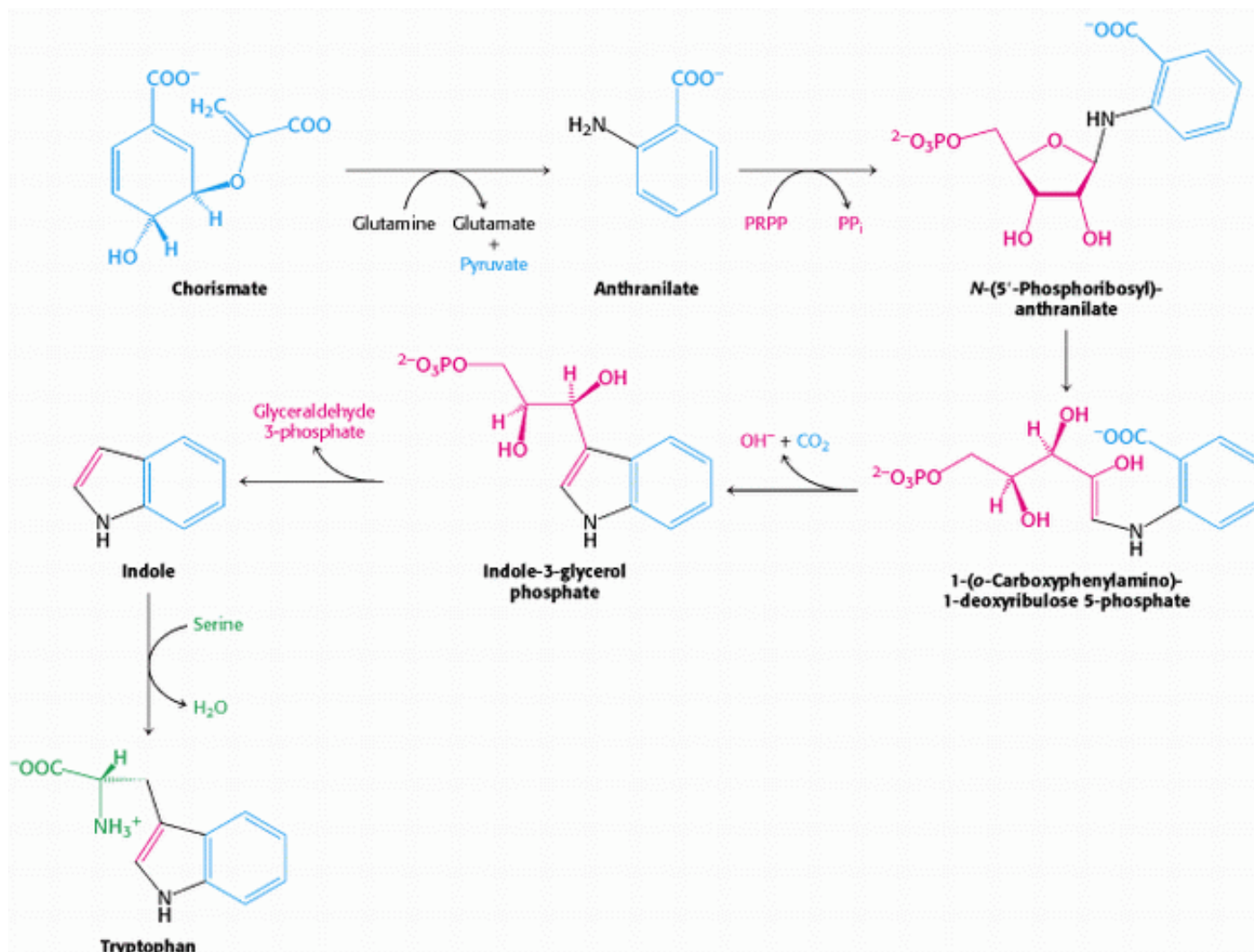


Figure 24.18. Synthesis of Tryptophan. Chorismate can be converted into anthranilate, which is subsequently converted into tryptophan.

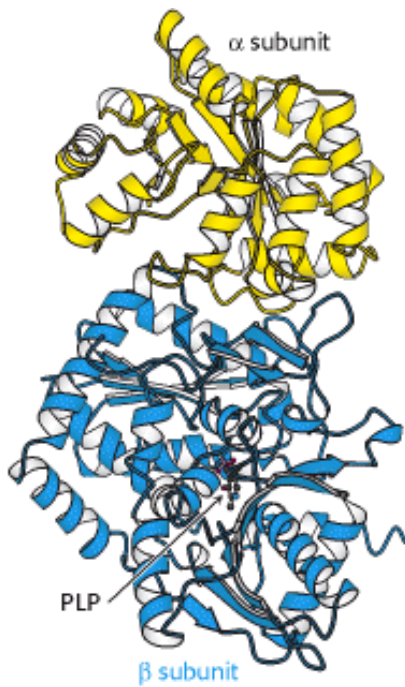


Figure 24.19. Structure of Tryptophan Synthetase. The structure of the complex formed by one α subunit and one β subunit. PLP is bound to the β subunit.

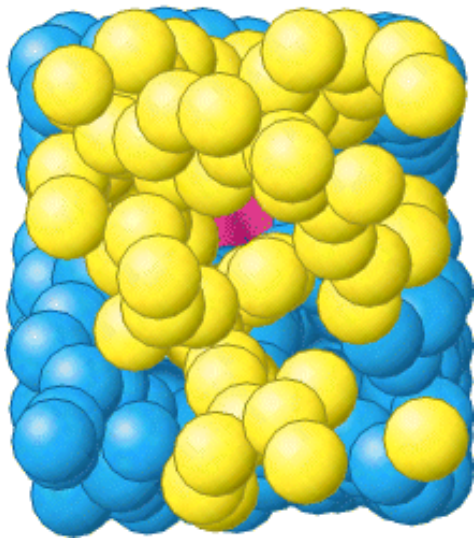


Figure 24.20. Substrate Channeling. A 25-Å tunnel runs from the active site of the α subunit of tryptophan synthetase (yellow) to the PLP cofactor (red) in the active site of the β subunit (blue).

24.3. Amino Acid Biosynthesis Is Regulated by Feedback Inhibition

The rate of synthesis of amino acids depends mainly on the *amounts* of the biosynthetic enzymes and on their *activities*. We now consider the control of enzymatic activity. The regulation of enzyme synthesis will be discussed in [Chapter 31](#).

In a biosynthetic pathway, the first irreversible reaction, called the *committed step*, is usually an important regulatory site. *The final product of the pathway (Z) often inhibits the enzyme that catalyzes the committed step ($A \rightarrow B$).*



This kind of control is essential for the conservation of building blocks and metabolic energy. Consider the biosynthesis of serine (Section 24.2.5). The committed step in this pathway is the oxidation of 3-phosphoglycerate, catalyzed by the enzyme *3-phosphoglycerate dehydrogenase*. The *E. coli* enzyme is a tetramer of four identical subunits, each comprising a catalytic domain and a serine-binding regulatory domain (Figure 24.21). The regulatory domains of two subunits interact to form a dimeric serine-binding regulatory unit so that the tetrameric enzyme contains two such regulatory units. Each unit is capable of binding two serine molecules. The binding of serine to a regulatory site reduces the value of V_{\max} for the enzyme; an enzyme bound to four molecules of serine is essentially inactive. Thus, if serine is abundant in the cell, the enzyme activity is inhibited, and so 3-phosphoglycerate, a key building block that can be used for other processes, is not wasted.

24.3.1. Branched Pathways Require Sophisticated Regulation

The regulation of branched pathways is more complicated because the concentration of two products must be accounted for. In fact, several intricate feedback mechanisms have been found in branched biosynthetic pathways.

Feedback Inhibition and Activation.

Consider, for example, the biosynthesis of the amino acids valine, leucine, and isoleucine. A common intermediate, hydroxyethyl thiamine pyrophosphate (hydroxyethyl-TPP; Section 17.1.1), initiates the pathways leading to all three of these amino acids. Hydroxyethyl-TPP can react with α -ketobutyrate in the initial step for the synthesis of isoleucine. Alternatively, hydroxyethyl-TPP can react with pyruvate in the committed step for the pathways leading to valine and leucine. Thus, the relative concentrations of α -ketobutyrate and pyruvate determine how much isoleucine is produced compared with valine and leucine. *Threonine deaminase*, the PLP enzyme that catalyzes the formation of α -ketobutyrate, is allosterically inhibited by isoleucine (Figure 24.22). This enzyme is also allosterically activated by valine. Thus, this enzyme is inhibited by the product of the pathway that it initiates and is activated by the end product of a competitive pathway. This mechanism balances the amounts of different amino acids that are synthesized.

The regulatory domain in threonine deaminase is very similar in structure to the dimeric regulatory domain in 3-phosphoglycerate dehydrogenase (Figure 24.23). In this case, the two half regulatory domains are fused into a single unit with two differentiated amino acid-binding sites, one for isoleucine and the other for valine. Sequence analysis shows that similar regulatory domains are present in other amino acid biosynthetic enzymes. *The similarities suggest that feedback-inhibition processes may have evolved by the linkage of specific regulatory domains to the catalytic domains of biosynthetic enzymes.*

Enzyme Multiplicity.

Sophisticated regulation can also evolve by duplication of the genes encoding the biosynthetic enzymes. For example, the phosphorylation of aspartate is the committed step in the biosynthesis of threonine, methionine, and lysine. Three distinct aspartokinases catalyze this reaction in *E. coli*, an example of a regulatory mechanism called *enzyme multiplicity*. (Figure 24.24). The catalytic domains of these enzymes show approximately 30% sequence identity. Although the mechanisms of catalysis are essentially identical, their activities are regulated differently: one enzyme is not subject to feedback inhibition, another is inhibited by threonine, and the third is inhibited by lysine.

Cumulative Feedback Inhibition

The regulation of glutamine synthetase in *E. coli* is a striking example of *cumulative feedback inhibition*. Recall that glutamine is synthesized from glutamate, NH_4^+ , and ATP. *Glutamine synthetase* consists of 12 identical 50-kd subunits arranged in two hexagonal rings that face each other (Figure 24.25). Earl Stadtman showed that this enzyme regulates the flow of nitrogen and hence plays a key role in controlling bacterial metabolism. The amide group of glutamine is a source of nitrogen in the biosyntheses of a variety of compounds, such as tryptophan, histidine, carbamoyl phosphate, glucosamine 6-phosphate, cytidine triphosphate, and adenosine monophosphate. Glutamine synthetase is cumulatively inhibited by each of these final products of glutamine metabolism, as well as by alanine and glycine. *In cumulative inhibition, each inhibitor can reduce the activity of the enzyme, even when other inhibitors are bound at saturating levels*. The enzymatic activity of glutamine synthetase is switched off almost completely when all final products are bound to the enzyme.

24.3.2. The Activity of Glutamine Synthetase Is Modulated by an Enzymatic Cascade

The activity of glutamine synthetase is also controlled by *reversible covalent modification* —the attachment of an AMP unit by a phosphodiester bond to the hydroxyl group of a specific tyrosine residue in each subunit (Figure 24.26). *This adenylylated enzyme is less active and more susceptible to cumulative feedback inhibition than is the deadenylylated form*. The covalently attached AMP unit is removed from the adenylylated enzyme by phosphorolysis. The attachment of an AMP unit is the final step in an enzymatic cascade that is initiated several steps back by reactants and immediate products in glutamine synthesis.

The adenylation and phosphorolysis reactions are catalyzed by the same enzyme, *adenylyl transferase*. Sequence analysis indicates that this adenylyl transferase comprises two homologous halves, suggesting that one half catalyzes the adenylation reaction and the other half the phospholytic de-adenylation reaction. What determines whether an AMP unit is added or removed? The specificity of adenylyl transferase is controlled by a *regulatory protein* (designated P or P_{II}), a trimeric protein that can exist in two forms, P_A and P_D (Figure 24.27). The complex of P_A and adenylyl transferase catalyzes the attachment of an AMP unit to glutamine synthetase, which reduces its activity. Conversely, the complex of P_D and adenylyl transferase removes AMP from the adenylylated enzyme.

This brings us to another level of reversible covalent modification. P_A is converted into P_D by the attachment of uridine monophosphate to a specific tyrosine residue (Figure 24.28). This reaction, which is catalyzed by *uridylyl transferase*, is stimulated by ATP and α -ketoglutarate, whereas it is inhibited by glutamine. In turn, the UMP units on P_D are removed by hydrolysis, a reaction promoted by glutamine and inhibited by α -ketoglutarate. These opposing catalytic activities are present on a single polypeptide chain, homologous to adenylyl transferase, and are controlled so that the enzyme does not simultaneously catalyze uridylylation and hydrolysis.

Why is an enzymatic cascade used to regulate glutamine synthetase? One advantage of a cascade is that it *amplifies signals*, as in blood clotting and the control of glycogen metabolism (Sections 10.5.5 and 21.3.1). Another advantage is that the *potential for allosteric control is markedly increased when each enzyme in the cascade is an independent target for regulation*. The integration of nitrogen metabolism in a cell requires that a large number of input signals be detected and processed. In addition, the regulatory protein P also participates in regulating the transcription of genes for glutamine synthetase and other enzymes taking part in nitrogen metabolism. The evolution of a cascade provided many more regulatory sites and made possible a finer tuning of the flow of nitrogen in the cell.



Figure 24.21. Structure of 3-Phosphoglycerate Dehydrogenase. This enzyme, which catalyzes the committed step in the serine biosynthetic pathway, includes a serine-binding regulatory domain. Serine binding to this domain reduces the activity of the enzyme.

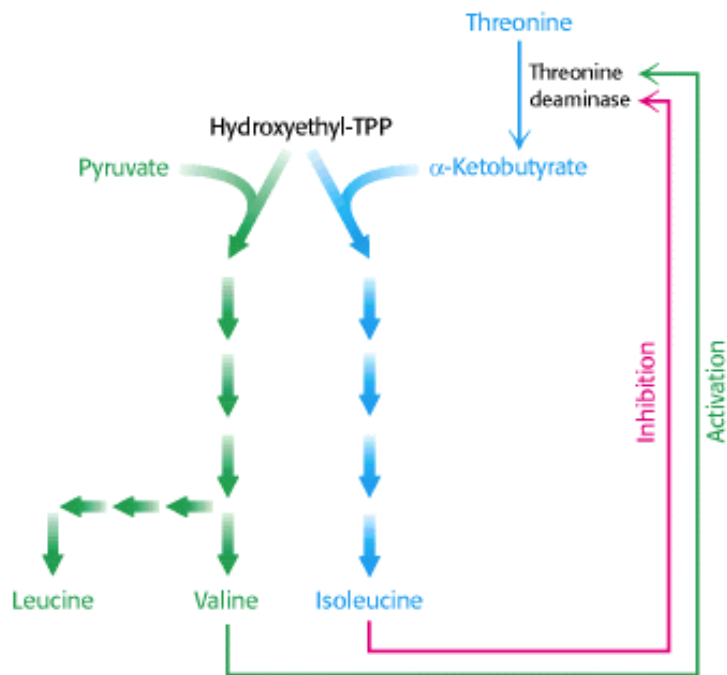


Figure 24.22. Regulation of Threonine Deaminase. Threonine is converted into α -ketobutyrate in the committed step leading to the synthesis of isoleucine. The enzyme that catalyzes this step, threonine deaminase, is inhibited by isoleucine and activated by valine, the product of a parallel pathway.



Figure 24.23. A Recurring Regulatory Domain. The regulatory domain formed by two subunits of 3-phosphoglycerate dehydrogenase is structurally related to the single-chain regulatory domain of threonine deaminase. Sequence analyses have revealed this amino acid-binding regulatory domain to be present in other enzymes as well.

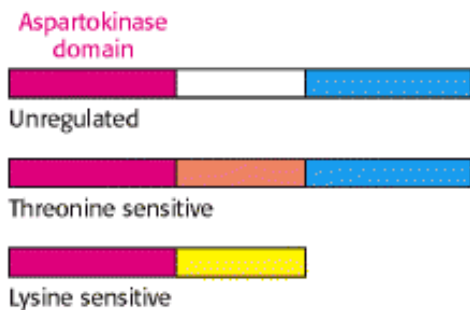


Figure 24.24. Domain Structures of Three Aspartokinases. Each catalyzes the committed step in the biosynthesis of a different amino acid: (top) methionine, (middle) threonine, and (bottom) lysine. They have a catalytic domain in common but differ in their regulatory domains.

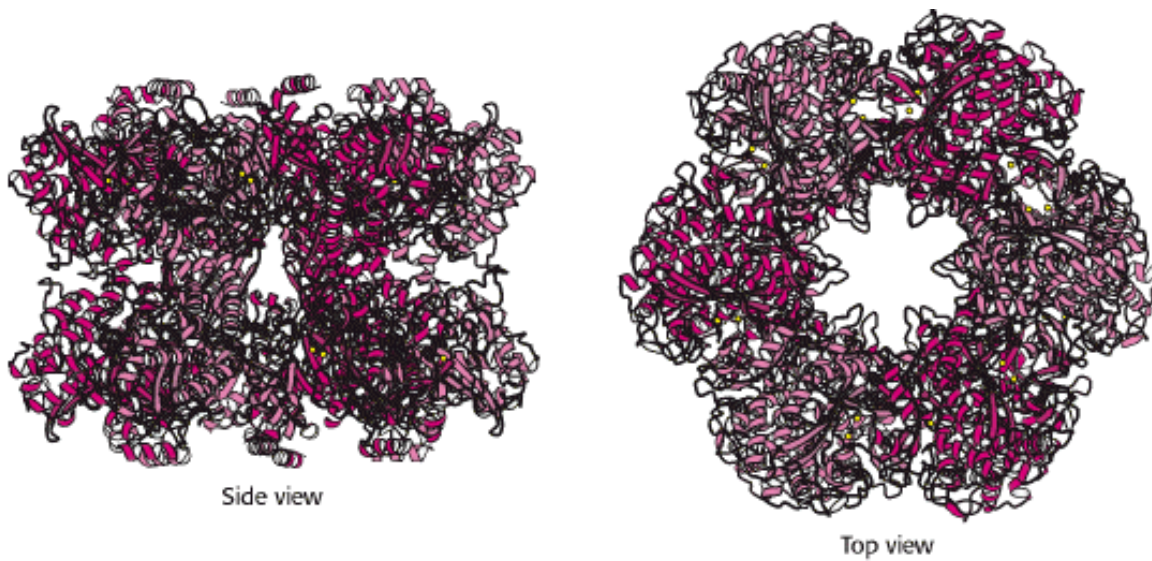


Figure 24.25. Structure of Glutamine Synthetase. Glutamine synthetase consists of 12 identical subunits arranged in two rings of six subunits. The active sites are indicated by the presence of manganese ions (two yellow spheres).

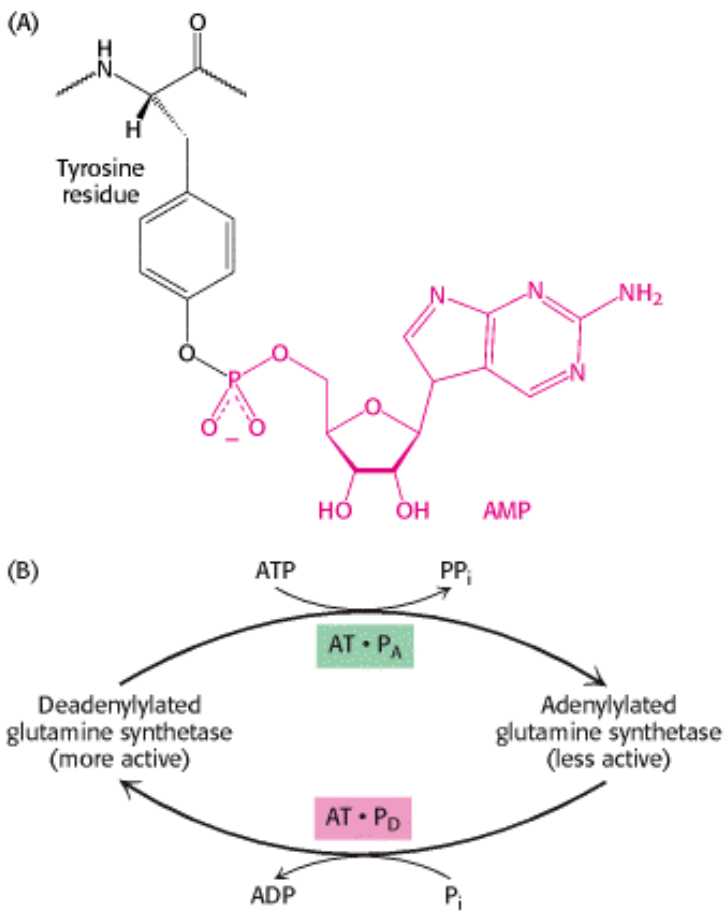


Figure 24.26. Regulation by Adenylation. (A) A specific tyrosine residue in each subunit in glutamine synthetase is modified by adenylation. (B) Adenylation of tyrosine is catalyzed by a complex of adenylyl transferase (AT) and one form of a regulatory protein (P_A). The same enzyme catalyzes deadenylation when it is complexed with the other form (P_D) of the regulatory protein.



Figure 24.27. Structure of the Regulatory Protein P. This trimeric regulatory protein controls the modification of glutamine synthetase.

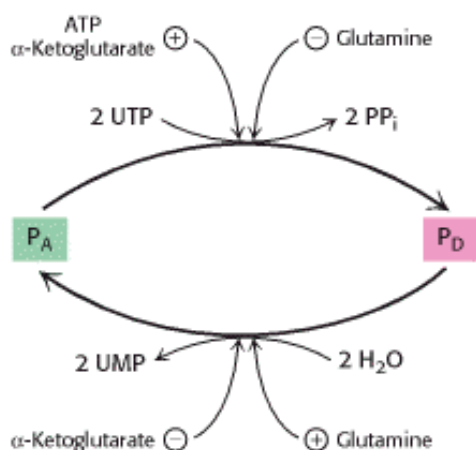


Figure 24.28. A Higher Level in the Regulatory Cascade of Glutamine Synthetase. P_A and P_D , the regulatory proteins that control the specificity of adenylyl transferase, are interconvertible. P_A is converted into P_D by uridylylation, which is reversed by hydrolysis. The enzymes catalyzing these reactions are regulated by the concentrations of metabolic intermediates.

24.4. Amino Acids Are Precursors of Many Biomolecules

In addition to being the building blocks of proteins and peptides, amino acids serve as precursors of many kinds of small molecules that have important and diverse biological roles. Let us briefly survey some of the biomolecules that are derived from amino acids (Figure 24.29).

Purines and *pyrimidines* are derived largely from amino acids. The biosynthesis of these precursors of DNA, RNA, and numerous coenzymes will be discussed in detail in [Chapter 25](#). The reactive terminus of *sphingosine*, an intermediate in the synthesis of sphingolipids, comes from serine. *Histamine*, a potent vasodilator, is derived from histidine by decarboxylation. Tyrosine is a precursor of the hormones *thyroxine* (tetraiodothyronine) and *epinephrine* and of *melanin*, a complex polymeric pigment. The neurotransmitter *serotonin* (5-hydroxytryptamine) and the *nicotinamide ring* of NAD^+ are synthesized from tryptophan. Let us now consider in more detail three particularly important biochemicals derived from amino acids.

24.4.1. Glutathione, a Gamma-Glutamyl Peptide, Serves as a Sulfhydryl Buffer and an Antioxidant

Glutathione, a tripeptide containing a sulfhydryl group, is a highly distinctive amino acid derivative with several important roles (Figure 24.30). For example, glutathione, present at high levels (≈ 5 mM) in animal cells, protects red cells from oxidative damage by serving as a sulfhydryl buffer (Section 20.5.1). It cycles between a reduced thiol form (GSH) and an oxidized form (GSSG) in which two tripeptides are linked by a disulfide bond. GSSG is reduced to GSH by *glutathione reductase*, a flavoprotein that uses NADPH as the electron source. The ratio of GSH to GSSG in most cells is greater than 500. *Glutathione plays a key role in detoxification by reacting with hydrogen peroxide and organic peroxides, the harmful by-products of aerobic life.*



Glutathione peroxidase, the enzyme catalyzing this reaction, is remarkable in having a modified amino acid containing a *selenium* (Se) atom (Figure 24.31). Specifically, its active site contains the selenium analog of cysteine, in which selenium has replaced sulfur. The selenolate (E-Se⁻) form of this residue reduces the peroxide substrate to an alcohol and is in turn oxidized to selenenic acid (E-SeOH). Glutathione then comes into action by forming a selenosulfide adduct (E-Se-S-G). A second molecule of glutathione then regenerates the active form of the enzyme by attacking the selenosulfide to form oxidized glutathione (Figure 24.32).

24.4.2. Nitric Oxide, a Short-Lived Signal Molecule, Is Formed from Arginine

Nitric oxide (NO) is an important messenger in many vertebrate signal-transduction processes. This free-radical gas is produced endogenously from *arginine* in a complex reaction that is catalyzed by *nitric oxide synthase*. NADPH and O₂ are required for the synthesis of nitric oxide (Figure 24.33). Nitric oxide acts by binding to and activating soluble guanylate cyclase, an important enzyme in signal transduction (Section 32.3.3). This enzyme is homologous to adenylate cyclase but includes a heme-containing domain that binds NO.

24.4.3. Mammalian Porphyrins Are Synthesized from Glycine and Succinyl Coenzyme A

The involvement of an amino acid in the biosynthesis of the porphyrin rings of hemes and chlorophylls was first revealed by the results of isotopic labeling experiments carried out by David Shemin and his colleagues. In 1945, they showed that the nitrogen atoms of heme were labeled after the feeding of [¹⁵N]glycine to human subjects (of whom Shemin was the first), whereas the ingestion of [¹⁵N]glutamate resulted in very little labeling.

15N labeling: A pioneer's account

"Myself as a Guinea Pig

". . . in 1944, I undertook, together with David Rittenberg, an investigation on the turnover of blood proteins of man. To this end I synthesized 66 g of glycine labeled with 35 percent ¹⁵N at a cost of \$1000 for the ¹⁵N. On 12 February 1945, I started the ingestion of the labeled glycine. Since we did not know the effect of relatively large doses of the stable isotope of nitrogen and since we believed that the maximum incorporation into the proteins could be achieved by the administration of glycine in some continual manner, I ingested 1 g samples of glycine at hourly intervals for the next 66

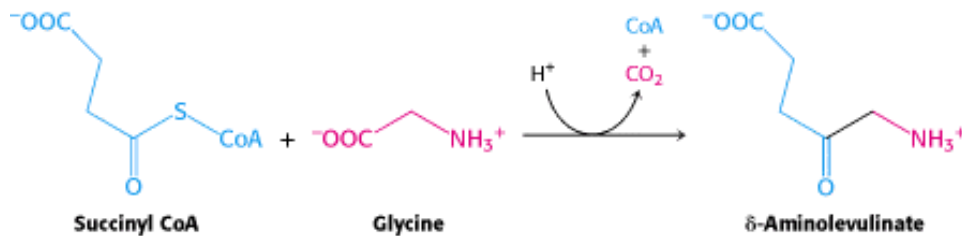
hours. . . At stated intervals, blood was withdrawn and after proper preparation the ^{15}N concentrations of different blood proteins were determined."

-David Shemin

Bioessays 10(1989):30

Using ^{14}C , which had just become available, they discovered that 8 of the carbon atoms of heme in nucleated duck erythrocytes are derived from the α -carbon atom of glycine and none from the carboxyl carbon atom. The results of subsequent studies demonstrated that the other 26 carbon atoms of heme can arise from acetate. Moreover, the ^{14}C in methyl-labeled acetate emerged in 24 of these carbons, whereas the ^{14}C in carboxyl-labeled acetate appeared only in the other 2 ([Figure 24.34](#)).

This highly distinctive labeling pattern led Shemin to propose that a heme precursor is formed by the condensation of glycine with an activated succinyl compound. In fact, *the first step in the biosynthesis of porphyrins in mammals is the condensation of glycine and succinyl CoA to form δ -aminolevulinate*.




This reaction is catalyzed by δ -aminolevulinate synthase, a PLP enzyme present in mitochondria. Two molecules of δ -aminolevulinate condense to form *porphobilinogen*, the next intermediate. Four molecules of porphobilinogen then condense head to tail to form a linear *tetrapyrrole* in a reaction catalyzed by *porphobilinogen deaminase*. The enzyme-bound linear tetrapyrrole then cyclizes to form uroporphyrinogen III, which has an asymmetric arrangement of side chains. This reaction requires a *cosynthase*. In the presence of synthase alone, uroporphyrinogen I, the nonphysiologic symmetric isomer, is produced. Uroporphyrinogen III is also a key intermediate in the synthesis of vitamin B_{12} by bacteria and that of chlorophyll by bacteria and plants ([Figure 24.35](#)).

The porphyrin skeleton is now formed. Subsequent reactions alter the side chains and the degree of saturation of the porphyrin ring (see [Figure 24.34](#)). *Coproporphyrinogen III* is formed by the decarboxylation of the acetate side chains. The desaturation of the porphyrin ring and the conversion of two of the propionate side chains into vinyl groups yield *protoporphyrin IX*. The chelation of iron finally gives *heme*, the prosthetic group of proteins such as myoglobin, hemoglobin, catalase, peroxidase, and cytochrome *c*. The insertion of the *ferrous* form of iron is catalyzed by *ferrochelatase*. Iron is transported in the plasma by *transferrin*, a protein that binds two ferric ions, and stored in tissues inside molecules of *ferritin*. The large internal cavity ($\approx 80 \text{ \AA}$ in diameter) of ferritin can hold as many as 4500 ferric ions ([Section 31.4.2](#)).

The normal human erythrocyte has a life span of about 120 days, as was first shown by the time course of ^{15}N in Shemin's own hemoglobin after he ingested ^{15}N -labeled glycine. The first step in the degradation of the heme group is the cleavage of its α -methene bridge to form the green pigment *biliverdin*, a linear tetrapyrrole. The central methene bridge of biliverdin is then reduced by *biliverdin reductase* to form *bilirubin*, a red pigment ([Figure 24.36](#)). The changing color of a bruise is a highly graphic indicator of these degradative reactions.

24.4.4. Porphyrins Accumulate in Some Inherited Disorders of Porphyrin Metabolism

 **Porphyrias** are inherited or acquired disorders caused by a deficiency of enzymes in the heme biosynthetic pathway. Porphyrin is synthesized in both the erythroblasts and the liver, and either one may be the site of a disorder. *Congenital erythropoietic porphyria*, for example, prematurely destroys erythrocytes. This disease results from insufficient uroporphyrinogen III cosynthase. In this porphyria, the synthesis of the required amount of uroporphyrinogen III is accompanied by the formation of very large quantities of uroporphyrinogen I, the useless symmetric isomer. Uroporphyrin I, coproporphyrin I, and other symmetric derivatives also accumulate. The urine of patients having this disease is red because of the excretion of large amounts of uroporphyrin I. Their teeth exhibit a strong red fluorescence under ultraviolet light because of the deposition of porphyrins. Furthermore, their *skin is usually very sensitive to light* because photoexcited porphyrins are quite reactive. *Acute intermittent porphyria* is the most prevalent of the porphyrias affecting the liver. This porphyria is characterized by the overproduction of porphobilinogen and δ -aminolevulinic acid, which results in severe abdominal pain and neurological dysfunction. The "madness" of George III, King of England during the American Revolution, is believed to have been due to this porphyria.

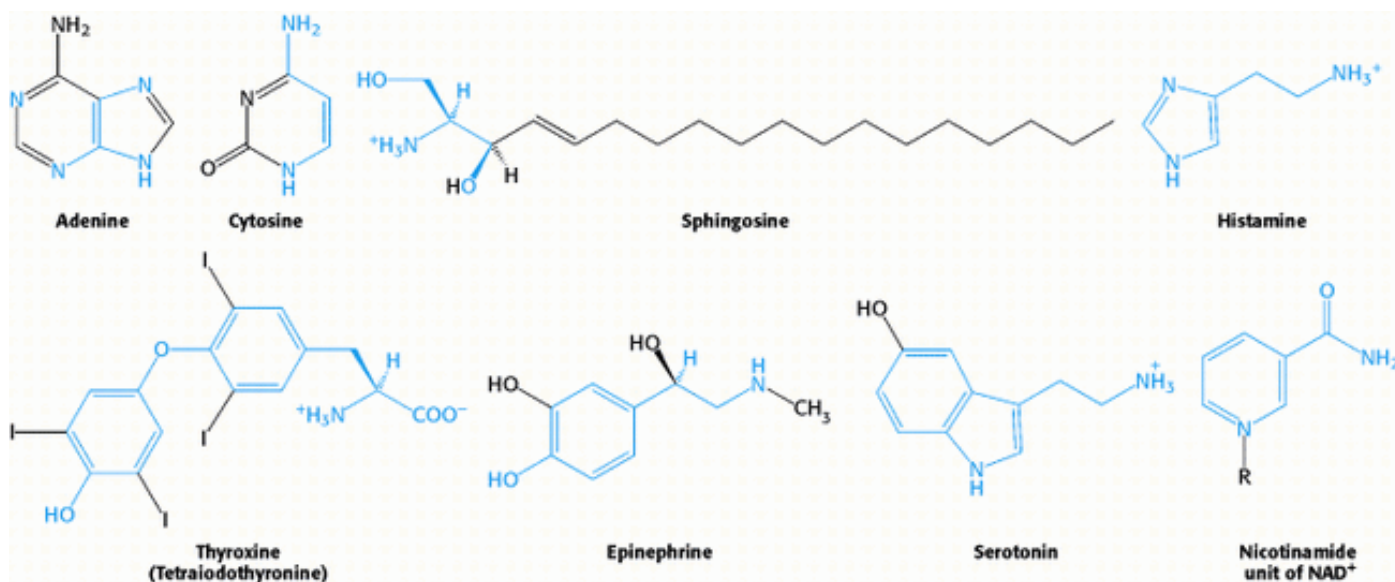


Figure 24.29. Selected Biomolecules Derived from Amino Acids. The atoms contributed by amino acids are shown in blue.

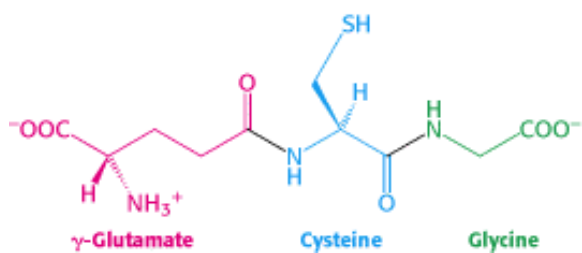


Figure 24.30. Glutathione. This tripeptide consists of a cysteine residue flanked by a glycine residue and a glutamate residue that is linked to cysteine by an isopeptide bond between glutamate's side-chain carboxylate group and cysteine's amino group.

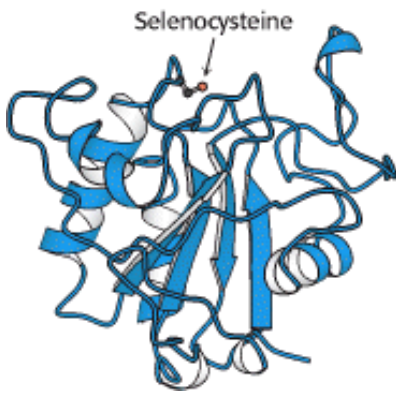


Figure 24.31. Structure of Glutathione Peroxidase. This enzyme, which has a role in peroxide detoxification, contains a selenocysteine residue in its active site.



Figure 24.32. Catalytic Cycle of Glutathione Peroxidase. [After O. Epp, R. Ladenstein, and A. Wendel. *Eur. J. Biochem.* 133(1983):51.]

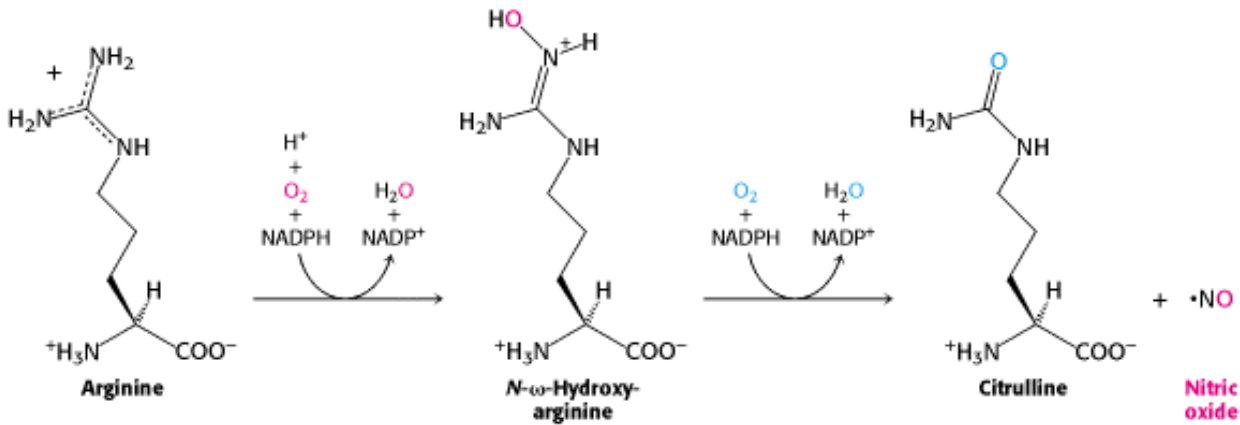


Figure 24.33. Formation of Nitric Oxide. NO is generated by the oxidation of arginine.

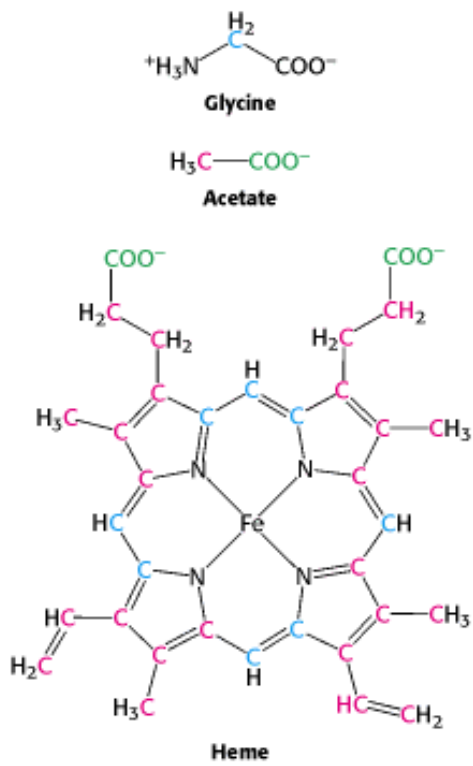


Figure 24.34. Heme Labeling. The origins of atoms in heme revealed by the results of isotopic labeling studies.

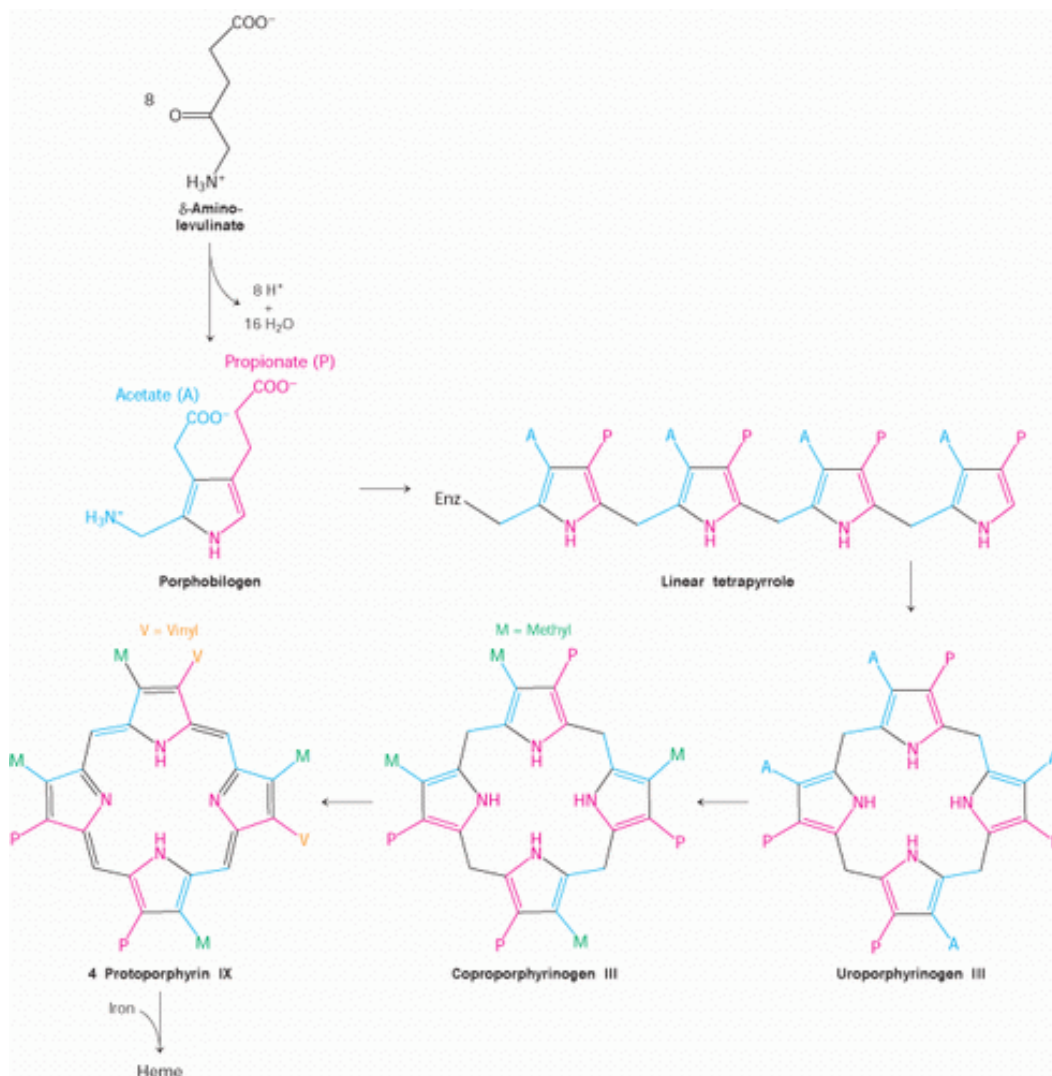




Figure 24.35. Heme Biosynthetic Pathway. The pathway for the formation of heme starts with eight molecules of δ -aminolevulinate.

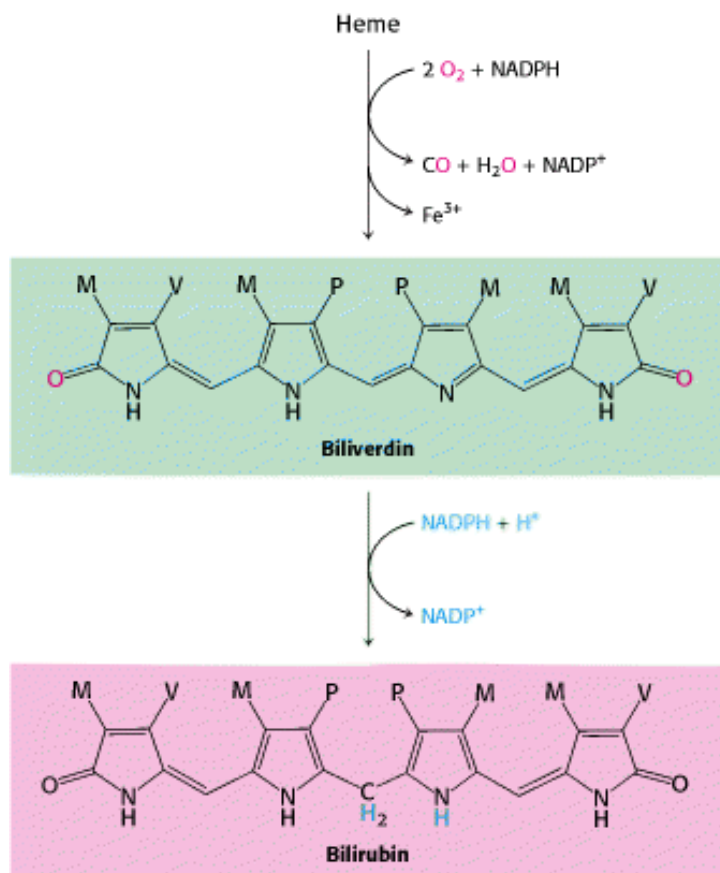


Figure 24.36. Heme Degradation. The formation of the heme-degradation products biliverdin and bilirubin is responsible for the color of bruises. Abbreviations: M, methyl; V, vinyl.

Summary

Nitrogen Fixation: Microorganisms Use ATP and a Powerful Reductant to Reduce Atmospheric Nitrogen to Ammonia

Microorganisms use ATP and reduced ferredoxin, a powerful reductant, to reduce N_2 to NH_3 . An iron-molybdenum cluster in nitrogenase deftly catalyzes the fixation of N_2 , a very inert molecule. Higher organisms consume the fixed nitrogen to synthesize amino acids, nucleotides, and other nitrogen-containing biomolecules. The major points of entry of NH_4^+ into metabolism are glutamine or glutamate.

Amino Acids Are Made from Intermediates of the Citric Acid Cycle and Other Major Pathways

Human beings can synthesize 11 of the basic set of 20 amino acids. These amino acids are called nonessential, in contrast with the essential amino acids, which must be supplied in the diet. The pathways for the synthesis of nonessential amino acids are quite simple. Glutamate dehydrogenase catalyzes the reductive amination of α -

ketoglutarate to glutamate. A transamination reaction takes place in the synthesis of most amino acids. At this step, the chirality of the amino acid is established. Alanine and aspartate are synthesized by the transamination of pyruvate and oxaloacetate, respectively. Glutamine is synthesized from NH_4^+ and glutamate, and asparagine is synthesized similarly. Proline and arginine are derived from glutamate. Serine, formed from 3-phosphoglycerate, is the precursor of glycine and cysteine. Tyrosine is synthesized by the hydroxylation of phenylalanine, an essential amino acid. The pathways for the biosynthesis of essential amino acids are much more complex than those for the nonessential ones.

Tetrahydrofolate, a carrier of activated one-carbon units, plays an important role in the metabolism of amino acids and nucleotides. This coenzyme carries one-carbon units at three oxidation states, which are interconvertible: most reduced—methyl; intermediate—methylene; and most oxidized—formyl, formimino, and methenyl. The major donor of activated methyl groups is *S*-adenosylmethionine, which is synthesized by the transfer of an adenosyl group from ATP to the sulfur atom of methionine. *S*-Adenosylhomocysteine is formed when the activated methyl group is transferred to an acceptor. It is hydrolyzed to adenosine and homocysteine, the latter of which is then methylated to methionine to complete the activated methyl cycle.

Amino Acid Biosynthesis Is Regulated by Feedback Inhibition

Most of the pathways of amino acid biosynthesis are regulated by feedback inhibition, in which the committed step is allosterically inhibited by the final product. Branched pathways require extensive interaction among the branches that includes both negative and positive regulation. The regulation of glutamine synthetase from *E. coli* is a striking demonstration of cumulative feedback inhibition and of control by a cascade of reversible covalent modifications.

Amino Acids Are Precursors of Many Biomolecules

Amino acids are precursors of a variety of biomolecules. Glutathione (γ -Glu-Cys-Gly) serves as a sulfhydryl buffer and detoxifying agent. Glutathione peroxidase, a selenoenzyme, catalyzes the reduction of hydrogen peroxide and organic peroxides by glutathione. Nitric oxide, a short-lived messenger, is formed from arginine. Porphyrins are synthesized from glycine and succinyl CoA, which condense to give δ -aminolevulinate. Two molecules of this intermediate become linked to form porphobilinogen. Four molecules of porphobilinogen combine to form a linear tetrapyrrole, which cyclizes to uroporphyrinogen III. Oxidation and side-chain modifications lead to the synthesis of protoporphyrin IX, which acquires an iron atom to form heme.

Key Terms

nitrogen fixation

nitrogenase complex

essential amino acids

nonessential amino acids

pyridoxal phosphate

tetrahydrofolate

S-adenosylmethionine (SAM)

activated methyl cycle

substrate channeling

committed step

enzyme multiplicity

cumulative feedback inhibition

Problems

1. *From sugar to amino acid.* Write a balanced equation for the synthesis of alanine from glucose.

See answer

2. *From air to blood.* What are the intermediates in the flow of nitrogen from N_2 to heme?

See answer

3. *One-carbon transfers.* Which derivative of folate is a reactant in the conversion of

(a) glycine into serine?

(b) homocysteine into methionine?

See answer

4. *Telltale tag.* In the reaction catalyzed by glutamine synthetase, an oxygen atom is transferred from the side chain of glutamate to orthophosphate, as shown by the results of ^{18}O -labeling studies. Account for this finding.

See answer

5. *Therapeutic glycine.* Isovaleric acidemia is an inherited disorder of leucine metabolism caused by a deficiency of isovaleryl CoA dehydrogenase. Many infants having this disease die in the first month of life. The administration of large amounts of glycine sometimes leads to marked clinical improvement. Propose a mechanism for the therapeutic action of glycine.

See answer

6. *Deprived algae.* Blue-green algae (cyanobacteria) form *heterocysts* when deprived of ammonia and nitrate. In this form, the algae lack nuclei and are attached to adjacent vegetative cells. Heterocysts have photosystem I activity but are entirely devoid of photosystem II activity. What is their role?

See answer

7. *Cysteine and cystine.* Most cytosolic proteins lack disulfide bonds, whereas extracellular proteins usually contain them. Why?

See answer

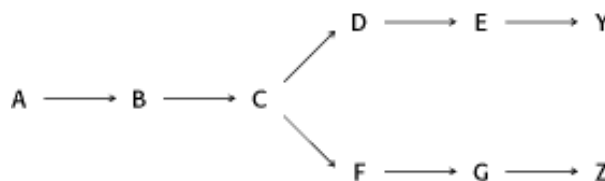
8. *To and fro.* The synthesis of δ -aminolevulinate takes place in the mitochondrial matrix, whereas the formation of porphobilinogen takes place in the cytosol. Propose a reason for the mitochondrial location of the first step in heme synthesis.

[See answer](#)

9. *Direct synthesis.* Which of the 20 amino acids can be synthesized directly from a common metabolic intermediate by a transamination reaction?

[See answer](#)

10. *Lines of communication.* For the following example of a branched pathway, propose a feedback inhibition scheme that would result in the production of equal amounts of Y and Z.



[See answer](#)

11. *Cumulative feedback inhibition.* Consider the branched pathway above. The first common step ($A \rightarrow B$) is partly inhibited by both of the final products, each acting independently of the other. Suppose that a high level of Y alone decreased the rate of the $A \rightarrow B$ step from 100 to 60 s^{-1} and that a high level of Z alone decreased the rate from 100 to 40 s^{-1} . What would the rate be in the presence of high levels of both Y and Z?

[See answer](#)

Mechanism Problems

12. *Ethylene formation.* Propose a mechanism for the conversion of S-adenosylmethionine into 1-aminocyclopropane-1-carboxylate (ACC) by ACC synthase, a PLP enzyme. What is the other product?

[See answer](#)

13. *Mirror-image serine.* Brain tissue contains substantial amounts of d-serine which is generated from l-serine by serine racemase, a PLP enzyme. Propose a mechanism for the interconversion of l- and d-serine. What is the equilibrium constant for the reaction $l\text{-serine} \rightleftharpoons d\text{-serine}$?

[See answer](#)

Chapter Integration Problems

14. *Connections.* How might increased synthesis of aspartate and glutamate affect energy production in a cell? How would the cell respond to such an effect?

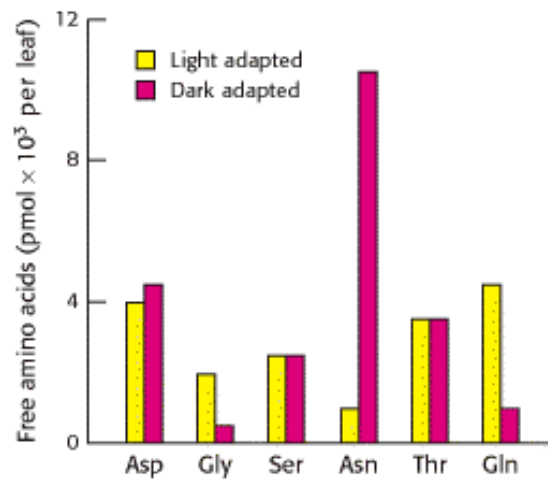
[See answer](#)

15. *Protection required.* Suppose that a mutation in bacteria resulted in diminished activity of methionine adenosyltransferase, the enzyme responsible for the synthesis of SAM from methionine and ATP. Predict how this might affect the stability of the mutated bacteria's DNA.

See answer

Chapter Integration and Data Interpretation Problem

16. *Light effects.* The adjoining graph shows the concentration of several free amino acids in light- and dark-adapted plants.



[After B. B. Buchanan, W. Gruissem, and R. L. Jones, *Biochemistry and Molecular Biology of Plants*. (American Society of Plant Physiology, 2000), [Figure 8.3](#), p. 363.]

- (a) Of the amino acids shown, which are most affected by light-dark adaptation?
- (b) Suggest a plausible biochemical explanation for the difference observed.
- (c) White asparagus, a culinary delicacy, is the result of growing asparagus plants in the dark. What chemical might you think enhances the taste of white asparagus?

See answer

Selected Readings

Where to start

- J. Kim and D.C. Rees. 1989. Nitrogenase and biological nitrogen fixation *Biochemistry* 33: 389-397. ([PubMed](#))
- P. Christen, R. Jaussi, N. Juretic, P.K. Mehta, T.I. Hale, and M. Ziak. 1990. Evolutionary and biosynthetic aspects of aspartate aminotransferase isoenzymes and other aminotransferases *Ann. N. Y. Acad. Sci.* 585: 331-338. ([PubMed](#))
- G. Schneider, H. Kack, and Y. Lindqvist. 2000. The manifold of vitamin B6 dependent enzymes *Structure Fold Des.* 8: R1-R6. ([PubMed](#))
- S.G. Rhee, P.B. Chock, and E.R. Stadtman. 1989. Regulation of *Escherichia coli* glutamine synthetase *Adv. Enzymol. Mol. Biol.* 62: 37-92. ([PubMed](#))

D. Shemin. 1989. An illustration of the use of isotopes: The biosynthesis of porphyrins *Bioessays* 10: 30-35. ([PubMed](#))

Books

Bender, D. A., 1985. *Amino Acid Metabolism* (2d ed.). Wiley.

Jordan, P. M. (Ed.), 1991. *Biosynthesis of Tetrapyrroles*. Elsevier.

Scriver, C. R., Beaudet, A. L., Sly, W. S., and Valle, D. (Eds.), 1995. *The Metabolic Basis of Inherited Disease* (7th ed.). McGraw-Hill

Meister, A., 1965. *Biochemistry of the Amino Acids* (vols. 1 and 2, 2d ed.). Academic Press.

Blakley, R. L., and Benkovic, S. J., 1989. *Folates and Pterins* (vol. 2). Wiley.

Walsh, C., 1979. *Enzymatic Reaction Mechanisms*. W. H. Freeman and Company.

Nitrogen fixation

C.M. Halbleib and P.W. Ludden. 2000. Regulation of biological nitrogen fixation *J. Nutr.* 130: 1081-1084. ([PubMed](#))

S.M. Mayer, D.M. Lawson, C.A. Gormal, S.M. Roe, and B.E. Smith. 1999. New insights into structure-function relationships in nitrogenase: A 1.6 Å resolution X-ray crystallographic study of *Klebsiella pneumoniae* MoFe-protein *J. Mol. Biol.* 292: 871-891. ([PubMed](#))

J.W. Peters, K. Fisher, and D.R. Dean. 1995. Nitrogenase structure and function: A biochemical-genetic perspective *Annu. Rev. Microbiol.* 49: 335-366. ([PubMed](#))

G.J. Leigh. 1995. The mechanism of dinitrogen reduction by molybdenum nitrogenases *Eur. J. Biochem.* 229: 14-20. ([PubMed](#))

M.K. Chan, J. Kim, and D.C. Rees. 1993. The nitrogenase FeMo-cofactor and P-cluster pair: 2.2 Å resolution studies *Science* 260: 792-794. ([PubMed](#))

M.M. Georgiadis, H. Komiya, P. Chakrabarti, D. Woo, J.J. Kornuc, and D.C. Rees. 1992. Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii* *Science* 257: 1653-1659. ([PubMed](#))

Regulation of amino acid biosynthesis

D. Eisenberg, H.S. Gill, G.M. Pfluegl, and S.H. Rotstein. 2000. Structure-function relationships of glutamine synthetases *Biochim. Biophys. Acta* 1477: 122-145. ([PubMed](#))

D.L. Purich. 1998. Advances in the enzymology of glutamine synthesis *Adv. Enzymol. Relat. Areas Mol. Biol.* 72: 9-42. ([PubMed](#))

M.M. Yamashita, R.J. Almassy, C.A. Janson, D. Cascio, and D. Eisenberg. 1989. Refined atomic model of glutamine synthetase at 3.5 Å resolution *J. Biol. Chem.* 264: 17681-17690. ([PubMed](#))

D.J. Schuller, G.A. Grant, and L.J. Banaszak. 1995. The allosteric ligand site in the V_{\max} -type cooperative enzyme phosphoglycerate dehydrogenase *Nat. Struct. Biol.* 2: 69-76. ([PubMed](#))

S.G. Rhee, R. Park, P.B. Chock, and E.R. Stadtman. 1978. Allosteric regulation of monocyclic interconvertible enzyme cascade systems: Use of *Escherichia coli* glutamine synthetase as an experimental model *Proc. Natl. Acad. Sci. USA* 75: 3138-3142. ([PubMed](#))

P.M. Wessel, E. Graciet, R. Douce, and R. Dumas. 2000. Evidence for two distinct effector-binding sites in threonine

deaminase by site-directed mutagenesis, kinetic, and binding experiments *Biochemistry* 39: 15136-15143. ([PubMed](#))

Y. Xu, P.D. Carr, T. Huber, S.G. Vasudevan, and D.L. Ollis. 2001. The structure of the PII-ATP complex *Eur. J. Biochem.* 268: 2028- 2037. ([PubMed](#))

Aromatic amino acid biosynthesis

P. Pan, E. Woehl, and M.F. Dunn. 1997. Protein architecture, dynamics and allostery in tryptophan synthase channeling *Trends Biochem. Sci.* 22: 22-27. ([PubMed](#))

A. Sachpatzidis, C. Dealwis, J.B. Lubetsky, P.H. Liang, K.S. Anderson, and E. Lolis. 1999. Crystallographic studies of phosphonate-based alpha-reaction transition-state analogues complexed to tryptophan synthase *Biochemistry* 38: 12665-12674. ([PubMed](#))

M. Weyand and I. Schlichting. 1999. Crystal structure of wild-type tryptophan synthase complexed with the natural substrate indole-3-glycerol phosphate *Biochemistry* 38: 16469-16480. ([PubMed](#))

I.P. Crawford. 1989. Evolution of a biosynthetic pathway: The tryptophan paradigm *Annu. Rev. Microbiol.* 43: 567-600. ([PubMed](#))

E.P. Carpenter, A.R. Hawkins, J.W. Frost, and K.A. Brown. 1998. Structure of dehydroquinase synthase reveals an active site capable of multistep catalysis *Nature* 394: 299-302. ([PubMed](#))

I. Schlichting, X.J. Yang, E.W. Miles, A.Y. Kim, and K.S. Anderson. 1994. Structural and kinetic analysis of a channel-impaired mutant of tryptophan synthase *J. Biol. Chem.* 269: 26591-26593. ([PubMed](#))

Glutathione

R. Edwards, D.P. Dixon, and V. Walbot. 2000. Plant glutathione *S*-transferases: Enzymes with multiple functions in sickness and in health *Trends Plant Sci.* 5: 193-198. ([PubMed](#))

S.C. Lu. 2000. Regulation of glutathione synthesis *Curr. Top. Cell Regul.* 36: 95-116. ([PubMed](#))

J.B. Schulz, J. Lindenau, J. Seyfried, and J. Dichgans. 2000. Glutathione, oxidative stress and neurodegeneration *Eur. J. Biochem.* 267: 4904-4911. ([PubMed](#))

S.C. Lu. 1999. Regulation of hepatic glutathione synthesis: Current concepts and controversies *FASEB J.* 13: 1169-1183. ([PubMed](#))

A.E. Salinas and M.G. Wong. 1991. Glutathione *S*-transferases: A review *Curr. Med. Chem.* 6: 279-309. ([PubMed](#))

Ethylene and nitric oxide

J. Haendeler, A.M. Zeiher, and S. Dimmeler. 1999. Nitric oxide and apoptosis *Vitam. Horm.* 57: 49-77. ([PubMed](#))

G. Capitani, E. Hohenester, L. Feng, P. Storici, J.F. Kirsch, and J.N. Jansonius. 1999. Structure of 1-aminocyclopropane-1-carboxylate synthase, a key enzyme in the biosynthesis of the plant hormone ethylene *J. Mol. Biol.* 294: 745-756. ([PubMed](#))

A.J. Hobbs, A. Higgs, and S. Moncada. 1999. Inhibition of nitric oxide synthase as a potential therapeutic target *Annu. Rev. Pharmacol. Toxicol.* 39: 191-220. ([PubMed](#))

D.J. Stuehr. 1999. Mammalian nitric oxide synthases *Biochim. Biophys. Acta* 1411: 217-230. ([PubMed](#))

C. Chang and J.A. Shockey. 1999. The ethylene-response pathway: Signal perception to gene regulation *Curr. Opin. Plant Biol.* 2: 352-358. ([PubMed](#))

P.R. Johnson and J.R. Ecker. 1998. The ethylene gas signal transduction pathway: A molecular perspective *Annu. Rev. Genet.* 32: 227-254. ([PubMed](#))

A. Theologis. 1992. One rotten apple spoils the whole bushel: The role of ethylene in fruit ripening *Cell* 70: 181-184. ([PubMed](#))

Biosynthesis of porphyrins

F.J. Leeper. 1989. The biosynthesis of porphyrins, chlorophylls, and vitamin B₁₂ *Nat. Prod. Rep.* 6: 171-199. ([PubMed](#))

R.J. Porra and H.-U. Meisch. 1984. The biosynthesis of chlorophyll *Trends Biochem. Sci.* 9: 99-104.

25. Nucleotide Biosynthesis

An ample supply of nucleotides is essential for many life processes. First, nucleotides are the *activated precursors of nucleic acids*. As such, they are necessary for the replication of the genome and the transcription of the genetic information into RNA. Second, an adenine nucleotide, ATP, is *the universal currency of energy*. A guanine nucleotide, GTP, also serves as an energy source for a more select group of biological processes. Third, nucleotide derivatives such as UDP-glucose *participate in biosynthetic processes* such as the formation of glycogen. Fourth, nucleotides are *essential components of signal-transduction pathways*. Cyclic nucleotides such as cyclic AMP and cyclic GMP are second messengers that transmit signals both within and between cells. ATP acts as the donor of phosphoryl groups transferred by protein kinases.

In this chapter, we continue along the path begun in [Chapter 24](#), which described the incorporation of nitrogen into amino acids from inorganic sources such as nitrogen gas. The amino acids glycine and aspartate are the scaffolds on which the ring systems present in nucleotides are assembled. Furthermore, aspartate and the side chain of glutamine serve as sources of NH₂ groups in the formation of nucleotides.

Nucleotide biosynthetic pathways are tremendously important as intervention points for therapeutic agents. Many of the most widely used drugs in the treatment of cancer block steps in nucleotide biosynthesis, particularly steps in the synthesis of DNA precursors.

25.0.1. Overview of Nucleotide Biosynthesis and Nomenclature

The pathways for the biosynthesis of nucleotides fall into two classes: *de novo* pathways and *salvage* pathways ([Figure 25.1](#)). In *de novo* (from scratch) pathways, the nucleotide bases are assembled from simpler compounds. The framework for a *pyrimidine* base is assembled first and then attached to ribose. In contrast, the framework for a *purine* base is synthesized piece by piece directly onto a ribose-based structure. These pathways comprise a small number of elementary reactions that are repeated with variation to generate different nucleotides, as might be expected for pathways that appeared very early in evolution. In salvage pathways, preformed bases are recovered and reconnected to a ribose unit.

Both *de novo* and salvage pathways lead to the synthesis of *ribonucleotides*. However, DNA is built from *deoxyribonucleotides*. Consistent with the notion that RNA preceded DNA in the course of evolution, all deoxyribonucleotides are synthesized from the corresponding ribonucleotides. The deoxyribose sugar is generated by the reduction of ribose within a fully formed nucleotide. Furthermore, the methyl group that distinguishes the thymine of DNA from the uracil of RNA is added at the last step in the pathway.

The nomenclature of nucleotides and their constituent units was presented earlier ([Section 5.1.2](#)). Recall that a *nucleoside* consists of a purine or pyrimidine base linked to a sugar and that a *nucleotide* is a phosphate ester of a nucleoside. The names of the major bases of RNA and DNA, and of their nucleoside and nucleotide derivatives, are

given in [Table 25.1](#).

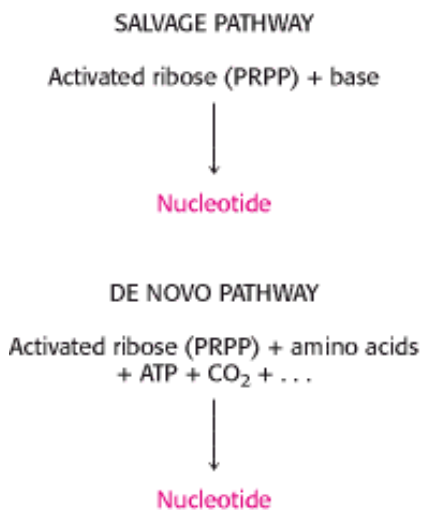
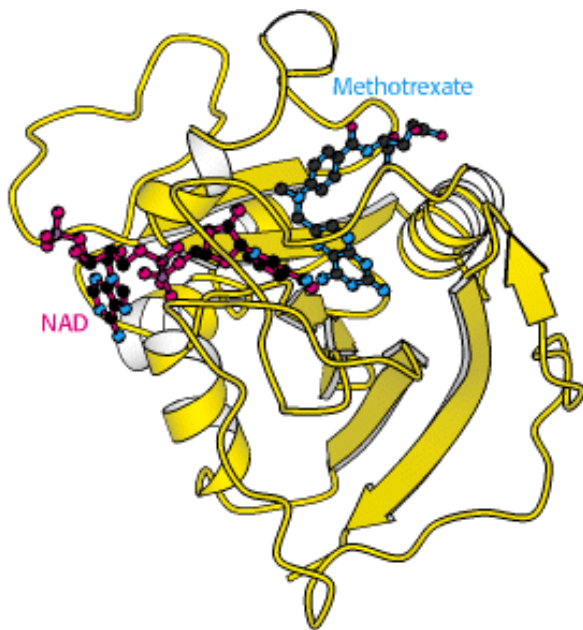


Figure 25.1. Salvage and de Novo Pathways. In a salvage pathway, a base is reattached to a ribose, activated in the form of 5-phosphoribosyl-1-pyrophosphate (PRPP). In de novo synthesis, the base itself is synthesized from simpler starting materials, including amino acids. ATP hydrolysis is required for de novo synthesis.

Table 25.1. Nomenclature of bases, nucleosides, and nucleotides

RNA		
Base	Ribonucleoside	Ribonucleotide (5 ^ʹ -monophosphate)
Adenine (A)	Adenosine	Adenylate (AMP)
Guanine (G)	Guanosine	Guanylate (GMP)
Uracil (U)	Uridine	Uridylate (UMP)
Cytosine (C)	Cytidine	Cytidylate (CMP)
DNA		
Base	Deoxyribonucleoside	Deoxyribonucleotide (5 ^ʹ -monophosphate)
Adenine (A)	Deoxyadenosine	Deoxyadenylate (dAMP)
Guanine (G)	Deoxyguanosine	Deoxyguanylate (dGMP)
Thymine (T)	Thymidine	Thymidylate (TMP)
Cytosine (C)	Deoxycytidine	Deoxycytidylate (dCMP)



Nucleotides are required for cell growth and replication. A key enzyme for the synthesis of one nucleotide is dihydrofolate reductase (right). Cells grown in the presence of methotrexate, a reductase inhibitor, respond by increasing the number of copies of the reductase gene. The bright yellow regions visible on three of the chromosomes in the fluorescence micrograph (left), which were grown in the presence of methotrexate, contain hundreds of copies of the reductase gene. [(Left) Courtesy of Dr. Barbara Trask and Dr. Joyce Hamlin.]

25.1. In de Novo Synthesis, the Pyrimidine Ring Is Assembled from Bicarbonate, Aspartate, and Glutamine

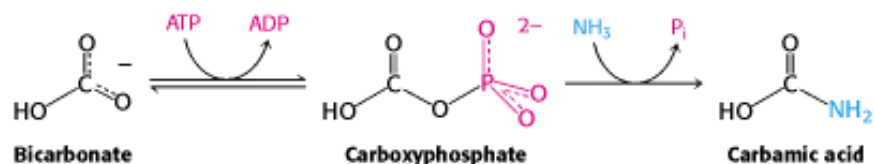
In de novo synthesis of pyrimidines, the ring is synthesized first and then it is attached to ribose to form a *pyrimidine nucleotide* (Figure 25.2). Pyrimidine rings are assembled from bicarbonate, aspartic acid, and ammonia. Although ammonia can be used directly, it is usually produced from the hydrolysis of the side chain of glutamine.

25.1.1. Bicarbonate and Other Oxygenated Carbon Compounds Are Activated by

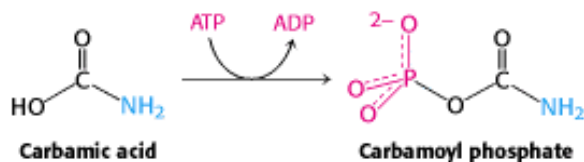
Phosphorylation

The first step in de novo pyrimidine biosynthesis is the synthesis of *carbamoyl phosphate* from bicarbonate and ammonia in a multistep process, requiring the cleavage of two molecules of ATP. This reaction is catalyzed by *carbamoyl phosphate synthetase (CPS)* (Section 23.4.1). Analysis of the structure of CPS reveals two homologous domains, each of which catalyzes an ATP-dependent step (Figure 25.3).

In the first step of the carbamoyl phosphate synthesis pathway, bicarbonate is phosphorylated by ATP to form carboxyphosphate and ADP. Ammonia then reacts with carboxyphosphate to form carbamic acid and inorganic phosphate.



The active site for this reaction lies in a domain formed by the aminoterminal third of CPS. This domain forms a structure, called an *ATP-grasp fold*, that surrounds ATP and holds it in an orientation suitable for nucleophilic attack at the γ phosphoryl group. Proteins containing ATP-grasp folds catalyze the formation of carbon-nitrogen bonds through acyl-phosphate intermediates and are widely used in nucleotide biosynthesis. In the final step catalyzed by carbamoyl phosphate synthetase, carbamic acid is phosphorylated by another molecule of ATP to form carbamoyl phosphate.



This reaction takes place in a second ATP-grasp domain within the enzyme. The active sites leading to carbamic acid formation and carbamoyl phosphate formation are very similar, revealing that this enzyme evolved by a gene duplication event. Indeed, duplication of a gene encoding an ATP-grasp domain followed by specialization was central to the evolution of nucleotide biosynthetic processes (Section 25.2.3).

25.1.2. The Side Chain of Glutamine Can Be Hydrolyzed to Generate Ammonia

Carbamoyl phosphate synthetase primarily uses glutamine as a source of ammonia. In this case, a second polypeptide component of the carbamoyl phosphate synthetase enzyme hydrolyzes glutamine to form ammonia and glutamate. The active site of the glutamine-hydrolyzing component of carbamoyl phosphate synthetase contains a catalytic dyad comprising a cysteine and a histidine residue (Figure 25.4). Such a catalytic dyad, reminiscent of the active site of cysteine proteases (Section 9.1.6), is conserved in a family of amidotransferases, including CTP synthetase (Section 25.1.6) and GMP synthetase (Section 25.2.4).

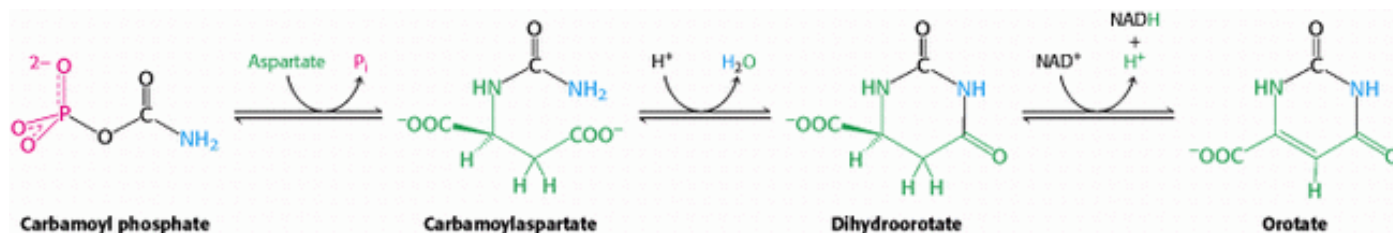
25.1.3. Intermediates Can Move Between Active Sites by Channeling

Carbamoyl phosphate synthetase contains three different active sites (see Figure 25.3), separated from one another by a total of 80 Å (Figure 25.5). Intermediates generated at one site move to the next without leaving the enzyme; that is, they move by means of substrate channeling, similar to the process described for tryptophan synthetase (Section 24.2.11). The ammonia generated in the glutamine-hydrolysis active site travels 45 Å through a channel within the enzyme to reach the

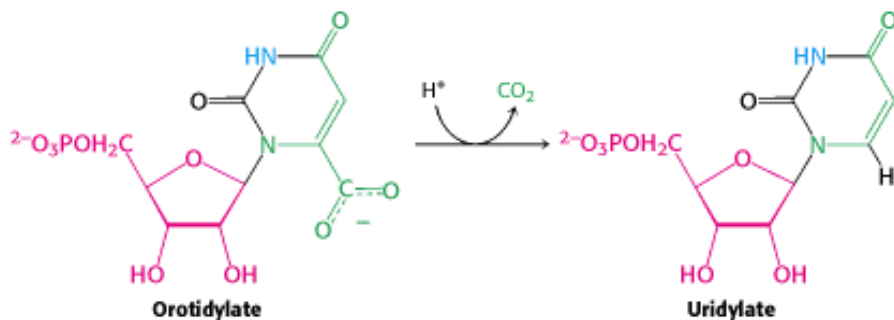
site at which carboxyphosphate has been generated. The carbamic acid generated at this site diffuses an additional 35 Å through an extension of the channel to reach the site at which carbamoyl phosphate is generated. This channeling serves two roles: (1) intermediates generated at one active site are captured with no loss caused by diffusion; and (2) labile intermediates, such as carboxyphosphate and carbamic acid (which decompose in less than 1 s at pH 7), are protected from hydrolysis.

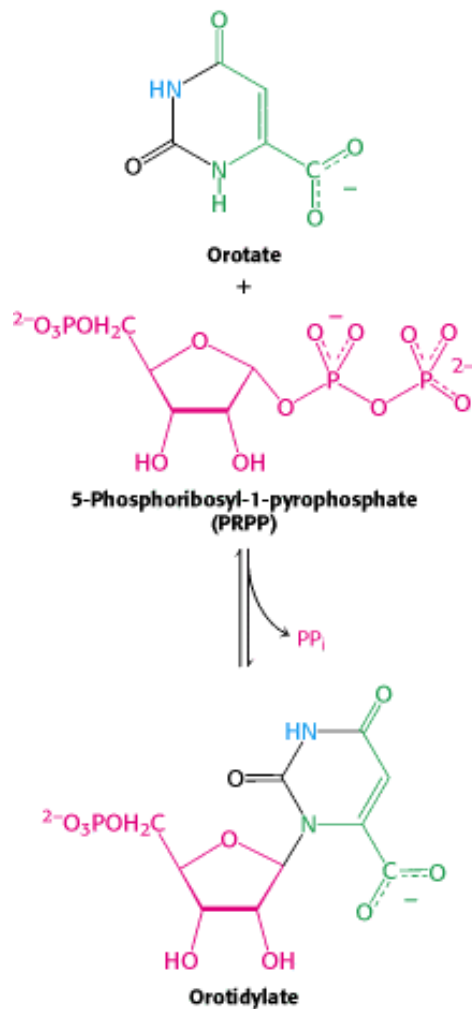
25.1.4. Orotate Acquires a Ribose Ring from PRPP to Form a Pyrimidine Nucleotide and Is Converted into Uridylate

Carbamoyl phosphate reacts with aspartate to form carbamoylaspartate in a reaction catalyzed by *aspartate transcarbamoylase* (Section 10.1). Carbamoylaspartate then cyclizes to form dihydroorotate which is then oxidized by NAD^+ to form orotate.



At this stage, orotate couples to ribose, in the form of *5-phosphoribosyl-1-pyrophosphate (PRPP)*, a form of ribose activated to accept nucleotide bases. PRPP is synthesized from ribose-5-phosphate, formed by the pentose phosphate pathway, by the addition of pyrophosphate from ATP. Orotate reacts with PRPP to form *orotidylate*, a pyrimidine nucleotide. This reaction is driven by the hydrolysis of pyrophosphate. The enzyme that catalyzes this addition, *pyrimidine phosphoribosyltransferase*, is homologous to a number of other phosphoribosyltransferases that add different groups to PRPP to form the other nucleotides. Orotidylate is then decarboxylated to form *uridylate (UMP)*, a major pyrimidine nucleotide that is a precursor to RNA. This reaction is catalyzed by *orotidylate decarboxylase*.





This enzyme is one of the most proficient enzymes known. In its absence, decarboxylation is extremely slow and is estimated to take place once every 78 million years; with the enzyme present, it takes place approximately once per second, a rate enhancement of 10^{17} -fold!

25.1.5. Nucleotide Mono-, Di-, and Triphosphates Are Interconvertible

How is the other major pyrimidine ribonucleotide, cytidine, formed? It is synthesized from the uracil base of UMP, but UMP is converted into UTP before the synthesis can take place. Recall that the diphosphates and triphosphates are the active forms of nucleotides in biosynthesis and energy conversions. Nucleoside monophosphates are converted into nucleoside triphosphates in stages. First, nucleoside monophosphates are converted into diphosphates by specific *nucleoside monophosphate kinases* that utilize ATP as the phosphoryl-group donor (Section 9.4). For example, UMP is phosphorylated to UDP by *UMP kinase*.



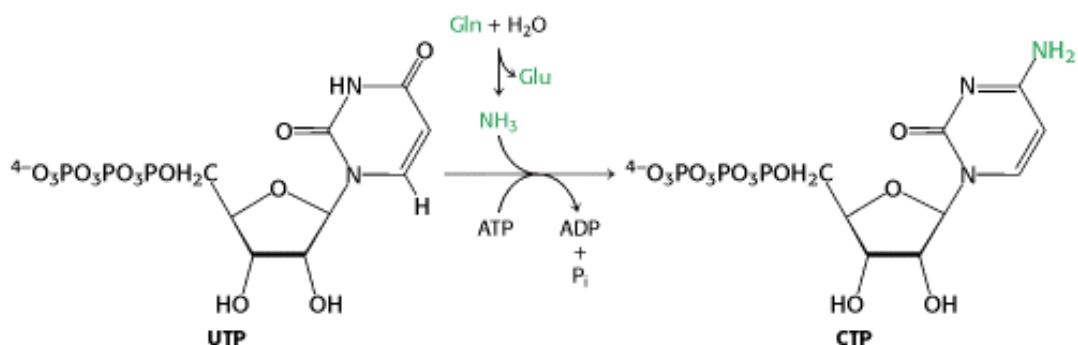
Nucleoside diphosphates and triphosphates are interconverted by *nucleoside diphosphate kinase*, an enzyme that has broad specificity, in contrast with the monophosphate kinases. X and Y can represent any of several ribonucleosides or even deoxyribonucleosides.



25.1.6. CTP Is Formed by Amination of UTP

After uridine triphosphate has been formed, it can be transformed into *cytidine triphosphate* by the replacement of a

carbonyl group by an amino group.



Like the synthesis of carbamoyl phosphate, this reaction requires ATP and uses glutamine as the source of the amino group. The reaction proceeds through an analogous mechanism in which the O-4 atom is phosphorylated to form a reactive intermediate, and then the phosphate is displaced by ammonia, freed from glutamine by hydrolysis. CTP can then be used in many biochemical processes, including RNA synthesis.

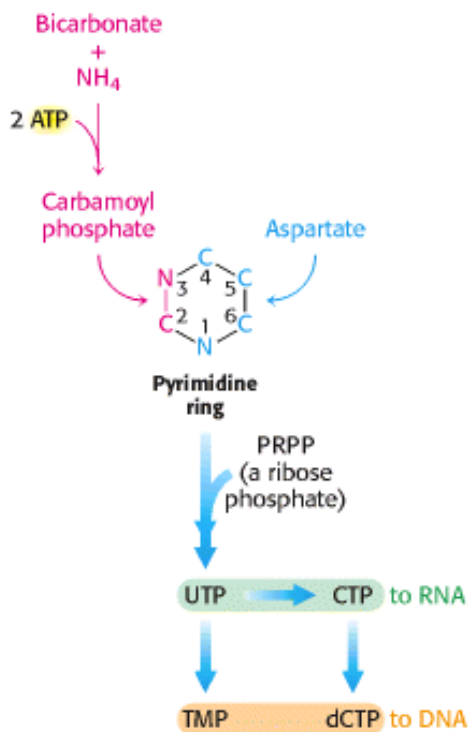


Figure 25.2. de Novo Pathway for Pyrimidine Nucleotide Synthesis. The C-2 and N-3 atoms in the pyrimidine ring come from carbamoyl phosphate, whereas the other atoms of the ring come from aspartate.



Figure 25.3. Structure of Carbamoyl Phosphate Synthetase. This enzyme consists of two chains. The smaller chain (yellow) contains a site for glutamine hydrolysis to generate ammonia. The larger chain includes two ATP-grasp domains (blue and red). In one ATP-grasp domain (blue), bicarbonate is phosphorylated to carboxyphosphate, which then reacts with ammonia to generate carbamic acid. In the other ATP-grasp domain, the carbamic acid is phosphorylated to produce carbamoyl phosphate.

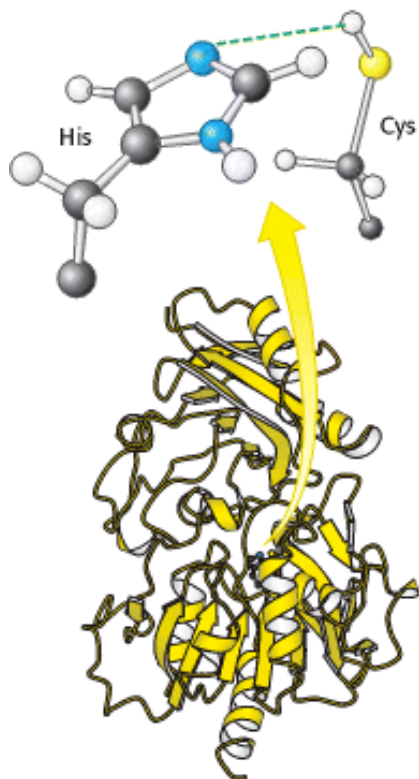



Figure 25.4. Ammonia-Generation Site. The smaller domain of carbamoyl phosphate synthetase contains an active site

 for the hydrolysis of the side chain carboxamide of glutamine to generate ammonia. Key residues in this active site include a cysteine residue and a histidine residue.

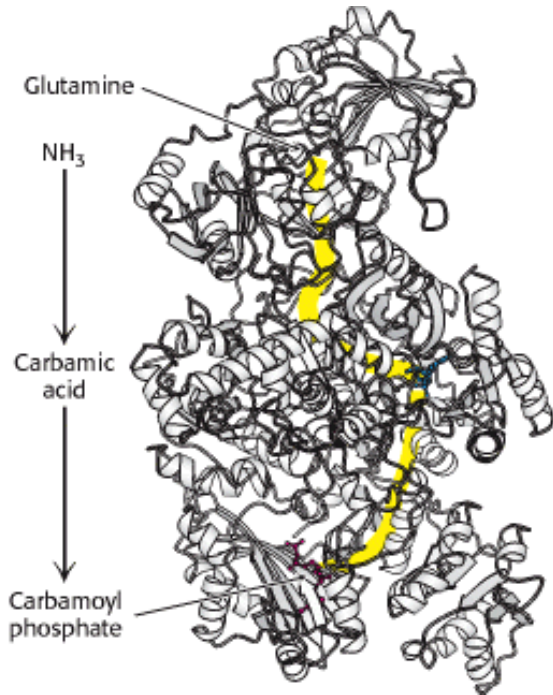


Figure 25.5. Substrate Channeling. The three active sites of carbamoyl phosphate synthetase are linked by a channel (yellow) through which intermediates pass. Glutamine enters one active site, and carbamoyl phosphate, which includes the nitrogen atom from the glutamine side chain, leaves another 80 Å away.

25.2. Purine Bases Can Be Synthesized de Novo or Recycled by Salvage Pathways

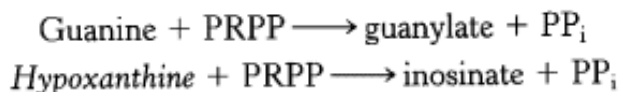
Purine nucleotides can be synthesized in two distinct pathways. First, purines are synthesized de novo, beginning with simple starting materials such as amino acids and bicarbonate (Figure 25.6). Unlike the case for pyrimidines, the purine bases are assembled already attached to the ribose ring. Alternatively, purine bases, released by the hydrolytic degradation of nucleic acids and nucleotides, can be salvaged and recycled. Purine salvage pathways are especially noted for the energy that they save and the remarkable effects of their absence (Section 25.6.2).

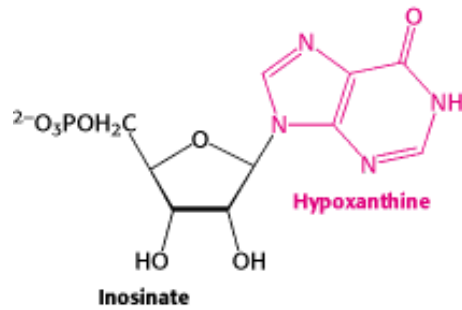
25.2.1. Salvage Pathways Economize Intracellular Energy Expenditure

Free purine bases, derived from the turnover of nucleotides or from the diet, can be attached to PRPP to form purine nucleoside monophosphates, in a reaction analogous to the formation of orotidylate. Two salvage enzymes with different specificities recover purine bases. *Adenine phosphoribosyltransferase* catalyzes the formation of adenylate



whereas *hypoxanthine-guanine phosphoribosyltransferase (HGPRT)* catalyzes the formation of guanylate as well as *inosinate* (inosine monophosphate, IMP), a precursor of guanylate and adenylate (Section 25.2.4).

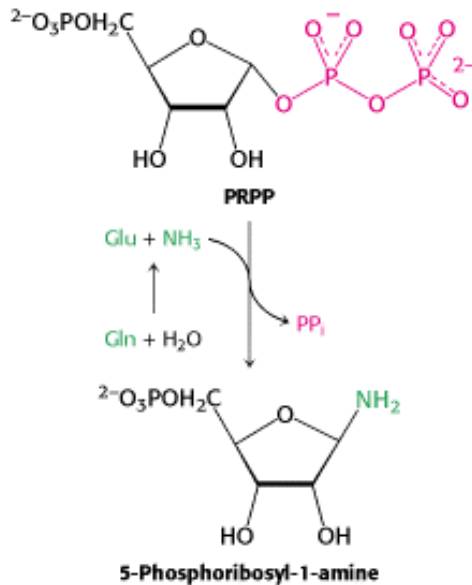




Similar salvage pathways exist for pyrimidines. Pyrimidine phosphoribosyltransferase will reconnect uracil, but not cytosine, to PRPP.


25.2.2. The Purine Ring System Is Assembled on Ribose Phosphate

De novo purine biosynthesis, like pyrimidine biosynthesis, requires PRPP, but for purines, PRPP provides the foundation on which the bases are constructed step by step. The initial committed step is the displacement of pyrophosphate by ammonia, rather than by a preassembled base, to produce *5-phosphoribosyl-1-amine*, with the amine in the β configuration.

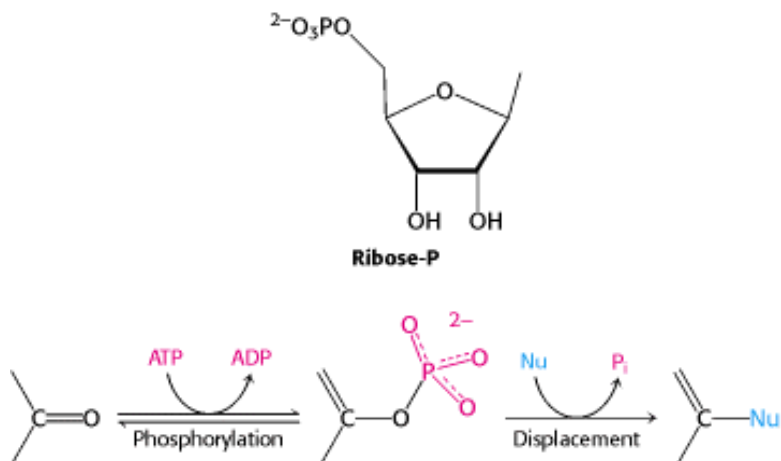


Glutamine phosphoribosyl amidotransferase catalyzes this reaction. This enzyme comprises two domains: the first is homologous to the phosphoribosyltransferases in salvage pathways, whereas the second produces ammonia from glutamine by hydrolysis. However, this glutamine-hydrolysis domain is distinct from the domain that performs the same function in carbamoyl phosphate synthetase. In glutamine phosphoribosyl amidotransferase, a cysteine residue located at the amino terminus facilitates glutamine hydrolysis. To prevent wasteful hydrolysis of either substrate, the amidotransferase assumes the active configuration only on binding of both PRPP and glutamine. As is the case with carbamoyl phosphate synthetase, the ammonia generated at the glutamine-hydrolysis active site passes through a channel to reach PRPP without being released into solution.

25.2.3. The Purine Ring Is Assembled by Successive Steps of Activation by Phosphorylation Followed by Displacement

 Nine additional steps are required to assemble the purine ring. Remarkably, the first six steps are analogous reactions. Most of these steps are catalyzed by enzymes with ATP-grasp domains that are homologous to those in carbamoyl phosphate synthetase. *Each step consists of the activation of a carbon-bound oxygen atom (typically a*

carbonyl oxygen atom) by phosphorylation, followed by the displacement of a phosphoryl group by ammonia or an amine group acting as a nucleophile (Nu).



De novo purine biosynthesis proceeds as follows (Figure 25.7).

1. The carboxylate group of a glycine residue is activated by phosphorylation and then coupled to the amino group of phosphoribosylamine. A new amide bond is formed while the amino group of glycine is free to act as a nucleophile in the next step.
2. Formate is activated and then added to this amino group to form formylglycinamide ribonucleotide. In some organisms, two distinct enzymes can catalyze this step. One enzyme transfers the formyl group from N^{10} -formyltetrahydrofolate (Section 24.2.6). The other enzyme activates formate as formyl phosphate, which is added directly to the glycine amino group.
3. The inner amide group is activated and then converted into an amidine by the addition of ammonia derived from glutamine.
4. The product of this reaction, formylglycinamidine ribonucleotide, cyclizes to form the five-membered imidazole ring found in purines. Although this cyclization is likely to be favorable thermodynamically, a molecule of ATP is consumed to ensure irreversibility. The familiar pattern is repeated: a phosphoryl group from the ATP molecule activates the carbonyl group and is displaced by the nitrogen atom attached to the ribose molecule. Cyclization is thus an intramolecular reaction in which the nucleophile and phosphate-activated carbon atom are present within the same molecule.
5. Bicarbonate is activated by phosphorylation and then attacked by the exocyclic amino group. The product of the reaction in step 5 rearranges to transfer the carboxylate group to the imidazole ring. Interestingly, mammals do not require ATP for this step; bicarbonate apparently attaches directly to the exocyclic amino group and is then transferred to the imidazole ring.
6. The imidazole carboxylate group is phosphorylated again and the phosphate group is displaced by the amino group of aspartate. Thus, a six-step process links glycine, formate, ammonia, bicarbonate, and aspartate to form an intermediate that contains all but two of the atoms necessary for the formation of the purine ring.

Three more steps complete the ring construction (Figure 25.8). Fumarate, an intermediate in the citric acid cycle, is eliminated, leaving the nitrogen atom from aspartate joined to the imidazole ring. The use of aspartate as an amino-group donor and the concomitant release of fumarate are reminiscent of the conversion of citrulline into arginine in the urea cycle and these steps are catalyzed by homologous enzymes in the two pathways (Section 23.4.2). A formyl group from N^{10} -formyltetrahydrofolate is added to this nitrogen atom to form a final intermediate that cyclizes with the loss of water to form inosinate.

25.2.4. AMP and GMP Are Formed from IMP

A few steps convert inosinate into either AMP or GMP (Figure 25.9). *Adenylate* is synthesized from inosinate by the substitution of an amino group for the carbonyl oxygen atom at C-6. Again, the addition of aspartate followed by the elimination of fumarate contributes the amino group. GTP, rather than ATP, is the phosphoryl-group donor in the synthesis of the adenylosuccinate intermediate from inosinate and aspartate. In accord with the use of GTP, the enzyme that promotes this conversion, *adenylsuccinate synthase*, is structurally related to the G-protein family and does not contain an ATP-grasp domain. The same enzyme catalyzes the removal of fumarate from adenylosuccinate in the synthesis of adenylate and from 5-aminoimidazole-4-*N*-succinocarboxamide ribonucleotide in the synthesis of inosinate.

Guanylate (GMP) is synthesized by the oxidation of inosinate to xanthylate (XMP), followed by the incorporation of an amino group at C-2. NAD^+ is the hydrogen acceptor in the oxidation of inosinate. Xanthylate is activated by the transfer of an AMP group (rather than a phosphoryl group) from ATP to the oxygen atom in the newly formed carbonyl group. Ammonia, generated by the hydrolysis of glutamine, then displaces the AMP group to form guanylate, in a reaction catalyzed by *GMP synthetase*. Note that the synthesis of adenylate requires GTP, whereas the synthesis of guanylate requires ATP. This reciprocal use of nucleotides by the pathways creates an important regulatory opportunity (Section 25.4).

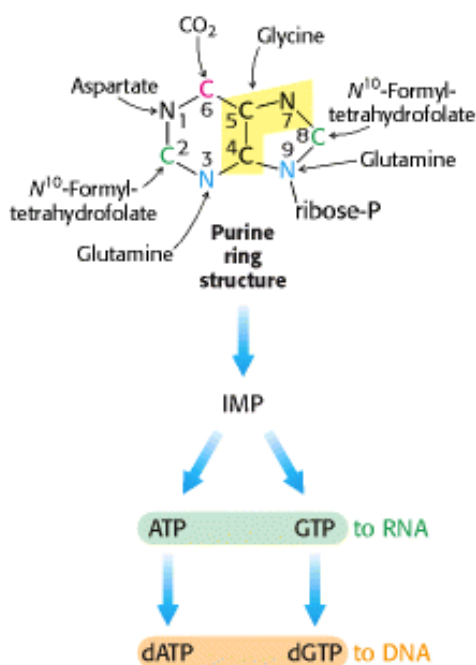


Figure 25.6. de Novo Pathway for Purine Nucleotide Synthesis. The origins of the atoms in the purine ring are indicated.

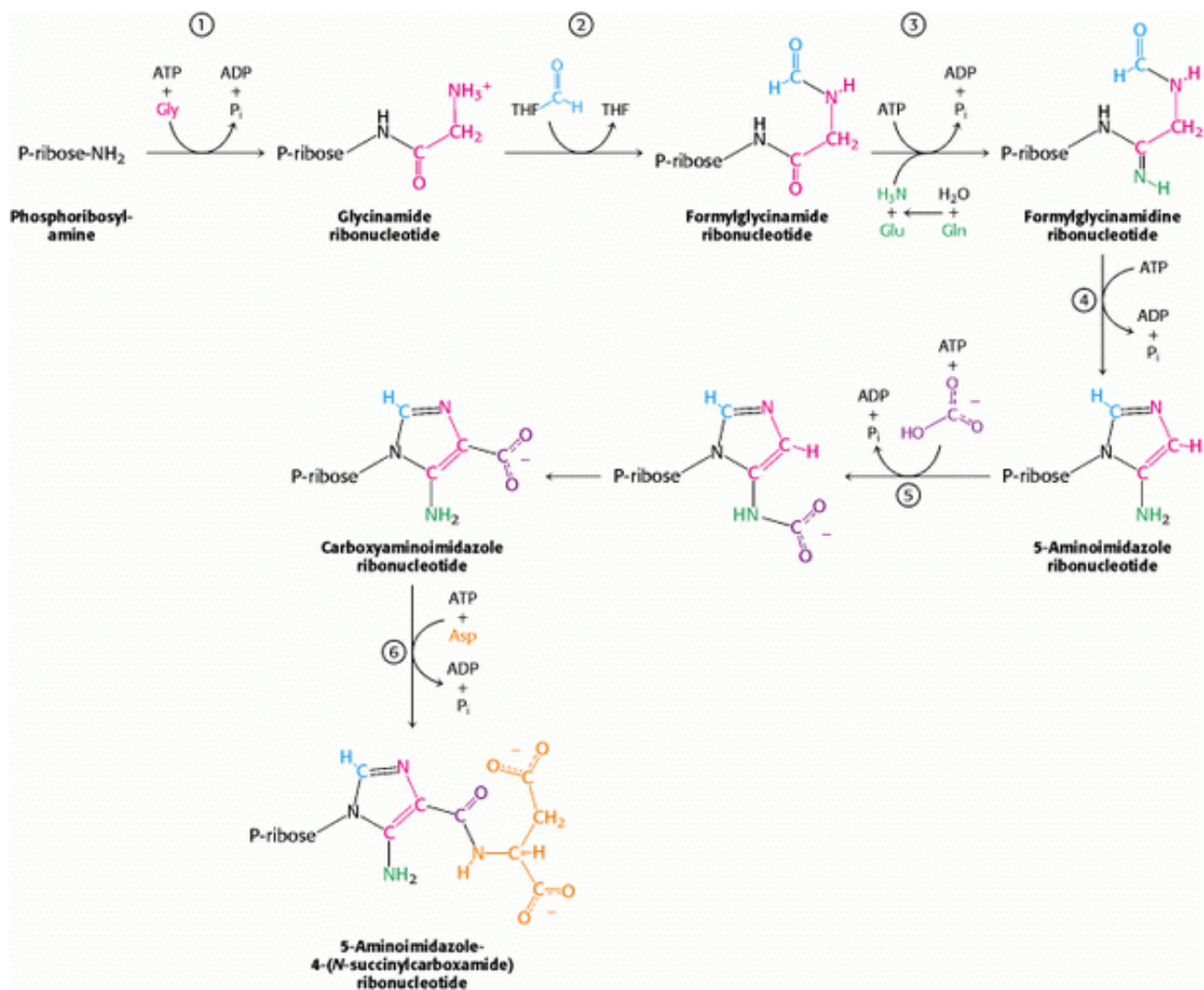


Figure 25.7. de Novo Purine Biosynthesis. 1. Glycine is coupled to the amino group of phosphoribosylamine. 2. N^{10} -Formyltetrahydrofolate transfers a formyl group to the amino group of the glycine residue. 3. The inner amide group is phosphorylated and converted into an amidine by the addition of ammonia derived from glutamine. 4. An intramolecular coupling reaction forms the five-membered imidazole ring. 5. Bicarbonate adds first to the exocyclic amino group and then to a carbon atom of the imidazole ring. 6. The imidazole carboxylate is phosphorylated, and the phosphate is displaced by the amino group of aspartate.

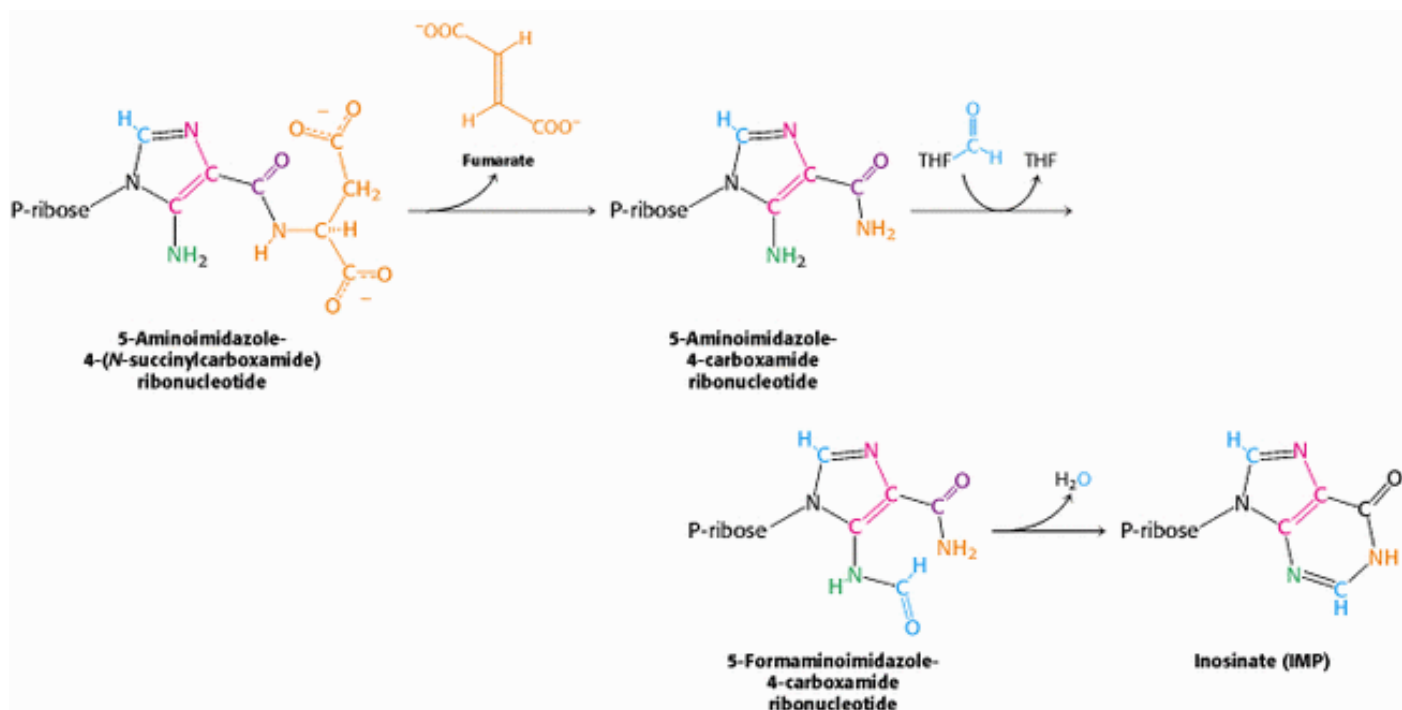


Figure 25.8. Inosinate Formation. The removal of fumarate, the addition of a second formyl group from N^{10} -formyltetrahydrofolate, and cyclization completes the synthesis of inosinate (IMP), a purine nucleotide.

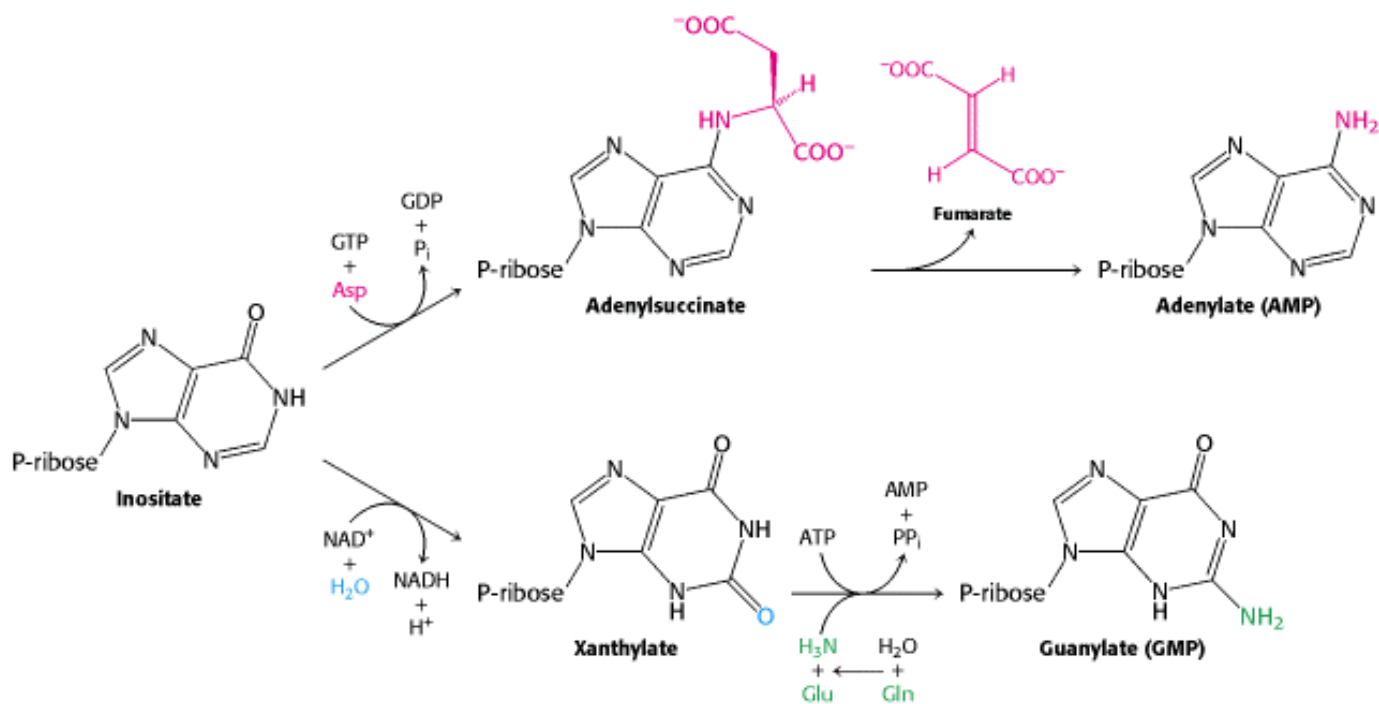


Figure 25.9. Generating AMP and GMP. Inosinate is the precursor of AMP and GMP. AMP is formed by the addition of aspartate followed by the release of fumarate. GMP is generated by the addition of water, dehydrogenation by NAD^+ , and the replacement of the carbonyl oxygen atom by $-NH_2$ derived by the hydrolysis of glutamine.

25.3. Deoxyribonucleotides Synthesized by the Reduction of Ribonucleotides Through a Radical Mechanism

We turn now to the synthesis of deoxyribonucleotides. These precursors of DNA are formed by the reduction of ribonucleotides—specifically, the 2'-hydroxyl group on the ribose moiety is replaced by a hydrogen atom. The substrates are ribonucleoside diphosphates or triphosphates, and the ultimate reductant is NADPH. The enzyme *ribonucleotide reductase* is responsible for the reduction reaction for all four ribonucleotides. The ribonucleotide reductases of different organisms are a remarkably diverse set of enzymes. The results of detailed studies have revealed that they have a common reaction mechanism, and their three-dimensional structural features indicate that these enzymes are homologous. We will focus on the best understood of these enzymes, that of *E. coli* living aerobically. This ribonucleotide reductase consists of two subunits: R1 (an 87-kd dimer) and R2 (a 43-kd dimer).


The R1 subunit contains the active site as well as two allosteric control sites (Section 25.4). This subunit includes three conserved cysteine residues and a glutamate residue, all four of which participate in the reduction of ribose to deoxyribose (Figure 25.10).

The R2 subunit's role in catalysis is to generate a remarkable free radical in each of its two chains. Each R2 chain contains a stable *tyrosyl radical* with an unpaired electron delocalized onto its aromatic ring (Figure 25.11). This very unusual free radical is generated by a nearby *iron center* consisting of two ferric (Fe^{3+}) ions bridged by an oxide (O^{2-}) ion.

In the synthesis of a deoxyribonucleotide, the hydroxyl group bonded to C-2' of the ribose ring is replaced by H, with retention of the configuration at the C-2' carbon atom (Figure 25.12).

1. The reaction begins with the transfer of an electron from a cysteine residue on R1 to the tyrosyl radical on R2. The loss of an electron generates a highly reactive *cysteine thiyl radical* within the active site of R1.
2. This radical then abstracts a hydrogen atom from C-3' of the ribose unit, generating a radical at that carbon atom.
3. The radical at C-3' promotes the release of the hydroxide ion on the carbon-2 atom. Protonated by a second cysteine residue, the departing hydroxide ion leaves as a water molecule.
4. A hydride ion (a proton on two electrons) is then transferred from a third cysteine residue to complete the reduction of the C-2' position, form a disulfide bond, and reform a C-3' radical.
5. This C-3' radical recaptures the same hydrogen atom originally abstracted by the first cysteine residue, and the deoxyribonucleotide is free to leave the enzyme.
6. The disulfide bond generated in the enzyme's active site is then reduced by specific disulfide-containing proteins, such as thioredoxin, to regenerate the active enzyme.

To complete the overall reaction, the oxidized thioredoxin generated by this process is reduced by NADH in a reaction catalyzed by *thioredoxin reductase*.

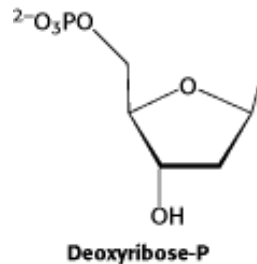
 Ribonucleotide reductases that do not contain tyrosyl radicals have been characterized in other organisms. Instead, these enzymes contain other stable radicals that are generated by other processes. For example, in one class of reductases, the coenzyme adenosylcobalamin is the radical source. Despite differences in the stable radical employed, the active sites of these enzymes are similar to that of the *E. coli* ribonucleotide reductase, and they appear to act by the same mechanism, based on the exceptional reactivity of cysteine radicals. Thus, these enzymes have a common ancestor but evolved a range of mechanisms for generating stable radical species that function well under different growth

conditions. It appears that the primordial enzymes were inactivated by oxygen, whereas enzymes such as the *E. coli* enzyme make use of oxygen to generate the initial tyrosyl radical. Note that the reduction of ribonucleotides to deoxyribonucleotides is a difficult reaction chemically, likely to require a sophisticated catalyst. The existence of a common protein enzyme framework for this process strongly suggests that proteins joined the RNA world before the evolution of DNA as a stable storage form for genetic information.

25.3.1. Thymidylate Is Formed by the Methylation of Deoxyuridylate

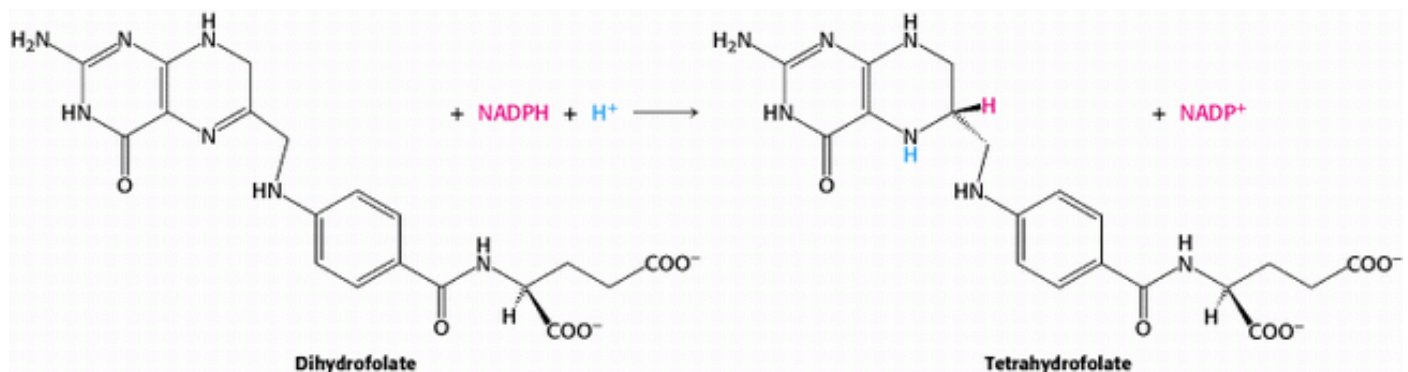
Uracil, produced by the pyrimidine synthesis pathway, is not a component of DNA. Rather, DNA contains *thymine*, a methylated analog of uracil. Another step is required to generate thymidylate from uracil. *Thymidylate synthase* catalyzes this finishing touch: deoxyuridylate (dUMP) is methylated to thymidylate (TMP). As will be discussed in [Chapter 27](#), the methylation of this nucleotide facilitates the identification of DNA damage for repair and, hence, helps preserve the integrity of the genetic information stored in DNA. The methyl donor in this reaction is N^5,N^{10} -methylenetetrahydrofolate rather than *S*-adenosylmethionine.

The methyl group becomes attached to the C-5 atom within the aromatic ring of dUMP, but this carbon atom is not a good nucleophile and cannot itself attack the appropriate group on the methyl donor. Thymidylate synthase promotes the methylation by adding a thiolate from a cysteine side chain to this ring to generate a nucleophilic species that can attack the methylene group of N^5,N^{10} -methylenetetrahydrofolate ([Figure 25.13](#)). This methylene group, in turn, is activated by distortions imposed by the enzyme that favor opening the open five-membered ring. The activated UMP's attack on the methylene group forms the new carbon-carbon bond. The intermediate formed is then converted into product: a hydride ion is transferred from the tetrahydrofolate ring to transform the methylene group into a methyl group, and a proton is abstracted from the carbon atom bearing the methyl group to eliminate the cysteine and regenerate the aromatic ring. Thus, the tetrahydrofolate derivative loses both its methylene group and a hydride ion and, hence, is oxidized to dihydrofolate. For the synthesis of more thymidylate, tetrahydrofolate must be regenerated.




25.3.2. Dihydrofolate Reductase Catalyzes the Regeneration of Tetrahydrofolate, a One-Carbon Carrier

Tetrahydrofolate is regenerated from the dihydrofolate that is produced in the synthesis of thymidylate. This regeneration is accomplished by *dihydrofolate reductase* with the use of NADPH as the reductant.



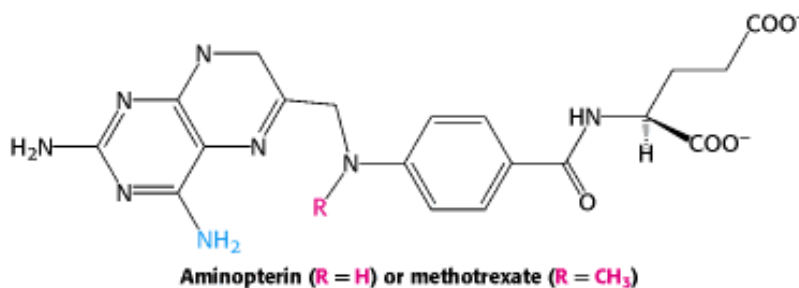
A hydride ion is directly transferred from the nicotinamide ring of NADPH to the pteridine ring of dihydrofolate. The bound dihydrofolate and NADPH are held in close proximity to facilitate the hydride transfer.

25.3.3. Several Valuable Anticancer Drugs Block the Synthesis of Thymidylate

 Rapidly dividing cells require an abundant supply of thymidylate for the synthesis of DNA. The vulnerability of these cells to the inhibition of TMP synthesis has been exploited in cancer chemotherapy. Thymidylate synthase and dihydrofolate reductase are choice targets of chemotherapy (Figure 25.14).

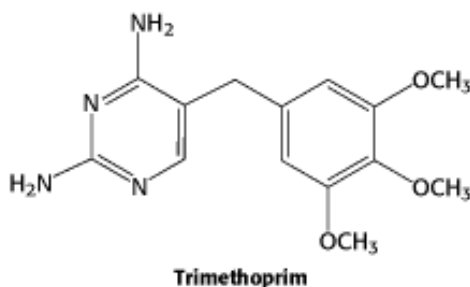
Fluorouracil, a clinically useful anticancer drug, is converted in vivo into *fluorodeoxyuridylate* (F-dUMP). This analog of dUMP irreversibly inhibits thymidylate synthase after acting as a normal substrate through part of the catalytic cycle. Recall that the formation of TMP requires the removal of a proton (H^+) from C-5 of the bound nucleotide (see Figure 25.13). However, the enzyme cannot abstract F^+ from F-dUMP, and so catalysis is blocked at the stage of the covalent complex formed by F-dUMP, methylenetetrahydrofolate, and the sulfhydryl group of the enzyme (Figure 25.15). We see here an example of *suicide inhibition*, in which an enzyme converts a substrate into a reactive inhibitor that halts the enzyme's catalytic activity (Section 8.5.2).

The synthesis of TMP can also be blocked by inhibiting the regeneration of tetrahydrofolate. Analogs of dihydrofolate, such as *aminopterin* and *methotrexate* (amethopterin), are potent competitive inhibitors ($K_i < 1$ nM) of dihydrofolate reductase.



Methotrexate is a valuable drug in the treatment of many rapidly growing tumors, such as those in acute leukemia and choriocarcinoma, a cancer derived from placental cells. However, methotrexate kills rapidly replicating cells whether they are malignant or not. Stem cells in bone marrow, epithelial cells of the intestinal tract, and hair follicles are vulnerable to the action of this folate antagonist, accounting for its toxic side effects, which include weakening of the immune system, nausea, and hair loss.

Folate analogs such as *trimethoprim* have potent antibacterial and antiprotozoal activity. Trimethoprim binds 10⁵-fold less tightly to mammalian dihydrofolate reductase than it does to reductases of susceptible microorganisms. Small differences in the active-site clefts of these enzymes account for its highly selective antimicrobial action. The combination of trimethoprim and sulfamethoxazole (an inhibitor of folate synthesis) is widely used to treat infections.



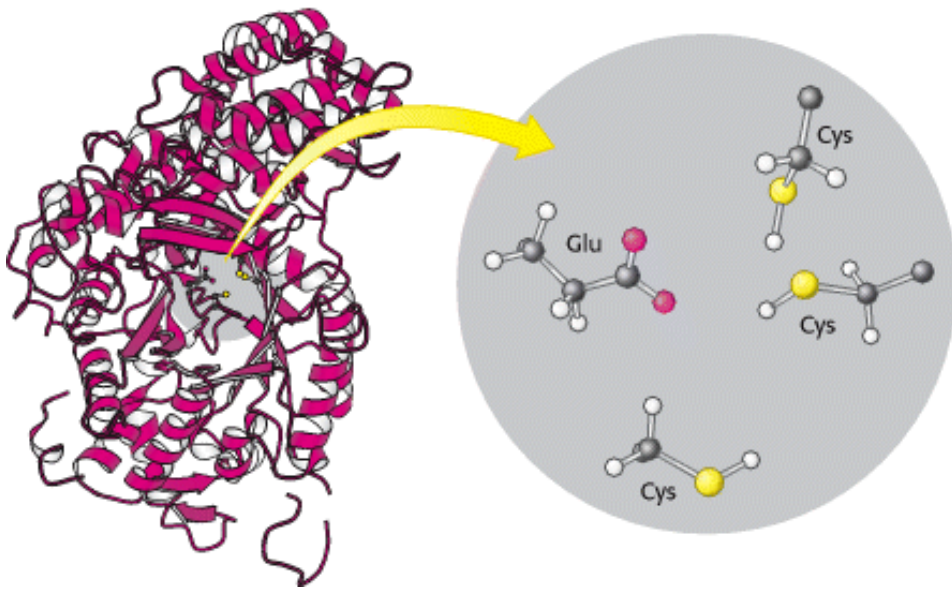


Figure 25.10. Ribonucleotide Reductase R1 Subunit. Ribonucleotide reductase reduces ribonucleotides to deoxyribonucleotides in its R1 subunit in an active site that contains three key cysteine residues and one glutamate residue. Two R1 subunits come together to form a dimer.

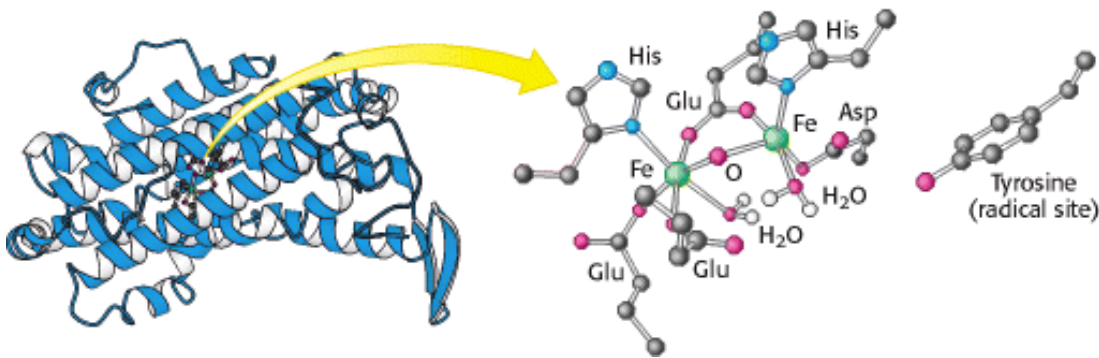


Figure 25.11. Ribonucleotide Reductase R2 Subunit. This subunit contains a stable free radical on a tyrosine residue. This radical is generated by the reaction of oxygen at a nearby site containing two iron atoms. Two R2 subunits come together to form a dimer.

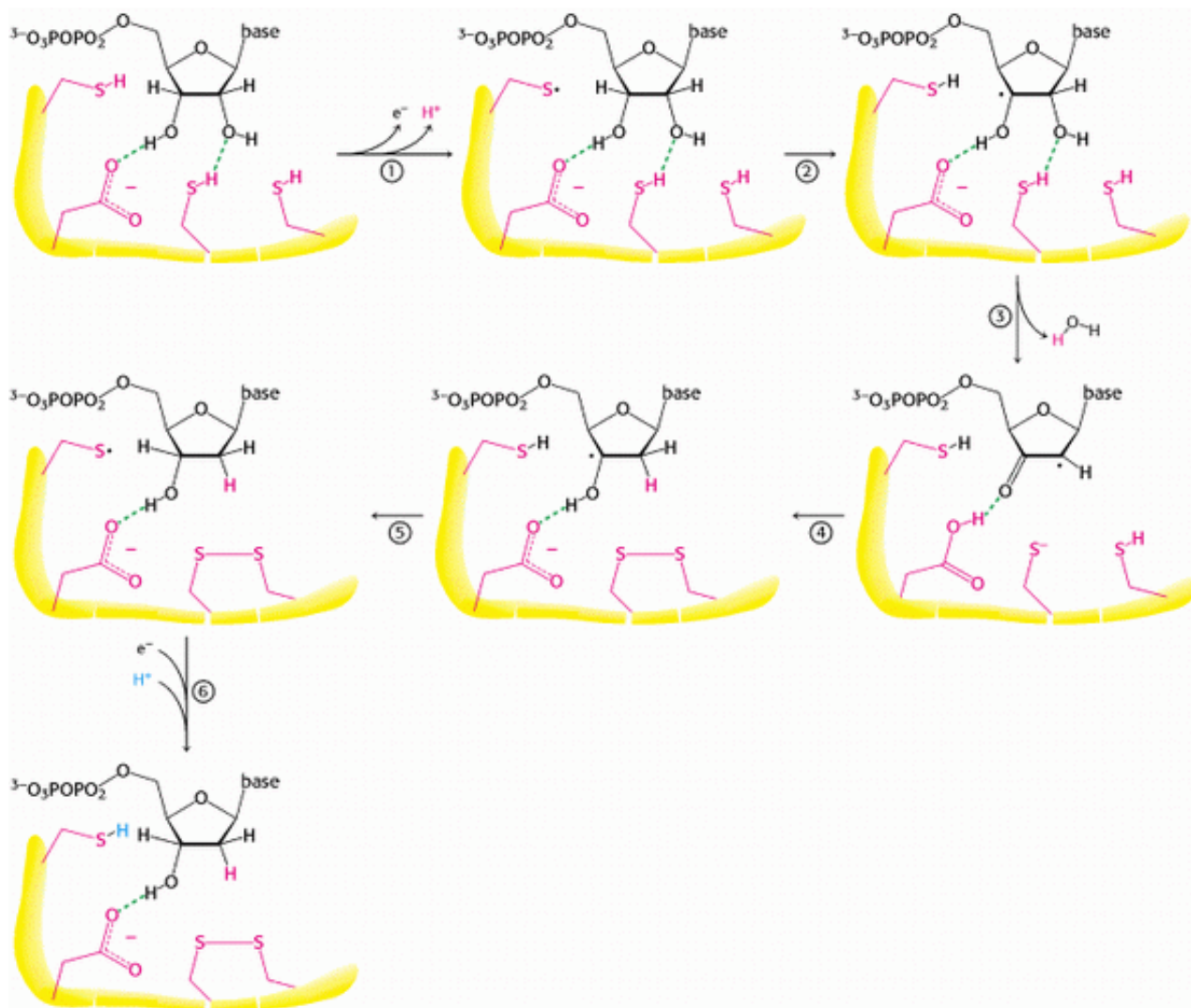


Figure 25.12. Ribonucleotide Reductase Mechanism. 1. An electron is transferred from a cysteine residue on R1 to a tyrosine radical on R2, generating a highly reactive cysteine thiyl radical. 2. This radical abstracts a hydrogen atom from C-3' of the ribose unit. 3. The radical at C-3' causes the removal of the hydroxide ion from the C-2' carbon atom. Combined with a hydrogen atom from a second cysteine residue, the hydroxide ion is eliminated as water. 4. A hydroxide ion is transferred from a third cysteine residue. 5. The C-3' radical recaptures the originally abstracted hydrogen atom. 6. An electron is transferred from R2 to reduce the thiyl radical. The deoxyribonucleotide is free to leave R1. The disulfide formed in the active site must be reduced to begin another reaction cycle.

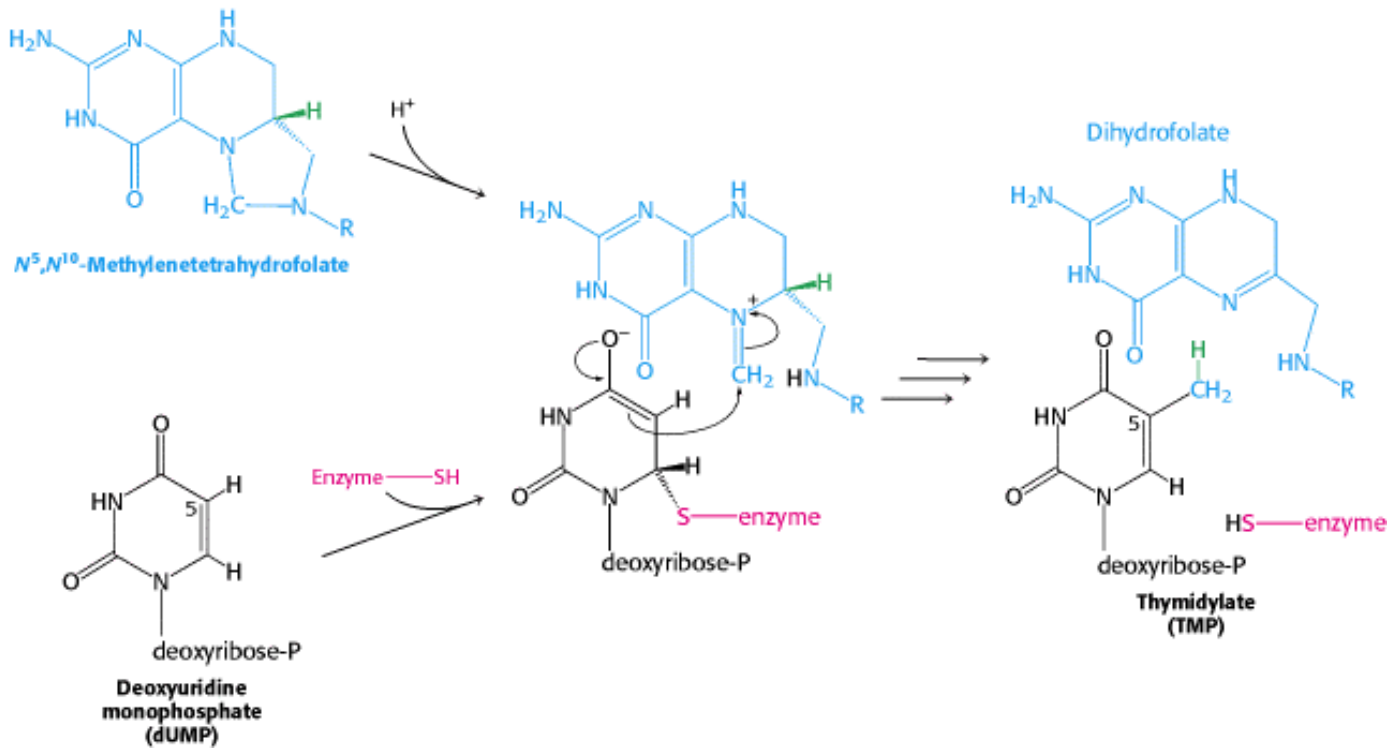


Figure 25.13. Thymidylate Synthesis. Thymidylate synthase catalyzes the addition of a methyl group (derived from *N*⁵, *N*¹⁰-methylenetetrahydrofolate to dUMP to form TMP. The addition of a thiolate from the enzyme activates dUMP. Opening the five-membered ring of the THF derivative prepares the methylene group for nucleophilic attack by the activated dUMP. The reaction is completed by the transfer of a hydride ion to form dihydrofolate.

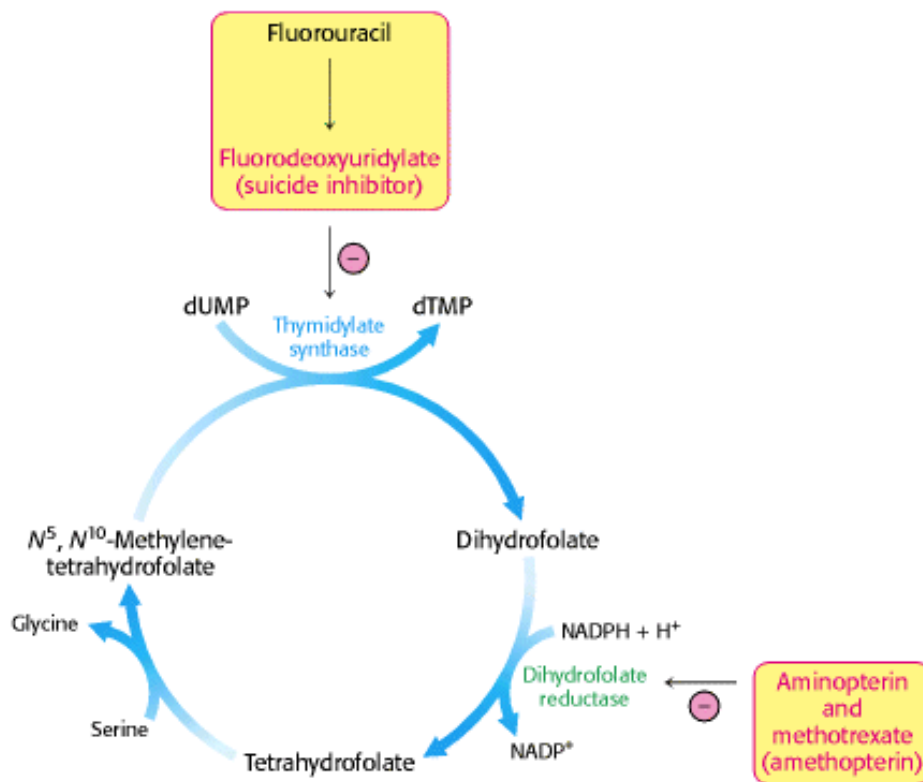


Figure 25.14. Anticancer Drug Targets. Thymidylate synthase and dihydrofolate reductase are choice targets in cancer chemotherapy because the generation of large quantities of precursors for DNA synthesis is required for rapidly dividing cancer cells.

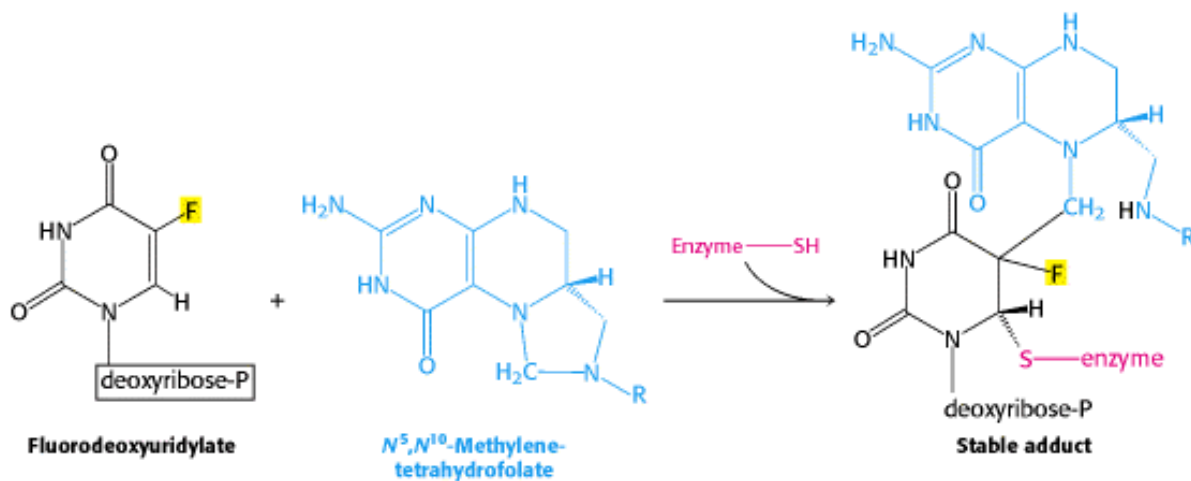


Figure 25.15. Suicide Inhibition. Fluorodeoxyuridylate (generated from fluorouracil) traps thymidylate synthase in a form that cannot proceed down the reaction pathway.

25.4. Key Steps in Nucleotide Biosynthesis Are Regulated by Feedback Inhibition

Nucleotide biosynthesis is regulated by feedback inhibition in a manner similar to the regulation of amino acid biosynthesis (Section 24.3). Indeed, aspartate transcarbamoylase, one of the key enzymes for the regulation of pyrimidine biosynthesis in bacteria, was described in detail in Chapter 10. Recall that *ATCase* is inhibited by CTP, the final product of pyrimidine biosynthesis, and stimulated by ATP. Carbamoyl phosphate synthetase is a site of feedback inhibition in both prokaryotes and eukaryotes.

The synthesis of purine nucleotides is controlled by feedback inhibition at several sites (Figure 25.16).

1. The committed step in purine nucleotide biosynthesis is the conversion of PRPP into phosphoribosylamine by *glutamine phosphoribosyl amidotransferase*. This important enzyme is feedback-inhibited by many purine ribonucleotides. It is noteworthy that AMP and GMP, the final products of the pathway, are synergistic in inhibiting the amidotransferase.
2. Inosinate is the branch point in the synthesis of AMP and GMP. *The reactions leading away from inosinate are sites of feedback inhibition.* AMP inhibits the conversion of inosinate into adenylosuccinate, its immediate precursor. Similarly, GMP inhibits the conversion of inosinate into xanthylate, its immediate precursor.
3. As already noted, GTP is a substrate in the synthesis of AMP, whereas ATP is a substrate in the synthesis of GMP. This *reciprocal substrate relation* tends to balance the synthesis of adenine and guanine ribonucleotides.

The reduction of ribonucleotides to deoxyribonucleotides is precisely controlled by allosteric interactions. Each polypeptide of the R1 subunit of the aerobic *E. coli* ribonucleotide reductase contains two allosteric sites: one of them controls the *overall activity* of the enzyme, whereas the other regulates *substrate specificity*. The overall catalytic activity of ribonucleotide reductase is diminished by the binding of dATP, which signals an abundance of deoxyribonucleotides. The binding of ATP reverses this feedback inhibition. The binding of dATP or ATP to the substrate-specificity control sites enhances the reduction of UDP and CDP, the pyrimidine nucleotides. The binding of thymidine triphosphate (TTP) promotes the reduction of GDP and inhibits the further reduction of pyrimidine ribonucleotides. The subsequent increase in the level of dGTP stimulates the reduction of ATP to dATP. This complex pattern of regulation supplies the appropriate balance of the four deoxyribonucleotides needed for the synthesis of DNA.

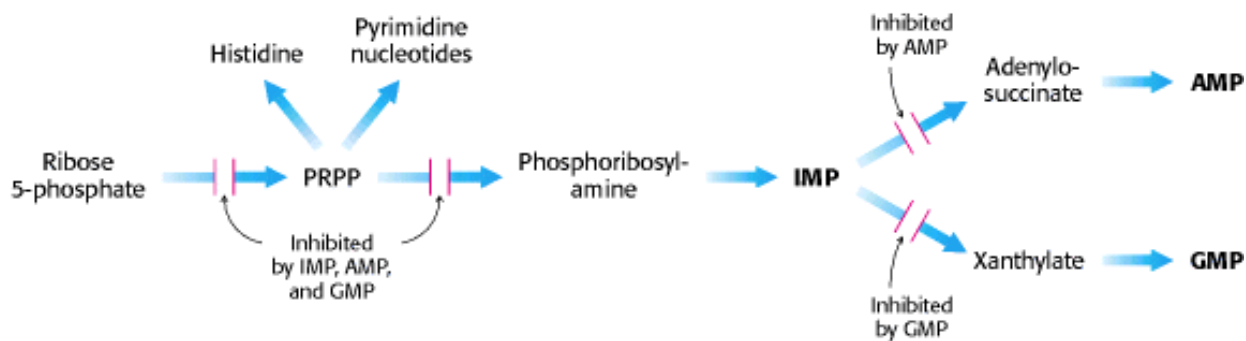
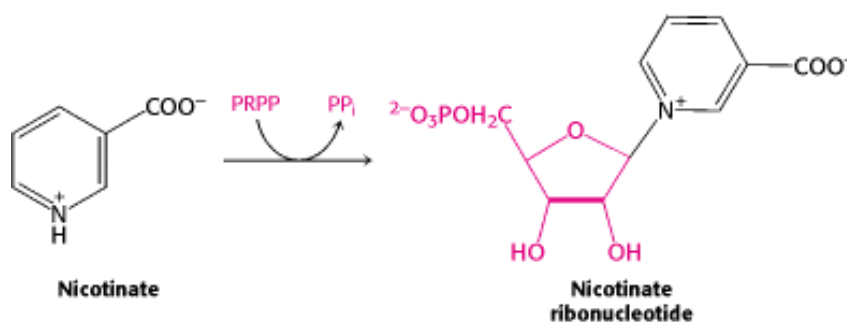



Figure 25.16. Control of Purine Biosynthesis. Feedback inhibition controls both the overall rate of purine biosynthesis and the balance between AMP and GMP production.

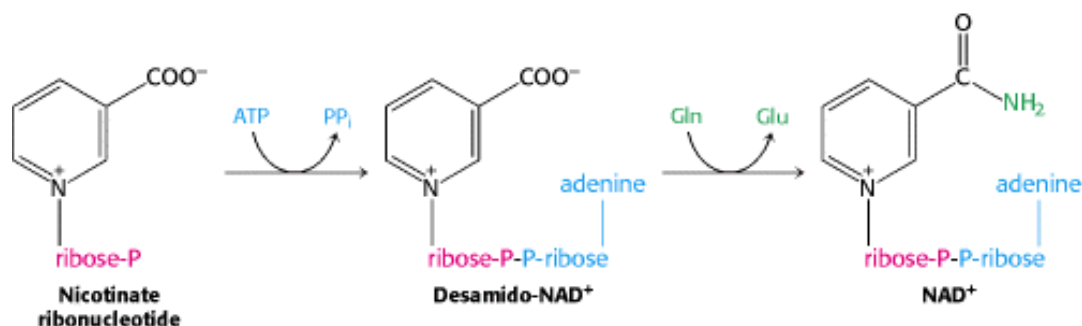
25.5. NAD⁺, FAD, and Coenzyme A Are Formed from ATP

Nucleotides are important constituents not only of RNA and DNA, but also of a number of key biomolecules considered many times in our study of biochemistry. NAD⁺ and NADP⁺, coenzymes that function in oxidation-reduction reactions, are metabolites of ATP. The first step in the synthesis of *nicotinamide adenine dinucleotide* (NAD⁺) is the formation of *nicotinate ribonucleotide* from nicotinate and PRPP.



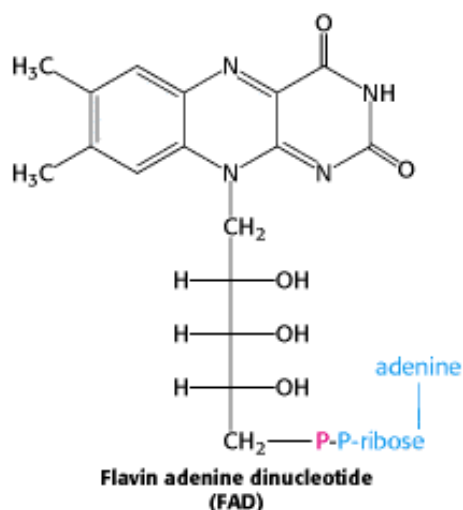
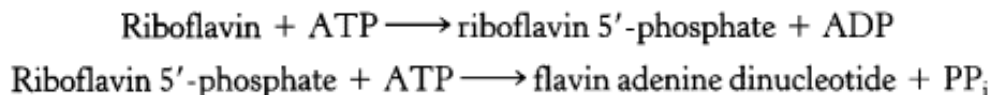
 *Nicotinate* (also called niacin or vitamin B₆) is derived from tryptophan. Human beings can synthesize the required amount of nicotinate if the supply of tryptophan in the diet is adequate. However, nicotinate must be obtained directly if the dietary intake of tryptophan is low. A dietary deficiency of tryptophan and nicotinate can lead to pellagra, a disease characterized by dermatitis, diarrhea, and dementia. An endocrine tumor that consumes large amounts of tryptophan in synthesizing the hormone and neurotransmitter serotonin (5-hydroxytryptamine) can lead to pellagra-like symptoms.

An AMP moiety is transferred from ATP to nicotinate ribonucleotide to form *desamido-NAD⁺*. The final step is the transfer of the ammonia generated from the amide group of glutamine to the nicotinate carboxyl group to form NAD⁺.



NAD⁺ is derived from NAD⁺ by phosphorylation of the 2'-hydroxyl group of the adenine ribose moiety. This transfer of a phosphoryl group from ATP is catalyzed by *NAD⁺ kinase*.

Flavin adenine dinucleotide (FAD) is synthesized from riboflavin and two molecules of ATP. Riboflavin is phosphorylated by ATP to give *riboflavin 5'-phosphate* (also called flavin mononucleotide, FMN). FAD is then formed from FMN by the transfer of an AMP moiety from a second molecule of ATP.




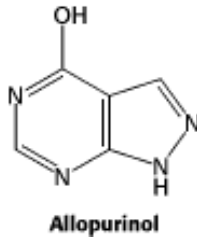
The AMP moiety of coenzyme A also comes from ATP. A common feature of the biosyntheses of NAD⁺, FAD, and CoA is the *transfer of the AMP moiety of ATP to the phosphate group of a phosphorylated intermediate*. The pyrophosphate formed in these condensations is then hydrolyzed to orthophosphate. As in many other biosyntheses, *much of the thermodynamic driving force comes from the hydrolysis of the released pyrophosphate*.


25.6. Disruptions in Nucleotide Metabolism Can Cause Pathological Conditions

Nucleotides are vital to a host of biochemical processes. It is not surprising, then, that disruption in nucleotide metabolism would have a variety of physiological effects.


25.6.1. Purines Are Degraded to Urate in Human Beings

 The nucleotides of a cell undergo continual turnover. Nucleotides are hydrolytically degraded to nucleosides by *nucleotidases*. The phosphorolytic cleavage of nucleosides to free bases and ribose 1-phosphate (or deoxyribose 1-phosphate) is catalyzed by *nucleoside phosphorylases*. Ribose 1-phosphate is isomerized by *phosphoribomutase* to ribose 5-phosphate, a substrate in the synthesis of PRPP. Some of the bases are reused to form nucleotides by salvage pathways. Others are degraded to products that are excreted (Figure 25.17). For example, AMP is degraded to the free base hypoxanthine through deamination and hydrolytic cleavage of the glycosidic bond. *Xanthine oxidase*, a molybdenum- and iron-containing flavoprotein, oxidizes hypoxanthine to *xanthine* and then to *uric acid*. Molecular oxygen, the oxidant in both reactions, is reduced to H₂O₂, which is decomposed to H₂O and O₂ by catalase. Uric acid loses a proton at physiological pH to form *urate*. *In human beings, urate is the final product of purine degradation and is excreted in the urine*. High serum levels of urate induce gout, a disease in which salts of urate crystallize and damage joints and kidneys (Figure 25.18). Allopurinol, an inhibitor of xanthine oxidase, is used to treat gout in some cases.



 The average serum level of urate in humans is close to the solubility limit. In contrast, prosimians (such as lemurs) have tenfold lower levels. A striking increase in urate levels occurred in the evolution of primates. What is the selective advantage of a urate level so high that it teeters on the brink of gout in many people? It turns out that urate has a markedly beneficial action. Urate is a highly effective scavenger of reactive oxygen species. Indeed, urate is about as effective as ascorbate (vitamin C) as an antioxidant. The increased level of urate in humans compared with prosimians and other lower primates may contribute significantly to the longer life span of humans and to lowering the incidence of human cancer.

25.6.2. Lesch-Nyhan Syndrome Is a Dramatic Consequence of Mutations in a Salvage-Pathway Enzyme

 Mutations in genes that encode nucleotide biosynthetic enzymes can reduce levels of needed nucleotides and can lead to an accumulation of intermediates. A nearly total absence of hypoxanthine-guanine phosphoribosyltransferase has unexpected and devastating consequences. The most striking expression of this inborn error of metabolism, called the *Lesch-Nyhan syndrome*, is *compulsive self-destructive behavior*. At age 2 or 3, children with this disease begin to bite their fingers and lips and will chew them off if unrestrained. These children also behave aggressively toward others. *Mental deficiency* and *spasticity* are other characteristics of the Lesch-Nyhan syndrome. Elevated levels of urate in the serum lead to the formation of kidney stones early in life, followed by the symptoms of gout years later. The disease is inherited as a sex-linked recessive disorder.

The biochemical consequences of the virtual absence of hypoxanthine-guanine phosphoribosyl transferase are *an elevated concentration of PRPP, a marked increase in the rate of purine biosynthesis by the de novo pathway, and an overproduction of urate*. The relation between the absence of the transferase and the bizarre neurologic signs is an enigma. Specific cells in the brain may be dependent on the salvage pathway for the synthesis of IMP and GMP. Indeed, transporters of the neurotransmitter dopamine are present at lower levels in affected individuals. Alternatively, cells may be damaged by the accumulation of intermediates to abnormal levels. The Lesch-Nyhan syndrome demonstrates that the salvage pathway for the synthesis of IMP and GMP is not gratuitous. Moreover, the Lesch-Nyhan syndrome reveals that *abnormal behavior such as self-mutilation and extreme hostility can be caused by the absence of a single enzyme*. Psychiatry will no doubt benefit from the unraveling of the molecular basis of such mental disorders.

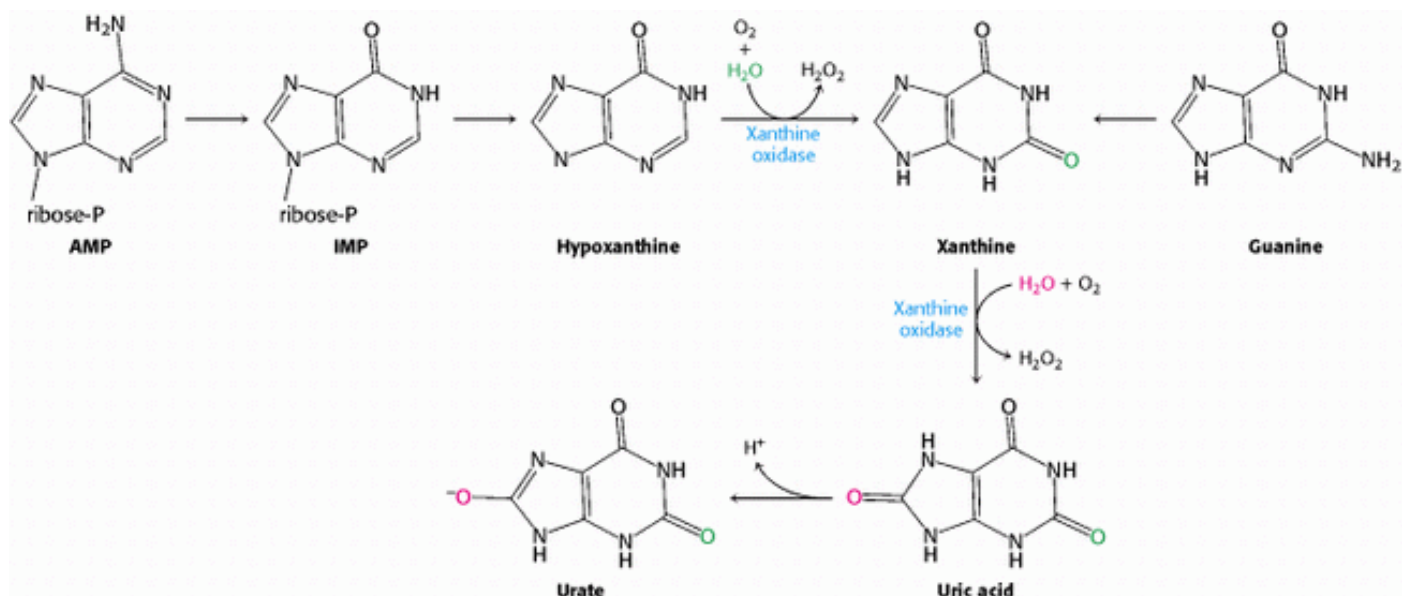


Figure 25.17. Purine Catabolism. Purine bases are converted first into xanthine and then into urate for excretion. Xanthine oxidase catalyzes two steps in this process.

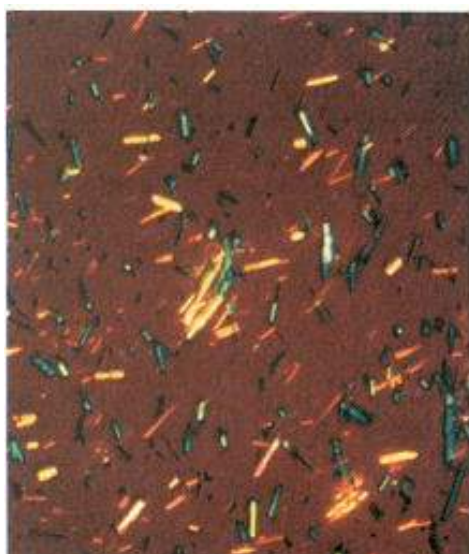


Figure 25.18. Urate Crystals. Micrograph of sodium urate crystals. Joints and kidneys are damaged by these crystals in gout. [Courtesy of Dr. James McGuire.]

Summary

In de Novo Synthesis, the Pyrimidine Ring Is Assembled from Bicarbonate, Aspartate, and Glutamine

The pyrimidine ring is assembled first and then linked to ribose phosphate to form a pyrimidine nucleotide. PRPP is the donor of the ribose phosphate moiety. The synthesis of the pyrimidine ring starts with the formation of carbamoylaspartate from carbamoyl phosphate and aspartate, a reaction catalyzed by aspartate transcarbamoylase. Dehydration, cyclization, and oxidation yield orotate, which reacts with PRPP to give orotidylate. Decarboxylation of this pyrimidine nucleotide yields UMP. CTP is then formed by the amination of UTP.

Purines Bases Can Be Synthesized de Novo or Recycled by Salvage Pathways

The purine ring is assembled from a variety of precursors: glutamine, glycine, aspartate, N^{10} -formyltetrahydrofolate, and CO_2 . The committed step in the de novo synthesis of purine nucleotides is the formation of 5-phosphoribosylamine from 5-phosphoribosyl-1-pyrophosphate (PRPP) and glutamine. The purine ring is assembled on ribose phosphate, in contrast with the de novo synthesis of pyrimidine nucleotides. The addition of glycine, followed by formylation, amination, and ring closure, yields 5-aminoimidazole ribonucleotide. This intermediate contains the completed five-membered ring of the purine skeleton. The addition of CO_2 , the nitrogen atom of aspartate, and a formyl group, followed by ring closure, yields inosinate (IMP), a purine ribonucleotide. AMP and GMP are formed from IMP. Purine ribonucleotides can also be synthesized by a salvage pathway in which a preformed base reacts directly with PRPP.

Deoxyribonucleotides Are Synthesized by the Reduction of Ribonucleotides Through a Radical Mechanism

Deoxyribonucleotides, the precursors of DNA, are formed in *E. coli* by the reduction of ribonucleoside diphosphates. These conversions are catalyzed by ribonucleotide reductase. Electrons are transferred from NADPH to sulfhydryl groups at the active sites of this enzyme by thioredoxin or glutaredoxin. A tyrosyl free radical generated by an iron center in the reductase initiates a radical reaction on the sugar, leading to the exchange of H for OH at C-2'. TMP is formed by methylation of dUMP. The donor of a methylene group and a hydride in this reaction is N^5, N^{10} -methylenetetrahydrofolate, which is converted into dihydrofolate. Tetrahydrofolate is regenerated by the reduction of dihydrofolate by NADPH. Dihydrofolate reductase, which catalyzes this reaction, is inhibited by folate analogs such as aminopterin and methotrexate. These compounds and fluorouracil, an inhibitor of thymidylate synthase, are used as anticancer drugs.

Key Steps in Nucleotide Biosynthesis Are Regulated by Feedback Inhibition

Pyrimidine biosynthesis in *E. coli* is regulated by the feedback inhibition of aspartate transcarbamoylase, the enzyme that catalyzes the committed step. CTP inhibits and ATP stimulates this enzyme. The feedback inhibition of glutamine-PRPP amidotransferase by purine nucleotides is important in regulating their biosynthesis.

NAD⁺, FAD, and Coenzyme A Are Formed from ATP

Nucleotides are important constituents not only of RNA and DNA, but also of a number of other key biomolecules. Coenzymes NAD^+ and FAD, prominent in oxidation-reduction reactions, have ADP as an important constituent. The acyl-group activation compound, coenzyme A, is also derived from ATP.

Disruptions in Nucleotide Metabolism Can Cause Pathological Conditions

Purines are degraded to urate in human beings. Gout, a disease that affects joints and leads to arthritis, is associated with the excessive accumulation of urate. The Lesch-Nyhan syndrome, a genetic disease characterized by self-mutilation, mental deficiency, and gout, is caused by the absence of hypoxanthine-guanine phosphoribosyltransferase. This enzyme is essential for the synthesis of purine nucleotides by the salvage pathway.

Key Terms

nucleoside

nucleotide

pyrimidine

carbamoyl phosphate synthetase (CSP)

ATP-grasp fold

5-phosphoribosyl-1-pyrophosphate (PRPP)

orotidylate

purine

salvage pathway

inosinate

hypoxanthine

glutamine phosphoribosyl amidotransferase

ribonucleotide reductase

thymidylate synthase

dihydrofolate reductase

Problems

1. *Activated ribose phosphate.* Write a balanced equation for the synthesis of PRPP from glucose through the oxidative branch of the pentose phosphate pathway.

See answer

2. *Making a pyrimidine.* Write a balanced equation for the synthesis of orotate from glutamine, CO₂, and aspartate.

See answer

3. *Identifying the donor.* What is the activated reactant in the biosynthesis of each of these compounds?

(a) Phosphoribosylamine

(b) Carbamoylaspartate

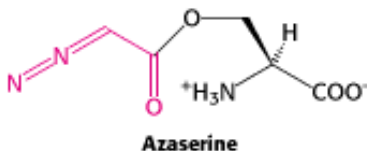
(c) Orotidylate (from orotate)

(d) Nicotinate ribonucleotide

(e) Phosphoribosylanthranilate

See answer

4. *Inhibiting purine biosynthesis.* Amidotransferases are inhibited by the antibiotic azaserine (*O*-diazooacetyl-L-serine), which is an analog of glutamine.



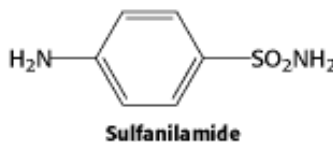
Which intermediates in purine biosynthesis would accumulate in cells treated with azaserine?

See answer

5. *The price of methylation.* Write a balanced equation for the synthesis of TMP from dUMP that is coupled to the conversion of serine into glycine.

See answer

6. *Sulfa action.* Bacterial growth is inhibited by sulfanilamide and related sulfa drugs, and there is a concomitant accumulation of 5-aminoimidazole-4-carboxamide ribonucleotide. This inhibition is reversed by the addition of *p*-aminobenzoate.



Propose a mechanism for the inhibitory effect of sulfanilamide.

See answer

7. *A generous donor.* What major biosynthetic reactions utilize PRPP?

See answer

8. *HAT medium.* Mutant cells unable to synthesize nucleotides by salvage pathways are very useful tools in molecular and cell biology. Suppose that cell A lacks thymidine kinase, the enzyme catalyzing the phosphorylation of thymidine to thymidylate, and that cell B lacks hypoxanthine-guanine phosphoribosyl transferase.

(a) Cell A and cell B do not proliferate in a *HAT* medium containing hypoxanthine, aminopterin or amethopterin (methotrexate), and thymine. However, cell C formed by the fusion of cells A and B grows in this medium. Why?

(b) Suppose that you wanted to introduce foreign genes into cell A. Devise a simple means of distinguishing between cells that have taken up foreign DNA and those that have not.

See answer

9. *Adjunct therapy.* Allopurinol is sometimes given to patients with acute leukemia who are being treated with anticancer drugs. Why is allopurinol used?

See answer

10. *A hobbled enzyme.* Both side-chain oxygen atoms of aspartate 27 at the active site of dihydrofolate reductase form hydrogen bonds with the pteridine ring of folates. The importance of this interaction was assessed by studying two mutants at this position, Asn 27 and Ser 27. The dissociation constant of methotrexate was 0.07 nM for the wild type, 1.9 nM for the Asn 27 mutant, and 210 nM for the Ser 27 mutant, at 25°C. Calculate the standard free energy of binding of methotrexate by these three proteins. What is the decrease in binding energy resulting from each mutation?

See answer

11. *Correcting deficiencies.* Suppose that a person is found who is deficient in an enzyme required for IMP synthesis. How might this person be treated?

See answer

12. *Labeled nitrogen.* Purine biosynthesis is allowed to take place in the presence of [¹⁵N]aspartate, and the newly synthesized GTP and ATP are isolated. What positions are labeled in the two nucleotides?

See answer

13. *Changed inhibitor.* Xanthine oxidase treated with allopurinol results in the formation of a new compound that is an extremely potent inhibitor of the enzyme. Propose a structure for this compound.

See answer

Mechanism Problems

14. *The same but not the same.* Write out mechanisms for the conversion of phosphoribosylamine into glycinamide ribonucleotide and of xanthylate into guanylate.

See answer

15. *Closing the ring.* Propose a mechanism for the conversion of 5-formamidoimidazole-4-carboxamide ribonucleotide into inosinate.

See answer

Chapter Integration Problems

16. *They're everywhere!* Nucleotides play a variety of roles in the cell. Give an example of a nucleotide that acts in each of the following roles or processes.

- (a) Second messenger
- (b) Phosphoryl-group transfer
- (c) Activation of carbohydrates
- (d) Activation of acetyl groups
- (e) Transfer of electrons
- (f) DNA sequencing
- (g) Chemotherapy
- (h) Allosteric effector

See answer

17. *Pernicious anemia.* Purine biosynthesis is impaired by vitamin B₁₂ deficiency. Why? How might fatty acid and amino acid metabolism also be affected by a vitamin B₁₂ deficiency?

See answer

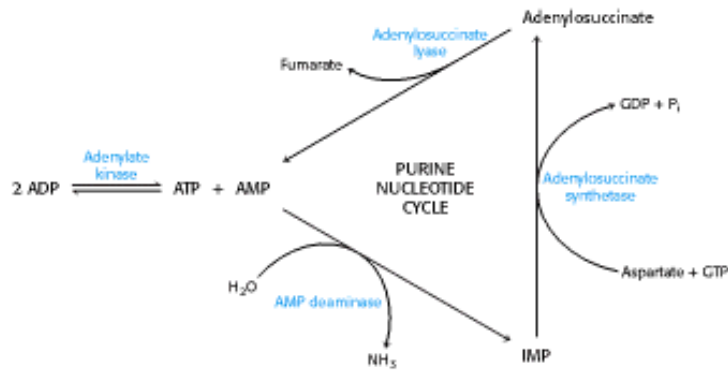
18. *Hyperuricemia.* Many patients with glucose 6-phosphatase deficiency have high serum levels of urate. Hyperuricemia can be induced in normal people by the ingestion of alcohol or by strenuous exercise. Propose a common mechanism that accounts for these findings.

See answer

19. *Labeled carbon.* Succinate uniformly labeled with ¹⁴C is added to cells actively engaged in pyrimidine biosynthesis. Propose a mechanism by which carbon atoms from succinate could be incorporated into a pyrimidine. At what positions is the pyrimidine labeled?

See answer

20. *Exercising muscle.* Some interesting reactions take place in muscle tissue to facilitate the generation of ATP for contraction.



In muscle contraction, ATP is converted into ADP. Adenylate kinase converts two molecules of ADP into a molecule of ATP and AMP.

- (a) Why is this reaction beneficial to contracting muscle?
- (b) Why is the equilibrium for the adenylate kinase approximately equal to 1?

Muscle can metabolize AMP by using the purine nucleotide cycle. The initial step in this cycle, catalyzed by AMP deaminase, is the conversion of AMP into IMP.

- (c) Why might the deamination of AMP facilitate ATP formation in muscle?
- (d) How does the purine nucleotide cycle assist the aerobic generation of ATP?

See answer

Selected Readings

Where to start

M.Y. Galperin and E.V. Koonin. 1997. A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity *Protein Sci.* 6: 2639-2643. ([PubMed](#))

A. Jordan and P. Reichard. 1998. Ribonucleotide reductases *Annu. Rev. Biochem.* 67: 71-98. ([PubMed](#))

J.E. Seegmiller. 1989. Contributions of Lesch-Nyhan syndrome to the understanding of purine metabolism *J. Inherited Metab. Dis.* 12: 184-196. ([PubMed](#))

Pyrimidine biosynthesis

F.M. Raushel, J.B. Thoden, G.D. Reinhart, and H.M. Holden. 1998. Carbamoyl phosphate synthetase: A crooked path from substrates to products *Curr. Opin. Chem. Biol.* 2: 624-632. ([PubMed](#))

T.P. Begley, T.C. Appleby, and S.E. Ealick. 2000. The structural basis for the remarkable proficiency of orotidine 5th-monophosphate decarboxylase *Curr. Opin. Struct. Biol.* 10: 711-718. ([PubMed](#))

T.W. Traut and B.R. Temple. 2000. The chemistry of the reaction determines the invariant amino acids during the evolution and divergence of orotidine 5th-monophosphate decarboxylase *J. Biol. Chem.* 275: 28675-28681. ([PubMed](#))

L. Lee, R.E. Kelly, S.C. Pastra-Landis, and D.R. Evans. 1985. Oligomeric structure of the multifunctional protein CAD that initiates pyrimidine biosynthesis in mammalian cells *Proc. Natl. Acad. Sci. USA* 82: 6802-6806. ([PubMed](#))

Purine biosynthesis

J.B. Thoden, S. Firestine, A. Nixon, S.J. Benkovic, and H.M. Holden. 2000. Molecular structure of *Escherichia coli* PurT-encoded glycinamide ribonucleotide transformylase *Biochemistry* 39: 8791-8802. ([PubMed](#))

F.M. McMillan, M. Cahoon, A. White, L. Hedstrom, G.A. Petsko, and D. Ringe. 2000. Crystal structure at 2.4 Å resolution of *Borrelia burgdorferi* inosine 5[′]-monophosphate dehydrogenase: Evidence of a substrate-induced hinged-lid motion by loop 6 *Biochemistry* 39: 4533-4542. ([PubMed](#))

E.J. Mueller, S. Oh, E. Kavalerchik, T.J. Kappock, E. Meyer, C. Li, S.E. Ealick, and J. Stubbe. 1999. Investigation of the ATP binding site of *Escherichia coli* aminoimidazole ribonucleotide synthetase using affinity labeling and site-directed mutagenesis *Biochemistry* 38: 9831-9839. ([PubMed](#))

V.M. Levdikov, V.V. Barynin, A.I. Grebenko, W.R. Melik-Adamyanyan, V.S. Lamzin, and K.S. Wilson. 1998. The structure of SAICAR synthase: An enzyme in the de novo pathway of purine nucleotide biosynthesis *Structure* 6: 363-376. ([PubMed](#))

J.L. Smith, E.J. Zaluzec, J.P. Wery, L. Niu, R.L. Switzer, H. Zalkin, and Y. Satow. 1994. Structure of the allosteric regulatory enzyme of purine biosynthesis *Science* 264: 1427-1433. ([PubMed](#))

G. Weber, M. Nagai, Y. Natsumeda, S. Ichikawa, H. Nakamura, J.N. Eble, H.N. Jayaram, W.N. Zhen, E. Paulik, and R. Hoffman. 1991. Regulation of de novo and salvage pathways in chemo-therapy *Adv. Enzyme Regul.* 31: 45-67. ([PubMed](#))

Ribonucleotide reductases

P. Reichard. 1997. The evolution of ribonucleotide reduction *Trends Biochem. Sci.* 22: 81-85. ([PubMed](#))

J. Stubbe. 2000. Ribonucleotide reductases: The link between an RNA and a DNA world? *Curr. Opin. Struct. Biol.* 10: 731-736. ([PubMed](#))

D.T. Logan, J. Andersson, B.M. Sjoberg, and P. Nordlund. 1999. A glycy radical site in the crystal structure of a class III ribonucleotide reductase *Science* 283: 1499-1504. ([PubMed](#))

A. Tauer and S.A. Benner. 1997. The B₁₂-dependent ribonucleotide reductase from the archaebacterium *Thermoplasma acidophila*: An evolutionary solution to the ribonucleotide reductase conundrum *Proc. Natl. Acad. Sci. USA* 94: 53-58. ([PubMed](#)) ([Full Text in PMC](#))

A. Jordan, E. Torrents, C. Jeanthon, R. Eliasson, U. Hellman, C. Wernstedt, J. Barbe, I. Gibert, and P. Reichard. 1997. B₁₂-dependent ribonucleotide reductases from deeply rooted eubacteria are structurally related to the aerobic enzyme from *Escherichia coli* *Proc. Natl. Acad. Sci. USA* 94: 13487-13492. ([PubMed](#)) ([Full Text in PMC](#))

J. Stubbe and P. Riggs-Gelasco. 1998. Harnessing free radicals: Formation and function of the tyrosyl radical in ribonucleotide reductase *Trends Biochem. Sci.* 23: 438-443. ([PubMed](#))

J.A. Stubbe. 1989. Protein radical involvement in biological catalysis? *Annu. Rev. Biochem* 58: 257-285. ([PubMed](#))

Thymidylate synthase and dihydrofolate reductase

R. Li, R. Sirawaraporn, P. Chitnumsub, W. Sirawaraporn, J. Wooden, F. Athappilly, S. Turley, and W.G. Hol. 2000. Three-dimensional structure of *M. tuberculosis* dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs *J. Mol. Biol.* 295: 307-323. ([PubMed](#))

P.H. Liang and K.S. Anderson. 1998. Substrate channeling and domain-domain interactions in bifunctional thymidylate synthase-dihydrofolate reductase *Biochemistry* 37: 12195-12205. ([PubMed](#))

G.P. Miller and S.J. Benkovic. 1998. Stretching exercises: Flexibility in dihydrofolate reductase catalysis *Chem. Biol.* 5: R105-R113. ([PubMed](#))

R.L. Blakley. 1995. Eukaryotic dihydrofolate reductase *Adv. Enzymol. Relat. Areas Mol. Biol.* 70: 23-102. ([PubMed](#))

C.W. Carreras and D.V. Santi. 1995. The catalytic mechanism and structure of thymidylate synthase *Annu. Rev. Biochem.* 64: 721-762. ([PubMed](#))

B.I. Schweitzer, A.P. Dicker, and J.R. Bertino. 1990. Dihydrofolate reductase as a therapeutic target *FASEB J.* 4: 2441-2452. ([PubMed](#))

K.A. Brown and J. Kraut. 1992. Exploring the molecular mechanism of dihydrofolate reductase *Faraday Discuss.* 1992: 217-224.

C. Bystroff, S.J. Oatley, and J. Kraut. 1990. Crystal structures of *Escherichia coli* dihydrofolate reductase: The NADP⁺ holoenzyme and the folate NADP⁺ ternary complex — Substrate binding and a model for the transition state *Biochemistry* 29: 3263-3277. ([PubMed](#))

Genetic diseases

Striver, C. R., Beaudet, A. L., Sly, W. S., Valle, D., Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S. (Eds.), 1995. *The Metabolic Basis of Inherited Diseases* (7th ed., pp. 1655 – 1840). McGraw-Hill.

W.L. Nyhan. 1997. The recognition of Lesch-Nyhan syndrome as an inborn error of purine metabolism *J. Inherited Metab. Dis.* 20: 171-178. ([PubMed](#))

D.F. Wong, J.C. Harris, S. Naidu, F. Yokoi, S. Marenco, R.F. Dannals, H.T. Ravert, M. Yaster, A. Evans, O. Rousset, R. N. Bryan, A. Gjedde, M.J. Kuhar, and G.R. Breese. 1996. Dopamine transporters are markedly reduced in Lesch-Nyhan disease in vivo *Proc. Natl. Acad. Sci. USA* 93: 5539-5543. ([PubMed](#)) ([Full Text in PMC](#))

R. Resta and L.F. Thompson. 1997. SCID: The role of adenosine deaminase deficiency *Immunol. Today* 18: 371-374. ([PubMed](#))

B.L. Davidson, M. Pashmforoush, W.N. Kelley, and T.D. Paella. 1989. Human hypoxanthine-guanine phosphoribosyltransferase deficiency: The molecular defect in a patient with gout (HPRTAshville) *J. Biol. Chem.* 264: 520-525. ([PubMed](#))

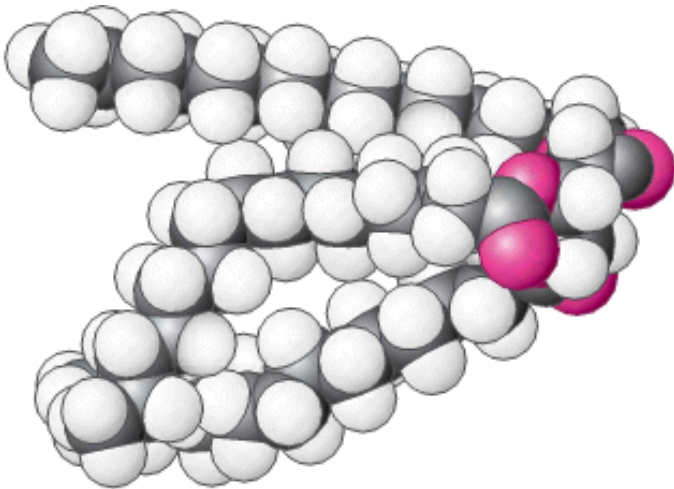
D.G. Sculley, P.A. Dawson, B.T. Emerson, and R.B. Gordon. 1992. A review of the molecular basis of hypoxanthine-guanine phosphoribosyltransferase (HPRT) deficiency *Hum. Genet.* 90: 195-207. ([PubMed](#))

26. The Biosynthesis of Membrane Lipids and Steroids

This chapter examines the biosynthesis of three important components of biological membranes—phospholipids, sphingolipids, and cholesterol ([Chapter 12](#)). Triacylglycerols also are considered here because the pathway for their synthesis overlaps that of phospholipids. Cholesterol is of interest both as a membrane component and as a precursor of many signal molecules, including the steroid hormones progesterone, testosterone, estrogen, and cortisol. The biosynthesis of cholesterol exemplifies a fundamental mechanism for the assembly of extended carbon skeletons from five-carbon units.

The transport of cholesterol in blood by the low-density lipoprotein and its uptake by a specific receptor on the cell surface vividly illustrate a recurring mechanism for the entry of metabolites and signal molecules into cells. The absence of this receptor in people with *familial hypercholesterolemia*, a genetic disease, leads to markedly elevated cholesterol

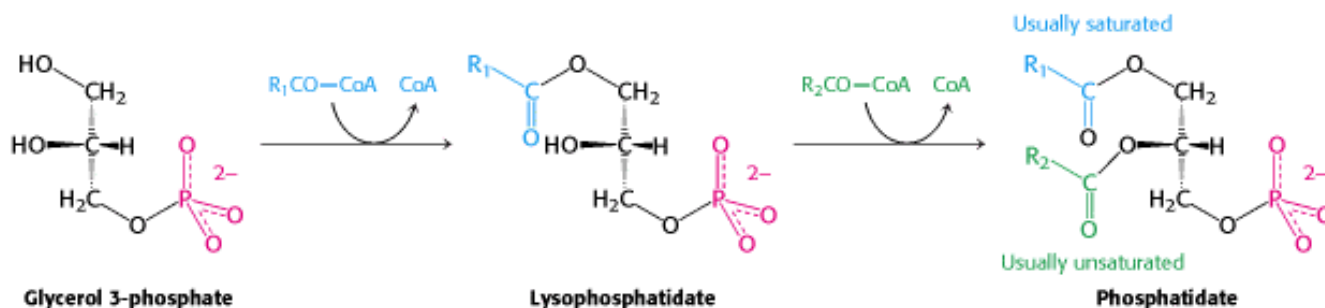
levels in the blood, cholesterol deposits on blood vessels, and childhood heart attacks. Indeed, cholesterol is implicated in the development of atherosclerosis in individuals without genetic defects. Thus, the regulation of cholesterol synthesis and transport can be a source of especially clear insight into the role that our understanding of biochemistry plays in medicine.



Fats such as the triacylglycerol molecule (below) are widely used to store excess energy for later use and to fulfill other purposes, illustrated by the insulating blubber of whales. The natural tendency of fats to exist in nearly water-free forms makes these molecules well-suited for these roles. [(Left) François Cohier/Photo Researchers.]

26.1. Phosphatidate Is a Common Intermediate in the Synthesis of Phospholipids and Triacylglycerols

The first step in the synthesis of both phospholipids for membranes and triacylglycerols for energy storage is the synthesis of *phosphatidate* (diacylglycerol 3-phosphate). In mammalian cells, phosphatidate is synthesized in the endoplasmic reticulum and the outer mitochondrial membrane. It is formed by the addition of two fatty acids to *glycerol 3-phosphate*, which in turn is formed primarily by the reduction of dihydroxyacetone phosphate, a glycolytic intermediate, and to a lesser extent by the phosphorylation of glycerol. Glycerol 3-phosphate is acylated by acyl CoA to form *lysophosphatidate*, which is again acylated by acyl CoA to yield phosphatidate.



These acylations are catalyzed by *glycerol phosphate acyltransferase*. In most phosphatidates, the fatty acyl chain attached to the C-1 atom is saturated, whereas the one attached to the C-2 atom is unsaturated.

The pathways diverge at phosphatidate. In the synthesis of triacylglycerols, phosphatidate is hydrolyzed by a specific phosphatase to give a *diacylglycerol (DAG)*. This intermediate is acylated to a *triacylglycerol* in a reaction that is catalyzed by *diglyceride acyltransferase*. Both enzymes are associated in a *triacylglycerol synthetase complex* that is bound to the endoplasmic reticulum membrane.



The liver is the primary site of triacylglycerol synthesis. From the liver, the triacylglycerols are transported to the muscles for energy conversion or to the adipocytes for storage.

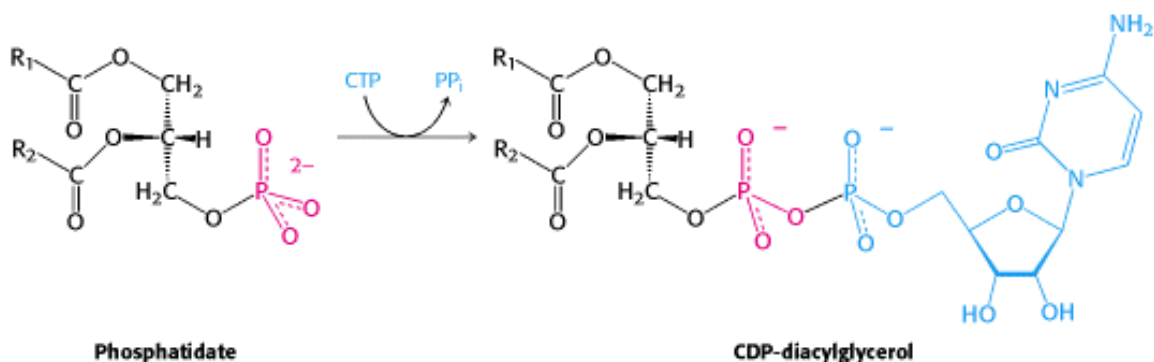
26.1.1. The Synthesis of Phospholipids Requires an Activated Intermediate

Phospholipid synthesis requires the combination of a diacylglyceride with an alcohol. As in most anabolic reactions, one of the components must be activated. In this case, either of the two components may be activated, depending on the source of the reactants.

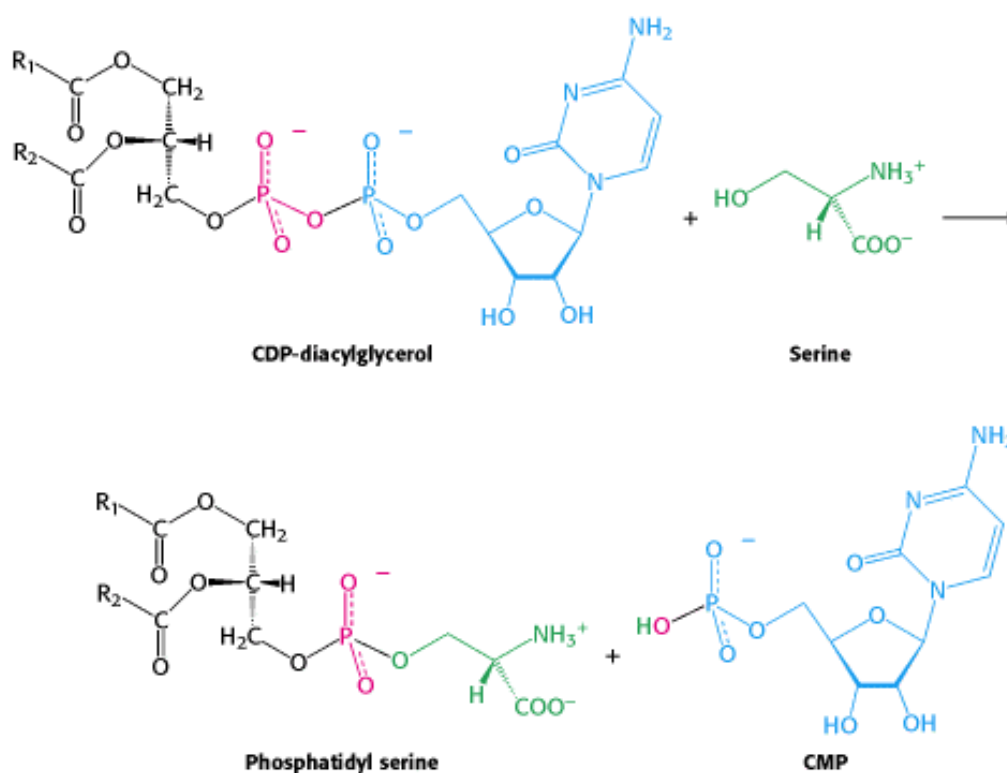
Synthesis from an Activated Diacylglycerol.

The de novo pathway starts with the reaction of phosphatidate with cytidine triphosphate (CTP) to form *cytidine diphosphodiacylglycerol (CDP-diacylglycerol)* (Figure 26.1). This reaction, like those of many biosyntheses, is driven

forward by the hydrolysis of pyrophosphate.



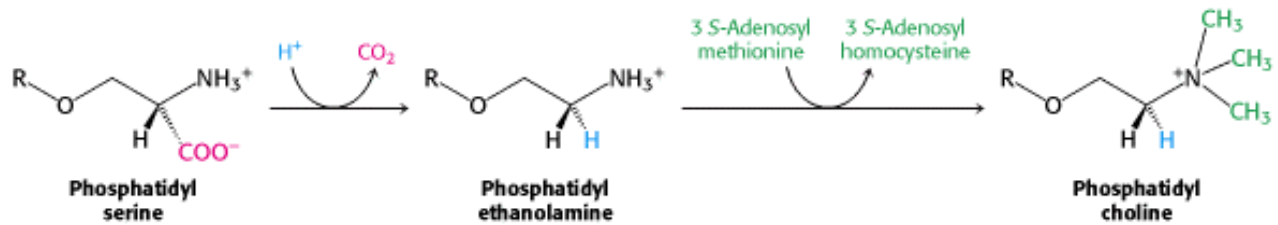
The activated phosphatidyl unit then reacts with the hydroxyl group of an alcohol to form a phosphodiester linkage. If the alcohol is serine, the products are *phosphatidyl serine* and cytidine monophosphate (CMP).



Likewise, phosphatidyl inositol is formed by the transfer of a diacylglycerol phosphate unit from CDP-diacylglycerol to inositol. Subsequent phosphorylations catalyzed by specific kinases lead to the synthesis of *phosphatidyl inositol 4,5-bisphosphate*, an important molecule in signal transduction. Recall that hormonal and sensory stimuli activate *phospholipase C*, an enzyme that hydrolyzes this phospholipid to form two intracellular messengers—diacylglycerol and inositol 1,4,5-trisphosphate (Section 15.2).

The fatty acid components of a phospholipid may vary, and thus phosphatidyl serine, as well as most other phospholipids, represents a class of molecules rather than a single species. As a result, a single mammalian cell may contain thousands of distinct phospholipids. Phosphatidyl inositol is unusual in that it has a nearly fixed fatty acid composition. Stearic acid usually occupies the C-1 position and arachidonic acid (Section 22.6.2) the C-2 position.

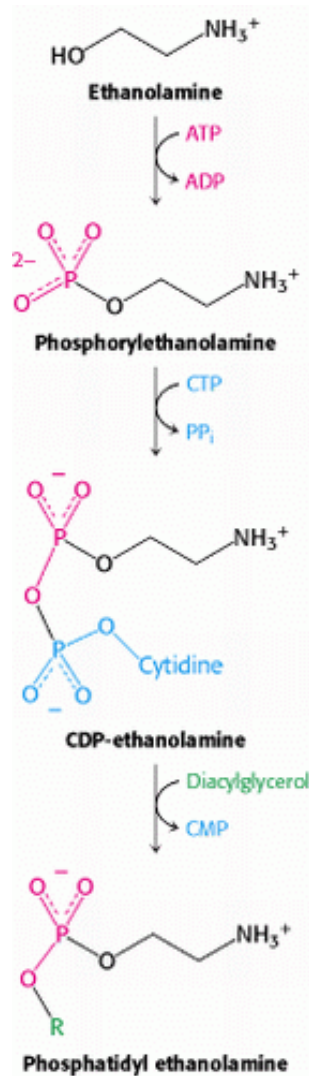
In bacteria, the decarboxylation of phosphatidyl serine by a pyridoxal phosphate-dependent enzyme yields *phosphatidyl ethanolamine*, another common phospholipid. The amino group of this phosphoglyceride is then methylated three times to form *phosphatidyl choline*. *S-Adenosylmethionine* is the methyl donor.



In mammals, phosphatidyl ethanolamine can be formed from phosphatidyl serine by the enzyme-catalyzed exchange of ethanolamine for the serine moiety of the phospholipid.

Synthesis from an Activated Alcohol.

In mammals, phosphatidyl ethanolamine can also be synthesized from ethanolamine through the formation of CDP-ethanolamine. In this case, the alcohol ethanolamine is phosphorylated by ATP to form the precursor, *phosphorylethanolamine*. This precursor then reacts with CTP to form the activated alcohol, *CDP-ethanolamine*. The phosphorylethanolamine unit of CDP-ethanolamine is then transferred to a diacylglycerol to form *phosphatidyl ethanolamine*.



In mammals, a pathway that utilizes choline obtained from the diet ends in the synthesis of phosphatidyl choline, the most common phospholipid in these organisms. In this case, choline is activated in a series of reactions analogous to those in the activation of ethanolamine. Interestingly, the liver possesses an enzyme, *phosphatidyl ethanolamine methyltransferase*, that synthesizes phosphatidyl choline from phosphatidyl ethanolamine, through the successive

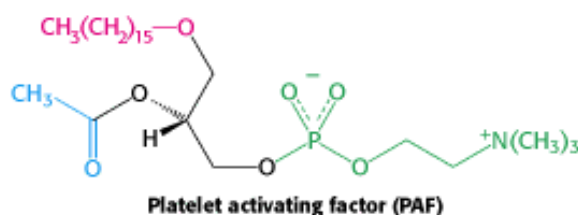
methylation of ethanolamine. Thus, phosphatidyl choline can be produced by two distinct pathways, ensuring that this phospholipid can be synthesized even if the components for one pathway are in limited supply.

Note that a cytidine nucleotide plays the same role in the synthesis of these phosphoglycerides as a uridine nucleotide does in the formation of glycogen (Section 21.4.1). In all of these biosyntheses, an activated intermediate (UDP-glucose, CDP-diacylglycerol, or CDP-alcohol) is formed from a phosphorylated substrate (glucose 1-phosphate, phosphatidate, or a phosphorylalcohol) and a nucleoside triphosphate (UTP or CTP). The activated intermediate then reacts with a hydroxyl group (the terminus of glycogen, the side chain of serine, or a diacylglycerol).

26.1.2. Plasmalogens and Other Ether Phospholipids Are Synthesized from Dihydroxyacetone Phosphate

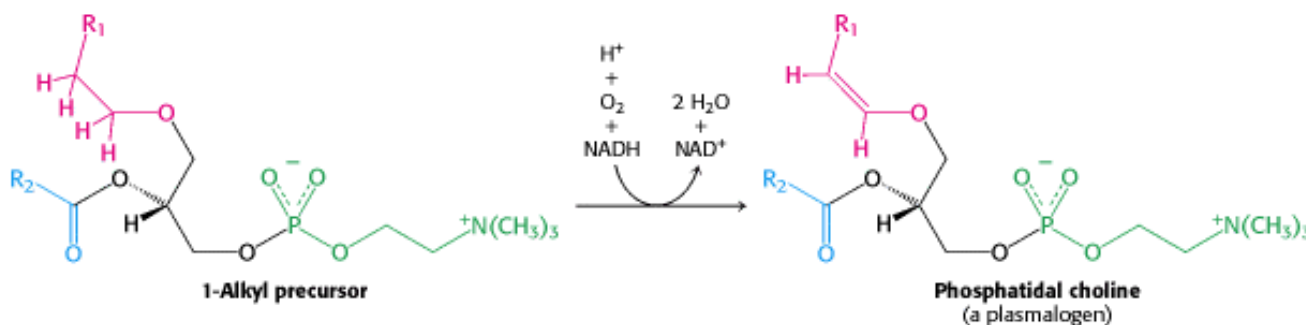
Glycerol ether phospholipids contain an ether unit instead of an acyl unit at C-1 and are synthesized starting with dihydroxyacetone phosphate rather than glycerol 3-phosphate (Figure 26.2). Acylation by a fatty acyl CoA yields a 1-acyl derivative that exchanges with a long-chain alcohol to form an ether at C-1. NADPH reduces the keto group at C-2, and the resulting alcohol is acylated by a long-chain acyl CoA. Removal of the 3-phosphate group yields 1-alkyl-2-acylglycerol, which reacts with CDP-choline to form the ether analog of phosphatidyl choline.

 *Platelet-activating factor (PAF)* is an ether phospholipid implicated in a number of allergic and inflammatory responses.



Subnanomolar concentrations of this 1-alkyl-2-acetyl ether analog of phosphatidyl choline induce the aggregation of blood platelets, smooth muscle contraction, and the activation of cells of the immune system. It is also a mediator of anaphylactic shock, a severe and often fatal allergic response. The presence of an acetyl group rather than a long-chain acyl group at C-2 increases the water solubility of this lipid, enabling it to function in the aqueous environment of the blood. PAF functions through a 7-TM receptor.

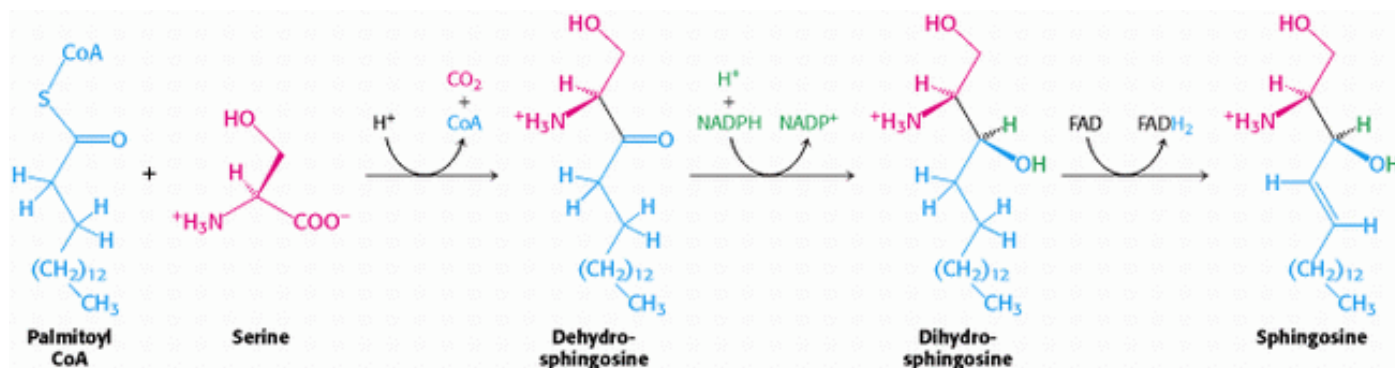
Plasmalogens are phospholipids containing an α,β -unsaturated ether at C-1. Phosphatidal choline, the plasmalogen corresponding to phosphatidyl choline, is formed by desaturation of a 1-alkyl precursor.



The desaturase catalyzing this final step in the synthesis of a plasmalogen is an endoplasmic reticulum enzyme akin to the one that introduces double bonds into long-chain fatty acyl CoA molecules. In both cases, O_2 and NADH are reactants, and cytochrome b_5 participates in catalysis (Section 22.6).

26.1.3. Sphingolipids Are Synthesized from Ceramide

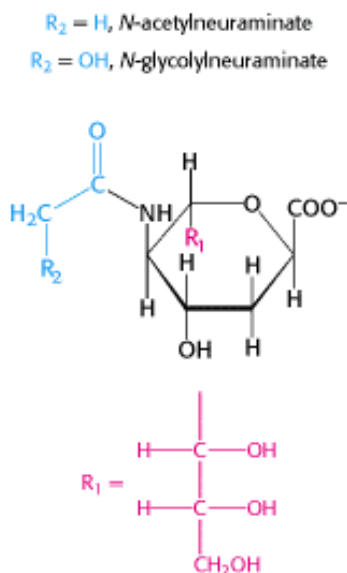
We turn now from glycerol-based phospholipids to another class of membrane lipid—the *sphingolipids*. These lipids are found in the plasma membranes of all eukaryotic cells, although the concentration is highest in the cells of the central nervous system. The backbone of a sphingolipid is *sphingosine*, rather than glycerol (Section 12.3.1). Palmitoyl CoA and serine condense to form dehydrosphingosine, which is then converted into sphingosine. The enzyme catalyzing this reaction requires pyridoxal phosphate, revealing again the dominant role of this cofactor in transformations that include amino acids.



In all sphingolipids, the amino group of sphingosine is acylated: a long-chain acyl CoA reacts with sphingosine to form *ceramide* (*N*-acyl sphingosine) (Figure 26.3). The terminal hydroxyl group also is substituted. In *sphingomyelin*, a component of the myelin sheath covering many nerve fibers, the substituent is phosphorylcholine, which comes from phosphatidyl choline. In a *cerebroside*, the substituent is glucose or galactose. UDP-glucose or UDP-galactose is the sugar donor. In a *ganglioside*, an oligosaccharide is linked to the terminal hydroxyl group of ceramide by a glucose residue (Figure 26.4).

26.1.4. Gangliosides Are Carbohydrate-Rich Sphingolipids That Contain Acidic Sugars

In gangliosides, the most complex sphingolipids, an *oligosaccharide chain* attached to the ceramide contains at least one acidic sugar. The acidic sugar is *N*-acetylneuraminate or *N*-glycolylneuraminate. These acidic sugars are called *sialic acids*. Their nine-carbon backbones are synthesized from phosphoenolpyruvate (a three-carbon unit) and *N*-acetylmannosamine 6-phosphate (a six-carbon unit).




Gangliosides are synthesized by the ordered, step-by-step addition of sugar residues to ceramide. The synthesis of these

complex lipids requires the activated sugars UDP-glucose, UDP-galactose, and UDP-*N*-acetylgalactosamine, as well as the CMP derivative of *N*-acetylneuraminic acid. CMP-*N*-acetylneuraminic acid is synthesized from CTP and *N*-acetylneuraminic acid. The structure of the resulting ganglioside is determined by the specificity of the glycosyltransferases in the cell. More than 60 different gangliosides have been characterized (see [Figure 26.4](#) for the structure of ganglioside G_{M1}).

26.1.5. Sphingolipids Confer Diversity on Lipid Structure and Function

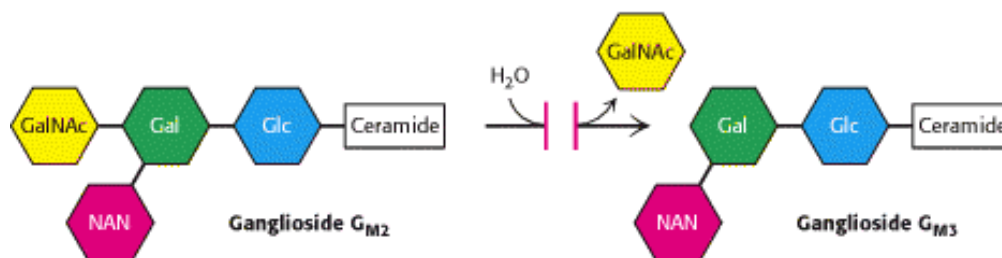
The structures of sphingolipids and the more abundant glycerophospholipids are very similar. Given the structural similarity of these two types of lipids, why are sphingolipids required at all? Indeed, the prefix "sphingo" was applied to capture the "sphinxlike" properties of this enigmatic class of lipids. Although the precise role of sphingolipids is not firmly established, progress toward solving the riddle of their function is being made. The most notable function attributed to sphingolipids is their role as a source of second messengers. For instance, ceramide derived from a sphingolipid may initiate programmed cell death in some cell types.

26.1.6. Respiratory Distress Syndrome and Tay-Sachs Disease Result from the Disruption of Lipid Metabolism

 *Respiratory distress syndrome* is a pathological condition resulting from a failure in the biosynthetic pathway for dipalmitoyl phosphatidyl choline. This phospholipid, in conjunction with specific proteins and other phospholipids, is found in the extracellular fluid that surrounds the alveoli of the lung, where it decreases the surface tension of the fluid to prevent lung collapse at the end of the expiration phase of breathing. Premature infants may suffer from respiratory distress syndrome because their immature lungs do not synthesize enough dipalmitoyl phosphatidyl choline.

Tay-Sachs disease is caused by a failure of lipid degradation: an inability to degrade gangliosides. Gangliosides are found in highest concentration in the nervous system, particularly in gray matter, where they constitute 6% of the lipids. Gangliosides are normally degraded inside lysosomes by the sequential removal of their terminal sugars but, in Tay-Sachs disease, this degradation does not occur. As a consequence, neurons become enormously swollen with lipid-filled lysosomes ([Figure 26.5](#)). An affected infant displays weakness and retarded psychomotor skills before 1 year of age. The child is demented and blind by age 2 and usually dead before age 3.

The ganglioside content of the brain of an infant with Tay-Sachs disease is greatly elevated. *The concentration of ganglioside G_{M2} is many times as high as normal because its terminal N-acetylgalactosamine residue is removed very slowly or not at all.* The missing or deficient enzyme is a specific β -*N*-acetylhexosaminidase.



Tay-Sachs disease can be diagnosed in the course of fetal development. Amniotic fluid is obtained by amniocentesis and assayed for β -*N*-acetylhexosaminidase activity.

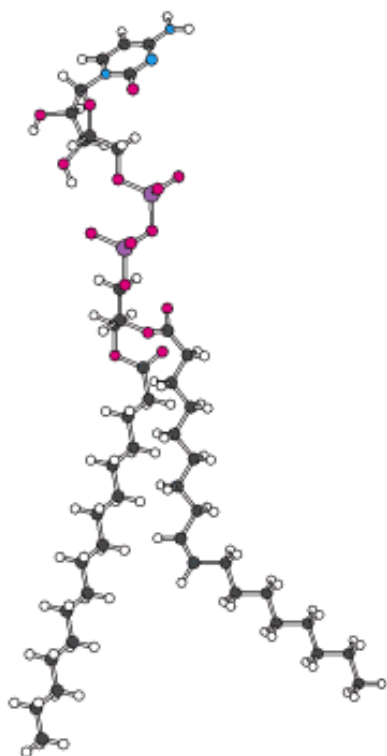


Figure 26.1. Structure of CDP-Diacylglycerol. A key intermediate in the synthesis of phospholipids consists of phosphatidate and CMP joined by a pyrophosphate linkage.

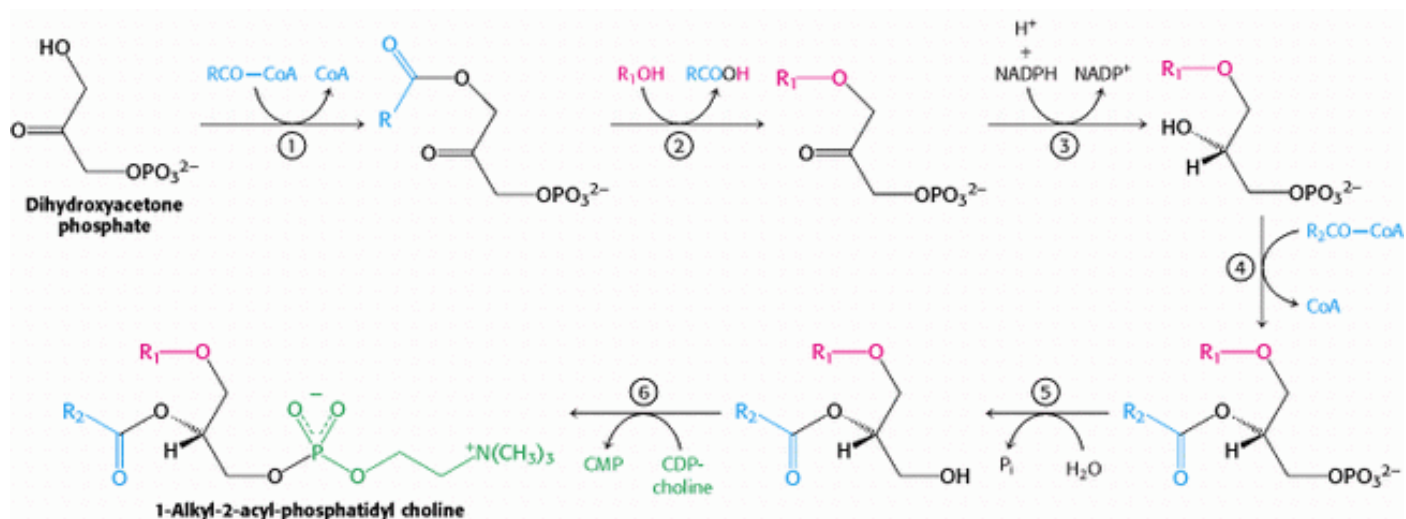


Figure 26.2. Synthesis of an Ether Phospholipid. Steps in the synthesis include (1) acylation of dihydroxyacetone phosphate by acyl CoA, (2) exchange of an alcohol for the carboxylic acid, (3) reduction by NADPH, (4) acylation by a second acyl CoA, (5) hydrolysis of the phosphate ester, and (6) transfer of a phosphocholine moiety.

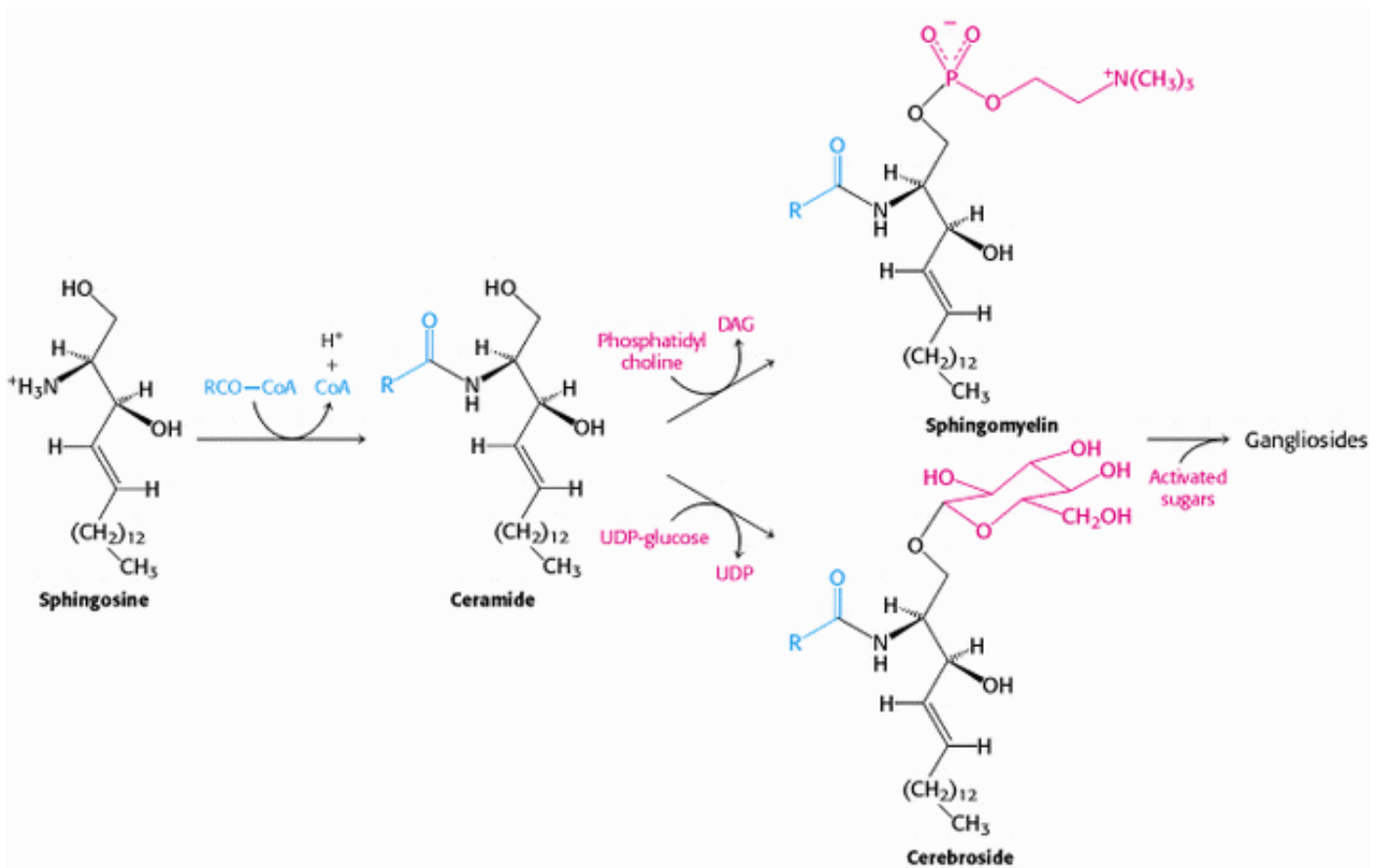


Figure 26.3. Synthesis of Sphingolipids. Sphingosine is converted into ceramide, which is an intermediate in the formation of sphingomyelin and gangliosides.

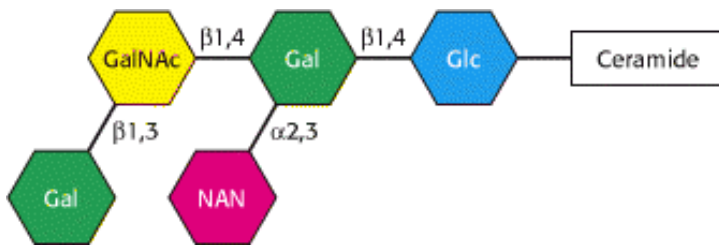


Figure 26.4. Ganglioside $\text{G}_{\text{M}1}$. This ganglioside consists of five monosaccharides linked to ceramide: one glucose (Glc) molecule, two galactose (Gal) molecules, one *N*-acetylgalactosamine (GalNAc) molecule, and one *N*-acetylneuraminic acid (NAN) molecule. The structures of the linkages are indicated.



Figure 26.5. Lysosome with Lipids. An electron micrograph of a lysosome containing an abnormal amount of lipid. [Courtesy of Dr. George Palade.]

26.2. Cholesterol Is Synthesized from Acetyl Coenzyme A in Three Stages

We now turn our attention to the synthesis of the fundamental lipid *cholesterol*. This steroid modulates the fluidity of animal cell membranes ([Section 12.6.2](#)) and is the precursor of steroid hormones such as progesterone, testosterone, estradiol, and cortisol. *All 27 carbon atoms of cholesterol are derived from acetyl CoA* in a three-stage synthetic process ([Figure 26.6](#)).

Cholesterol-

"Cholesterol is the most highly decorated small molecule in biology. Thirteen Nobel Prizes have been awarded to scientists who devoted major parts of their careers to cholesterol. Ever since it was isolated from gallstones in 1784, cholesterol has exerted an almost hypnotic fascination for scientists from the most diverse areas of science and medicine. . . . Cholesterol is a Janus-faced molecule. The very property that makes it useful in cell membranes, namely its absolute insolubility in water, also makes it lethal."

-Michael Brown and Joseph Goldstein

Nobel Lectures (1985)

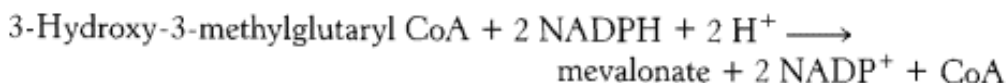
© The Nobel Foundation, 1985

1. Stage one is the synthesis of isopentenyl pyrophosphate, an activated isoprene unit that is the key building block of cholesterol.
2. Stage two is the condensation of six molecules of isopentenyl pyrophosphate to form squalene.
3. In stage three, squalene cyclizes in an astounding reaction and the tetracyclic product is subsequently converted into cholesterol.

26.2.1. The Synthesis of Mevalonate, Which Is Activated as Isopentenyl Pyrophosphate, Initiates the Synthesis of Cholesterol

The first stage in the synthesis of cholesterol is the formation of isopentenyl pyrophosphate from acetyl CoA. This set of reactions, which takes place in the cytosol, starts with the formation of 3-hydroxy-3-methylglutaryl CoA (HMG CoA) from acetyl CoA and acetoacetyl CoA. This intermediate is reduced to *mevalonate* for the synthesis of cholesterol (Figure 26.7). Recall that mitochondrial 3-hydroxy-3-methylglutaryl CoA is processed to form ketone bodies (Section 22.3.5).

The synthesis of mevalonate is the committed step in cholesterol formation. The enzyme catalyzing this irreversible step, 3-hydroxy-3-methylglutaryl CoA reductase (HMG-CoA reductase), is an important control site in cholesterol biosynthesis, as will be discussed shortly.

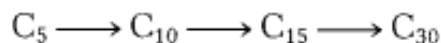


HMG-CoA reductase is an integral membrane protein in the endoplasmic reticulum.

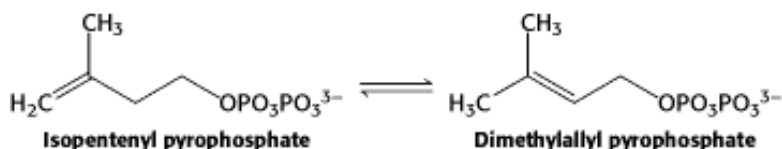
Mevalonate is converted into 3-isopentenyl pyrophosphate in three consecutive reactions requiring ATP (Figure 26.8). Decarboxylation yields isopentenyl pyrophosphate, an activated isoprene unit that is a key building block for many important biomolecules throughout the kingdoms of life. We will return to a discussion of this molecule later in the chapter.

26.2.2. Squalene (C₃₀) Is Synthesized from Six Molecules of Isopentenyl Pyrophosphate (C₅)

Squalene is synthesized from isopentenyl pyrophosphate by the reaction sequence



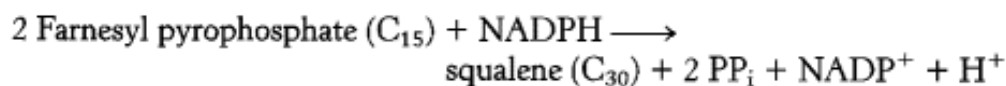
This stage in the synthesis of cholesterol starts with the isomerization of *isopentenyl pyrophosphate* to *dimethylallyl pyrophosphate*.



These isomeric C₅ units condense to form a C₁₀ compound: isopentenyl pyrophosphate attacks an allylic carbonium ion formed from dimethylallyl pyrophosphate to yield *geranyl pyrophosphate* (Figure 26.9). The same kind of reaction takes place again: geranyl pyrophosphate is converted into an allylic carbonium ion, which is attacked by isopentenyl pyrophosphate. The resulting C₁₅ compound is called *farnesyl pyrophosphate*. The same enzyme, *geranyl transferase*, catalyzes each of these condensations.

The last step in the synthesis of *squalene* is a reductive tail-to-tail condensation of two molecules of farnesyl

pyrophosphate catalyzed by the endoplasmic reticulum enzyme *squalene synthase*.



The reactions leading from C₅ units to squalene, a C₃₀ isoprenoid, are summarized in [Figure 26.10](#).

26.2.3. Squalene Cyclizes to Form Cholesterol

The final stage of cholesterol biosynthesis starts with the cyclization of squalene ([Figure 26.11](#)). Squalene is first activated by conversion into squalene epoxide (2,3-oxidosqualene) in a reaction that uses O₂ and NADPH. Squalene epoxide is then cyclized to *lanosterol* by *oxidosqualene cyclase* ([Figure 26.12](#)). This remarkable transformation proceeds in a concerted fashion. The enzyme holds squalene epoxide in an appropriate conformation and initiates the reaction by protonating the epoxide oxygen. The carbocation formed spontaneously rearranges to produce lanosterol. Lanosterol is converted into cholesterol in a multistep process by the removal of three methyl groups, the reduction of one double bond by NADPH, and the migration of the other double bond ([Figure 26.13](#)).

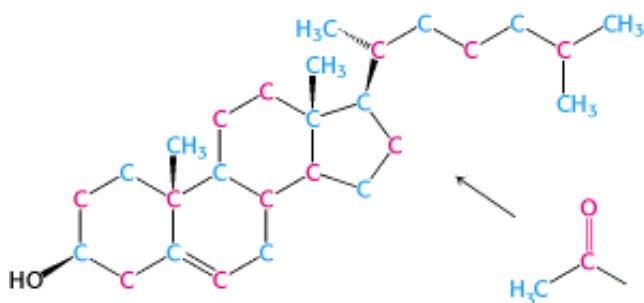


Figure 26.6. Labeling of Cholesterol. The results of isotope-labeling experiments reveal the source of carbon atoms in cholesterol synthesized from acetate labeled in its methyl group (blue) or carboxylate atom (red).

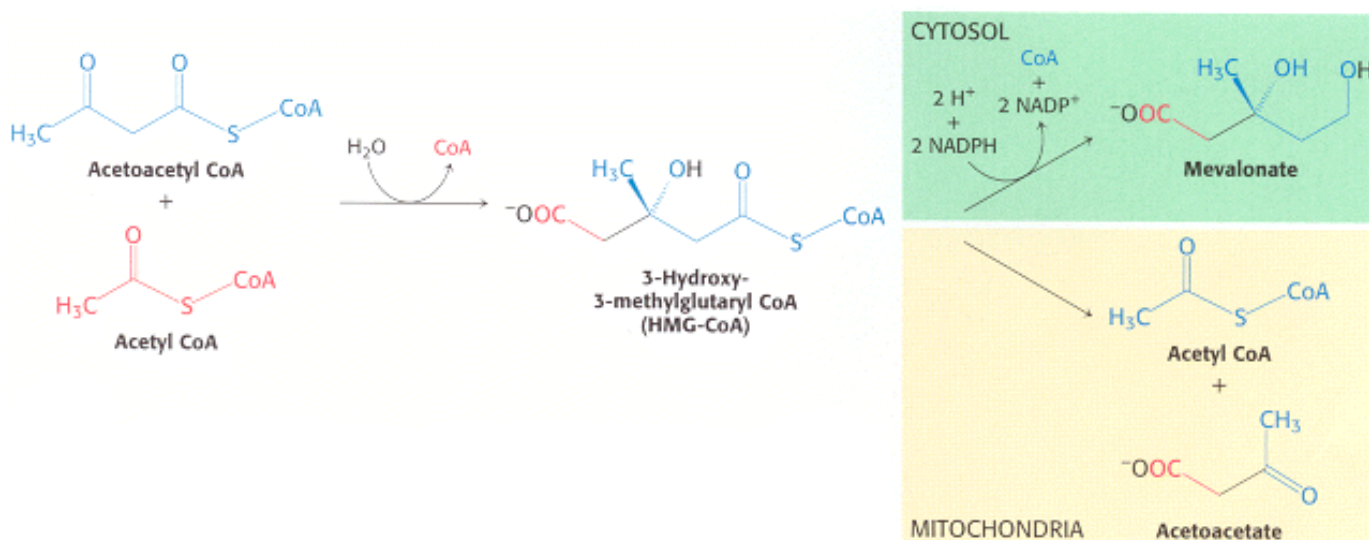


Figure 26.7. Fates of 3-Hydroxy-3-Methylglutaryl CoA. In the cytosol, HMG-CoA is converted into mevalonate. In mitochondria, it is converted into acetyl CoA and acetoacetate.

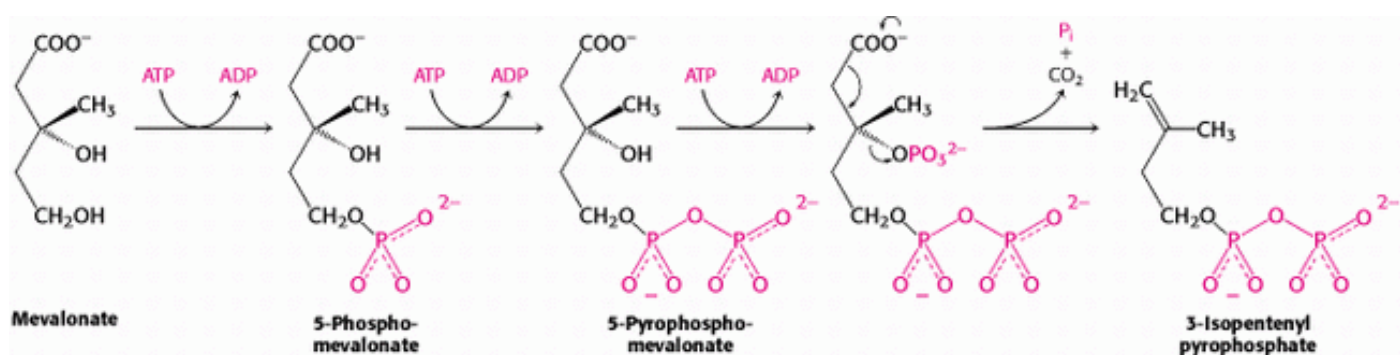


Figure 26.8. Synthesis of Isopentenyl Pyrophosphate. This activated intermediate is formed from mevalonate in three steps, the last of which includes a decarboxylation.

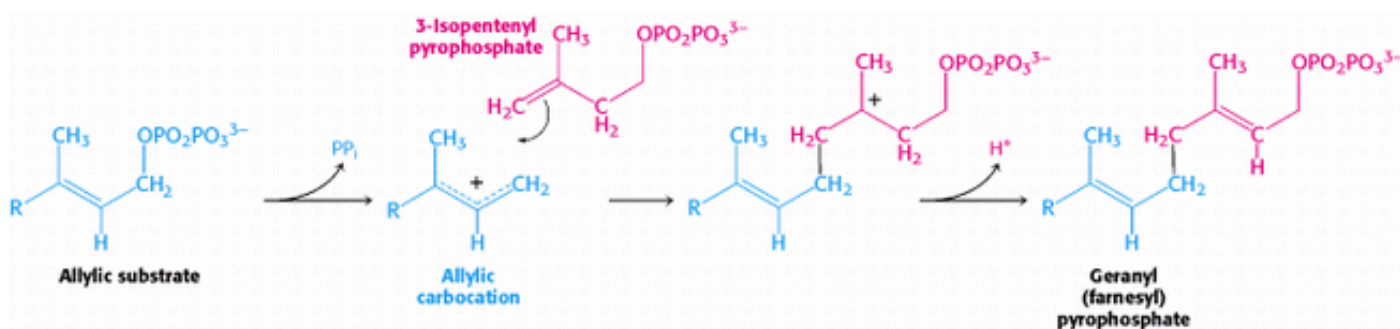


Figure 26.9. Condensation Mechanism in Cholesterol Synthesis. The mechanism for joining dimethylallyl pyrophosphate and isopentenyl pyrophosphate to form geranyl pyrophosphate. The same mechanism is used to add an additional isopentenyl pyrophosphate to form farnesyl pyrophosphate.

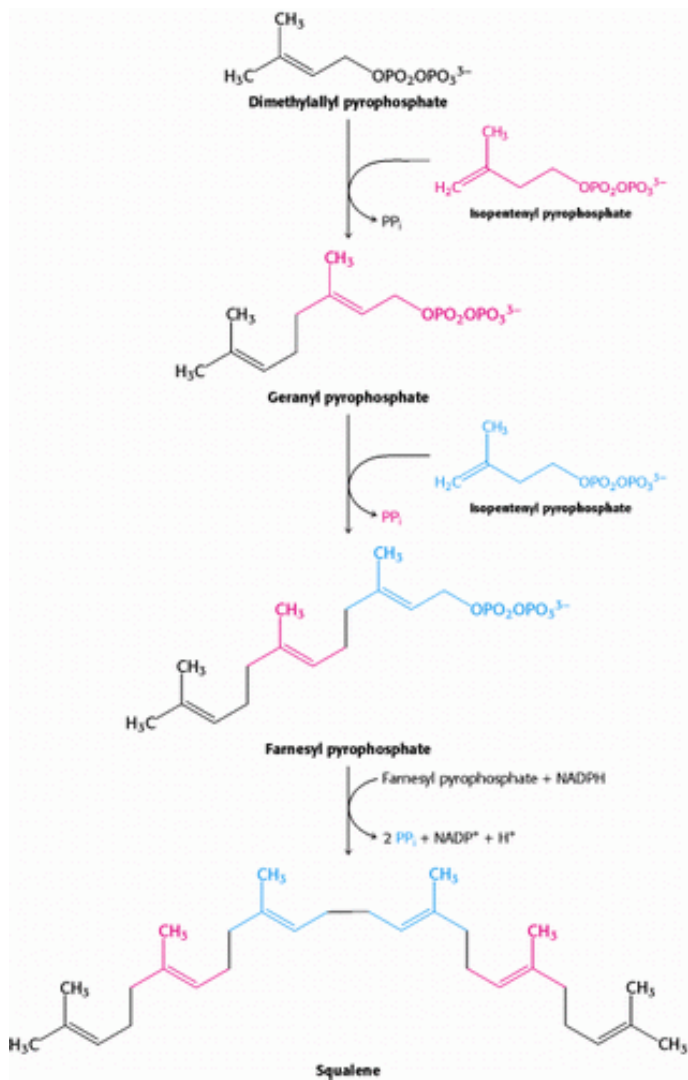


Figure 26.10. Squalene Synthesis. One molecule of dimethylallyl pyrophosphate and two molecules of isopentenyl pyrophosphate condense to form farnesyl pyrophosphate. The tail-to-tail coupling of two molecules of farnesyl pyrophosphate yields squalene.

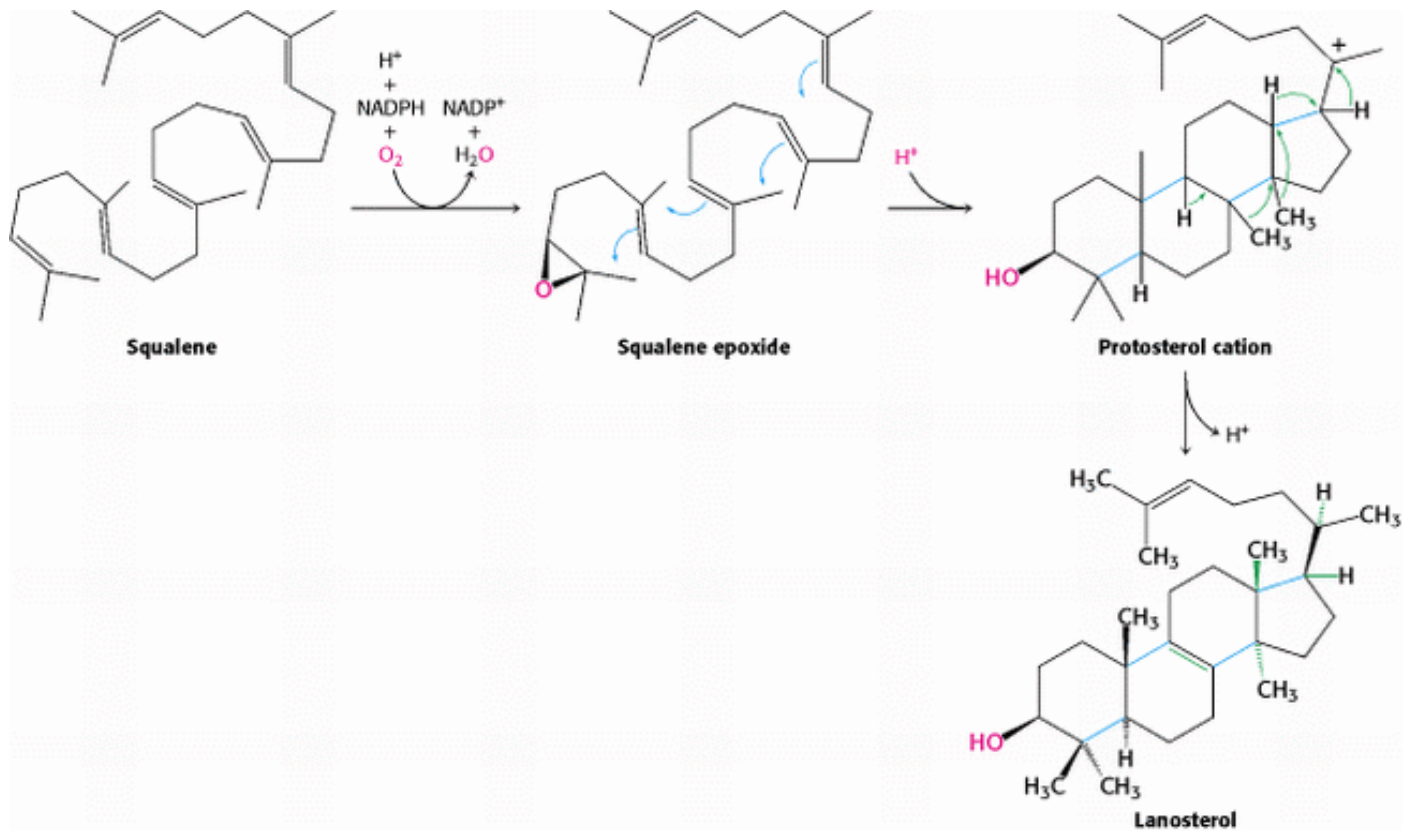


Figure 26.11. Squalene Cyclization. The formation of the steroid nucleus from squalene begins with the formation of squalene epoxide. This intermediate is protonated to form a carbocation that cyclizes to form a tetracyclic structure, which rearranges to form lanosterol.



Figure 26.12. Oxidosqualene Cyclase. The structure of an enzyme homologous to oxidosqualene cyclase shows a central cavity lined primarily with hydrophobic side chains (shown in red) in which the cyclization reaction takes place.

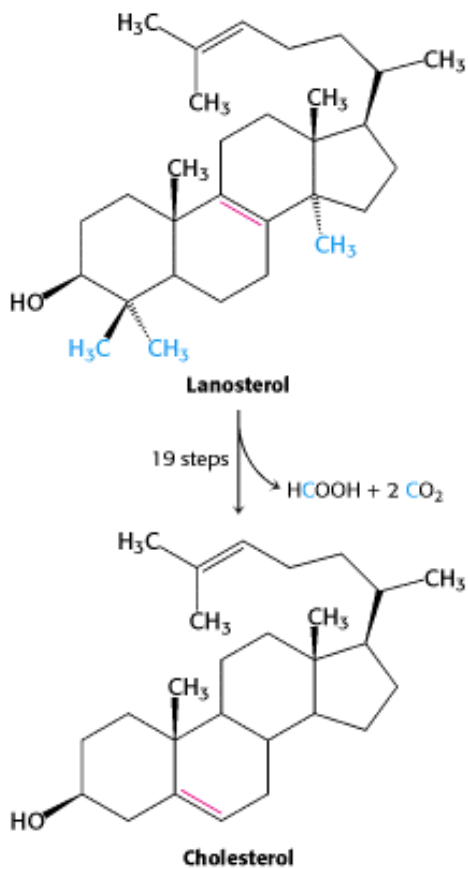


Figure 26.13. Cholesterol Formation. Lanosterol is converted into cholesterol in a complex process.

26.3. The Complex Regulation of Cholesterol Biosynthesis Takes Place at Several Levels

Cholesterol can be obtained from the diet or it can be synthesized *de novo*. An adult on a low-cholesterol diet typically synthesizes about 800 mg of cholesterol per day. The liver is the major site of cholesterol synthesis in mammals, although the intestine also forms significant amounts. The rate of cholesterol formation by these organs is highly responsive to the cellular level of cholesterol. *This feedback regulation is mediated primarily by changes in the amount and activity of 3-hydroxy-3-methylglutaryl CoA reductase (Figure 26.14).* As discussed in Section 26.2.1, this enzyme catalyzes the formation of mevalonate, the committed step in cholesterol biosynthesis. HMG CoA reductase is controlled in multiple ways:

1. The rate of *synthesis of reductase mRNA* is controlled by the *sterol regulatory element binding protein (SREBP)*. This transcription factor binds to a short DNA sequence called the *sterol regulatory element (SRE)* on the 5' side of the reductase gene. In its inactive state, the SREBP is anchored to the endoplasmic reticulum or nuclear membrane. When cholesterol levels fall, the amino-terminal domain is released from its association with the membrane by two specific proteolytic cleavages. The released protein migrates to the nucleus and binds the SRE of the HMG-CoA reductase gene, as well as several other genes in the cholesterol biosynthetic pathway, to enhance transcription. When cholesterol levels rise, the proteolytic release of the SREBP is blocked, and the SREBP in the nucleus is rapidly degraded. These two events halt the transcription of the genes of the cholesterol biosynthetic pathways.
2. The rate of *translation of reductase mRNA* is inhibited by nonsterol metabolites derived from mevalonate as well as by dietary cholesterol.

3. The *degradation of the reductase* is stringently controlled. The enzyme is bipartite: its cytosolic domain carries out catalysis and *its membrane domain senses signals that lead to its degradation*. The membrane domain may undergo a change in its oligomerization state *in response to increasing concentrations of sterols such as cholesterol*, making the enzyme more susceptible to proteolysis. Homologous sterol-sensing regions are present in the protease that activates SREBP. The reductase may be further degraded by ubiquitination and targeting to the 26S proteasome under some conditions. A combination of these three regulatory devices can regulate the amount of enzyme over a 200-fold range.

4. *Phosphorylation decreases the activity of the reductase*. This enzyme, like acetyl CoA carboxylase (which catalyzes the committed step in fatty acid synthesis, [Section 22.5](#)), is switched off by an AMP-activated protein kinase. Thus, cholesterol synthesis ceases when the ATP level is low.

As we will see shortly, all four regulatory mechanisms are modulated by receptors that sense the presence of cholesterol in the blood.

26.3.1. Lipoproteins Transport Cholesterol and Triacylglycerols Throughout the Organism

Cholesterol and triacylglycerols are transported in body fluids in the form of *lipoprotein particles*. Each particle consists of a core of hydrophobic lipids surrounded by a shell of more polar lipids and apoproteins. The protein components of these macromolecular aggregates have two roles: *they solubilize hydrophobic lipids and contain cell-targeting signals*. Lipoprotein particles are classified according to increasing density ([Table 26.1](#)): *chylomicrons*, *chylomicron remnants*, *very low density lipoproteins (VLDL)*, *intermediate-density lipoproteins (IDL)*, *low-density lipoproteins (LDL)*, and *high-density lipoproteins (HDL)*. Ten principal apoproteins have been isolated and characterized. They are synthesized and secreted by the liver and the intestine.


Triacylglycerols, cholesterol, and other lipids obtained from the diet are carried away from the intestine in the form of large *chylomicrons* (180– 500 nm in diameter; [Section 22.1.2](#)). These particles have a very low density ($d < 0.94 \text{ g cm}^{-3}$) because triacylglycerols constitute ~99% of their content. Apolipoprotein B-48 (apo B-48), a large protein (240 kd), forms an amphipathic spherical shell around the fat globule; the external face of this shell is hydrophilic. The triacylglycerols in chylomicrons are released through hydrolysis by *lipoprotein lipases*. These enzymes are located on the lining of blood vessels in muscle and other tissues that use fatty acids as fuels and in the synthesis of fat. The liver then takes up the cholesterol-rich residues, known as *chylomicron remnants*.

The liver is a major site of triacylglycerol and cholesterol synthesis ([Figure 26.15](#)). Triacylglycerols and cholesterol in excess of the liver's own needs are exported into the blood in the form of very low density lipoproteins ($d < 1.006 \text{ g cm}^{-3}$). These particles are stabilized by two lipoproteins—apo B-100 and apo E (34 kd). Apo B-100, one of the largest proteins known (513 kd), is a longer version of apo B-48. Both apo B proteins are encoded by the same gene and produced from the same initial RNA transcript. In the intestine, RNA editing ([Section 28.3.2](#)) modifies the transcript to generate the mRNA for apo B-48, the truncated form. Triacylglycerols in very low density lipoproteins, as in chylomicrons, are hydrolyzed by lipases on capillary surfaces. The resulting remnants, which are rich in cholesteryl esters, are called *intermediate-density lipoproteins* ($1.006 < d < 1.019 \text{ g cm}^{-3}$). These particles have two fates. Half of them are taken up by the liver for processing, and half are converted into low-density lipoprotein ($1.019 < d < 1.063 \text{ g cm}^{-3}$) by the removal of more triacylglycerol.

Low-density lipoprotein is the major carrier of cholesterol in blood. This lipoprotein particle has a diameter of 22 nm and a mass of about 3 million daltons ([Figure 26.16](#)). It contains a core of some 1500 esterified cholesterol molecules; the most common fatty acyl chain in these esters is linoleate, a polyunsaturated fatty acid. A shell of phospholipids and unesterified cholesterols surrounds this highly hydrophobic core. The shell also contains a single copy of apo B-100, which is recognized by target cells. *The role of LDL is to transport cholesterol to peripheral tissues and regulate de novo cholesterol synthesis at these sites*, as described in [Section 26.3.3](#). A different purpose is served by *high-density lipoprotein* ($1.063 < d < 1.21 \text{ g cm}^{-3}$), which picks up cholesterol released into the plasma from dying cells and from membranes undergoing turnover. An acyltransferase in HDL esterifies these cholesterols, which are then either rapidly

shuttled to VLDL or LDL by a specific transfer protein or returned by HDL to the liver.

26.3.2. The Blood Levels of Certain Lipoproteins Can Serve Diagnostic Purposes

 High serum levels of cholesterol cause disease and death by contributing to the formation of atherosclerotic plaques in arteries throughout the body. This excess cholesterol is present in the form of the low density lipoprotein particle, so-called "bad cholesterol." The ratio of cholesterol in the form of high density lipoprotein, sometimes referred to as "good cholesterol," to that in the form of LDL can be used to evaluate susceptibility to the development of heart disease. For a healthy person, the LDL/HDL ratio is 3.5.

High-density lipoprotein functions as a shuttle that moves cholesterol throughout the body. HDL binds and esterifies cholesterol released from the peripheral tissues and then transfers cholesteryl esters to the liver or to tissues that use cholesterol to synthesize steroid hormones. A specific receptor mediates the docking of the HDL to these tissues. The exact nature of the protective effect of HDL levels is not known; however, a possible mechanism is discussed in [Section 26.3.5](#).

26.3.3. Low-Density Lipoproteins Play a Central Role in Cholesterol Metabolism

Cholesterol metabolism must be precisely regulated to prevent atherosclerosis. The mode of control in the liver, the primary site of cholesterol synthesis, has already been discussed: dietary cholesterol reduces the activity and amount of 3-hydroxy-3-methylglutaryl CoA reductase, the enzyme catalyzing the committed step. The results of studies by Michael Brown and Joseph Goldstein are sources of insight into the control of cholesterol metabolism in nonhepatic cells. In general, cells outside the liver and intestine obtain cholesterol from the plasma rather than synthesizing it *de novo*. Specifically, *their primary source of cholesterol is the low-density lipoprotein*. The process of LDL uptake, called *receptor-mediated endocytosis*, serves as a paradigm for the uptake of many molecules.

The steps in the receptor-mediated endocytosis of LDL are as follows (see [Figure 12.40](#)).

1. Apolipoprotein B-100 on the surface of an LDL particle binds to a specific receptor protein on the plasma membrane of nonhepatic cells. The receptors for LDL are localized in specialized regions called *coated pits*, which contain a specialized protein called *clathrin*.
2. The receptor-LDL complex is internalized by *endocytosis*, that is, the plasma membrane in the vicinity of the complex invaginates and then fuses to form an endocytic vesicle ([Figure 26.17](#)).
3. These vesicles, containing LDL, subsequently fuse with *lysosomes*, acidic vesicles that carry a wide array of degradative enzymes. The protein component of the LDL is hydrolyzed to free amino acids. The cholesteryl esters in the LDL are hydrolyzed by a lysosomal acid lipase. The LDL receptor itself usually returns unscathed to the plasma membrane. The round-trip time for a receptor is about 10 minutes; in its lifetime of about a day, it may bring many LDL particles into the cell.
4. *The released unesterified cholesterol can then be used for membrane biosynthesis*. Alternatively, it can be *reesterified for storage inside the cell*. In fact, free cholesterol activates *acyl CoA:cholesterol acyltransferase (ACAT)*, the enzyme catalyzing this reaction. Reesterified cholesterol contains mainly oleate and palmitoleate, which are monounsaturated fatty acids, in contrast with the cholesterol esters in LDL, which are rich in linoleate, a polyunsaturated fatty acid (see [Table 24.1](#)). It is imperative that the cholesterol be reesterified. High concentrations of unesterified cholesterol disrupt the integrity of cell membranes.


The synthesis of LDL receptor is itself subject to feedback regulation. The results of studies of cultured fibroblasts show that, *when cholesterol is abundant inside the cell, new LDL receptors are not synthesized, and so the uptake of additional cholesterol from plasma LDL is blocked*. The gene for the LDL receptor, like that for the reductase, is regulated by SREBP, which binds to a sterol regulatory element that controls the rate of mRNA synthesis.

26.3.4. The LDL Receptor Is a Transmembrane Protein Having Five Different Functional Regions

The amino acid sequence of the human LDL receptor reveals the mosaic structure of this 115-kd protein, which is composed of six different types of domain (Figure 26.18). The amino-terminal region of the mature receptor consists of a cysteine-rich sequence of about 40 residues that is repeated, with some variation, seven times to form the LDL-binding domain (Figure 26.19). A set of conserved acidic side chains in this domain bind calcium ion; this metal ion lies at the center of each domain and, along with disulfide bonds formed from the conserved cysteine residues, stabilizes the three-dimensional structure. Protonation of these glutamate and aspartate side chains of the receptor in lysosomes leads to the release of calcium and hence to structural disruption and the release of LDL from its receptor. A second region of the LDL receptor includes two types of recognizable domains, three domains homologous to epidermal growth factor and six repeats that are similar to the blades of the transducin β subunit (Section 15.2.2). The six repeats form a propeller-like structure that packs against one of the EGF-like domains (Figure 26.20). An aspartate residue forms hydrogen bonds that hold each blade to the rest of the structure. These interactions, too, would most likely be disrupted at the low pH in the lysosome.

The third region contains a single domain that is very rich in serine and threonine residues and contains *O*-linked sugars. These oligosaccharides may function as struts to keep the receptor extended from the membrane so that the LDL-binding domain is accessible to LDL. The fourth region contains the fifth type of domain, which consists of 22 hydrophobic residues that span the plasma membrane. The final region contains the sixth type of domain; it consists of 50 residues and emerges on the cytosolic side of the membrane, where it controls the interaction of the receptor with coated pits and participates in endocytosis. The gene for the LDL receptor consists of 18 exons, which correspond closely to the structural units of the protein. *The LDL receptor is a striking example of a mosaic protein encoded by a gene that was assembled by exon shuffling.*


26.3.5. The Absence of the LDL Receptor Leads to Hypercholesteremia and Atherosclerosis

 The results of Brown and Goldstein's pioneering studies of *familial hypercholesterolemia* revealed the physiologic importance of the LDL receptor. The total concentration of cholesterol and LDL in the plasma is markedly elevated in this genetic disorder, which results from a mutation at a single autosomal locus. The cholesterol level in the plasma of homozygotes is typically 680 mg dl⁻¹, compared with 300 mg dl⁻¹ in heterozygotes (clinical assay results are often expressed in milligrams per deciliter, which is equal to milligrams per 100 milliliters). A value of < 200 mg dl⁻¹ is regarded as desirable, but many people have higher levels. *In familial hypercholesterolemia, cholesterol is deposited in various tissues because of the high concentration of LDL cholesterol in the plasma.* Nodules of cholesterol called *xanthomas* are prominent in skin and tendons. Of particular concern is the oxidation of the excess blood LDL to form oxidized LDL (oxLDL). The oxLDL is taken up by immune-system cells called macrophages, which become engorged to form foam cells. These foam cells become trapped in the walls of the blood vessels and contribute to the formation of atherosclerotic plaques that cause arterial narrowing and lead to heart attacks (Figure 26.21). In fact, *most homozygotes die of coronary artery disease in childhood.* The disease in heterozygotes (1 in 500 people) has a milder and more variable clinical course. A serum esterase that degrades oxidized lipids is found in association with HDL. Possibly, the HDL-associated protein destroys the oxLDL, accounting for HDL's ability to protect against coronary disease.

The molecular defect in most cases of familial hypercholesterolemia is an absence or deficiency of functional receptors for LDL. Receptor mutations that disrupt each of the stages in the endocytotic pathway have been identified.

Homozygotes have almost no functional receptors for LDL, whereas heterozygotes have about half the normal number. Consequently, the entry of LDL into liver and other cells is impaired, leading to an increased plasma level of LDL. Furthermore, less IDL enters liver cells because IDL entry, too, is mediated by the LDL receptor. Consequently, IDL stays in the blood longer in familial hypercholesterolemia, and more of it is converted into LDL than in normal people. All deleterious consequences of an absence or deficiency of the LDL receptor can be attributed to the ensuing elevated level of LDL cholesterol in the blood.

26.3.6. The Clinical Management of Cholesterol Levels Can Be Understood at a Biochemical Level

 Homozygous familial hypercholesterolemia can be treated only by a liver transplant. A more generally applicable therapy is available for heterozygotes and others with high levels of cholesterol. *The goal is to reduce the amount of cholesterol in the blood by stimulating the single normal gene to produce more than the customary number of LDL receptors.* We have already observed that the production of LDL receptors is controlled by the cell's need for cholesterol. Therefore, in essence, the strategy is to deprive the cell of ready sources of cholesterol. When cholesterol is required, the amount of mRNA for the LDL receptor rises and more receptor is found on the cell surface. This state can be induced by a two-pronged approach. First, the intestinal reabsorption of bile salts is inhibited. Bile salts are cholesterol derivatives that promote the absorption of dietary cholesterol and dietary fats (Section 22.1.1). Second, de novo synthesis of cholesterol is blocked.

The reabsorption of bile is impeded by oral administration of positively charged polymers, such as cholestyramine, that bind negatively charged bile salts and are not themselves absorbed. Cholesterol synthesis can be effectively blocked by a class of compounds called *statins* (e.g., lovastatin, which is also called mevacor; Figure 26.22). These compounds are potent competitive inhibitors ($K_i < 1$ nM) of HMG-CoA reductase, the essential control point in the biosynthetic pathway. Plasma cholesterol levels decrease by 50% in many patients given both lovastatin and inhibitors of bile-salt reabsorption. Lovastatin and other inhibitors of HMG-CoA reductase are widely used to lower the plasma cholesterol level in people who have atherosclerosis, which is the leading cause of death in industrialized societies.

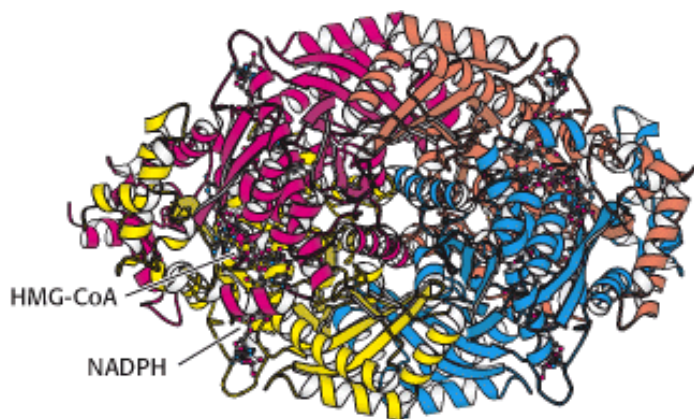


Figure 26.14. HMG-CoA Reductase. The structure of a portion of the tetrameric enzyme is shown.



Table 26.1. Properties of plasma lipoproteins

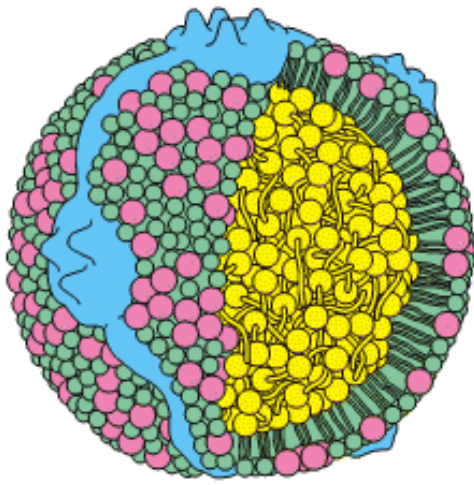
Lipoproteins	Major core lipids	Apoproteins	Mechanism of lipid delivery
Chylomicron	Dietary triacylglycerols	B-48, C, E	Hydrolysis by lipoprotein lipase
Chylomicron remnant	Dietary cholesterol esters	B-48, E	Receptor-mediated endocytosis by liver
Very low density lipoprotein (VLDL)	Endogenous triacylglycerols	B-100, C, E	Hydrolysis by lipoprotein lipase

Intermediate-density lipoprotein (IDL)	Endogenous cholesterol esters	B-100, E	Receptor-mediated endocytosis by liver and conversion into LDL
Low-density lipoprotein (LDL)	Endogenous cholesterol esters	B-100	Receptor-mediated endocytosis by liver and other tissues
High-density lipoprotein (HDL)	Endogenous cholesterol esters	A	Transfer of cholesterol esters to IDL and LDL

Source: After M. S. Brown and J. L. Goldstein, *The Pharmacological Basis of Therapeutics*. 7th ed., A. G. Gilman, L. S. Goodman, T. W. Rall, and F. Murad, Eds. (Macmillan, 1985), p. 828.



Figure 26.15. Site of Cholesterol Synthesis. Electron micrograph of a part of a liver cell actively engaged in the synthesis and secretion of very low density lipoprotein (VLDL). The arrow points to a vesicle that is releasing its content of VLDL particles. [Courtesy of Dr. George Palade.]



- Unesterified cholesterol
- Phospholipid
- Cholesteryl ester
- Apoprotein B-100

Figure 26.16. Schematic Model of Low-Density Lipoprotein. The LDL particle is approximately 22 nm (220 Å) in diameter.

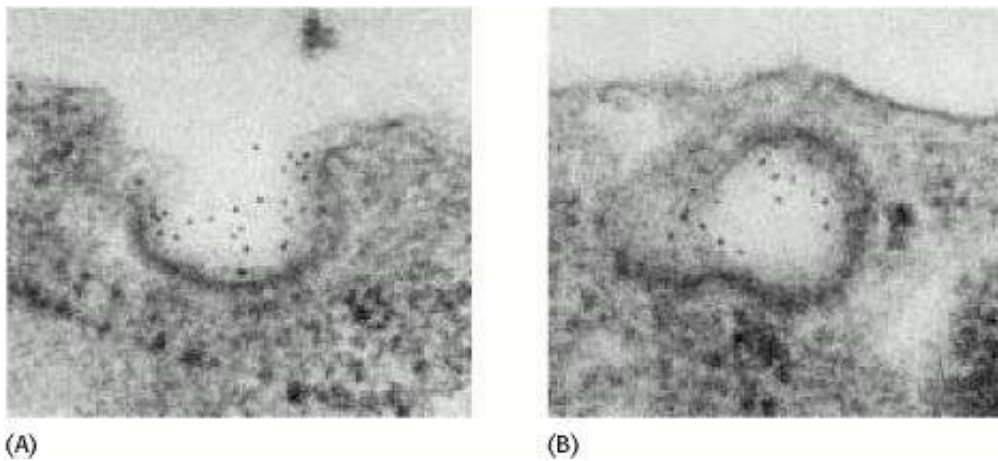


Figure 26.17. Endocytosis of LDL Bound to Its Receptor. (A) Electron micrograph showing LDL (conjugated to ferritin for visualization, dark spots) bound to a coated-pit region on the surface of a cultured human fibroblast cell. (B) Micrograph showing this region invaginating and fusing to form an endocytic vesicle [From R. G. W. Anderson, M. S. Brown, and J. L. Goldstein. *Cell* 10 (1977): 351.]

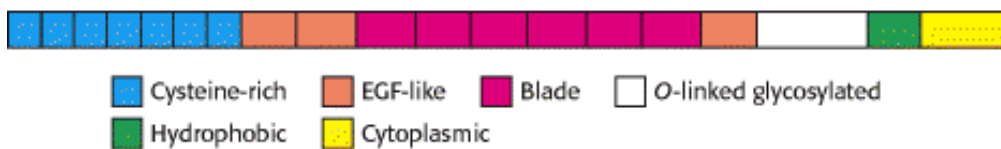


Figure 26.18. LDL Receptor Domains. A schematic representation of the amino acid sequence of the LDL receptor showing six types of domain.

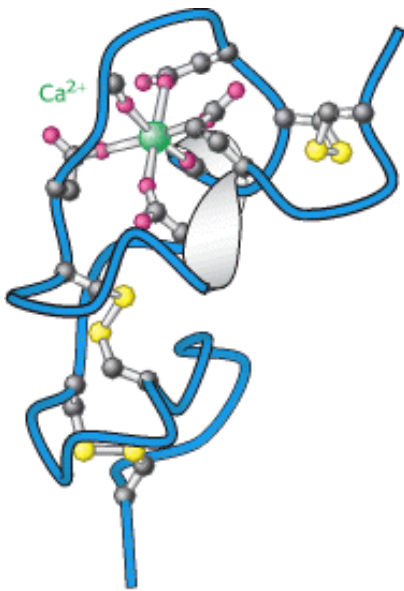


Figure 26.19. Structure of Cysteine-Rich Domain. This calcium-binding cysteine-rich domain is repeated seven times at the amino terminus of the LDL receptor.

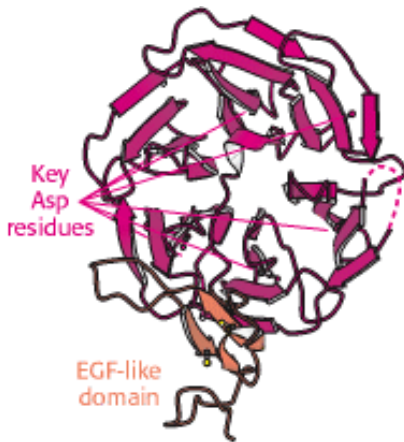


Figure 26.20. Structure of Propeller Domain. The six-bladed propeller domain and an adjacent EGF-like domain of the LDL receptor.

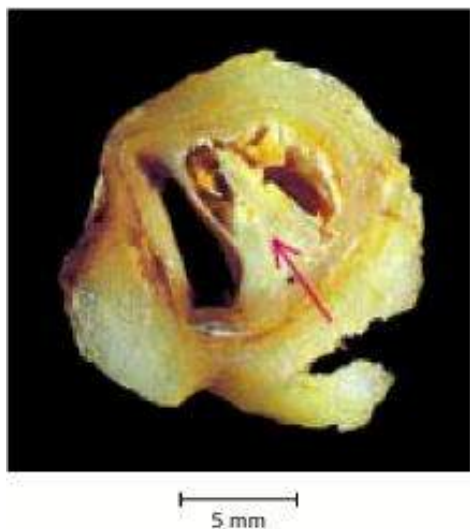


Figure 26.21. An Atherosclerotic Plaque. A plaque (marked by an arrow) blocks most of the lumen of this blood vessel. The plaque is rich in cholesterol. [Courtesy of Dr. Jeffrey Sklar.]

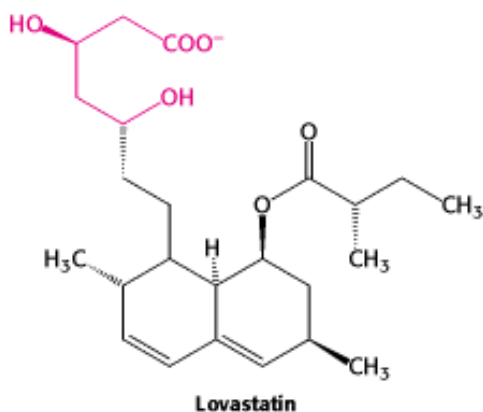


Figure 26.22. Lovastatin, a Competitive Inhibitor of HMG-CoA Reductase. The part of the structure that resembles the 3-hydroxy-3-methylglutaryl moiety is shown in red.

26.4. Important Derivatives of Cholesterol Include Bile Salts and Steroid Hormones

Cholesterol is a precursor for other important steroid molecules: the bile salts, steroid hormones, and vitamin D.

Bile Salts.

As polar derivatives of cholesterol, *bile salts* are highly effective *detergents* because they contain both polar and nonpolar regions. Bile salts are synthesized in the liver, stored and concentrated in the gall bladder, and then released into the small intestine. Bile salts, the major constituent of bile, *solubilize dietary lipids* (Section 22.1.1). Solubilization increases in the effective surface area of lipids with two consequences: more surface area is exposed to the digestive action of lipases and lipids are more readily absorbed by the intestine. Bile salts are also the major breakdown products of cholesterol.

Cholesterol is converted into trihydroxycoprostanate and then into *cholyl CoA*, the activated intermediate in the synthesis of most bile salts (Figure 26.23). The activated carboxyl carbon of cholyl CoA then reacts with the amino group of glycine to form *glycocholate* or it reacts with the amino group of taurine ($\text{H}_2\text{NCH}_2\text{CH}_2\text{SO}_3^-$), derived from cysteine, to form *taurocholate*. *Glycocholate is the major bile salt.*

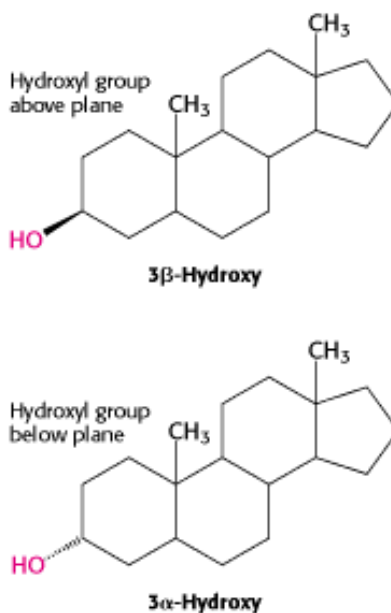
Steroid Hormones.

Cholesterol is the precursor of the five major classes of *steroid hormones*: progestagens, glucocorticoids, mineralocorticoids, androgens, and estrogens (Figure 26.24). These hormones are powerful signal molecules that regulate a host of organismal functions. *Progesterone*, a *progestagen*, prepares the lining of the uterus for implantation of an ovum. Progesterone is also essential for the maintenance of pregnancy. *Androgens* of male secondary sex characteristics, whereas *estrogens* (such as *estrone*) are required for the development of female secondary sex characteristics. Estrogens, along with progesterone, also participate in the ovarian cycle. *Glucocorticoids* (such as *cortisol*) promote gluconeogenesis and the formation of glycogen, enhance the degradation of fat and protein, and inhibit the inflammatory response. They enable animals to respond to stress—indeed, the absence of glucocorticoids can be fatal. *Mineralocorticoids* (primarily *aldosterone*) act on the distal tubules of the kidney to increase the reabsorption of Na^+ and the excretion of K^+ and H^+ , which leads to an increase in blood volume and blood pressure. The major sites of synthesis of these classes of hormones are the corpus luteum, for progestagens; the ovaries, for estrogens; the testes, for androgens; and the adrenal cortex, for glucocorticoids and mineralocorticoids.

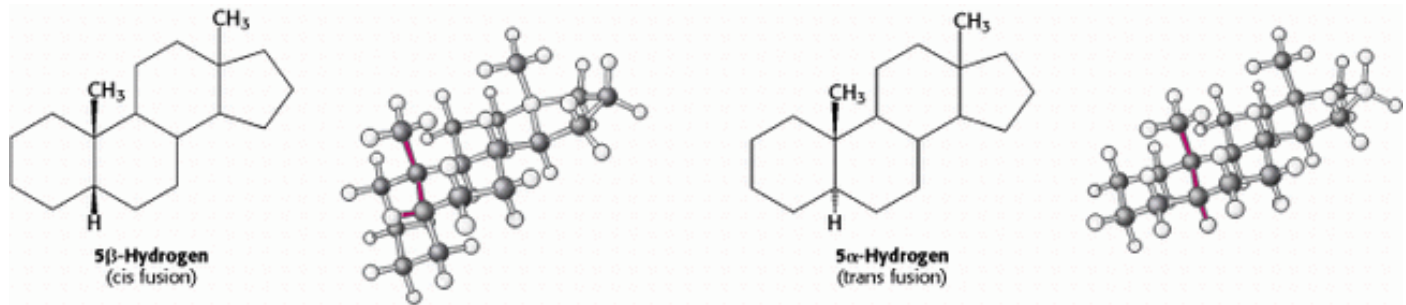
Steroid hormones bind to and activate receptor molecules that serve as transcription factors to regulate gene expression (Section 31.3.1). These small, relatively similar molecules are able to have greatly differing effects because the slight structural differences among them allow interactions with specific receptor molecules.

26.4.1. The Nomenclature of Steroid Hormones

Carbon atoms in steroids are numbered as shown for cholesterol in (Figure 26.25). The rings in steroids are denoted by the letters A, B, C, and D. Cholesterol contains two angular methyl groups: the C-19 methyl group is attached to C-10, and the C-18 methyl group is attached to C-13. The C-18 and C-19 methyl groups of cholesterol lie *above* the plane containing the four rings. A substituent that is above the plane is termed β oriented, whereas a substituent that is below the plane is α oriented.



If a hydrogen atom is attached to C-5, it can be either α or β oriented. The A and B steroid rings are fused in a *trans* conformation if the C-5 hydrogen is α oriented, and *cis* if it is β oriented. The absence of a Greek letter for the C-5 hydrogen atom on the steroid nucleus implies a *trans* fusion. The C-5 hydrogen atom is α oriented in all steroid hormones that contain a hydrogen atom in that position. In contrast, bile salts have a β -oriented hydrogen atom at C-5. Thus, a *cis* fusion is characteristic of the bile salts, whereas a *trans* fusion is characteristic of all steroid hormones that possess a hydrogen atom at C-5. A *trans* fusion yields a nearly planar structure, whereas a *cis* fusion gives a buckled structure.



26.4.2. Steroids Are Hydroxylated by Cytochrome P450 Monooxygenases That Utilize NADPH and O₂

Hydroxylation reactions play a very important role in the synthesis of cholesterol from squalene and in the conversion of cholesterol into steroid hormones and bile salts. All these hydroxylations require *NADPH and O₂*. The oxygen atom of the incorporated hydroxyl group comes from O₂ rather than from H₂O. While one oxygen atom of the O₂ molecule goes into the substrate, the other is reduced to water. The enzymes catalyzing these reactions are called *monooxygenases* (or *mixed-function oxygenases*). Recall that a monooxygenase also participates in the hydroxylation of aromatic amino acids (Section 23.5.7).




Hydroxylation requires the activation of oxygen. In the synthesis of steroid hormones and bile salts, activation is accomplished by a cytochrome P450, a family of cytochromes that absorb light maximally at 450 nm when complexed in vitro with exogenous carbon monoxide. These membraneanchored proteins (~50 kd) contain a heme prosthetic group. Because the hydroxylation reactions promoted by P450 enzymes are oxidation reactions, it is at first glance surprising that they also consume the reductant NADPH. NADPH transfers its high-potential electrons to a flavoprotein, which transfers them, one at a time, to *adrenodoxin*, a nonheme iron protein. Adrenodoxin transfers one electron to reduce the ferric (Fe³⁺) form of P450 to the ferrous (Fe²⁺) form (Figure 26.26). Without the addition of this electron, P450 will not bind oxygen. Recall that only the ferrous form of hemoglobin binds oxygen (Section 10.2.1). The binding of O₂ to the heme is followed by the acceptance of a second electron from adrenodoxin. The acceptance of this second electron leads to cleavage of the O–O bond. One of the oxygen atoms is then protonated and released as water. The remaining oxygen atom forms a highly reactive ferryl (Fe = O) intermediate. This intermediate abstracts a hydrogen atom from the substrate RH to form R•. This transient free radical captures the OH group from the iron atom to form ROH, the hydroxylated product, returning the iron atom to the ferric state.

26.4.3. The Cytochrome P450 System Is Widespread and Performs a Protective Function

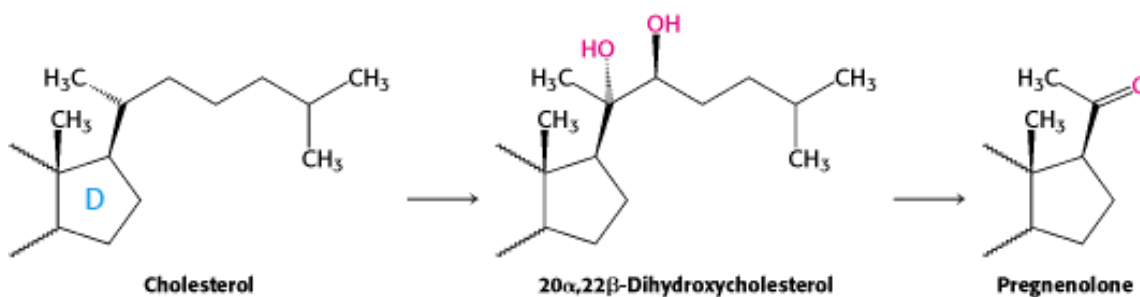
The cytochrome P450 system, which in mammals is located primarily in the endoplasmic reticulum of the liver and small intestine, is also important in the *detoxification of foreign substances* (xenobiotic compounds) by oxidative metabolism. For example, the hydroxylation of phenobarbital, a barbiturate, *increases its solubility and facilitates its excretion*. Likewise, polycyclic aromatic hydrocarbons are hydroxylated by P450, providing sites for conjugation with highly polar units (e.g., glucuronate or sulfate), which markedly increase the solubility of the modified aromatic molecule. One of the most relevant functions of the cytochrome P450 system to human beings is its role in drug metabolism. Drugs such as caffeine and ibuprofen are oxidatively metabolized by these monooxygenases. Indeed, the duration of action of many medications depends on their rate of inactivation by the P450 system. Despite its general protective role in the removal of foreign chemicals, the action of the P450 system is not always beneficial. *Some of the most powerful carcinogens are generated from harmless compounds by the P450 system in vivo* in the process of *metabolic activation*. In plants, the cytochrome P450 system plays a role in the synthesis of toxic compounds as well as

the pigments of flowers.

 The cytochrome P450 system is a ubiquitous superfamily of monooxygenases that is present in plants, animals, and prokaryotes. The human genome encodes more than 50 members of the family, whereas the genome of the plant *Arabidopsis* encodes more than 250 members. All members of this large family arose by gene duplication followed by subsequent divergence that generated a range of substrate specificity. Indeed, the specificity of these enzymes is encoded in delimited regions of the primary structure, and the substrate specificity of closely related members is often defined by a few critical residues or even a single amino acid.

26.4.4. Pregnenolone, a Precursor for Many Other Steroids, Is Formed from Cholesterol by Cleavage of Its Side Chain

Steroid hormones contain 21 or fewer carbon atoms, whereas cholesterol contains 27. Thus, the first stage in the synthesis of steroid hormones is the removal of a six-carbon unit from the side chain of cholesterol to form *pregnenolone*. The side chain of cholesterol is hydroxylated at C-20 and then at C-22, and the bond between these carbon atoms is subsequently cleaved by *desmolase*. Three molecules of NADPH and three molecules of O₂ are consumed in this remarkable six-electron oxidation.



Adrenocorticotrophic hormone (ACTH, or corticotropin), a polypeptide synthesized by the anterior pituitary gland, stimulates the conversion of cholesterol into pregnenolone, the precursor of all steroid hormones.

26.4.5. The Synthesis of Progesterone and Corticosteroids from Pregnenolone

Progesterone is synthesized from pregnenolone in two steps. The 3-hydroxyl group of pregnenolone is oxidized to a 3-keto group, and the Δ^5 double bond is isomerized to a Δ^4 double bond (Figure 26.27). *Cortisol*, the major glucocorticoid, is synthesized from progesterone by hydroxylations at C-17, C-21, and C-11; C-17 must be hydroxylated before C-21 is, whereas C-11 can be hydroxylated at any stage. The enzymes catalyzing these hydroxylations are highly specific, as shown by some inherited disorders. The initial step in the synthesis of *aldosterone*, the major mineralocorticoid, is the hydroxylation of progesterone at C-21. The resulting deoxycorticosterone is hydroxylated at C-11. The oxidation of the C-18 angular methyl group to an aldehyde then yields aldosterone.


26.4.6. The Synthesis of Androgens and Estrogens from Pregnenolone

Androgens and estrogens also are synthesized from pregnenolone through the intermediate progesterone. Androgens contain 19 carbon atoms. The synthesis of androgens (Figure 26.28) starts with the hydroxylation of progesterone at C-17. The side chain consisting of C-20 and C-21 is then cleaved to yield *androstenedione*, an androgen. *Testosterone*, another androgen, is formed by the reduction of the 17-keto group of androstenedione. Testosterone, through its actions in the brain, is paramount in the development of male sexual behavior. It is also important for maintenance of the testes and development of muscle mass. Owing to the latter activity, testosterone is referred to as an *anabolic steroid*. Testosterone is reduced by *5 α -reductase* to yield *dihydrotestosterone (DHT)*, a powerful embryonic androgen that instigates the development and differentiation of the male phenotype. Estrogens are synthesized from androgens by the loss of the C-19 angular methyl group and the formation of an aromatic A ring. *Estrone*, an estrogen, is derived from

androstenedione, whereas *estradiol*, another estrogen, is formed from testosterone.

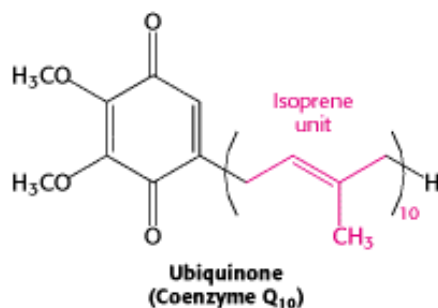
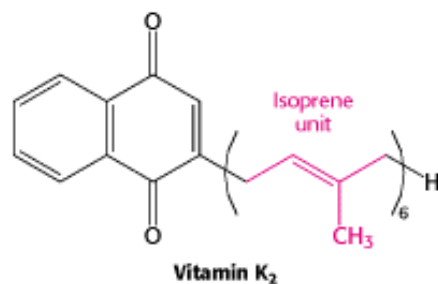
26.4.7. Vitamin D Is Derived from Cholesterol by the Ring-Splitting Activity of Light

Cholesterol is also the precursor of vitamin D, which plays an essential role in the control of calcium and phosphorus metabolism. *7-Dehydrocholesterol* (provitamin D_3) is photolyzed by the ultraviolet light of sunlight to previtamin D_3 , which spontaneously isomerizes to vitamin D_3 (Figure 26.29). Vitamin D_3 (cholecalciferol) is converted into *calcitriol* (1,25-dihydroxycholecalciferol), the active hormone, by hydroxylation reactions in the liver and kidneys. Although not a steroid, vitamin D acts in an analogous fashion. It binds to a receptor, structurally similar to the steroid receptors, to form a complex that functions as a transcription factor, regulating gene expression.

 Vitamin D deficiency in childhood produces *rickets*, a disease characterized by inadequate calcification of cartilage and bone. Rickets was so common in seventeenth-century England that it was called the "children's disease of the English." The 7-dehydrocholesterol in the skin of these children was not photolyzed to previtamin D_3 , because there was little sunlight for many months of the year. Furthermore, their diets provided little vitamin D, because most naturally occurring foods have a low content of this vitamin. Fish-liver oils are a notable exception. Cod-liver oil, abhorred by generations of children because of its unpleasant taste, was used in the past as a rich source of vitamin D. Today, the most reliable dietary sources of vitamin D are fortified foods. Milk, for example, is fortified to a level of 400 international units per quart (10 μg per quart). The recommended daily intake of vitamin D is 400 international units, irrespective of age. In adults, vitamin D deficiency leads to softening and weakening of bones, a condition called *osteomalacia*. The occurrence of osteomalacia in Bedouin Arab women who are clothed so that only their eyes are exposed to sunlight is a striking reminder that vitamin D is needed by adults as well as by children.

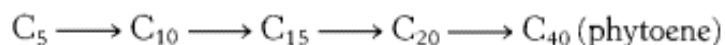
26.4.8. Isopentenyl Pyrophosphate Is a Precursor for a Wide Variety of Biomolecules

Before this chapter ends, we will revisit isopentenyl pyrophosphate, the activated precursor of cholesterol. The combination of isopentenyl pyrophosphate (C_5) units to form squalene (C_{30}) exemplifies a fundamental mechanism for the assembly of carbon skeletons of biomolecules. *A remarkable array of compounds is formed from isopentenyl pyrophosphate, the basic five-carbon building block.* The fragrances of many plants arise from volatile C_{10} and C_{15} compounds, which are called *terpenes*. For example, myrcene ($C_{10}H_{16}$) from bay leaves consists of two isoprene units, as does limonene ($C_{10}H_{16}$) from lemon oil (Figure 26.30). Zingiberene ($C_{15}H_{24}$), from the oil of ginger, is made up of three isoprene units. Some terpenes, such as geraniol from geraniums and menthol from peppermint oil, are alcohols; others, such as citronellal, are aldehydes. We shall see later (Chapter 32) how specialized sets of 7-TM receptors are responsible for the diverse and delightful odor and taste sensations that these molecules induce.



We have already encountered several molecules that contain isoprenoid side chains. The C_{30} hydrocarbon *side chain* of *vitamin K_2* , an important molecule in clotting ([Section 10.5.7](#)), is built from 6 isoprene (C_5) units. *Coenzyme Q_{10}* in the mitochondrial respiratory chain ([Section 18.3](#)) has a side chain made up of 10 isoprene units. Yet another example is the *phytol side chain* of *chlorophyll* ([Section 19.2](#)), which is formed from 4 isoprene units. Many proteins are targeted to membranes by the covalent attachment of a farnesyl (C_{15}) or a geranylgeranyl (C_{20}) unit to the carboxyl-terminal cysteine residue of the protein ([Section 12.5.3](#)). *The attachment of isoprenoid side chains confers hydrophobic character.*

Isoprenoids can delight by their color as well as by their fragrance. The color of tomatoes and carrots comes from *carotenoids*. These compounds absorb light because they contain extended networks of single and double bonds and are important pigments in photosynthesis ([Section 19.5.2](#)). Their C_{40} carbon skeletons are built by the successive addition of C_5 units to form *geranylgeranyl pyrophosphate*, a C_{20} intermediate, which then condenses tail-to-tail with another molecule of geranylgeranyl pyrophosphate.



Phytoene, the C_{40} condensation product, is dehydrogenated to yield lycopene. Cyclization of both ends of lycopene gives β -carotene, which is the precursor of retinal, the chromophore in all known visual pigments ([Section 32.3.2](#)). *These examples illustrate the fundamental role of isopentenyl pyrophosphate in the assembly of extended carbon skeletons of biomolecules.* It is evident that isoprenoids are ubiquitous in nature and have diverse significant roles, including the enhancement of the sensuality of life.

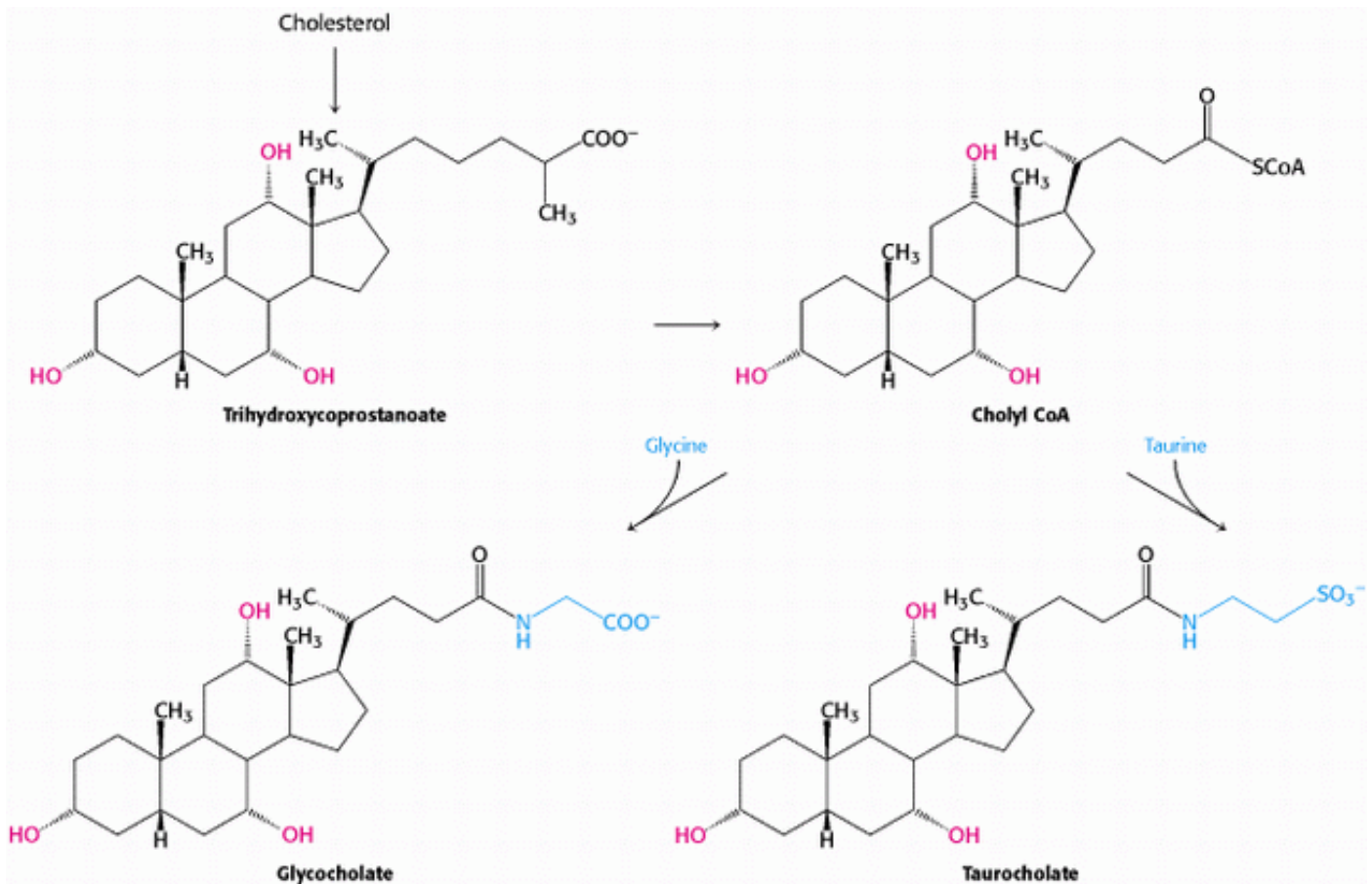


Figure 26.23. Synthesis of Bile Salts. Pathways for the formation of bile salts from cholesterol.

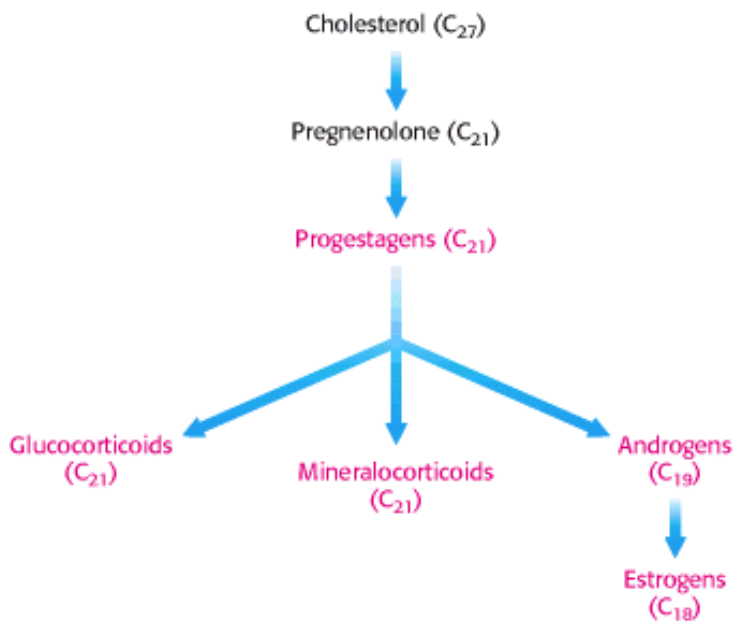


Figure 26.24. Biosynthetic Relations of Classes of Steroid Hormones and Cholesterol.

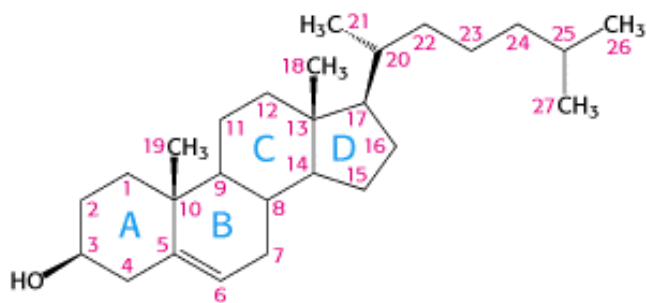


Figure 26.25. Cholesterol Carbon Numbering. The numbering scheme for the carbon atoms in cholesterol and other steroids.

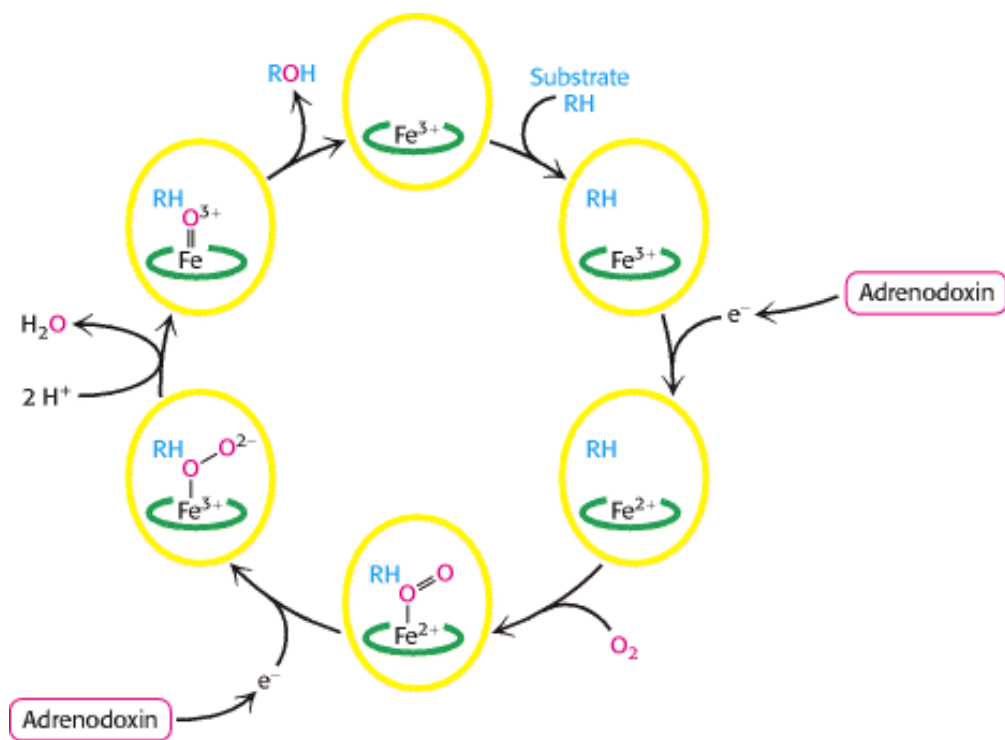


Figure 26.26. Cytochrome P450 Mechanism. These enzyme-bind O_2 and use one oxygen atom to hydroxylate their substrates.

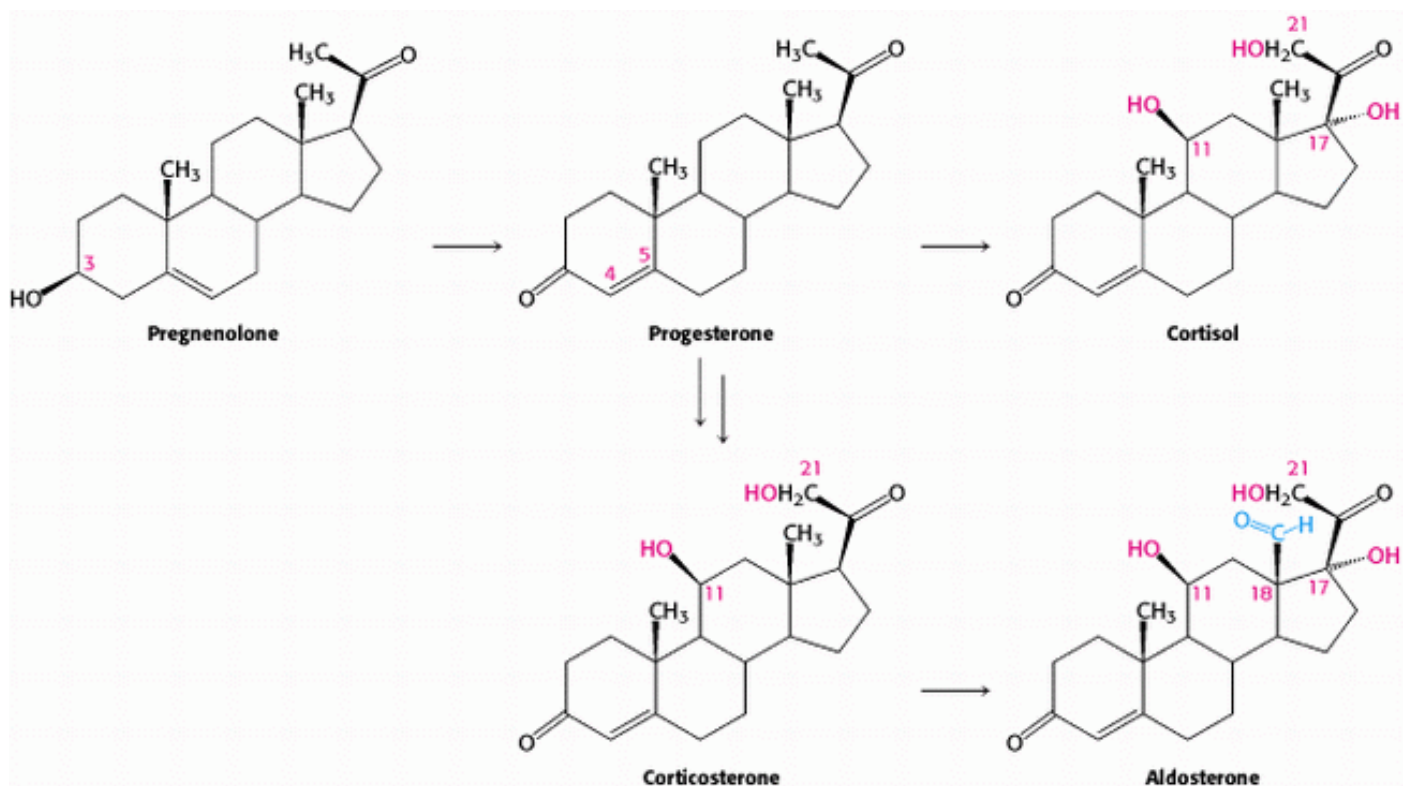


Figure 26.27. Pathways for the Formation of Progesterone, Cortisol, and Aldosterone.

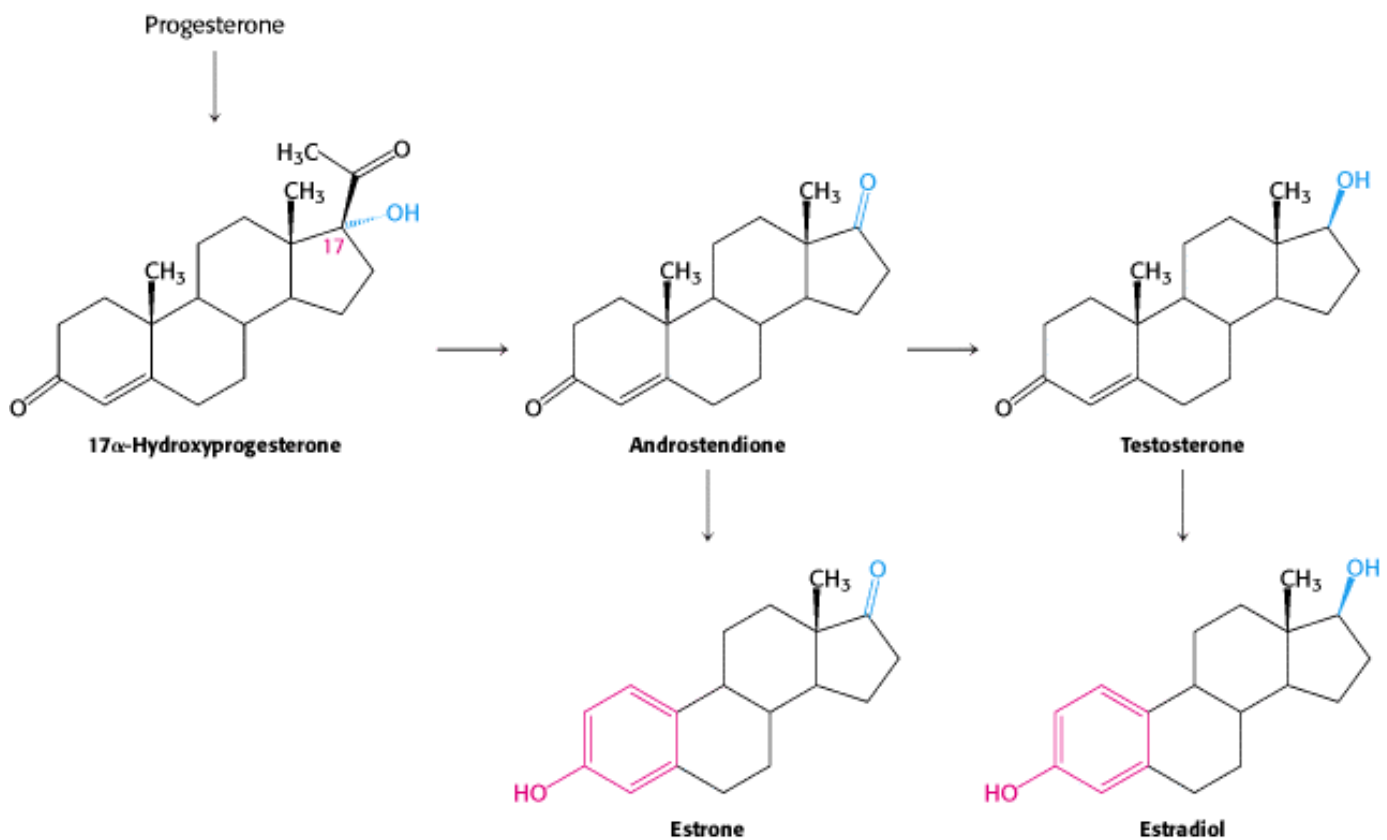


Figure 26.28. Pathways for the Formation of Androgens and Estrogens.

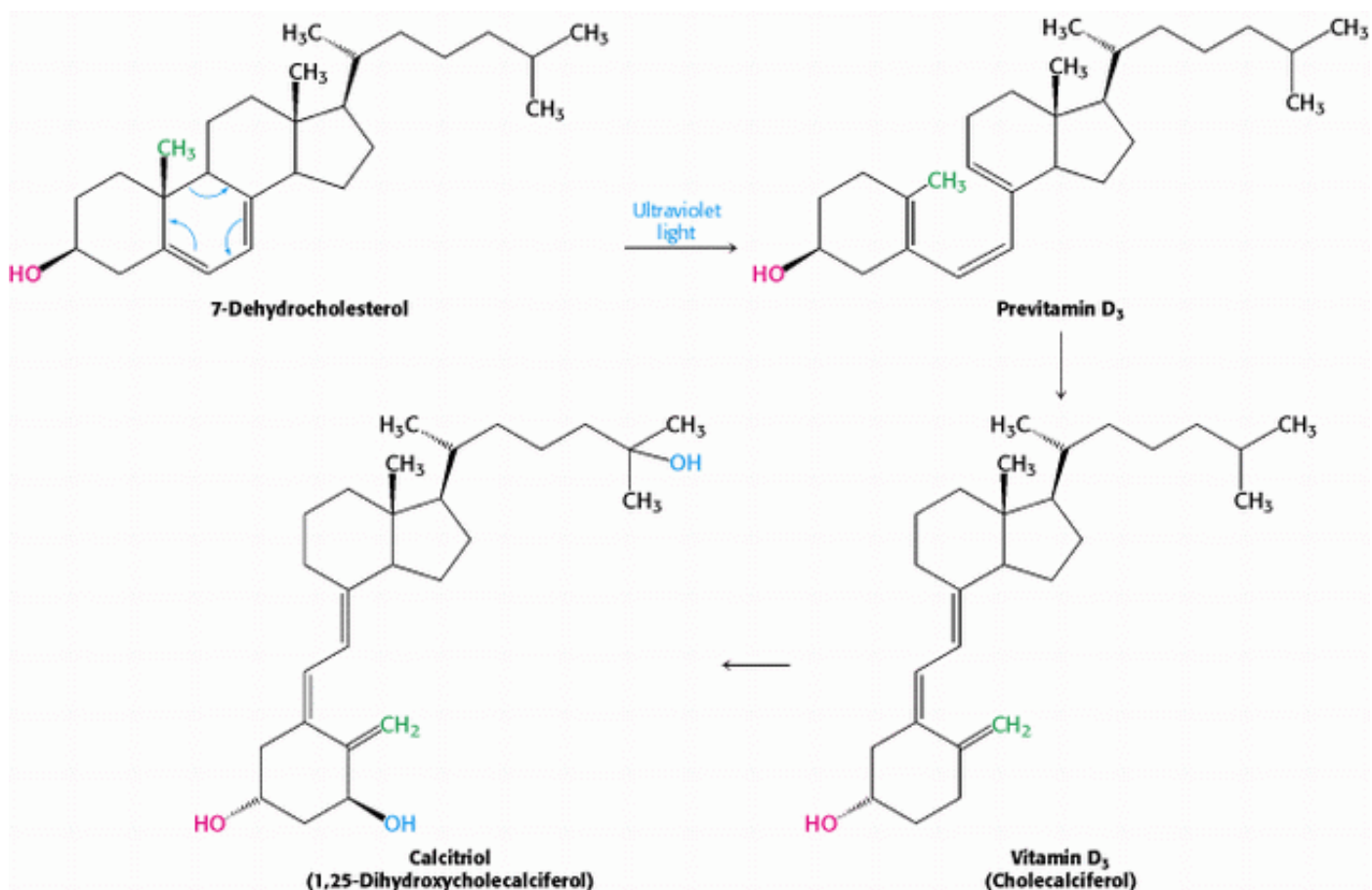


Figure 26.29. Vitamin D Synthesis. The pathway for the conversion of 7-dehydrocholesterol into vitamin D₃ and then into calcitriol, the active hormone.

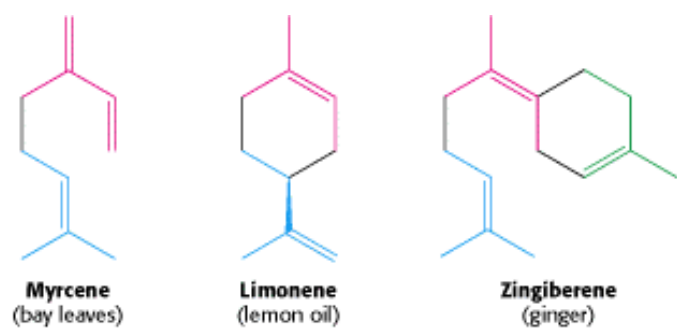


Figure 26.30. Three Isoprenoids from Familiar Sources.

Summary

Phosphatidate Is a Common Intermediate in the Synthesis of Phospholipids and Triacylglycerols

Phosphatidate is formed by successive acylations of glycerol 3-phosphate by acyl CoA. Hydrolysis of its phosphoryl group followed by acylation yields a triacylglycerol. CDP-diacylglycerol, the activated intermediate in the de novo synthesis of several phosphoglycerides, is formed from phosphatidate and CTP. The activated phosphatidyl unit is then transferred to the hydroxyl group of a polar alcohol, such as serine, to form a phospholipid such as phosphatidyl serine. In bacteria, decarboxylation of this phosphoglyceride yields phosphatidyl ethanolamine, which is methylated by *S*-adenosylmethionine to form phosphatidyl choline. In mammals, this phosphoglyceride is synthesized by a pathway that utilizes dietary choline. CDP-choline is the activated intermediate in this route.

Sphingolipids are synthesized from ceramide, which is formed by the acylation of sphingosine. Gangliosides are sphingolipids that contain an oligosaccharide unit having at least one residue of *N*-acetylneuraminate or a related sialic acid. They are synthesized by the step-by-step addition of activated sugars, such as UDP-glucose, to ceramide.

Cholesterol Is Synthesized from Acetyl Coenzyme A in Three Stages

Cholesterol is a steroid component of eukaryotic membranes and a precursor of steroid hormones. The committed step in its synthesis is the formation of mevalonate from 3-hydroxy-3-methylglutaryl CoA (derived from acetyl CoA and acetoacetyl CoA). Mevalonate is converted into isopentenyl pyrophosphate (C_5), which condenses with its isomer, dimethylallyl pyrophosphate (C_5), to form geranyl pyrophosphate (C_{10}). The addition of a second molecule of isopentenyl pyrophosphate yields farnesyl pyrophosphate (C_{15}), which condenses with itself to form squalene (C_{30}). This intermediate cyclizes to lanosterol (C_{30}), which is modified to yield cholesterol (C_{27}).

The Complex Regulation of Cholesterol Biosynthesis Takes Place at Several Levels

In the liver, cholesterol synthesis is regulated by changes in the amount and activity of 3-hydroxy-3-methylglutaryl CoA reductase. Transcription of the gene, translation of the mRNA, and degradation of the enzyme are stringently controlled. In addition, the activity of the reductase is regulated by phosphorylation.

Triacylglycerols exported by the intestine are carried by chylomicrons and then hydrolyzed by lipases lining the capillaries of target tissues. Cholesterol and other lipids in excess of those needed by the liver are exported in the form of very low density lipoprotein. After delivering its content of triacylglycerols to adipose tissue and other peripheral tissue, VLDL is converted into intermediate-density lipoprotein and then into low-density lipoprotein. IDL and LDL carry cholesteryl esters, primarily cholesteryl linoleate. Liver and peripheral tissue cells take up LDL by receptor-mediated endocytosis. The LDL receptor, a protein spanning the plasma membrane of the target cell, binds LDL and mediates its entry into the cell. Absence of the LDL receptor in the homozygous form of familial hypercholesterolemia leads to a markedly elevated plasma level of LDL-cholesterol, the deposition of cholesterol on blood-vessel walls, and heart attacks in childhood. Apolipoprotein B, a very large protein, is a key structural component of chylomicrons, VLDL, and LDL.

Important Derivatives of Cholesterol Include Bile Salts and Steroid Hormones

In addition to bile salts, which facilitate the digestion of lipids, five major classes of steroid hormones are derived from cholesterol: progestagens, glucocorticoids, mineralocorticoids, androgens, and estrogens. Hydroxylations by P450 monooxygenases that use NADPH and O_2 play an important role in the synthesis of steroid hormones and bile salts from

cholesterol. P450 enzymes, a large superfamily, also participate in the detoxification of drugs and other foreign substances.

Pregnenolone (C_{21}) is an essential intermediate in the synthesis of steroids. This steroid is formed by scission of the side chain of cholesterol. Progesterone (C_{21}), synthesized from pregnenolone, is the precursor of cortisol and aldosterone. Hydroxylation of progesterone and cleavage of its side chain yields androstenedione, an androgen (C_{19}). Estrogens (C_{18}) are synthesized from androgens by the loss of an angular methyl group and the formation of an aromatic A ring. Vitamin D, which is important in the control of calcium and phosphorus metabolism, is formed from a derivative of cholesterol by the action of light.

In addition to cholesterol and its derivatives, a remarkable array of biomolecules are synthesized from isopentenyl pyrophosphate, the basic five-carbon building block. The hydrocarbon side chains of vitamin K₂, coenzyme Q₁₀, and chlorophyll are extended chains constructed from this activated C₅ unit. Prenyl groups that are derived from this activated intermediate target many proteins to membranes.

Key Terms

phosphatidate

triacylglycerol

phospholipid

cytidine diphosphodiacylglycerol (CDP-diacylglycerol)

glyceryl ether phospholipid

sphingolipid

ceramide (*N*-acyl sphingosine)

cerebroside

ganglioside

cholesterol

mevalonate

3-hydroxy-3-methylglutaryl CoA reductase (HMG-CoA reductase)

3-isopentenyl pyrophosphate

sterol regulatory element binding protein (SREBP)

lipoprotein particles

low-density lipoprotein (LDL)

high-density lipoprotein (HDL)

receptor-mediated endocytosis

bile salts

steroid hormones

cytochrome P450 monooxygenase

pregnenolone

Problems

1. *Making fat.* Write a balanced equation for the synthesis of a triacylglycerol, starting from glycerol and fatty acids.

See answer

2. *Making phospholipid.* Write a balanced equation for the synthesis of phosphatidyl serine by the de novo pathway, starting from serine, glycerol, and fatty acids.

See answer

3. *Activated donors.* What is the activated reactant in each of the following biosyntheses?

(a) Phosphatidyl serine from serine

(b) Phosphatidyl ethanolamine from ethanolamine

(c) Ceramide from sphingosine

(d) Sphingomyelin from ceramide

(e) Cerebroside from ceramide

(f) Ganglioside G_{M1} from ganglioside G_{M2}

(g) Farnesyl pyrophosphate from geranyl pyrophosphate

See answer

4. *Telltale labels.* What is the distribution of isotopic labeling in cholesterol synthesized from each of the following precursors?

(a) Mevalonate labeled with ^{14}C in its carboxyl carbon atom

(b) Malonyl CoA labeled with ^{14}C in its carboxyl carbon atom

See answer

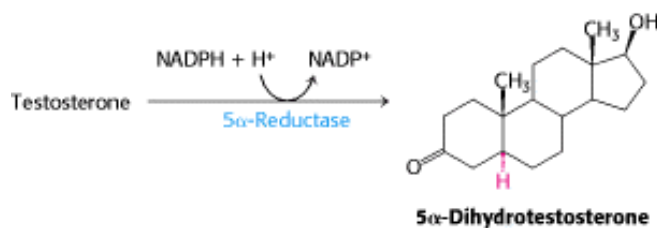
5. *Familial hypercholesterolemia.* Several classes of LDL-receptor mutations have been identified as causes of this disease. Suppose that you have been given cells from patients with different mutations, an antibody specific for the LDL receptor that can be seen with an electron microscope, and access to an electron microscope. What differences in antibody distribution might you expect to find in the cells from different patients?

See answer

6. *RNA editing.* A shortened version (apo B-48) of apolipoprotein B is formed by the intestine, whereas the full-length protein (apo B-100) is synthesized by the liver. A glutamine codon (CAA) is changed into a stop codon. Propose a simple mechanism for this change.

See answer

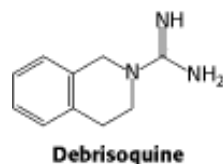
7. *Inspiration for drug design.* Some actions of androgens are mediated by dihydrotestosterone, which is formed by the reduction of testosterone. This finishing touch is catalyzed by an NADPH-dependent 5 α -reductase.



Chromosomal XY males with a genetic deficiency of this reductase are born with a male internal urogenital tract but predominantly female external genitalia. These people are usually reared as girls. At puberty, they masculinize because the testosterone level rises. The testes of these reductase-deficient men are normal, whereas their prostate glands remain small. How might this information be used to design a drug to treat *benign prostatic hypertrophy*, a common consequence of the normal aging process in men? A majority of men older than age 55 have some degree of prostatic enlargement, which often leads to urinary obstruction.

See answer

8. *Drug idiosyncrasies.* Debrisoquine, a β -adrenergic blocking agent, has been used to treat hypertension. The optimal dose varies greatly (20–400 mg daily) in a population of patients. The urine of most patients taking the drug contains a high level of 4-hydroxydebrisoquine. However, those most sensitive to the drug (about 8% of the group studied) excrete debrisoquine and very little of the 4-hydroxy derivative. Propose a molecular basis for this drug idiosyncrasy. Why should caution be exercised in giving other drugs to patients who are very sensitive to debrisoquine?

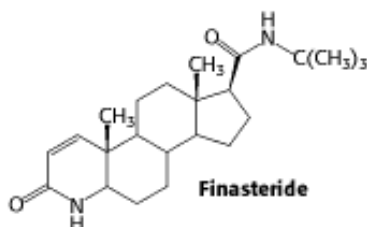


See answer

9. *Removal of odorants.* Many odorant molecules are highly hydrophobic and concentrate within the olfactory epithelium. They would give a persistent signal independent of their concentration in the environment if they were not rapidly modified. Propose a mechanism for converting hydrophobic odorants into water-soluble derivatives that can be rapidly eliminated.

See answer

10. *Development difficulties.* Propecia (finasteride) is a synthetic steroid that functions as a competitive and specific inhibitor of 5α -reductase, the enzyme responsible for the synthesis of dihydrotestosterone from testosterone.



It is now widely used to retard the development of male pattern hair loss. Pregnant women are advised to avoid handling this drug. Why is it vitally important that pregnant women avoid contact with Propecia?

See answer

11. *Life-style consequences.* Human beings and the plant *Arabidopsis* evolved from the same distant ancestor possessing a small number of cytochrome P450 genes. Human beings have approximately 50 such genes, whereas *Arabidopsis* has more than 250 of them. Propose a role for the large number of P450 isozymes in plants.

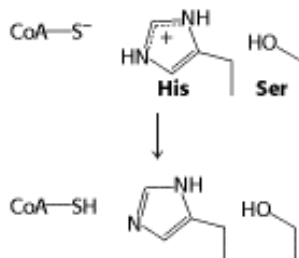
See answer

12. *Personalized medicine.* The cytochrome P450 system metabolizes many medically useful drugs. Although all human beings have the same number of P450 genes, individual polymorphisms exist that alter the specificity and efficiency of the proteins encoded by the genes. How could knowledge of individual polymorphisms be useful clinically?

See answer

Mechanism Problem

13. *An interfering phosphate.* In the course of the overall reaction catalyzed by HMG-CoA reductase, a histidine residue protonates a coenzyme A thiolate, CoAS^- , generated in a previous step.



The nearby serine residue can be phosphorylated by AMP-dependent kinase, which results in a loss of activity. Propose an explanation for why phosphorylation of the serine residue inhibits enzyme activity.

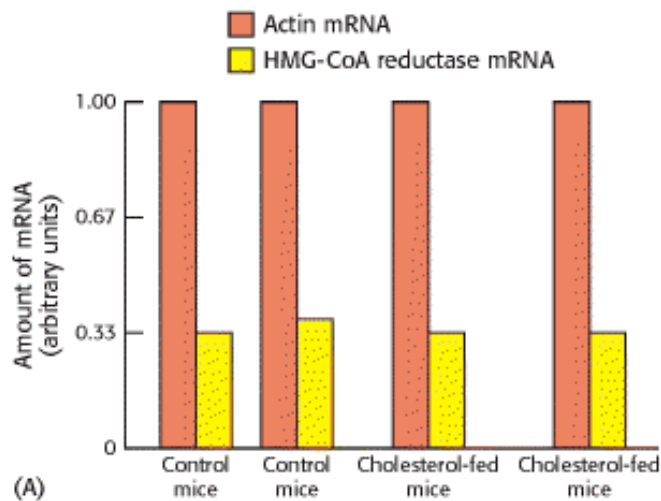
See answer

14. *Demethylation*. Methyl amines are often demethylated by cytochrome P450 enzymes. Propose a mechanism for the formation of methylamine from dimethylamine catalyzed by cytochrome P450. What is the other product?

See answer

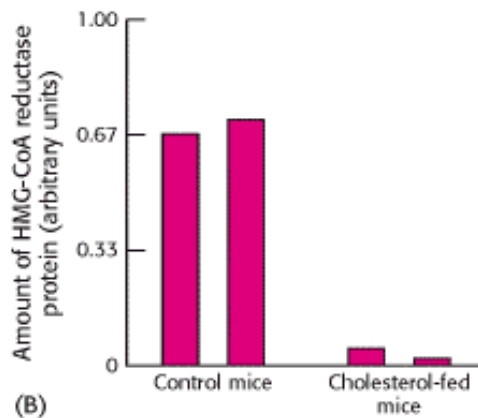
Data Interpretation and Chapter Integration Problems

15. *Cholesterol feeding*. Mice were divided into four groups, two of which were fed a normal diet and two of which were fed a cholesterol-rich diet. HMG-CoA reductase mRNA and protein from liver were then isolated and quantified. Graph A shows the results of the mRNA isolation.



- (a) What is the effect of cholesterol feeding on the amount of HMG-CoA reductase mRNA?
- (b) What is the purpose of also isolating the mRNA for the protein actin, which is not under the control of the sterol response element?

HMG-CoA reductase protein was isolated by precipitation with a monoclonal antibody to HMG-CoA reductase. The amount of HMG-CoA protein in each group is shown in graph B.



- (c) What is the effect of the cholesterol diet on the amount of HMG-CoA reductase protein?
- (d) Why is this result surprising in light of the results in graph A?

(e) Suggest possible explanations for the results shown in graph B.

See answer

Selected Readings

Where to start

D.E. Vance and H. Van den Bosch. 2000. Cholesterol in the year 2000 *Biochim. Biophys. Acta* 1529: 1-8. ([PubMed](#))

M.S. Brown and J.L. Goldstein. 1986. A receptor-mediated pathway for cholesterol homeostasis *Science* 232: 34-47. ([PubMed](#))

M.S. Brown and J.L. Goldstein. 1984. How LDL receptors influence cholesterol and atherosclerosis *Sci. Am.* 251: (5) 58-66.

L. Chan. 1992. Apolipoprotein B, the major protein component of triglyceride-rich and low density lipoproteins *J. Biol. Chem.* 267: 25621-25624. ([PubMed](#))

A. Endo. 1992. The discovery and development of HMG-CoA reductase inhibitors *J. Lipid Res.* 33: 1569-1582. ([PubMed](#))

S. Hakomori.. 1986. Glycosphingolipids *Sci. Am.* 254: (5) 44-53. ([PubMed](#))

Books

Vance, D. E., and Vance, J. E. (Eds.), 1996. *Biochemistry of Lipids, Lipoproteins and Membranes*. Elsevier.

Striver, C. R., Beaudet, A. L., Sly, W. S., Valle, D., Stanbury, J. B., Wyngaarden, J. B., and Fredrickson, D. S. (Eds.), 1995. *The Metabolic Basis of Inherited Diseases* (7th ed.). McGraw-Hill.

Phospholipids and sphingolipids

A. Huwiler, T. Kolterb, J. Pfeilschifter, and K. Sandhoff. 2000. Physiology and pathophysiology of sphingolipid metabolism and signaling *Biochim. Biophys. Acta* 1485: 63-99. ([PubMed](#))

A. Lykidis and S. Jackowski. 2000. Regulation of mammalian cell membrane biosynthesis *Prog. Nucleic Acid Res. Mol. Biol.* 65: 361-393. ([PubMed](#))

G.M. Carman and G.M. Zeimet. 1996. Regulation of phospholipid biosynthesis in the yeast *Saccharomyces cerevisiae* *J. Biol. Chem.* 271: 13293-13296. ([PubMed](#))

S.A. Henry and J.L. Patton-Vogt. 1998. Genetic regulation of phospholipid metabolism: Yeast as a model eukaryote *Prog. Nucleic Acid Res. Mol. Biol.* 61: 133-179. ([PubMed](#))

C. Kent. 1995. Eukaryotic phospholipid biosynthesis *Annu. Rev. Biochem.* 64: 315-343. ([PubMed](#))

S.M. Prescott, G.A. Zimmerman, D.M. Stafforini, and T.M. McIntyre. 2000. Platelet-activating factor and related lipid mediators *Annu. Rev. Biochem.* 69: 419-445. ([PubMed](#))

Biosynthesis of cholesterol and steroids

J.L. Goldstein and M.S. Brown. 1990. Regulation of the mevalonate pathway *Nature* 343: 425-430. ([PubMed](#))

R.G. Gardner, H. Shan, S.P.T. Matsuda, and R.Y. Hampton. 2001. An oxysterol-derived positive signal for 3-hydroxy-3-

methylglutaryl-CoA reductase degradation in yeast *J. Biol. Chem.* 276: 8681-8694. ([PubMed](#))

E.S. Istvan and J. Deisenhofer. 2001. Structural mechanism for statin inhibition of HMG-CoA reductase *Science* 292: 1160-1164. ([PubMed](#))

G.C. Ness and C.M. Chambers. 2000. Feedback and hormonal regulation of hepatic 3-hydroxy-3-methylglutaryl coenzyme A reductase: The concept of cholesterol buffering capacity *Proc. Soc. Exp. Biol. Med.* 224: 8-19. ([PubMed](#))

P. Libby, M. Aikawa, and U. Schonbeck. 2000. Cholesterol and atherosclerosis *Biochim. Biophys. Acta* 1529: 299-309. ([PubMed](#))

S. Yokoyama. 2000. Release of cellular cholesterol: Molecular mechanism for cholesterol homeostasis in cells and in the body *Biochim. Biophys. Acta* 1529: 231-244. ([PubMed](#))

S.R. Cronin, A. Khoury, D.K. Ferry, and R.Y. Hampton. 2000. Regulation of HMG-CoA reductase degradation requires the P-type ATPase Cod1p/Spf1p *J. Cell Biol.* 148: 915-924. ([PubMed](#))

P.A. Edwards, D. Tabor, H.R. Kast, and A. Venkateswaran. 2000. Regulation of gene expression by SREBP and SCAP *Biochim. Biophys. Acta* 1529: 103-113. ([PubMed](#))

E.S. Istvan, M. Palnitkar, S.K. Buchanan, and J. Deisenhofer. 2000. Crystal structure of the catalytic portion of human HMG-CoA reductase: Insights into regulation of activity and catalysis *EMBO J.* 19: 819-830. ([PubMed](#))

L. Tabernero, D.A. Bochar, V.W. Rodwell, and C.V. Stauffacher. 1999. Substrate-induced closure of the flap domain in the ternary complex structures provides insights into the mechanism of catalysis by 3-hydroxy-3-methylglutaryl-CoA reductase *Proc. Natl. Acad. Sci. USA.* 96: 7167-7171. ([PubMed](#)) ([Full Text in PMC](#))

D. Fass, S. Blacklow, P.S. Kim, and J.M. Berger. 1997. Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module *Nature* 388: 691-693. ([PubMed](#))

H. Jeon, W. Meng, J. Takagi, M.J. Eck, T.A. Springer, and S.C. Blacklow. 2001. Implications for familial hypercholesterolemia from the structure of the LDL receptor YWTD-EGF domain pair *Nat. Struct. Biol.* 8: 499-504. ([PubMed](#))

Lipoproteins and their receptors

C.G. Brouillette, G.M. Anantharamaiah, J.A. Engler, and D.W. Borhani. 2001. Structural models of human apolipoprotein A-I: A critical analysis and review *Biochem. Biophys. Acta* 1531: 4-46. ([PubMed](#))

T. Hevonoja, M.O. Pentikainen, M.T. Hyvonen, P.T. Kovanen, and M. Ala-Korpela. 2000. Structure of low density lipoprotein (LDL) particles: Basis for understanding molecular changes in modified LDL *Biochim. Biophys. Acta* 1488: 189-210. ([PubMed](#))

D.L. Silver, X.C. Jiang, T. Arai, C. Bruce, and A.R. Tall. 2000. Receptors and lipid transfer proteins in HDL metabolism *Ann. N. Y. Acad. Sci.* 902: 103-111. ([PubMed](#))

J. Nimpf and W.J. Schneider. 2000. From cholesterol transport to signal transduction: Low density lipoprotein receptor, very low density lipoprotein receptor, and apolipoprotein E receptor-2 *Biochim. Biophys. Acta* 1529: 287-298. ([PubMed](#))

D.W. Borhani, D.P. Rogers, J.A. Engler, and C.G. Brouillette. 1997. Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation *Proc. Natl. Acad. Sci. USA.* 94: 12291-12296. ([PubMed](#)) ([Full Text in PMC](#))

C. Wilson, M.R. Wardell, K.H. Weisgraber, R.W. Mahley, and D.A. Agard. 1991. Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E *Science* 252: 1817-1822. ([PubMed](#))

A.S. Plump, J.D. Smith, T. Hayek, K. Aalto-Setälä, A. Walsh, J.G. Verstuyft, E.M. Rubin, and J.L. Breslow. 1992. Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous

recombination in ES cells *Cell* 71: 343-353. ([PubMed](#))

T.C. Sudhof, J.L. Goldstein, M.S. Brown, and D.W. Russell. 1985. The LDL receptor gene: A mosaic of exons shared with different proteins *Science* 228: 815-822. ([PubMed](#))

Oxygen activation and P450 catalysis

M. Ingelman-Sundberg, M. Oscarson, and R.A. McLellan. 1999. Polymorphic human cytochrome P450 enzymes: An opportunity for individualized drug treatment *Trends Pharmacol. Sci.* 20: 342-349. ([PubMed](#))

D.R. Nelson. 1999. Cytochrome P450 and the individuality of species *Arch. Biochem. Biophys.* 369: 1-10. ([PubMed](#))

L.L. Wong. 1998. Cytochrome P450 monooxygenases *Curr. Opin. Chem. Biol.* 2: 263-268. ([PubMed](#))

M.S. Denison and J.P. Whitlock. 1995. Xenobiotic-inducible transcription of cytochrome P450 genes *J. Biol. Chem.* 270: 18175-18178. ([PubMed](#))

T.L. Poulos. 1995. Cytochrome P450 *Curr. Opin. Struct. Biol.* 5: 767-774. ([PubMed](#))

A.D. Vaz and M.J. Coon. 1994. On the mechanism of action of cytochrome P450: Evaluation of hydrogen abstraction in oxygen-dependent alcohol oxidation *Biochemistry* 33: 6442-6449. ([PubMed](#))

F.J. Gonzalez and D.W. Nebert. 1990. Evolution of the P450 gene superfamily: Animal-plant "warfare," molecular drive and human genetic differences in drug oxidation *Trends Genet.* 6: 182-186. ([PubMed](#))

27. DNA Replication, Recombination, and Repair

Perhaps the most exciting aspect of the structure of DNA deduced by Watson and Crick was, as expressed in their words, that the "specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." A double helix separated into two single strands can be replicated because each strand serves as a template on which its complementary strand can be assembled ([Figure 27.1](#)). Although this notion of how DNA is replicated is absolutely correct, the doublehelical structure of DNA poses a number of challenges to replication, as does the need for extremely faithful copying of the genetic information.

1. The two strands of the double helix have a tremendous affinity for one another, created by the cooperative effects of the many hydrogen bonds that hold adjacent base pairs together. Thus, a mechanism is required for separating the strands in a local region to provide access to the bases that act as templates. Specific proteins melt the double helix at specific sites to initiate DNA replication, and other enzymes, termed *helicases*, use the free energy of ATP hydrolysis to move this melted region along the double helix as replication progresses.

2. The DNA helix must be unwound to separate the two strands. The local unwinding in one region leads to stressful overwinding in surrounding regions ([Figure 27.2](#)). Enzymes termed *topoisomerases* introduce supercoils that release the strain caused by overwinding.

3. DNA replication must be highly accurate. As noted in [Chapter 5](#), the free energies associated with base pairing within the double helix suggest that approximately 1 in 10^4 bases incorporated will be incorrect. Yet, DNA replication has an error rate estimated to be 1 per 10^{10} nucleotides. As we shall see, additional mechanisms allow proofreading of the newly formed double helix.

4. DNA replication must be very rapid, given the sizes of the genomes and the rates of cell division. The *E. coli* genome contains 4.8 million base pairs and is copied in less than 40 minutes. Thus, 2000 bases are incorporated per second. We shall examine some of the properties of the macromolecular machines that replicate DNA with such high accuracy and

speed.

5. The enzymes that copy DNA polymerize nucleotides in the $5' \rightarrow 3'$ direction. The two polynucleotide strands of DNA run in opposite directions, yet both strands appear to grow in the same direction ([Figure 27.3](#)). Further analysis reveals that one strand is synthesized in a continuous fashion, whereas the opposite strand is synthesized in fragments in a discontinuous fashion. The synthesis of each fragment must be initiated in an independent manner, and then the fragments must be linked together. The DNA replication apparatus includes enzymes for these priming and ligation reactions.

6. The replication machinery alone cannot replicate the ends of linear DNA molecules, so a mechanism is required to prevent the loss of sequence information with each replication. Specialized structures called *telomeres* are added by another enzyme to maintain the information content at chromosome ends.

7. Most components of the DNA replication machinery serve to preserve the integrity of a DNA sequence to the maximum possible extent, yet a variety of biological processes require DNA formed by the exchange of material between two parent molecules. These processes range from the development of diverse antibody sequences in the immune system ([Chapter 33](#)) to the integration of viral genomes into host DNA. Specific enzymes, termed *recombinases*, facilitate these rearrangements.

8. After replication, ultraviolet light and a range of chemical species can damage DNA in a variety of ways. All organisms have enzymes for detecting and repairing harmful DNA modifications. Agents that introduce chemical lesions into DNA are key factors in the development of cancer, as are defects in the repair systems that correct these lesions.

We begin with a review of the structural properties of the DNA double helix.

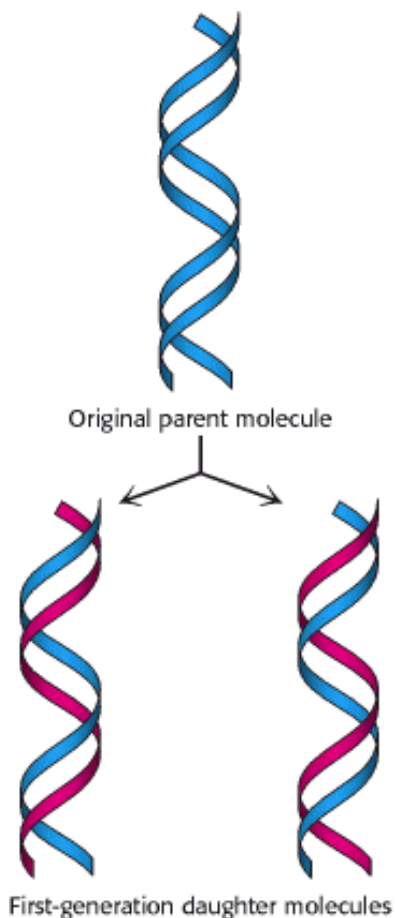


Figure 27.1. DNA Replication. The double-helical structure immediately suggests how DNA is replicated.



Figure 27.2. Consequences of Strand Separation. DNA must be locally unwound to expose single-stranded templates for replication. This unwinding puts a strain on the molecule by causing the overwinding of nearby regions.

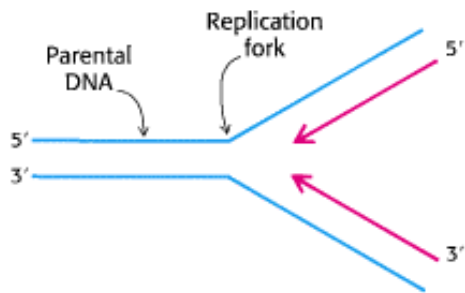
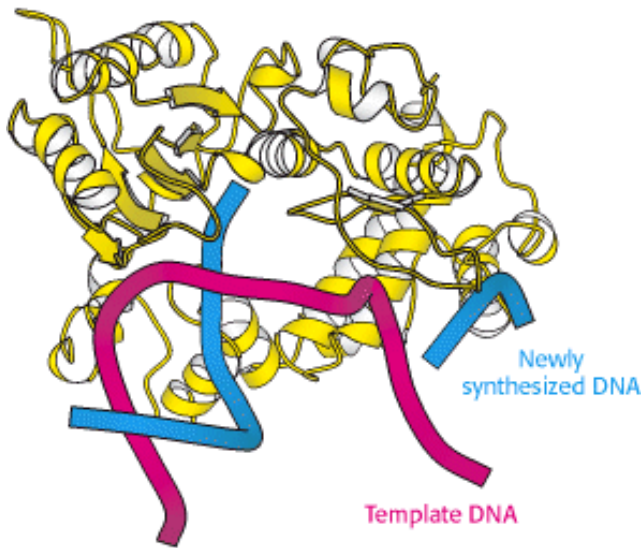


Figure 27.3. DNA Replication At Low Resolution. On cursory examination, both strands of a DNA template appear to replicate continuously in the same direction.





Faithful copying is essential to the storage of genetic information. With the precision of a diligent monk copying an illuminated manuscript, a DNA polymerase (below) copies DNA strands, preserving the precise sequence of bases with very few errors. [(Left) The Pierpont Morgan Library/ Art Resource.]

27.1. DNA Can Assume a Variety of Structural Forms

The double-helical structure of DNA deduced by Watson and Crick immediately suggested how genetic information is stored and replicated. As was discussed earlier ([Section 5.2.1](#)), the essential features of their model are:

1. Two polynucleotide chains running in opposite directions coil around a common axis to form a right-handed double helix.
2. Purine and pyrimidine bases are on the inside of the helix, whereas phosphate and deoxyribose units are on the outside.
3. Adenine (A) is paired with thymine (T), and guanine (G) with cytosine (C). An A-T base pair is held together by two hydrogen bonds, and that of a G-C base pair by three such bonds.

27.1.1. A-DNA Is a Double Helix with Different Characteristics from Those of the More Common B-DNA

Watson and Crick based their model (known as the *B-DNA helix*) on x-ray diffraction patterns of DNA fibers, which provided information about properties of the double helix that are averaged over its constituent residues. The results of x-ray diffraction studies of dehydrated DNA fibers revealed a different form called *A-DNA*, which appears when the relative humidity is reduced to less than about 75%. A-DNA, like B-DNA, is a right-handed double helix made up of antiparallel strands held together by Watson-Crick base-pairing. The A helix is wider and shorter than the B helix, and its base pairs are tilted rather than perpendicular to the helix axis ([Figure 27.4](#)).

Many of the structural differences between B-DNA and A-DNA arise from different puckerings of their ribose units ([Figure 27.5](#)). In A-DNA, C-3' lies out of the plane (a conformation referred to as C-3'-endo) formed by the other four atoms of the furanose ring; in B-DNA, C-2' lies out of the plane (a conformation called C-2'-endo). The C-3'-endo pucker in A-DNA leads to a 19-degree tilting of the base pairs away from the normal to the helix. The phosphates and other groups in the A helix bind fewer H₂O molecules than do those in B-DNA. Hence, dehydration favors the A form.

The A helix is not confined to dehydrated DNA. *Double-stranded regions of RNA and at least some RNA-DNA hybrids*

adopt a double-helical form very similar to that of A-DNA. The position of the 2'-hydroxyl group of ribose prevents RNA from forming a classic Watson-Crick B helix because of steric hindrance (Figure 27.6): the 2'-oxygen atom would come too close to three atoms of the adjoining phosphate group and one atom in the next base. In an A-type helix, in contrast, the 2'-oxygen projects outward, away from other atoms.

27.1.2. The Major and Minor Grooves Are Lined by Sequence-Specific Hydrogen-Bonding Groups

Double-helical nucleic acid molecules contain two grooves, called the *major groove* and the *minor groove*. These grooves arise because the glycosidic bonds of a base pair are not diametrically opposite each other (Figure 27.7). The minor groove contains the pyrimidine O-2 and the purine N-3 of the base pair, and the major groove is on the opposite side of the pair. The methyl group of thymine also lies in the major groove. In B-DNA, the major groove is wider (12 versus 6 Å) and deeper (8.5 versus 7.5 Å) than the minor groove (Figure 27.8).

Each groove is lined by potential hydrogen-bond donor and acceptor atoms that enable specific interactions with proteins (see Figure 27.7). In the minor groove, N-3 of adenine or guanine and O-2 of thymine or cytosine can serve as hydrogen acceptors, and the amino group attached to C-2 of guanine can be a hydrogen donor. In the major groove, N-7 of guanine or adenine is a potential acceptor, as are O-4 of thymine and O-6 of guanine. The amino groups attached to C-6 of adenine and C-4 of cytosine can serve as hydrogen donors. Note that the major groove displays more features that distinguish one base pair from another than does the minor groove. The larger size of the major groove in B-DNA makes it more accessible for interactions with proteins that recognize specific DNA sequences.

27.1.3. The Results of Studies of Single Crystals of DNA Revealed Local Variations in DNA Structure

X-ray analyses of single crystals of DNA oligomers had to await the development of techniques for synthesizing large amounts of DNA fragments with defined base sequences. X-ray analyses of single crystals of DNA at atomic resolution revealed that *DNA exhibits much more structural variability and diversity than formerly envisaged*.

The x-ray analysis of a crystallized DNA dodecamer by Richard Dickerson and his coworkers revealed that its overall structure is very much like a B-form Watson-Crick double helix. However, the dodecamer differs from the Watson-Crick model in not being uniform; there are rather large local deviations from the average structure. The Watson-Crick model has 10 residues per complete turn, and so a residue is related to the next along a chain by a rotation of 36 degrees. In Dickerson's dodecamer, the rotation angles range from 28 degrees (less tightly wound) to 42 degrees (more tightly wound). Furthermore, the two bases of many base pairs are not perfectly coplanar (Figure 27.9). Rather, they are arranged like the blades of a propeller. This deviation from the idealized structure, called *propeller twisting*, enhances the stacking of bases along a strand. These and other local variations of the double helix depend on base sequence. A protein searching for a specific target sequence in DNA may sense its presence through its effect on the precise shape of the double helix.

27.1.4. Z-DNA Is a Left-Handed Double Helix in Which Backbone Phosphates Zigzag

Alexander Rich and his associates discovered a third type of DNA helix when they solved the structure of dCGCGCG. They found that this hexanucleotide forms a duplex of antiparallel strands held together by Watson-Crick base-pairing, as expected. What was surprising, however, was that this double helix was *left-handed*, in contrast with the *right-handed* screw sense of the A and B helices. Furthermore, the phosphates in the backbone *zigzagged*; hence, they called this new form *Z-DNA* (Figure 27.10).

The Z-DNA form is adopted by short oligonucleotides that have *sequences of alternating pyrimidines and purines*. High salt concentrations are required to minimize electrostatic repulsion between the backbone phosphates, which are closer to each other than in A- and B-DNA. Under physiological conditions, *most DNA is in the B form*. Although the biological

role of Z-DNA is still under investigation, its existence graphically shows that DNA is a flexible, dynamic molecule. The properties of A-, B-, and Z-DNA are compared in [Table 27.1](#).

Primer—

The initial segment of a polymer that is to be extended on which elongation depends.

Template—

A sequence of DNA or RNA that directs the synthesis of a complementary sequence.

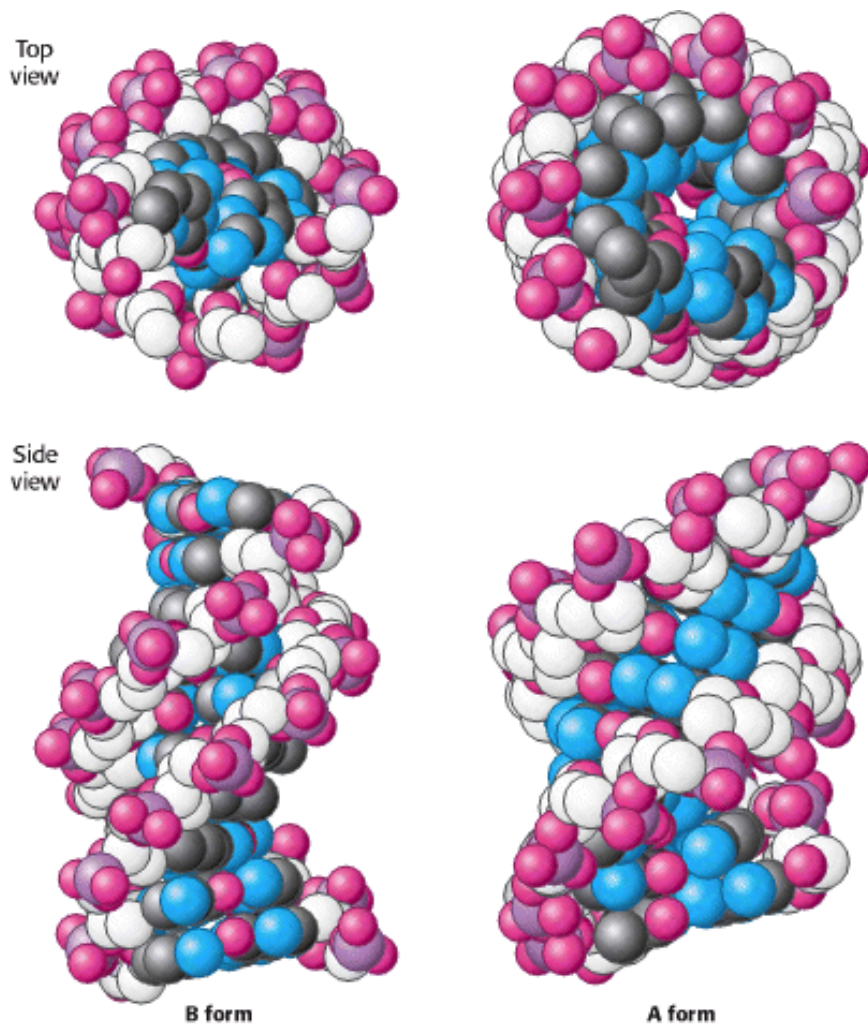


Figure 27.4. B-Form and A-Form DNA. Space-filling models of ten base pairs of B-form and A-form DNA depict their right-handed helical structures. The B-form helix is longer and narrower than the A-form helix. The carbon atoms of the backbone are shown in white.

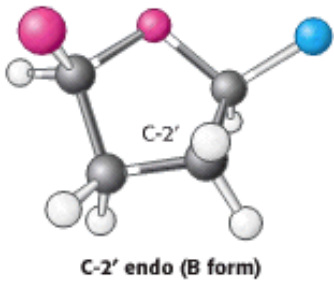
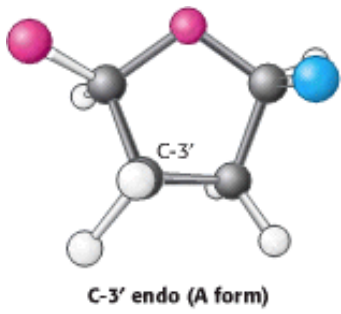


Figure 27.5. Sugar Puckers. In A-form DNA, the C-3' carbon atom lies above the approximate plane defined by the four other sugar nonhydrogen atoms (called C-3' endo). In B-form DNA, each ribose is in a C-2' endo conformation.

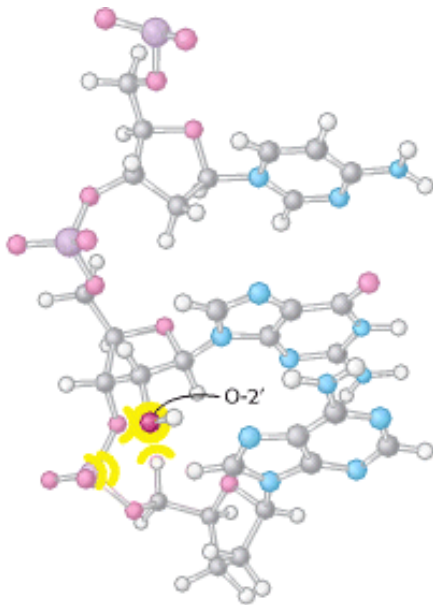


Figure 27.6. Steric Clash. The introduction of a 2'-hydroxyl group into a B-form structure leads to several steric clashes with nearby atoms.

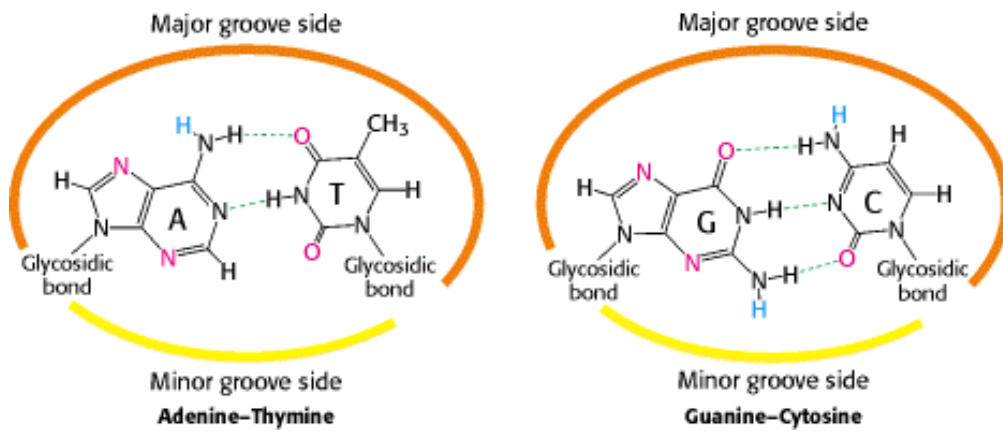


Figure 27.7. Major- and Minor-Groove Sides. Because the two glycosidic bonds are not diametrically opposite each other, each base pair has a larger side that defines the major groove and a smaller side that defines the minor groove. The grooves are lined by potential hydrogen-bond donors (blue) and acceptors (red).

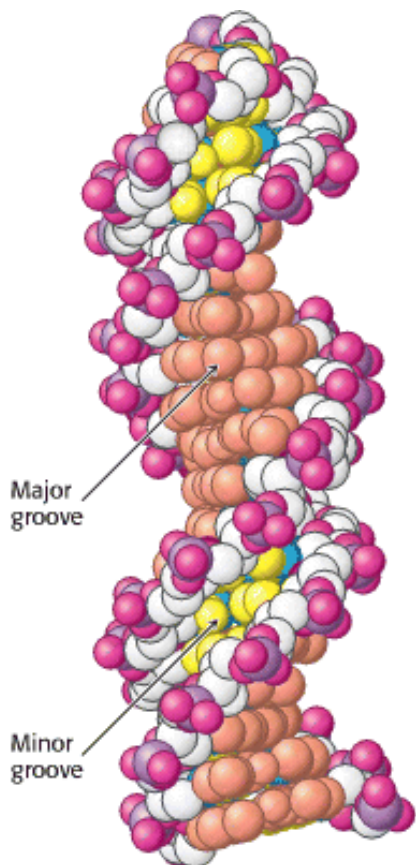


Figure 27.8. Major and Minor Grooves in B-Form DNA. The major groove is depicted in orange, and the minor groove is depicted in yellow. The carbon atoms of the backbone are shown in white.

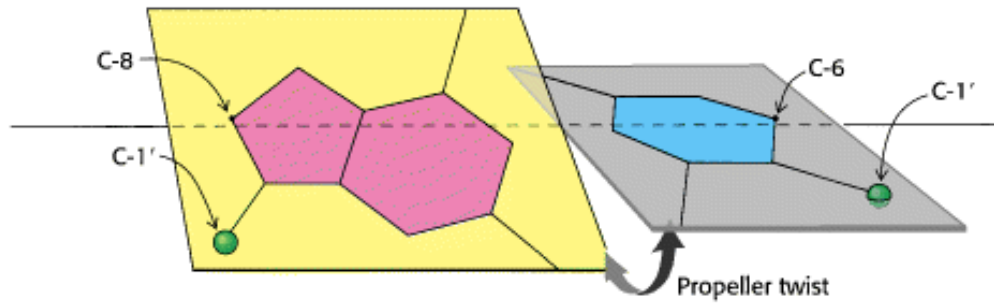
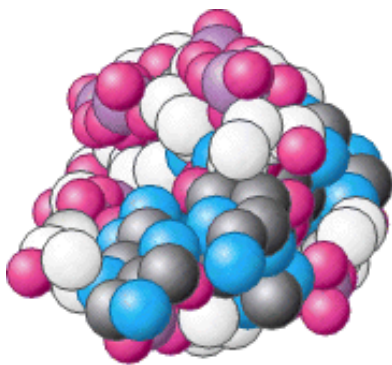
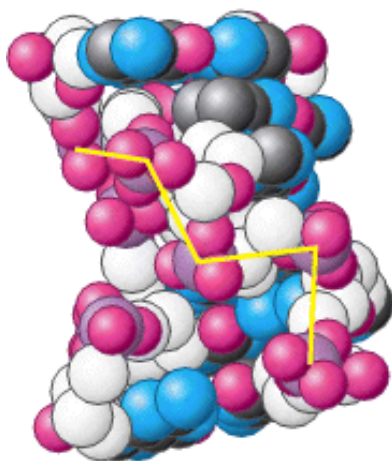


Figure 27.9. Propeller Twist. The bases of a DNA base pair are often not precisely coplanar. They are twisted with respect to each other, like the blades of a propeller.



Top view



Side view

Figure 27.10. Z-DNA. DNA oligomers such as dCGCGCG adopt an alternative conformation under some conditions.



This conformation is called Z-DNA because the phosphate groups zigzag along the backbone.

Table 27.1. Comparison of A-, B-, and Z-DNA

	Helix type		
	A	B	Z
Shape	Broadest	Intermediate	Narrowest
Rise per base pair	2.3 Å	3.4 Å	3.8 Å
Helix diameter	25.5 Å	23.7 Å	18.4 Å
Screw sense	Right-handed	Right-handed	Left-handed
Glycosidic bond	<i>anti</i>	<i>anti</i>	alternating <i>anti</i> and <i>syn</i>
Base pairs per turn of helix	11	10.4	12
Pitch per turn of helix	25.3 Å	35.4 Å	45.6 Å
Tilt of base pairs from normal to helix axis	19°	1°	9°
Major groove	Narrow and very deep	Wide and quite deep	Flat
Minor groove	Very broad and shallow	Narrow and quite deep	Very narrow and deep


27.2. DNA Polymerases Require a Template and a Primer

DNA polymerases catalyze the formation of polynucleotide chains through the addition of successive nucleotides derived from deoxynucleoside triphosphates. The polymerase reaction takes place only in the presence of an appropriate DNA template. Each incoming nucleoside triphosphate first forms an appropriate base pair with a base in this template. Only then does the DNA polymerase link the incoming base with the predecessor in the chain. Thus, *DNA polymerases are template-directed enzymes.*

DNA polymerases add nucleotides to the 3' end of a polynucleotide chain. The polymerase catalyzes the nucleophilic attack of the 3'-hydroxyl group terminus of the polynucleotide chain on the α -phosphate group of the nucleoside triphosphate to be added (see [Figure 5.22](#)). To initiate this reaction, DNA polymerases require a *primer* with a free 3'-hydroxyl group already base-paired to the template. They cannot start from scratch by adding nucleotides to a free single-stranded DNA template. RNA polymerase, in contrast, can initiate RNA synthesis without a primer ([Section 28.1.4](#)).

27.2.1. All DNA Polymerases Have Structural Features in Common

The three-dimensional structures of a number of DNA polymerase enzymes are known. The first such structure to be determined was that of the so-called Klenow fragment of DNA polymerase I from *E. coli* ([Figure 27.11](#)). This fragment comprises two main parts of the full enzyme, including the polymerase unit. This unit approximates the shape of a right hand with domains that are referred to as the fingers, the thumb, and the palm. In addition to the polymerase, the Klenow fragment includes a domain with 3' \rightarrow 5' *exonuclease* activity that participates in proofreading and correcting the polynucleotide product ([Section 27.2.4](#)).

 DNA polymerases are remarkably similar in overall shape, although they differ substantially in detail. At least five structural classes have been identified; some of them are clearly homologous, whereas others are probably the

products of convergent evolution. In all cases, the finger and thumb domains wrap around DNA and hold it across the enzyme's active site, which comprises residues primarily from the palm domain. Furthermore, all the polymerases catalyze the same polymerase reaction, which is dependent on two metal ions.

27.2.2. Two Bound Metal Ions Participate in the Polymerase Reaction

Like all enzymes with nucleoside triphosphate substrates, DNA polymerases require metal ions for activity. Examination of the structures of DNA polymerases with bound substrates and substrate analogs reveals the presence of two metal ions in the active site. One metal ion binds both the deoxynucleoside triphosphate (dNTP) and the 3'-hydroxyl group of the primer, whereas the other interacts only with the 3'-hydroxyl group (Figure 27.12). The two metal ions are bridged by the carboxylate groups of two aspartate residues in the palm domain of the polymerase. These side chains hold the metal ions in the proper position and orientation. The metal ion bound to the primer activates the 3'-hydroxyl group of the primer, facilitating its attack on the α -phosphate group of the dNTP substrate in the active site. The two metal ions together help stabilize the negative charge that accumulates on the pentacoordinate transition state. The metal ion initially bound to dNTP stabilizes the negative charge on the pyrophosphate product.

27.2.3. The Specificity of Replication Is Dictated by Hydrogen Bonding and the Complementarity of Shape Between Bases

DNA must be replicated with high fidelity. Each base added to the growing chain should with high probability be the Watson-Crick complement of the base in the corresponding position in the template strand. The binding of the NTP containing the proper base is favored by the formation of a base pair, which is stabilized by specific hydrogen bonds. The binding of a noncomplementary base is unlikely, because the interactions are unfavorable. The hydrogen bonds linking two complementary bases make a significant contribution to the fidelity of DNA replication. However, DNA polymerases replicate DNA more faithfully than these interactions alone can account for.

The examination of the crystal structures of various DNA polymerases indicated several additional mechanisms by which replication fidelity is improved. First, residues of the enzyme form hydrogen bonds with the minor-groove side of the base pair in the active site (Figure 27.13). In the minor groove, hydrogen-bond acceptors are present in the same positions for all Watson-Crick base pairs. These interactions act as a "ruler" that measures whether a properly spaced base pair has formed in the active site. Second, DNA polymerases close down around the incoming NTP (Figure 27.14). The binding of a nucleoside triphosphate into the active site of a DNA polymerase triggers a conformational change: the finger domain rotates to form a tight pocket into which only a properly shaped base pair will readily fit. The mutation of a conserved tyrosine residue at the top of the pocket results in a polymerase that is approximately 40 times as error prone as the parent polymerase.

27.2.4. Many Polymerases Proofread the Newly Added Bases and Excise Errors

Many polymerases further enhance the fidelity of replication by the use of proofreading mechanisms. As already noted, the Klenow fragment of *E. coli* DNA polymerase I includes an exonuclease domain that does not participate in the polymerization reaction itself. Instead, this domain removes mismatched nucleotides from the 3' end of DNA by hydrolysis. The exonuclease active site is 35 Å from the polymerase active site, yet it can be reached by the newly synthesized polynucleotide chain under appropriate conditions. The proofreading mechanism relies on the increased probability that the end of a growing strand with an incorrectly incorporated nucleotide will leave the polymerase site and transiently move to the exonuclease site (Figure 27.15).

How does the enzyme sense whether a newly added base is correct? First, an incorrect base will not pair correctly with the template strand. Its greater structural fluctuation, permitted by the weaker hydrogen bonding, will frequently bring the newly synthesized strand to the exonuclease site. Second, after the addition of a new nucleotide, the DNA translocates by one base pair into the enzyme. The newly formed base pair must be of the proper dimensions to fit into a tight binding site and participate in hydrogen-bonding interactions in the minor groove similar to those in the


polymerization site itself (see [Figure 27.13](#)). Indeed, the duplex DNA within the enzyme adopts an A-form structure, allowing clear access to the minor groove. If an incorrect base is incorporated, the enzyme stalls, and the pause provides additional time for the strand to migrate to the exonuclease site. There is a cost to this editing function, however: DNA polymerase I removes approximately 1 correct nucleotide in 20 by hydrolysis. Although the removal of correct nucleotides is slightly wasteful energetically, proofreading increases the accuracy of replication by a factor of approximately 1000.

27.2.5. The Separation of DNA Strands Requires Specific Helicases and ATP Hydrolysis

For a double-stranded DNA molecule to replicate, the two strands of the double helix must be separated from each other, at least locally. This separation allows each strand to act as a template on which a new polynucleotide chain can be assembled. For long double-stranded DNA molecules, the rate of spontaneous strand separation is negligibly low under physiological conditions. Specific enzymes, termed *helicases*, utilize the energy of ATP hydrolysis to power strand separation.

The detailed mechanisms of helicases are still under active investigation. However, the determination of the three-dimensional structures of several helicases has been a source of insight. For example, a bacterial helicase called *PcrA* comprises four domains, hereafter referred to as domains A1, A2, B1, and B2 ([Figure 27.16](#)). Domain A1 contains a P-loop NTPase fold, as was expected from amino acid sequence analysis. This domain participates in ATP binding and hydrolysis. Domain B1 is homologous to domain A1 but lacks a P-loop. Domains A2 and B2 have unique structures.

From an analysis of a set of helicase crystal structures bound to nucleotide analogs and appropriate double- and single-stranded DNA molecules, a mechanism for the action of these enzymes was proposed ([Figure 27.17](#)). Domains A1 and B1 are capable of binding single-stranded DNA. In the absence of bound ATP, both domains are bound to DNA. The binding of ATP triggers conformational changes in the P-loop and adjacent regions that lead to the closure of the cleft between these two domains. To achieve this movement, domain A1 releases the DNA and slides along the DNA strand, moving closer to domain B1. The enzyme then catalyzes the hydrolysis of ATP to form ADP and orthophosphate. On product release, the cleft between domains A and B springs open. In this state, however, domain A1 has a tighter grip on the DNA than does domain B1, so the DNA is pulled across domain B1 toward domain A1. The result is the translocation of the enzyme along the DNA strand in a manner similar to the way in which an inchworm moves. In regard to *PcrA*, the enzyme translocates in the $3' \rightarrow 5'$ direction. When the helicase encounters a region of double-stranded DNA, it continues to move along one strand and displaces the opposite DNA strand as it progresses. Interactions with specific pockets on the helicase help destabilize the DNA duplex, aided by ATP-induced conformational changes.

 Helicases constitute a large and diverse class of enzymes. Some of these enzymes move in a $5' \rightarrow 3'$ direction, whereas others unwind RNA rather than DNA and participate in processes such as RNA splicing and the initiation of mRNA translation. A comparison of the amino acid sequences of hundreds of these enzymes reveals seven regions of striking conservation ([Figure 27.18](#)). Mapping these regions onto the *PcrA* structure shows that they line the ATP-binding site and the cleft between the two domains, consistent with the notion that other helicases undergo conformational changes analogous to those found in *PcrA*. However, whereas *PcrA* appears to function as a monomer, other members of the helicase class function as oligomers. The hexameric structures of one important group are similar to that of the F_1 component of ATP synthase ([Section 18.4.1](#)), suggesting potential mechanistic similarities.

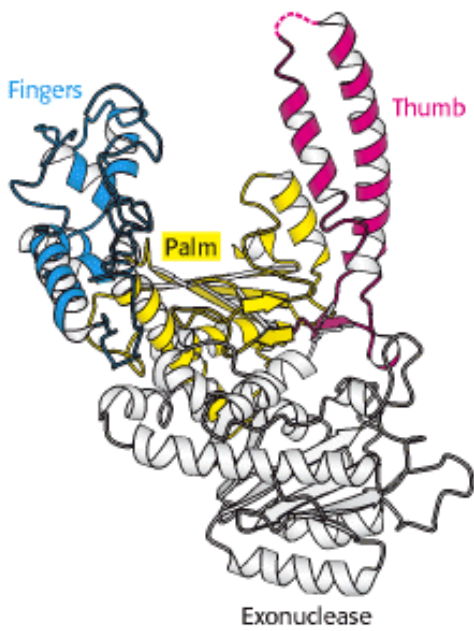


Figure 27.11. DNA Polymerase Structure. The first DNA polymerase structure determined was that of a fragment of *E. coli* DNA polymerase I called the Klenow fragment. Like other DNA polymerases, the polymerase unit resembles a right hand with fingers (blue), palm (yellow), and thumb (red). The Klenow fragment also includes an exonuclease domain.

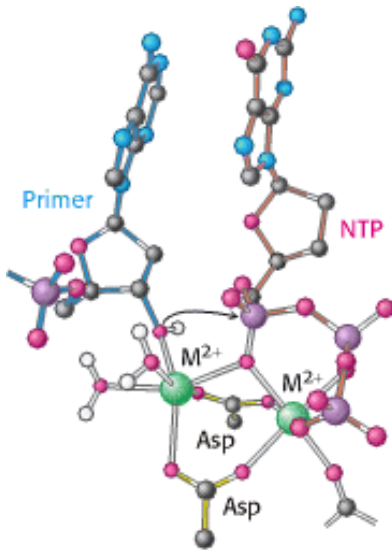


Figure 27.12. DNA Polymerase Mechanism. Two metal ions (typically, Mg^{2+}) participate in the DNA polymerase reaction. One metal ion coordinates the 3'-hydroxyl group of the primer, whereas the phosphate group of the nucleoside triphosphate bridges between the two metal ions. The hydroxyl group of the primer attacks the phosphate group to form a new O-P bond.

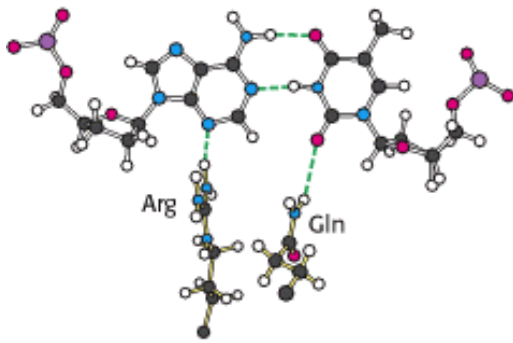


Figure 27.13. Minor-Groove Interactions. DNA polymerases donate two hydrogen bonds to base pairs in the minor groove. Hydrogen-bond acceptors are present in these two positions for all Watson-Crick base pairs including the A-T base pair shown.

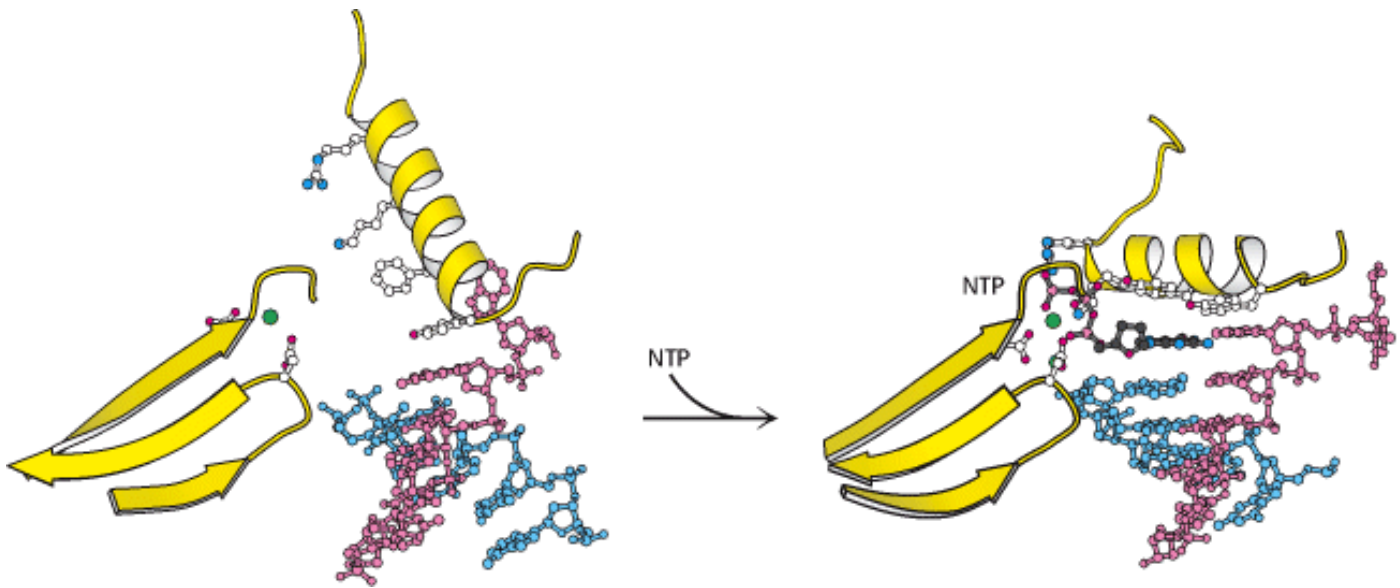


Figure 27.14. Shape Selectivity. The binding of a nucleoside triphosphate (NTP) to DNA polymerase induces a conformational change, generating a tight pocket for the base pair consisting of the NTP and its partner on the template strand. Such a conformational change is possible only when the NTP corresponds to the Watson-Crick partner of the template base.

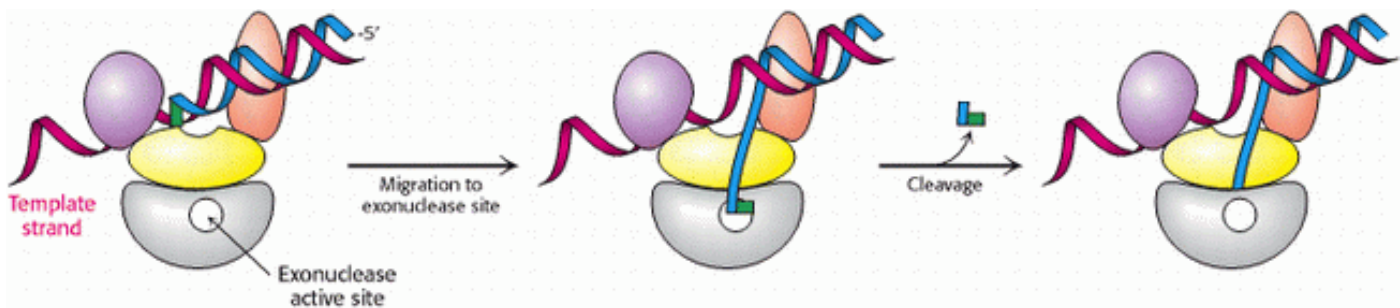


Figure 27.15. Proofreading. The growing polynucleotide chain occasionally leaves the polymerase site of DNA polymerase I and migrates to the exonuclease site. There, the last nucleotide added is removed by hydrolysis. Because mismatched bases are more likely to leave the polymerase site, this process serves to proofread the sequence of the DNA being synthesized.

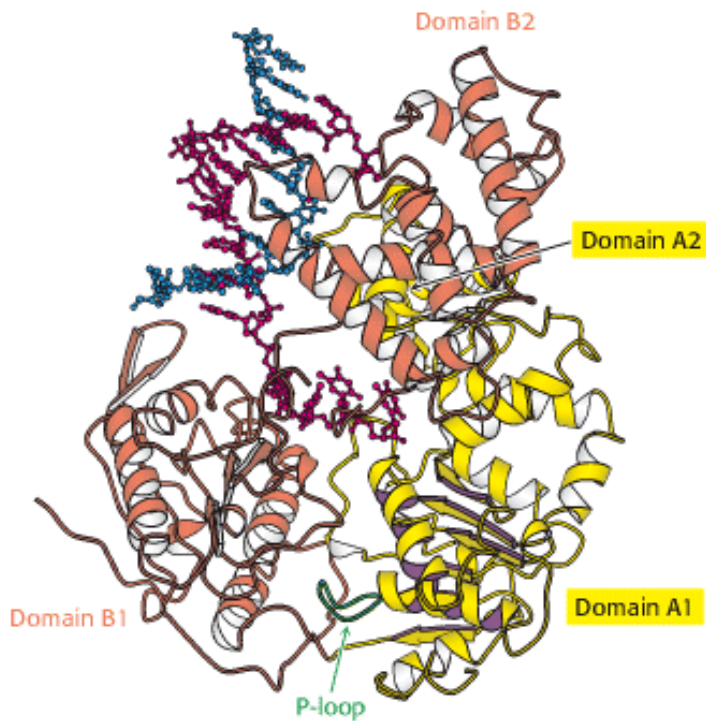


Figure 27.16. Helicase Structure. The bacterial helicase PcrA comprises four domains: A1, A2, B1, and B2. The A1 domain includes a P-loop NTPase fold, whereas the B1 domain has a similar overall structure but lacks a P-loop and does not bind nucleotides. Single-stranded DNA binds to the A1 and B1 domains near the interfaces with domains A2 and B2.

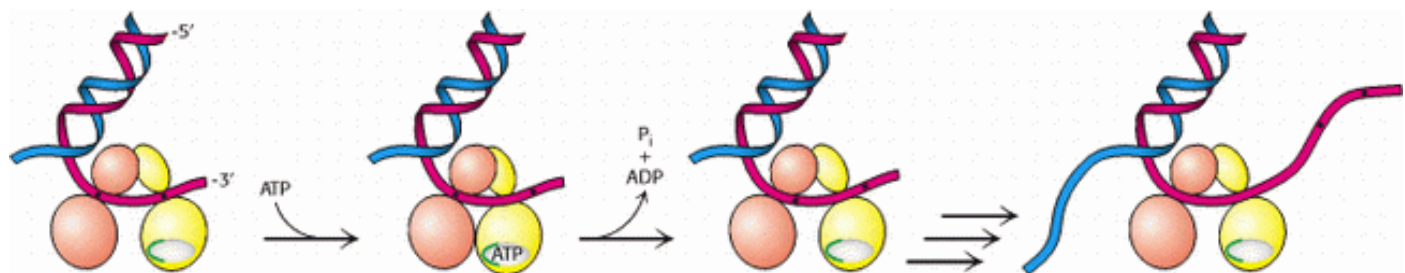


Figure 27.17. Helicase Mechanism. Initially, both domains A1 and B1 of PcrA bind single-stranded DNA. On binding of ATP, the cleft between these domains closes and domain A1 slides along the DNA. On ATP hydrolysis, the cleft opens up, pulling the DNA from domain B1 toward domain A1. As this process is repeated, double-stranded DNA is unwound.

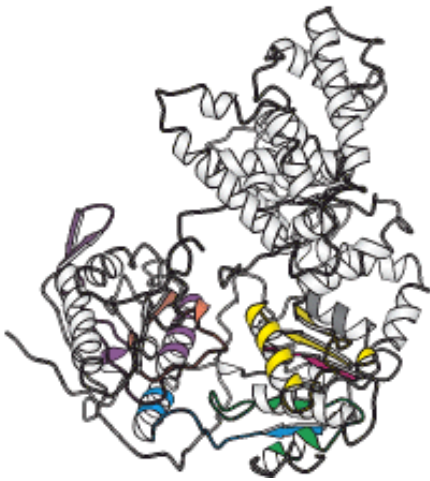


Figure 27.18. Conserved Residues among Helicases. A comparison of the amino acid sequences of hundreds of helicases revealed seven regions of strong sequence conservation (shown in color). When mapped onto the structure of PcrA, these conserved regions lie along the interface between the A1 and B1 domains and along the ATP binding surface.

27.3. Double-Stranded DNA Can Wrap Around Itself to Form Supercoiled Structures

The separation of the two strands of DNA in replication requires the local unwinding of the double helix. This local unwinding must lead either to the overwinding of surrounding regions of DNA or to supercoiling. To prevent the strain induced by overwinding, a specialized set of enzymes is present to introduce supercoils that favor strand separation.

27.3.1. The Linking Number of DNA, a Topological Property, Determines the Degree of Supercoiling

In 1963, Jerome Vinograd found that circular DNA from polyoma virus separated into two distinct species when it was centrifuged. In pursuing this puzzle, he discovered an important property of circular DNA not possessed by linear DNA with free ends. Consider a linear 260-bp DNA duplex in the B-DNA form (Figure 27.19). Because the number of residues per turn in an unstressed DNA molecule is 10.4, this linear DNA molecule has 25 (260/10.4) turns. The ends of this helix can be joined to produce a *relaxed* circular DNA (Figure 27.19B). A different circular DNA can be formed by unwinding the linear duplex by two turns before joining its ends (Figure 27.19C). What is the structural consequence of unwinding before ligation? Two limiting conformations are possible: the DNA can either fold into a structure containing 23 turns of B helix and an unwound loop (Figure 27.19D) or adopt a *supercoiled* structure with 25 turns of B helix and 2 turns of *right-handed* (termed *negative*) superhelix (Figure 27.19E).

Supercoiling markedly alters the overall form of DNA. A *supercoiled DNA molecule is more compact than a relaxed DNA molecule of the same length*. Hence, supercoiled DNA moves faster than relaxed DNA when analyzed by centrifugation or electrophoresis. The rapidly sedimenting DNA in Vinograd's experiment was supercoiled, whereas the slowly sedimenting DNA was relaxed because one of its strands was nicked. Unwinding will cause supercoiling in both circular DNA molecules and in DNA molecules that are constrained in closed configurations by other means.

27.3.2. Helical Twist and Superhelical Writhe Are Correlated with Each Other Through the Linking Number

Our understanding of the conformation of DNA is enriched by concepts drawn from topology, a branch of mathematics

dealing with structural properties that are unchanged by deformations such as stretching and bending. A key topological property of a circular DNA molecule is its *linking number* (Lk), which is equal to the number of times that a strand of DNA winds in the right-handed direction around the helix axis when the axis is constrained to lie in a plane. For the relaxed DNA shown in [Figure 27.19B](#), $Lk = 25$. For the partly unwound molecule shown in part D and the supercoiled one shown in part E, $Lk = 23$ because the linear duplex was unwound two complete turns *before* closure. Molecules differing only in linking number are *topological isomers* (*topoisomers*) of one another. *Topoisomers of DNA can be interconverted only by cutting one or both DNA strands and then rejoining them.*

The unwound DNA and supercoiled DNA shown in [Figure 27.19D](#) and [E](#) are topologically identical but geometrically different. They have the same value of Lk but differ in Tw (twist) and Wr (writhe). Although the rigorous definitions of twist and writhe are complex, twist is a measure of the helical winding of the DNA strands around each other, whereas writhe is a measure of the coiling of the axis of the double helix, which is called *super-coiling*. A right-handed coil is assigned a negative number (negative supercoiling) and a left-handed coil is assigned a positive number (positive supercoiling). Is there a relation between Tw and Wr ? Indeed, there is. Topology tells us that the sum of Tw and Wr is equal to Lk .

$$Lk = Tw + Wr$$

In [Figure 27.19](#), the partly unwound circular DNA has $Tw \sim 23$ and $Wr \sim 0$, whereas the supercoiled DNA has $Tw \sim 25$ and $Wr \sim -2$. These forms can be interconverted without cleaving the DNA chain because they have the same value of Lk ; namely, 23. The partitioning of Lk (which must be an integer) between Tw and Wr (which need not be integers) is determined by energetics. The free energy is minimized when about 70% of the change in Lk is expressed in Wr and 30% is expressed in Tw . Hence, the most stable form would be one with $Tw = 24.4$ and $Wr = -1.4$. Thus, *a lowering of Lk causes both right-handed (negative) supercoiling of the DNA axis and unwinding of the duplex.* Topoisomers differing by just 1 in Lk , and consequently by 0.7 in Wr , can be readily separated by agarose gel electrophoresis because their hydrodynamic volumes are quite different— *supercoiling condenses DNA* ([Figure 27.20](#)). Most naturally occurring DNA molecules are negatively supercoiled. What is the basis for this prevalence? As already stated, negative supercoiling arises from the unwinding or underwinding of the DNA. In essence, negative supercoiling prepares DNA for processes requiring separation of the DNA strands, such as replication or transcription. Positive supercoiling condenses DNA as effectively, but it makes strand separation more difficult.

27.3.3. Type I Topoisomerases Relax Supercoiled Structures

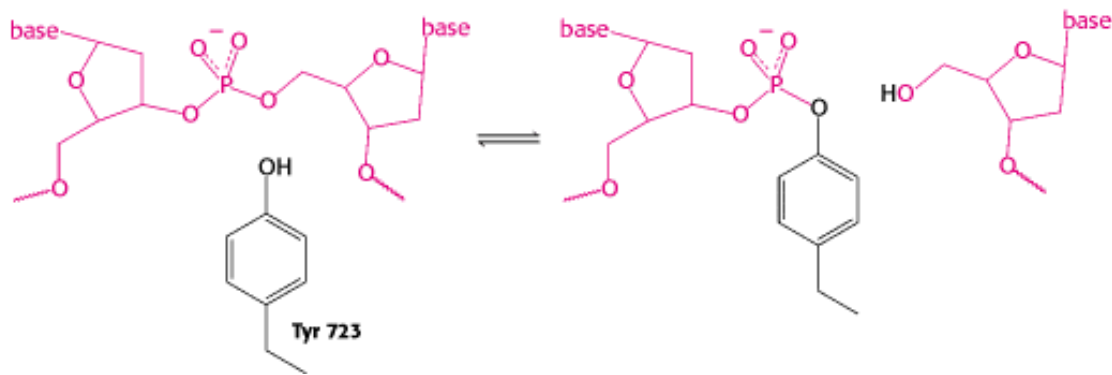
The interconversion of topoisomers of DNA is catalyzed by enzymes called *topoisomerases* which were discovered by James Wang and Martin Gellert. These enzymes alter the linking number of DNA by catalyzing a three-step process: (1) the *cleavage* of one or both strands of DNA, (2) the *passage* of a segment of DNA through this break, and (3) the *resealing* of the DNA break. Type I topoisomerases cleave just one strand of DNA, whereas type II enzymes cleave both strands. Both type I and type II topoisomerases play important roles in DNA replication and in transcription and recombination.

Type I topoisomerases catalyze the relaxation of supercoiled DNA, a thermodynamically favorable process. *Type II topoisomerases* utilize free energy from ATP hydrolysis to add negative supercoils to DNA. The two types of enzymes have several common features, including the use of key tyrosine residues to form covalent links to the polynucleotide backbone that is transiently broken.

The three-dimensional structures of several type I topoisomerases have been determined ([Figure 27.21](#)). These structures reveal many features of the reaction mechanism. Human type I topoisomerase comprises four domains, which are arranged around a central cavity having a diameter of 20 Å, just the correct size to accommodate a double-stranded DNA molecule. This cavity also includes a tyrosine residue (Tyr 723), which acts as a nucleophile to cleave the DNA backbone in the course of catalysis.

From analyses of these structures and the results of other studies, the relaxation of negatively supercoiled DNA

molecules are known to proceed in the following manner (Figure 27.22). First, the DNA molecule binds inside the cavity of the topoisomerase. The hydroxyl group of tyrosine 723 attacks a phosphate group on one strand of the DNA backbone to form a phosphodiester linkage between the enzyme and the DNA, cleaving the DNA and releasing a free 5'-hydroxyl group.



With the backbone of one strand cleaved, the DNA can now rotate around the remaining strand, driven by the release of the energy stored because of the supercoiling. The rotation of the DNA unwinds supercoils. The enzyme controls the rotation so that the unwinding is not rapid. The free hydroxyl group of the DNA attacks the phosphotyrosine residue to reseal the backbone and release tyrosine. The DNA is then free to dissociate from the enzyme. Thus, reversible cleavage of one strand of the DNA allows controlled rotation to partly relax supercoiled DNA.


27.3.4. Type II Topoisomerases Can Introduce Negative Supercoils Through Coupling to ATP Hydrolysis

Supercoiling requires an input of energy because a supercoiled molecule, in contrast with its relaxed counterpart, is torsionally stressed. The introduction of an additional supercoil into a 3000-bp plasmid typically requires about 7 kcal mol⁻¹.

Supercoiling is catalyzed by type II topoisomerases. These elegant molecular machines couple the binding and hydrolysis of ATP to the directed passage of one DNA double helix through another that has been temporarily cleaved. These enzymes have several mechanistic features in common with the type I topoisomerases.

The topoisomerase II from yeast is a heart-shaped dimer with a large central cavity (Figure 27.23). This cavity has gates at both the top and the bottom that are crucial to topoisomerase action. The reaction begins with the binding of one double helix (hereafter referred to as the G, for gate, segment) to the enzyme (Figure 27.24). Each strand is positioned next to a tyrosine residue, one from each monomer, capable of forming a covalent linkage with the DNA backbone. This complex then loosely binds a second DNA double helix (hereafter referred to as the T, for transported, segment). Each monomer of the enzyme has a domain that binds ATP; this ATP binding leads to a conformational change that strongly favors the coming together of the two domains. As these domains come closer together, they trap the bound T segment. This conformational change also forces the separation and cleavage of the two strands of the G segment. Each strand is joined to the enzyme by a tyrosine-phosphodiester linkage. Unlike the type I enzymes, the type II topoisomerases hold the DNA tightly so that it cannot rotate. The T segment then passes through the cleaved G segment and into the large central cavity. The ligation of the G segment leads to release of the T segment through the gate at the bottom of the enzyme. The hydrolysis of ATP and the release of ADP and orthophosphate allow the ATP-binding domains to separate, preparing the enzyme to bind another T segment. The overall process leads to a decrease in the linking number by two.

The degree of supercoiling of DNA is thus determined by the opposing actions of two enzymes. Negative supercoils are introduced by topoisomerase II and are relaxed by topoisomerase I. The amounts of these enzymes and their activities are regulated to maintain an appropriate degree of negative supercoiling.

 The bacterial topoisomerase II (often called DNA gyrase) is the target of several antibiotics that inhibit the prokaryotic enzyme much more than the eukaryotic one. *Novobiocin* blocks the binding of ATP to gyrase. *Nalidixic acid* and *ciprofloxacin*, in contrast, interfere with the breakage and rejoining of DNA chains. These two gyrase inhibitors are widely used to treat urinary tract and other infections. *Camptothecin*, an antitumor agent, inhibits human topoisomerase I by stabilizing the form of the enzyme covalently linked to DNA.

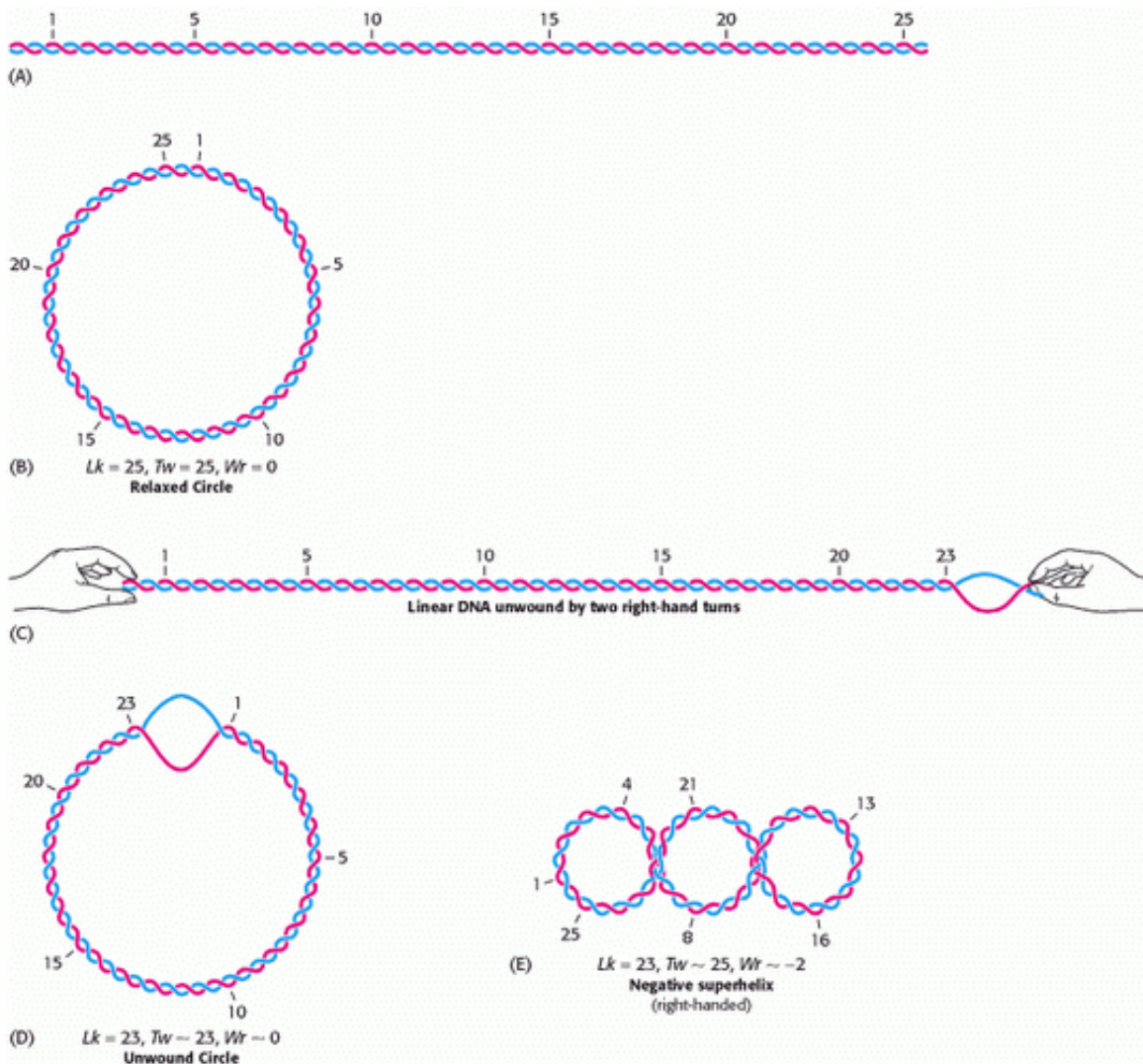
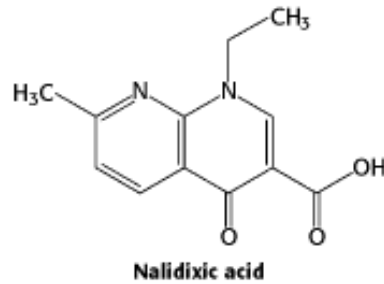


Figure 27.19. Linking Number. The relations between the linking number (Lk), twisting number (Tw), and writhing number (Wr) of a circular DNA molecule revealed schematically. [After W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, 1984), p. 452.]

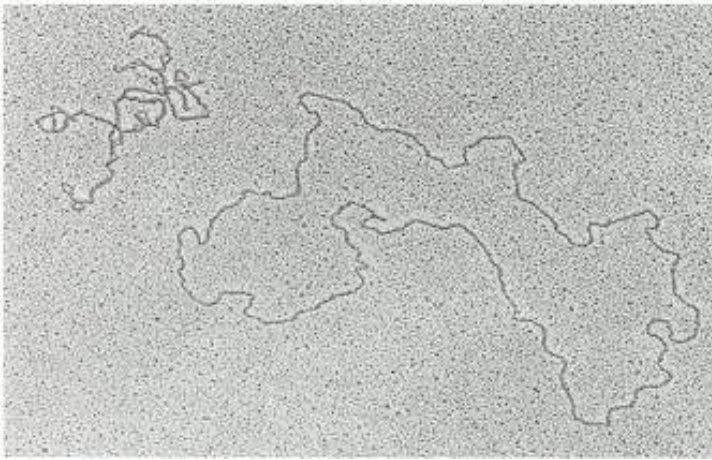


Figure 27.20. Topoisomers. An electron micrograph showing negatively supercoiled and relaxed DNA. [Courtesy of Dr. Jack Griffith.]

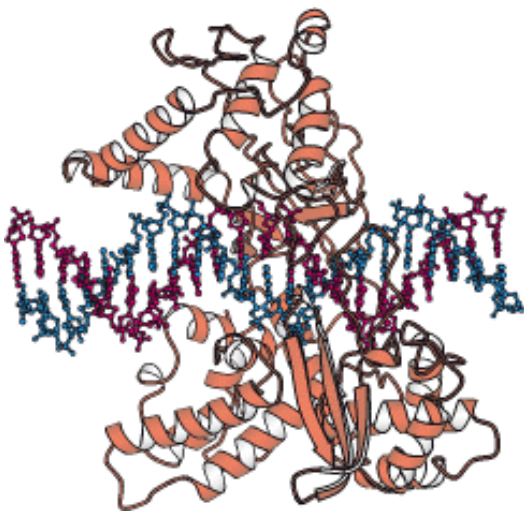


Figure 27.21. Structure of a Topoisomerase. The structure of a complex between a fragment of human topoisomerase I and DNA.

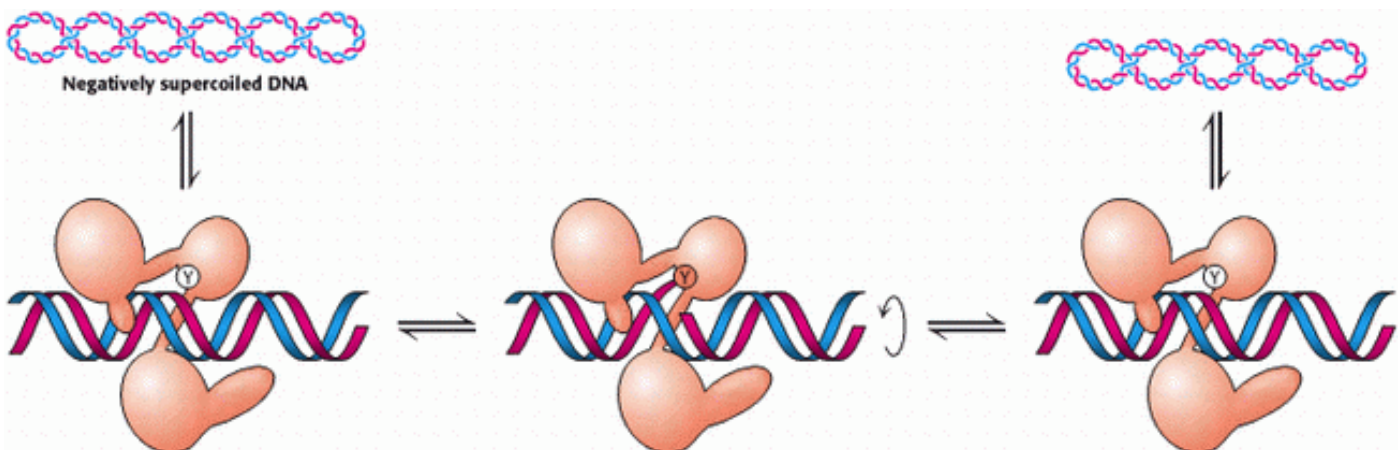


Figure 27.22. Topoisomerase I Mechanism. On binding to DNA, topoisomerase I cleaves one strand of the DNA through a tyrosine (Y) residue attacking a phosphate. When the strand has been cleaved, it rotates in a controlled manner

around the other strand. The reaction is completed by religation of the cleaved strand. This process results in partial or complete relaxation of a supercoiled plasmid.



Figure 27.23. Structure of Topoisomerase II. A composite structure of topoisomerase II formed from the amino-terminal ATP-binding domain of *E. coli* topoisomerase II (green) and the carboxyl-terminal fragment from yeast topoisomerase II (yellow). Both units form dimeric structures as shown.

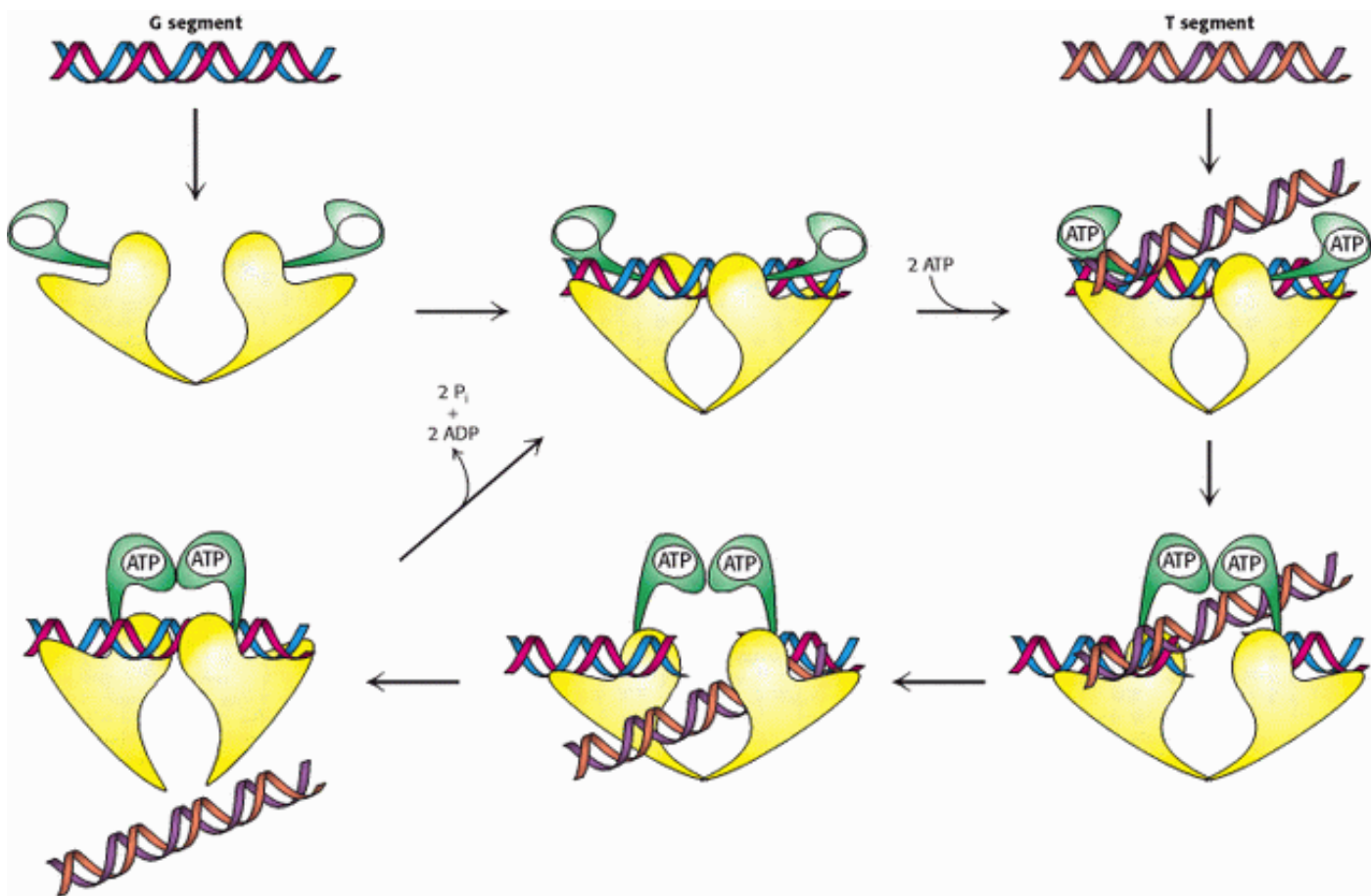


Figure 27.24. Topoisomerase II Mechanism. Topoisomerase II first binds one DNA duplex termed the G (for gate) segment. The binding of ATP to the two N-terminal domains brings these two domains together. This conformational change leads to the cleavage of both strands of the G segment and the binding of an additional DNA duplex, the T segment. This T segment then moves through the break in the G segment and out the bottom of the enzyme. The hydrolysis of ATP resets the enzyme with the G segment still bound.

27.4. DNA Replication of Both Strands Proceeds Rapidly from Specific Start Sites

So far, we have met many of the key players in DNA replication. Here, we ask, Where on the DNA molecule does replication begin, and how is the double helix manipulated to allow the simultaneous use of the two strands as templates? In *E. coli*, DNA replication starts at a unique site within the entire 4.8×10^6 bp genome. This *origin of replication*, called the *oriC locus*, is a 245-bp region that has several unusual features (Figure 27.25). The *oriC* locus contains four repeats of a sequence that together act as a binding site for an initiation protein called *dnaA*. In addition, the locus contains a tandem array of 13-bp sequences that are rich in A-T base pairs.

The binding of the *dnaA* protein to the four sites initiates an intricate sequence of steps leading to the unwinding of the template DNA and the synthesis of a primer. Additional proteins join *dnaA* in this process. The *dnaB* protein is a helicase that utilizes ATP hydrolysis to unwind the duplex. *The single-stranded regions are trapped by a single-stranded binding protein (SSB)*. The result of this process is the generation of a structure called the *prepriming complex*, which makes single-stranded DNA accessible for other enzymes to begin synthesis of the complementary strands.

27.4.1. An RNA Primer Synthesized by Primase Enables DNA Synthesis to Begin

Even with the DNA template exposed, new DNA cannot be synthesized until a primer is constructed. Recall that all known DNA polymerases require a primer with a free 3'-hydroxyl group for DNA synthesis. How is this primer formed? An important clue came from the observation that RNA synthesis is essential for the initiation of DNA synthesis. In fact,

RNA primes the synthesis of DNA. A specialized RNA polymerase called *primase* joins the prepriming complex in a multisubunit assembly called the *primosome*. Primase synthesizes a short stretch of RNA (~5 nucleotides) that is complementary to one of the template DNA strands (Figure 27.26). The primer is RNA rather than DNA because DNA polymerases cannot start chains de novo. Recall that, to ensure fidelity, DNA polymerase tests the correctness of the preceding base pair before forming a new phosphodiester bond (Section 27.2.4). RNA polymerases can start chains de novo because they do not examine the preceding base pair. Consequently, their error rates are orders of magnitude as high as those of DNA polymerases. The ingenious solution is to start DNA synthesis with a low-fidelity stretch of polynucleotide but mark it "temporary" by placing ribonucleotides in it. The RNA primer is removed by hydrolysis by a $5' \rightarrow 3'$ exonuclease; in *E. coli*, the exonuclease is present as an additional domain of DNA polymerase I, rather than being present in the Klenow fragment. Thus, the complete polymerase I has three distinct active sites: a $3' \rightarrow 5'$ exonuclease proofreading activity, a polymerase activity, and a $5' \rightarrow 3'$ exonuclease activity.

27.4.2. One Strand of DNA Is Made Continuously, Whereas the Other Strand Is Synthesized in Fragments

Both strands of parental DNA serve as templates for the synthesis of new DNA. The site of DNA synthesis is called the *replication fork* because the complex formed by the newly synthesized daughter strands arising from the parental duplex resembles a two-pronged fork. Recall that the two strands are antiparallel; that is, they run in opposite directions. As shown in Figure 27.3, both daughter strands appear to grow in the same direction on cursory examination. However, all known DNA polymerases synthesize DNA in the $5' \rightarrow 3'$ direction but not in the $3' \rightarrow 5'$ direction. How then does one of the daughter DNA strands appear to grow in the $3' \rightarrow 5'$ direction?

This dilemma was resolved by Reiji Okazaki, who found that *a significant proportion of newly synthesized DNA exists as small fragments*. These units of about a thousand nucleotides (called *Okazaki fragments*) are present briefly in the vicinity of the replication fork (Figure 27.27). As replication proceeds, these fragments become covalently joined through the action of DNA ligase (Section 27.4.3) to form one of the daughter strands. The other new strand is synthesized continuously. The strand formed from Okazaki fragments is termed the *lagging strand*, whereas the one synthesized without interruption is the *leading strand*. Both the Okazaki fragments and the leading strand are synthesized in the $5' \rightarrow 3'$ direction. *The discontinuous assembly of the lagging strand enables $5' \rightarrow 3'$ polymerization at the nucleotide level to give rise to overall growth in the $3' \rightarrow 5'$ direction.*

27.4.3. DNA Ligase Joins Ends of DNA in Duplex Regions

The joining of Okazaki fragments requires an enzyme that catalyzes the joining of the ends of two DNA chains. The existence of circular DNA molecules also points to the existence of such an enzyme. In 1967, scientists in several laboratories simultaneously discovered *DNA ligase*. *This enzyme catalyzes the formation of a phosphodiester bond between the $3'$ hydroxyl group at the end of one DNA chain and the $5'$ -phosphate group at the end of the other* (Figure 27.28). An energy source is required to drive this thermodynamically uphill reaction. In eukaryotes and archaea, ATP is the energy source. In bacteria, NAD^+ typically plays this role. We shall examine the mechanistic features that allow these two molecules to power the joining of two DNA chains.

DNA ligase cannot link two molecules of single-stranded DNA or circularize single-stranded DNA. Rather, *ligase seals breaks in double-stranded DNA molecules*. The enzyme from *E. coli* ordinarily forms a phosphodiester bridge only if there are at least several base pairs near this link. Ligase encoded by T4 bacteriophage can link two blunt-ended double-helical fragments, a capability that is exploited in recombinant DNA technology.

Let us look at the mechanism of joining, which was elucidated by I. Robert Lehman (Figure 27.29). ATP donates its activated AMP unit to DNA ligase to form a *covalent enzyme-AMP (enzyme-adenylate) complex* in which AMP is linked to the ϵ -amino group of a lysine residue of the enzyme through a phosphoamide bond. Pyrophosphate is concomitantly released. The activated AMP moiety is then transferred from the lysine residue to the phosphate group at the $5'$ terminus

of a DNA chain, forming a *DNA-adenylate complex*. The final step is a nucleophilic attack by the 3' hydroxyl group at the other end of the DNA chain on this activated 5' phosphorus atom.

In bacteria, NAD⁺ instead of ATP functions as the AMP donor. NMN is released instead of pyrophosphate. Two high-transfer-potential phosphoryl groups are spent in regenerating NAD⁺ from NMN and ATP when NAD⁺ is the adenylyate donor. Similarly, two high-transfer-potential phosphoryl groups are spent by the ATP-utilizing enzymes because the pyrophosphate released is hydrolyzed. The results of structural studies revealed that the ATP- and NAD⁺-utilizing enzymes are homologous even though this homology could not be deduced from their amino acid sequences alone.

27.4.4. DNA Replication Requires Highly Processive Polymerases

Enzyme activities must be highly coordinated to replicate entire genomes precisely and rapidly. A prime example is provided by *DNA polymerase III* holoenzyme, the enzyme responsible for DNA replication in *E. coli*. The hallmarks of this multisubunit assembly are its *very high catalytic potency, fidelity, and processivity*. *Processivity* refers to the ability of an enzyme to catalyze many consecutive reactions without releasing its substrate. The holoenzyme catalyzes the formation of many thousands of phosphodiester bonds before releasing its template, compared with only 20 for DNA polymerase I. DNA polymerase III holoenzyme has evolved to grasp its template and not let go until the template has been completely replicated. A second distinctive feature of the holoenzyme is its catalytic prowess: 1000 nucleotides are added per second compared with only 10 per second for DNA polymerase I. This acceleration is accomplished with no loss of accuracy. The greater catalytic prowess of polymerase III is largely due to its processivity; no time is lost in repeatedly stepping on and off the template.

Processive enzyme—

From the Latin *procedere*, "to go forward."

An enzyme that catalyzes multiple rounds of elongation or digestion of a polymer while the polymer stays bound. A *distributive enzyme*, in contrast, releases its polymeric substrate between successive catalytic steps.

These striking features of DNA polymerase III do not come cheaply. The holoenzyme consists of 10 kinds of polypeptide chains and has a mass of ~900 kd, nearly an order of magnitude as large as that of a single-chain DNA polymerase, such as DNA polymerase I. This replication complex is an *asymmetric dimer* (Figure 27.30). The holoenzyme is structured as a dimer to enable it to replicate both strands of parental DNA in the same place at the same time. It is asymmetric because the leading and lagging strands are synthesized differently. A τ_2 subunit is associated with one branch of the holoenzyme; γ_2 and $(\delta \delta' \chi \psi)_2$ are associated with the other. The core of each branch is the same, an $\alpha \epsilon \theta$ complex. The α subunit is the polymerase, and the ϵ subunit is the proofreading 3' \rightarrow 5' exonuclease. Each core is catalytically active but not processive. Processivity is conferred by β_2 and τ_2 .

The source of the processivity was revealed by the determination of the three-dimensional structure of the β_2 subunit (Figure 27.31). This unit has the form of a star-shaped ring. A 35-Å-diameter hole in its center can readily accommodate a duplex DNA molecule, yet leaves enough space between the DNA and the protein to allow rapid sliding and turning during replication. A catalytic rate of 1000 nucleotides polymerized per second requires the sliding of 100 turns of duplex DNA (a length of 3400 Å, or 0.34 μ m) through the central hole of β_2 per second. Thus, β_2 plays a key role in replication by serving as a sliding DNA clamp.

27.4.5. The Leading and Lagging Strands Are Synthesized in a Coordinated Fashion

The holoenzyme synthesizes the leading and lagging strands simultaneously at the replication fork (Figure 27.32). DNA polymerase III begins the synthesis of the leading strand by using the RNA primer formed by primase. The duplex DNA ahead of the polymerase is unwound by an ATP-driven helicase. Single-stranded binding protein again keeps the strands separated so that both strands can serve as templates. The leading strand is synthesized continuously by polymerase III, which does not release the template until replication has been completed. Topoisomerases II (DNA gyrase) concurrently introduces right-handed (negative) supercoils to avert a topological crisis.

The mode of synthesis of the lagging strand is necessarily more complex. As mentioned earlier, the lagging strand is synthesized in fragments so that $5' \rightarrow 3'$ polymerization leads to overall growth in the $3' \rightarrow 5'$ direction. A looping of the template for the lagging strand places it in position for $5' \rightarrow 3'$ polymerization (Figure 27.33). The looped lagging-strand template passes through the polymerase site in one subunit of a dimeric polymerase III in the same direction as that of the leading-strand template in the other subunit. DNA polymerase III lets go of the lagging-strand template after adding about 1000 nucleotides. A new loop is then formed, and primase again synthesizes a short stretch of RNA primer to initiate the formation of another Okazaki fragment.

The gaps between fragments of the nascent lagging strand are then filled by DNA polymerase I. This essential enzyme also uses its $5' \rightarrow 3'$ exonuclease activity to remove the RNA primer lying ahead of the polymerase site. The primer cannot be erased by DNA polymerase III, because the enzyme lacks $5' \rightarrow 3'$ editing capability. Finally, DNA ligase connects the fragments.

27.4.6. DNA Synthesis Is More Complex in Eukaryotes Than in Prokaryotes

Replication in eukaryotes is mechanistically similar to replication in prokaryotes but is more challenging for a number of reasons. One of them is sheer size: *E. coli* must replicate 4.8 million base pairs, whereas a human diploid cell must replicate 6 billion base pairs. Second, the genetic information for *E. coli* is contained on 1 chromosome, whereas, in human beings, 23 pairs of chromosomes must be replicated. Finally, whereas the *E. coli* chromosome is circular, human chromosomes are linear. Unless countermeasures are taken (Section 27.4.7), linear chromosomes are subject to shortening with each round of replication.

The first two challenges are met by the use of multiple origins of replication, which are located between 30 and 300 kbp apart. In human beings, replication requires about 30,000 origins of replication, with each chromosome containing several hundred. Each origin of replication represents a replication unit, or *replicon*. The use of multiple origins of replication requires mechanisms for ensuring that each sequence is replicated once and only once. The events of eukaryotic DNA replication are linked to the eukaryotic *cell cycle* (Figure 27.34). In the cell cycle, the processes of DNA synthesis and cell division (mitosis) are coordinated so that the replication of all DNA sequences is complete before the cell progresses into the next phase of the cycle. This coordination requires several *checkpoints* that control the progression along the cycle.

The origins of replication have not been well characterized in higher eukaryotes but, in yeast, the DNA sequence is referred to as an *autonomously replicating sequence* (ARS) and is composed of an AT-rich region made up of discrete sites. The ARS serves as a docking site for the *origin of replication complex* (ORC). The ORC is composed of six proteins with an overall mass of ~400 kd. The ORC recruits other proteins to form the prereplication complex. Several of the recruited proteins are called *licensing factors* because they permit the formation of the initiation complex. These proteins serve to ensure that each replicon is replicated once and only once in a cell cycle. How is this regulation achieved? After the licensing factors have established the initiation complex, these factors are marked for destruction by the attachment of ubiquitin and subsequently destroyed by proteasomal digestion (Section 23.2.2).

DNA helicases separate the parental DNA strands, and the single strands are stabilized by the binding of *replication protein A*, a single-stranded-DNA-binding protein. Replication begins with the binding of *DNA polymerase α* , which is the initiator polymerase. This enzyme has primase activity, used to synthesize RNA primers, as well as DNA polymerase activity, although it possesses no exonuclease activity. After a stretch of about 20 deoxynucleotides have been added to

the primer, another replication protein, called *protein replication factor C (RFC)*, displaces DNA polymerase α and attracts *proliferating cell nuclear antigen (PCNA)*. Homologous to the β_2 subunit of *E. coli* polymerase III, PCNA then binds to *DNA polymerase δ* . The association of polymerase δ with PCNA renders the enzyme highly processive and suitable for long stretches of replication. This process is called *polymerase switching* because polymerase δ has replaced polymerase α . Polymerase δ has $3' \rightarrow 5'$ exonuclease activity and can thus edit the replicated DNA. Replication continues in both directions from the origin of replication until adjacent replicons meet and fuse. RNA primers are removed and the DNA fragments are ligated by DNA ligase.

27.4.7. Telomeres Are Unique Structures at the Ends of Linear Chromosomes

Whereas the genomes of essentially all prokaryotes are circular, the chromosomes of human beings and other eukaryotes are linear. The free ends of linear DNA molecules introduce several complications that must be resolved by special enzymes. In particular, it is difficult to fully replicate DNA ends, because polymerases act only in the $5' \rightarrow 3'$ direction. The lagging strand would have an incomplete $5'$ end after the removal of the RNA primer. Each round of replication would further shorten the chromosome.

The first clue to how this problem is resolved came from sequence analyses of the ends of chromosomes, which are called *telomeres* (from the Greek *telos*, "an end"). Telomeric DNA contains hundreds of tandem repeats of a hexanucleotide sequence. One of the strands is G rich at the $3'$ end, and it is slightly longer than the other strand. In human beings, the repeating G-rich sequence is AGGGTT.

The structure adopted by telomeres has been extensively investigated. Recent evidence suggests that they may form large duplex loops (Figure 27.35). The single-stranded region at the very end of the structure has been proposed to loop back to form a DNA duplex with another part of the repeated sequence, displacing a part of the original telomeric duplex. This looplike structure is formed and stabilized by specific telomere-binding proteins. Such structures would nicely protect and mask the end of the chromosome.

27.4.8. Telomeres Are Replicated by Telomerase, a Specialized Polymerase That Carries Its Own RNA Template

How are the repeated sequences generated? An enzyme, termed *telomerase*, that executes this function has been purified and characterized. When a primer ending in GGTT is added to the human enzyme in the presence of deoxynucleoside triphosphates, the sequences GGTTAGGGTT and GGTTAGGGTTAGGGTT, as well as longer products, are generated. Elizabeth Blackburn and Carol Greider discovered that the enzyme contains an RNA molecule that serves as the template for elongation of the G-rich strand (Figure 27.36). Thus, the enzyme carries the information necessary to generate the telomere sequences. The exact number of repeated sequences is not crucial.

Subsequently, a protein component of telomerases also was identified. From its amino acid sequence, this component is clearly related to reverse transcriptases, enzymes first discovered in retroviruses that copy RNA into DNA. Thus, *telomerase is a specialized reverse transcriptase that carries its own template*. Telomeres may play important roles in cancer-cell biology and in cell aging.

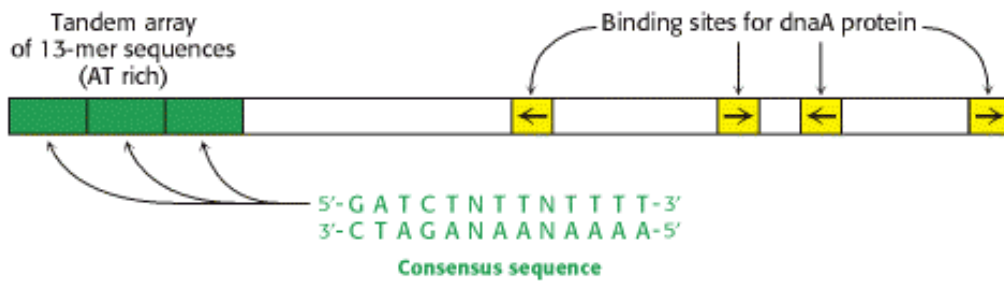


Figure 27.25. Origin of Replication in *E. coli*. *OriC* has a length of 245 bp. It contains a tandem array of three nearly identical 13-nucleotide sequences (green) and four binding sites (yellow) for the dnaA protein. The relative orientations of the four dnaA sites are denoted by arrows.

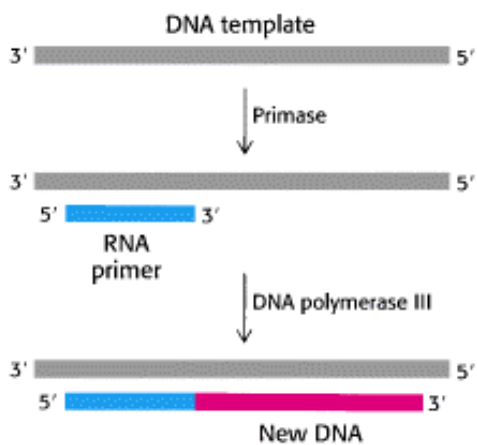


Figure 27.26. Priming. DNA replication is primed by a short stretch of DNA that is synthesized by primase, an RNA polymerase. The RNA primer is removed at a later stage of replication.

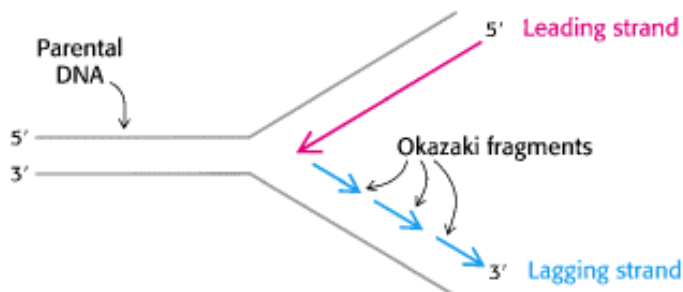


Figure 27.27. Okazaki Fragments. At a replication fork, both strands are synthesized in a 5' → 3' direction. The leading strand is synthesized continuously, whereas the lagging strand is synthesized in short pieces termed Okazaki fragments.

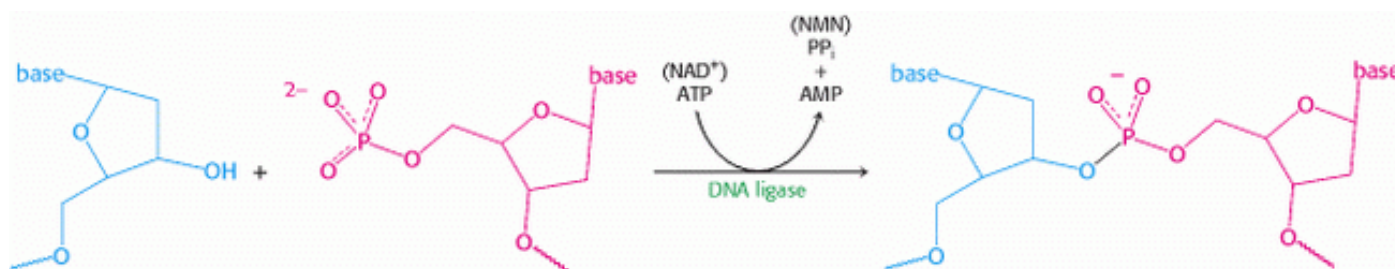


Figure 27.28. DNA Ligase Reaction. DNA ligase catalyzes the joining of one DNA strand with a free 3'-hydroxyl group to another with a free 5'-phosphate group. In eukaryotes and archaea, ATP is cleaved to AMP and PP_i to drive this reaction. In bacteria, NAD⁺ is cleaved to AMP and nicotinamide mononucleotide (NMN).

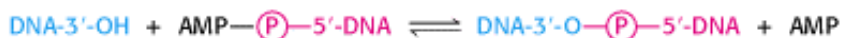


Figure 27.29. DNA Ligase Mechanism. DNA ligation proceeds by the transfer of an AMP unit first to a lysine side chain on DNA ligase and then to the 5'-phosphate group of the substrate. The AMP unit is released on formation of the phosphodiester linkage in DNA.

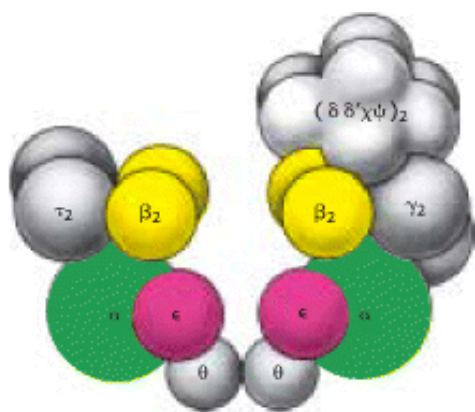


Figure 27.30. Proposed Architecture of DNA Polymerase III Holoenzyme. [After A. Kornberg and T. Baker, *DNA Replication*, 2d ed. (W. H. Freeman and Company, 1992).]



Figure 27.31. Structure of the Sliding Clamp. The dimeric β_2 subunit of DNA polymerase III forms a ring that surrounds the DNA duplex. It allows the polymerase enzyme to move without falling off the DNA substrate.

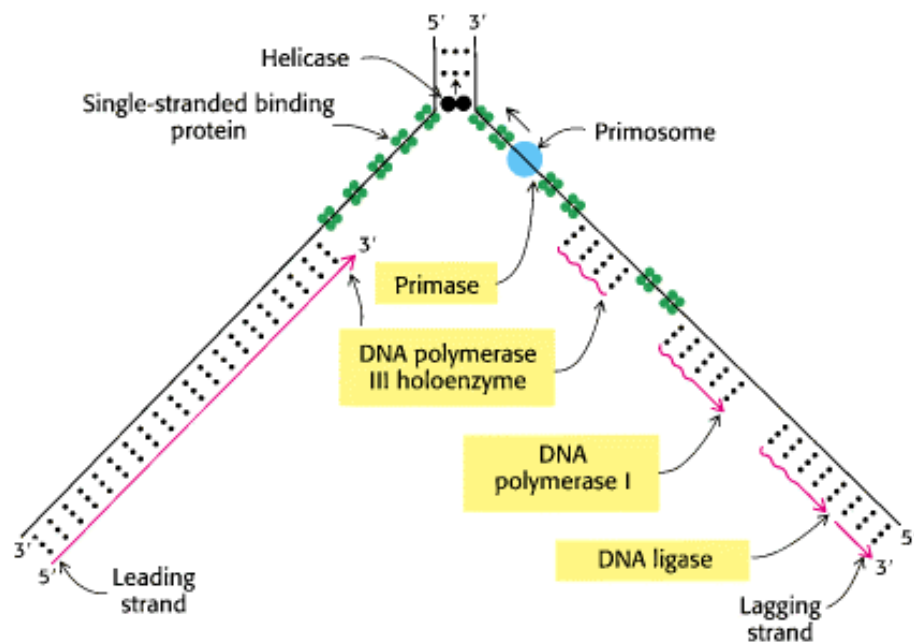


Figure 27.32. Replication Fork. Schematic representation of the enzymatic events at a replication fork in *E. coli*. Enzymes shaded in yellow catalyze chain initiation, elongation, and ligation. The wavy lines on the lagging strand denote RNA primers. [After A. Kornberg and T. Baker, *DNA Replication*, 2d ed. (W. H. Freeman and Company, 1992).]

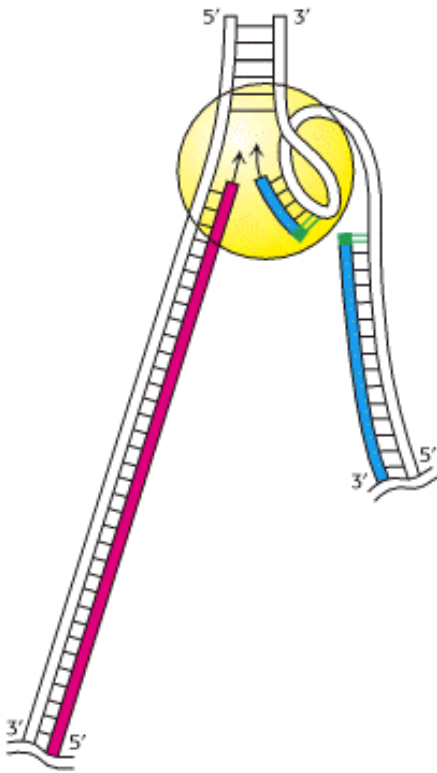


Figure 27.33. Coordination between the Leading and the Lagging Strands. The looping of the template for the lagging strand enables a dimeric DNA polymerase III holoenzyme to synthesize both daughter strands. The leading strand is shown in red, the lagging strand in blue, and the RNA primers in green. [Courtesy of Dr. Arthur Kornberg.]



Figure 27.34. Eukaryotic Cell Cycle. DNA replication and cell division must take place in a highly coordinated fashion in eukaryotes. Mitosis (M) takes place only after DNA synthesis (S). Two gaps (G_1 and G_2) in time separate the two processes.



Figure 27.35. Proposed Model for Telomeres. A single-stranded segment of the G-rich strand extends from the end of the telomere. In one model for telomeres, this single-stranded region invades the duplex to form a large duplex loop.

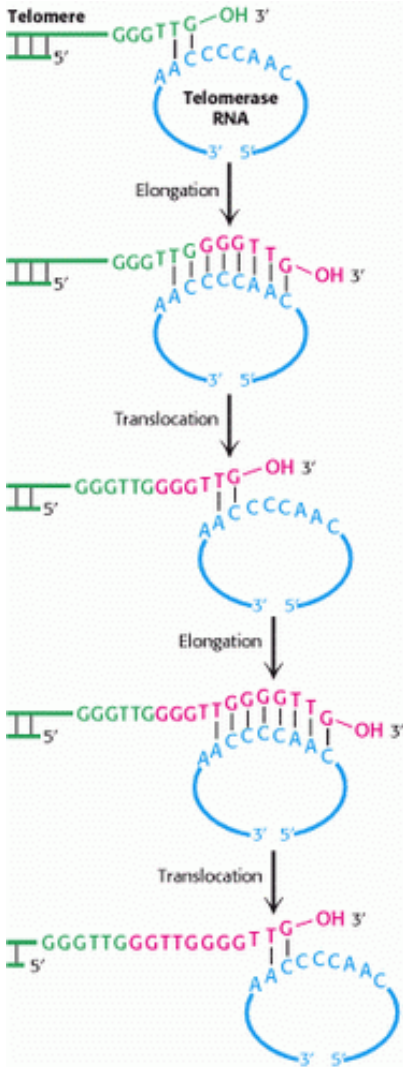


Figure 27.36. Telomere Formation. Mechanism of synthesis of the G-rich strand of telomeric DNA. The RNA template of telomerase is shown in blue and the nucleotides added to the G-rich strand of the primer are shown in red. [After E. H. Blackburn. *Nature* 350(1991):569.]

27.5. Double-Stranded DNA Molecules with Similar Sequences Sometimes Recombine

Most processes associated with DNA replication function to copy the genetic message as faithfully as possible. However, several biochemical processes require the *recombination* of genetic material between two DNA molecules. In genetic recombination, two daughter molecules are formed by the exchange of genetic material between two parent molecules (Figure 27.37).

1. In meiosis, the limited exchange of genetic material between paired chromosomes provides a simple mechanism for generating genetic diversity in a population.
2. As we shall see in Chapter 33, recombination plays a crucial role in generating molecular diversity for antibodies and some other molecules in the immune system.
3. Some viruses utilize recombination pathways to integrate their genetic material into the DNA of the host cell.
4. Recombination is used to manipulate genes in, for example, the generation of "gene knockout" mice (Section 6.3.5).

Recombination is most efficient between DNA sequences that are similar in sequence. Such processes are often referred to as *homologous recombination* reactions.

27.5.1. Recombination Reactions Proceed Through Holliday Junction Intermediates



The Structural Insights module for this chapter shows how a recombinase forms a Holliday junction from two DNA duplexes and suggests how this intermediate is resolved to produce recombinants.

Enzymes called *recombinases* catalyze the exchange of genetic material that takes place in recombination. By what pathway do these enzymes catalyze this exchange? An appealing scheme was proposed by Robin Holliday in 1964. A key intermediate in this mechanism is a crosslike structure, known as a *Holliday junction*, formed by four polynucleotide chains. Such intermediates have been characterized by a wide range of techniques including x-ray crystallography (Figure 27.38). Note that such intermediates can form only when the nucleotide sequences of the two parental duplexes are very similar or identical in the region of recombination because specific base pairs must form between the bases of the two parental duplexes.

How are such intermediates formed from the parental duplexes and resolved to form products? Many details for this process are now available, based largely on the results of studies of Cre recombinase from bacteriophage P1. This mechanism begins with the recombinase binding to the DNA substrates (Figure 27.39). Four molecules of the enzyme and their associated DNA molecules come together to form a *recombination synapse*. The reaction begins with the cleavage of one strand from each duplex. The 5'-hydroxyl group of each cleaved strand remains free, whereas the 3'-phosphoryl group becomes linked to a specific tyrosine residue in the recombinase. The free 5' ends invade the other duplex in the synapse and attack the DNA-tyrosine units to form new phosphodiester-bonds and free the tyrosine residues. These reactions result in the formation of a Holliday junction. This junction can then isomerize to form a structure in which the polynucleotide chains in the center of the structure are reoriented. From this junction, the processes of strand cleavage and phosphodiester-bond formation repeat. The result is a synapse containing the two recombined duplexes. Dissociation of this complex generates the final recombined products.

27.5.2. Recombinases Are Evolutionarily Related to Topoisomerases


 The intermediates that form in recombination reactions, with their tyrosine adducts possessing 3'-phosphoryl groups, are reminiscent of the intermediates that form in the reactions catalyzed by topoisomerases. This mechanistic similarity reflects deeper evolutionary relationships. Examination of the three-dimensional structures of recombinases and type I topoisomerases reveals that these proteins are related by divergent evolution despite little amino acid sequence similarity (Figure 27.40). From this perspective, the action of a recombinase can be viewed as an intermolecular topoisomerase reaction. In each case, a tyrosine-DNA adduct is formed. In a topoisomerase reaction, this adduct is resolved when the 5'-hydroxyl group of the same duplex attacks to reform the same phosphodiester bond that was initially cleaved. In a recombinase reaction, the attacking 5'-hydroxyl group comes from a DNA chain that was not initially linked to the phosphoryl group participating in the phosphodiester bond.



Figure 27.37. Recombination. Two DNA molecules can recombine with each other to form new DNA molecules that have segments from both parental molecules.

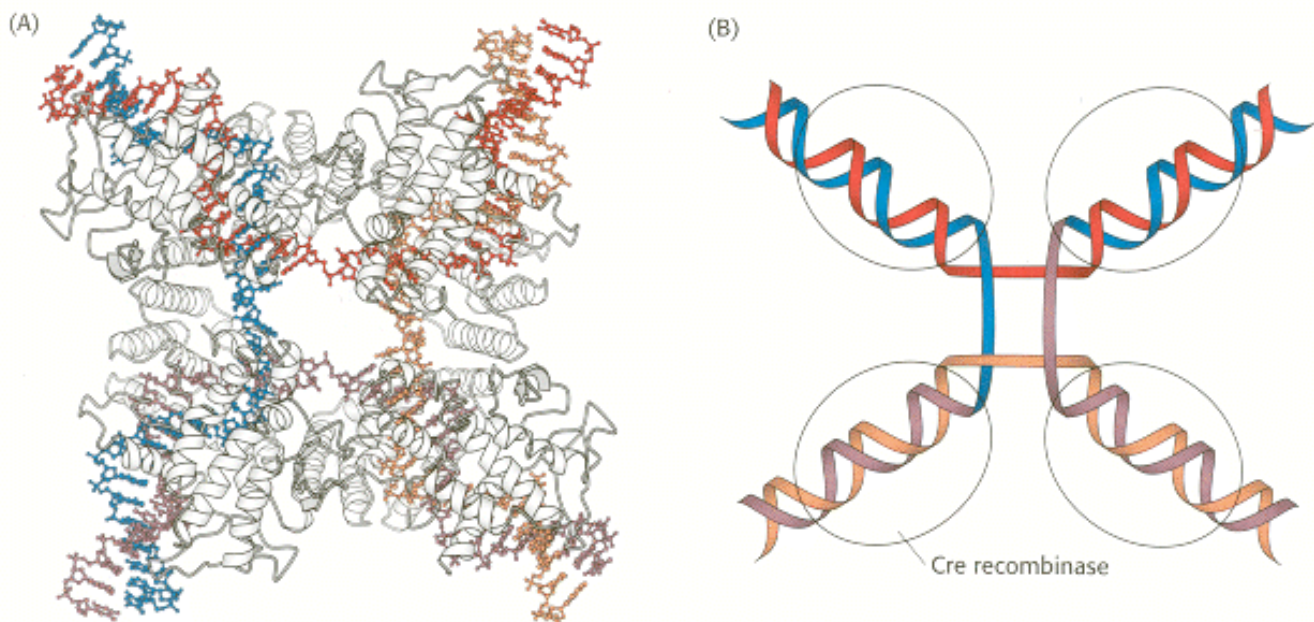


Figure 27.38. Holliday Junction. (A) The structure of a Holliday junction bound by Cre recombinase (gray), a bacteriophage protein. (B) A schematic view of a Holliday junction.

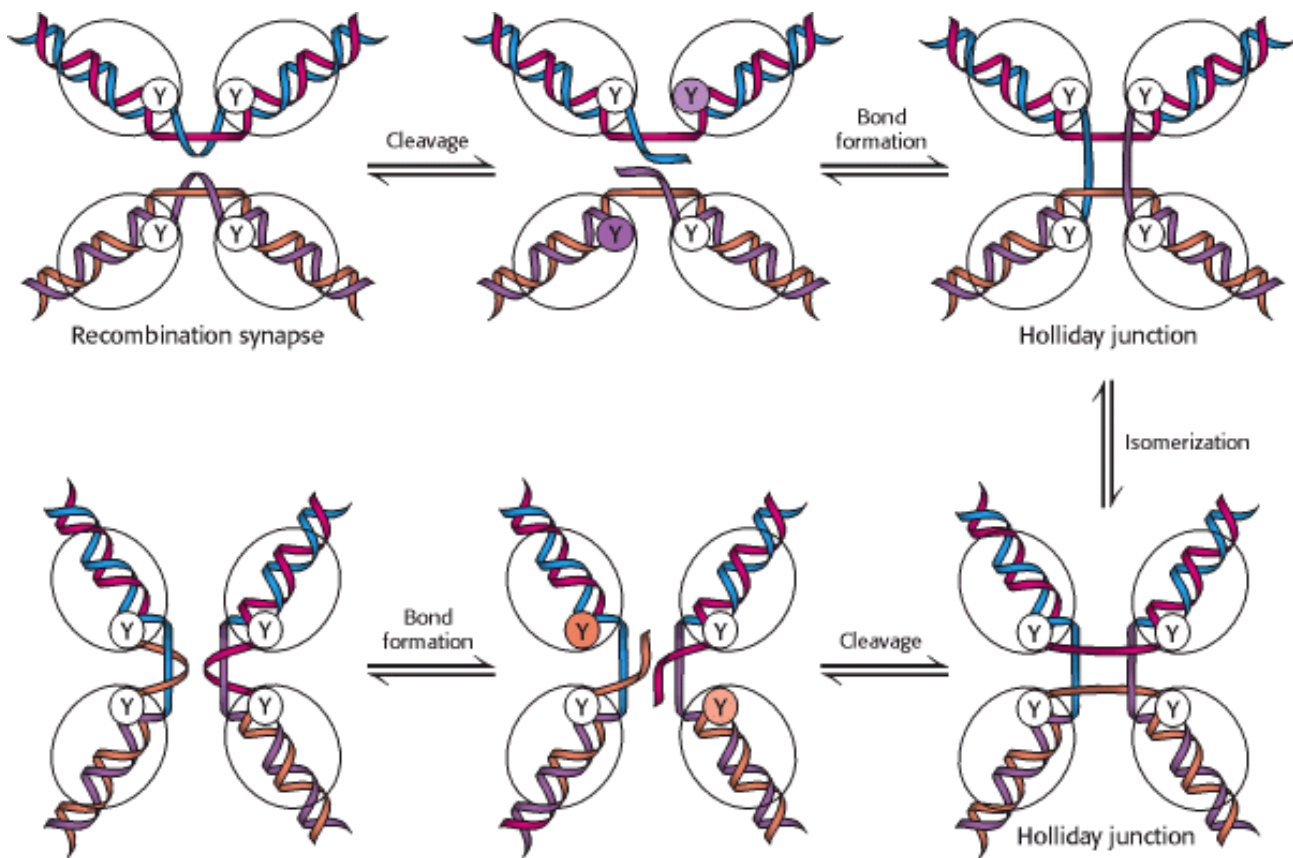


Figure 27.39. Recombination Mechanism. Recombination begins as two DNA molecules come together to form a recombination synapse. One strand from each duplex is cleaved by the recombinase enzyme; the 3rd end of each of the cleaved strands is linked to a tyrosine (Y) residue on the recombinase enzyme. New phosphodiester bonds are formed when a 5th end of the other cleaved strand in the complex attacks these tyrosine-DNA adducts. After isomerization, these steps are repeated to form the recombined products.

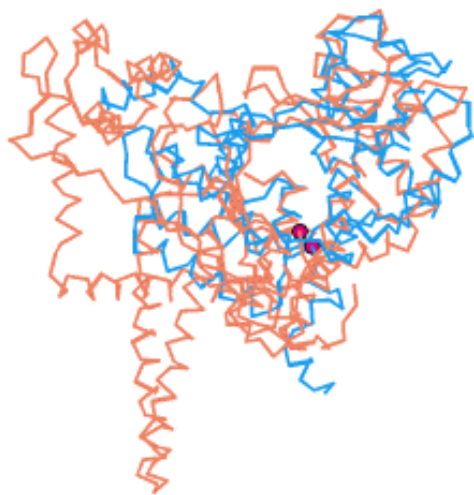
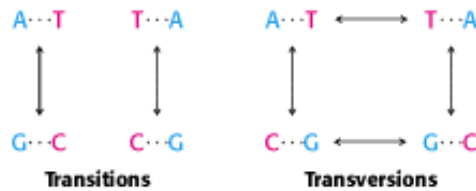


Figure 27.40. Recombinases and Topoisomerase I. A superposition of Cre recombinase (blue) and topoisomerase I (orange) reveals that these two enzymes have a common structural core. The positions of the tyrosine residues that participate in DNA cleavage reactions are shown as red spheres for both enzymes.

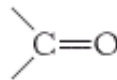
27.6. Mutations Involve Changes in the Base Sequence of DNA

We now turn from DNA replication to DNA mutations and repair. Several types of mutations are known: (1) the *substitution* of one base pair for another, (2) the *deletion* of one or more base pairs, and (3) the *insertion* of one or more base pairs. The spontaneous mutation rate of T4 phage is about 10^{-7} per base per replication. *E. coli* and *Drosophila melanogaster* have much lower mutation rates, of the order of 10^{-10} .

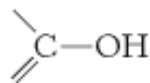


The substitution of one base pair for another is the a common type of mutation. Two types of substitutions are possible. A *transition* is the replacement of one purine by the other or that of one pyrimidine by the other. In contrast, a *transversion* is the replacement of a purine by a pyrimidine or that of a pyrimidine by a purine.

Watson and Crick suggested a mechanism for the spontaneous occurrence of transitions in a classic paper on the DNA double helix. They noted that some of the hydrogen atoms on each of the four bases can change their location to produce a *tautomer*. An amino group ($-\text{NH}_2$) can tautomerize to an imino form ($=\text{NH}$). Likewise, a keto group (



can tautomerize to an enol form



. The fraction of each type of base in the form of these imino and enol tautomers is about 10^{-4} . These transient tautomers

can form nonstandard base pairs that fit into a double helix. For example, the imino tautomer of adenine can pair with cytosine (Figure 27.41). This A*-C pairing (the asterisk denotes the imino tautomer) would allow C to become incorporated into a growing DNA strand where T was expected, and it would lead to a mutation if left uncorrected. In the next round of replication, A* will probably reautomerize to the standard form, which pairs as usual with thymine, but the cytosine residue will pair with guanine. Hence, one of the daughter DNA molecules will contain a G-C base pair in place of the normal A-T base pair.

Tautomerization—

The interconversion of two isomers that differ only in the position of protons (and, often, double bonds).

27.6.1. Some Chemical Mutagens Are Quite Specific

Base analogs such as 5-bromouracil and 2-aminopurine can be incorporated into DNA and are even more likely than normal nucleic acid bases to form transient tautomers that lead to transition mutations. 5-Bromouracil, an analog of thymine, normally pairs with adenine. However, the proportion of 5-bromouracil in the enol tautomer is higher than that of thymine because the bromine atom is more electronegative than is a methyl group on the C-5 atom. Thus, the incorporation of 5-bromouracil is especially likely to cause altered base-pairing in a subsequent round of DNA replication ([Figure 27.42](#)).

Other mutagens act by chemically modifying the bases of DNA. For example, nitrous acid (HNO_2) reacts with bases that contain amino groups. Adenine is oxidatively deaminated to hypoxanthine, cytosine to uracil, and guanine to xanthine. Hypoxanthine pairs with cytosine rather than with thymine ([Figure 27.43](#)). Uracil pairs with adenine rather than with guanine. Xanthine, like guanine, pairs with cytosine. Consequently, nitrous acid causes A-T \leftrightarrow G-C transitions.

A different kind of mutation is produced by flat aromatic molecules such as the acridines ([Figure 27.44](#)). These compounds intercalate in DNA—that is, they slip in between adjacent base pairs in the DNA double helix. Consequently, they lead to the insertion or deletion of one or more base pairs. The effect of such mutations is to alter the reading frame in translation, unless an integral multiple of three base pairs is inserted or deleted. In fact, the analysis of such mutants contributed greatly to the revelation of the triplet nature of the genetic code.

Some compounds are converted into highly active mutagens through the action of enzymes that normally play a role in detoxification. A striking example is aflatoxin B₁, a compound produced by molds that grows on peanuts and other foods. A cytochrome P450 enzyme ([Section 26.4.3](#)) converts this compound into a highly reactive epoxide ([Figure 27.45](#)). This agent reacts with the N-7 atom of guanosine to form an adduct that frequently leads to a G-C-to-T-A transversion.

27.6.2. Ultraviolet Light Produces Pyrimidine Dimers

The ultraviolet component of sunlight is a ubiquitous DNA-damaging agent. Its major effect is to covalently link adjacent pyrimidine residues along a DNA strand ([Figure 27.46](#)). Such a pyrimidine dimer cannot fit into a double helix, and so replication and gene expression are blocked until the lesion is removed.

27.6.3. A Variety of DNA-Repair Pathways Are Utilized

The maintenance of the integrity of the genetic message is key to life. Consequently, all cells possess mechanisms to repair damaged DNA. Three types of repair pathways are direct repair, base-excision repair, and nucleotide-excision repair ([Figure 27.47](#)).

An example of *direct repair* is the photochemical cleavage of pyrimidine dimers. Nearly all cells contain a *photoreactivating enzyme* called *DNA photolyase*. The *E. coli* enzyme, a 35-kd protein that contains bound *N*⁵,*N*¹⁰-methenyltetrahydrofolate and flavin adenine dinucleotide cofactors, binds to the distorted region of DNA. The enzyme uses light energy—specifically, the absorption of a photon by the *N*⁵,*N*¹⁰-methenyltetrahydrofolate coenzyme—to form an excited state that cleaves the dimer into its original bases.

The excision of modified bases such as 3-methyladenine by the *E. coli* enzyme *AlkA* is an example of *base-excision repair*. The binding of this enzyme to damaged DNA flips the affected base out of the DNA double helix and into the active site of the enzyme ([Figure 27.48](#)). Base flipping also occurs in the enzymatic addition of methyl groups to DNA bases ([Section 24.2.7](#)). The enzyme then acts as a glycosylase, cleaving the glycosidic bond to release the damaged base. At this stage, the DNA backbone is intact, but a base is missing. This hole is called an *AP site* because it is apurinic (devoid of A or G) or apyrimidinic (devoid of C or T). An AP endonuclease recognizes this defect and nicks the backbone adjacent to the missing base. *Deoxyribose phosphodiesterase* excises the residual deoxyribose phosphate unit,

and DNA polymerase I inserts an undamaged nucleotide, as dictated by the base on the undamaged complementary strand. Finally, the repaired strand is sealed by DNA ligase.


One of the best-understood examples of *nucleotide-excision repair* is the excision of a pyrimidine dimer. Three enzymatic activities are essential for this repair process in *E. coli* (Figure 27.49). First, an enzyme complex consisting of the proteins encoded by the *uvrABC* genes detects the distortion produced by the pyrimidine dimer. A specific *uvrABC* enzyme then cuts the damaged DNA strand at two sites, 8 nucleotides away from the dimer on the 5' side and 4 nucleotides away on the 3' side. The 12-residue oligonucleotide excised by this highly specific *excinuclease* (from the Latin *exci*, "to cut out") then diffuses away. DNA polymerase I enters the gap to carry out repair synthesis. The 3' end of the nicked strand is the primer, and the intact complementary strand is the template. Finally, the 3' end of the newly synthesized stretch of DNA and the original part of the DNA chain are joined by *DNA ligase*.

27.6.4. The Presence of Thymine Instead of Uracil in DNA Permits the Repair of Deaminated Cytosine

The presence in DNA of thymine rather than uracil was an enigma for many years. Both bases pair with adenine. The only difference between them is a methyl group in thymine in place of the C-5 hydrogen atom in uracil. Why is a methylated base employed in DNA and not in RNA? The existence of an active repair system to correct the deamination of cytosine provides a convincing solution to this puzzle.

Cytosine in DNA spontaneously deaminates at a perceptible rate to form uracil. The deamination of cytosine is potentially mutagenic because uracil pairs with adenine, and so one of the daughter strands will contain an U-A base pair rather than the original C-G base pair (Figure 27.50). This mutation is prevented by a repair system that recognizes uracil to be foreign to DNA. This enzyme, *uracil DNA glycosylase*, is homologous to AlkA. The enzyme hydrolyzes the glycosidic bond between the uracil and deoxyribose moieties but does not attack thymine-containing nucleotides. The AP site generated is repaired to reinsert cytosine. Thus, *the methyl group on thymine is a tag that distinguishes thymine from deaminated cytosine*. If thymine were not used in DNA, uracil correctly in place would be indistinguishable from uracil formed by deamination. The defect would persist unnoticed, and so a C-G base pair would necessarily be mutated to U-A in one of the daughter DNA molecules. This mutation is prevented by a repair system that searches for uracil and leaves thymine alone. *Thymine is used instead of uracil in DNA to enhance the fidelity of the genetic message*. In contrast, RNA is not repaired, and so uracil is used in RNA because it is a less-expensive building block.


27.6.5. Many Cancers Are Caused by Defective Repair of DNA

 As discussed in Chapter 15, cancers are caused by mutations in genes associated with growth control. Defects in DNA-repair systems are expected to increase the overall frequency of mutations and, hence, the likelihood of a cancer-causing mutation. *Xeroderma pigmentosum*, a rare human skin disease, is genetically transmitted as an autosomal recessive trait. The skin in an affected homozygote is extremely sensitive to sunlight or ultraviolet light. In infancy, severe changes in the skin become evident and worsen with time. The skin becomes dry, and there is a marked atrophy of the dermis. Keratoses appear, the eyelids become scarred, and the cornea ulcerates. Skin cancer usually develops at several sites. Many patients die before age 30 from metastases of these malignant skin tumors.

Ultraviolet light produces pyrimidine dimers in human DNA, as it does in *E. coli* DNA. Furthermore, the repair mechanisms are similar. Studies of skin fibroblasts from patients with xeroderma pigmentosum have revealed a biochemical defect in one form of this disease. In normal fibroblasts, half the pyrimidine dimers produced by ultraviolet radiation are excised in less than 24 hours. In contrast, almost no dimers are excised in this time interval in fibroblasts derived from patients with xeroderma pigmentosum. The results of these studies show that *xeroderma pigmentosum can be produced by a defect in the excinuclease that hydrolyzes the DNA backbone near a pyrimidine dimer*. *The drastic clinical consequences of this enzymatic defect emphasize the critical importance of DNA-repair processes*. The disease can also be caused by mutations in eight other genes for DNA repair, which attests to the complexity of repair processes.

Defects in other repair systems can increase the frequency of other tumors. For example, *hereditary nonpolyposis colorectal cancer (HNPCC, or Lynch syndrome)* results from defective DNA mismatch repair. HNPCC is not rare—as many as 1 in 200 people will develop this form of cancer. Mutations in two genes, called *hMSH2* and *hMLH1*, account for most cases of this hereditary predisposition to cancer. The striking finding is that these genes encode the human counterparts of MutS and MutL of *E. coli*. The MutS protein binds to mismatched base pairs (e.g., G-T) in DNA. An MutH protein, together with MutL, participates in cleaving one of the DNA strands in the vicinity of this mismatch to initiate the repair process (Figure 27.51). It seems likely that mutations in *hMSH2* and *hMLH1* lead to the accumulation of mutations throughout the genome. In time, genes important in controlling cell proliferation become altered, resulting in the onset of cancer.

27.6.6. Some Genetic Diseases Are Caused by the Expansion of Repeats of Three Nucleotides

 Some genetic diseases are caused by the presence of DNA sequences that are inherently prone to errors in the course of replication. A particularly important class of such diseases are characterized by the presence of long tandem arrays of repeats of three nucleotides. An example is *Hunt-ington disease*, an autosomal dominant neurological disorder with a variable age of onset. The mutated gene in this disease expresses a protein called huntingtin, which is expressed in the brain and contains a stretch of consecutive glutamine residues. These glutamine residues are encoded by a tandem array of CAG sequences within the gene. In unaffected persons, this array is between 6 and 31 repeats, whereas, in those with the disease, the array is between 36 and 82 repeats or longer. Moreover, the array tends to become longer from one generation to the next. The consequence is a phenomenon called *anticipation*: the children of an affected parent tend to show symptoms of the disease at an earlier age than did the parent.

The tendency of these *trinucleotide repeats* to expand is explained by the formation of alternative structures in DNA replication (Figure 27.52). Part of the array within the daughter strand can loop out without disrupting base-pairing outside this region. DNA polymerase extends this strand through the remainder of the array, leading to an increase in the number of copies of the trinucleotide sequence.

A number of other neurological diseases are characterized by expanding arrays of trinucleotide repeats. How do these long stretches of repeated amino acids cause disease? For huntingtin, it appears that the polyglutamine stretches become increasingly prone to aggregate as their length increases; the additional consequences of such aggregation are still under active investigation.

27.6.7. Many Potential Carcinogens Can Be Detected by Their Mutagenic Action on Bacteria

Many human cancers are caused by exposure to chemicals. These chemical carcinogens usually cause mutations, which suggests that *damage to DNA is a fundamental event in the origin of mutations and cancer*. It is important to identify these compounds and ascertain their potency so that human exposure to them can be minimized. Bruce Ames devised a simple and sensitive test for detecting chemical mutagens. In the *Ames test*, a thin layer of agar containing about 10^9 bacteria of a specially constructed tester strain of *Salmonella* is placed on a petri dish. These bacteria are unable to grow in the absence of histidine, because a mutation is present in one of the genes for the biosynthesis of this amino acid. The addition of a chemical mutagen to the center of the plate results in many new mutations. A small proportion of them reverse the original mutation, and histidine can be synthesized. These *revertants* multiply in the absence of an external source of histidine and appear as discrete colonies after the plate has been incubated at 37°C for 2 days (Figure 27.53). For example, 0.5 µg of 2-aminoanthracene gives 11,000 revertant colonies, compared with only 30 spontaneous revertants in its absence. A series of concentrations of a chemical can be readily tested to generate a dose-response curve. These curves are usually linear, which suggests that there is no threshold concentration for mutagenesis.

Some of the tester strains are responsive to *base-pair substitutions*, whereas others detect *deletions or additions of base pairs (frameshifts)*. The sensitivity of these specially designed strains has been enhanced by the genetic deletion of their excision-repair systems. Potential mutagens enter the tester strains easily because the lipopolysaccharide barrier that

normally coats the surface of *Salmonella* is incomplete in these strains.

A key feature of this detection system is the inclusion of a *mammalian liver homogenate* (Section 4.1.2). Recall that some potential carcinogens such as aflatoxin are converted into their active forms by enzyme systems in the liver or other mammalian tissues (Section 27.6.1). Bacteria lack these enzymes, and so the test plate requires a few milligrams of a liver homogenate to activate this group of mutagens.

The *Salmonella* test is extensively used to help evaluate the mutagenic and carcinogenic risks of a large number of chemicals. This rapid and inexpensive bacterial assay for mutagenicity complements epidemiological surveys and animal tests that are necessarily slower, more laborious, and far more expensive. The *Salmonella* test for mutagenicity is an outgrowth of studies of gene-protein relations in bacteria. It is a striking example of how fundamental research in molecular biology can lead directly to important advances in public health.

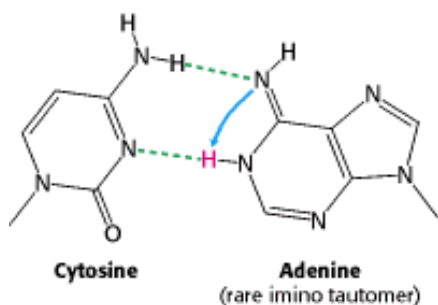


Figure 27.41. Base Pair with Mutagenic Tautomer. The bases of DNA can exist in rare tautomeric forms. The imino tautomer of adenine can pair with cytosine, eventually leading to a transition from A-T to G-C.

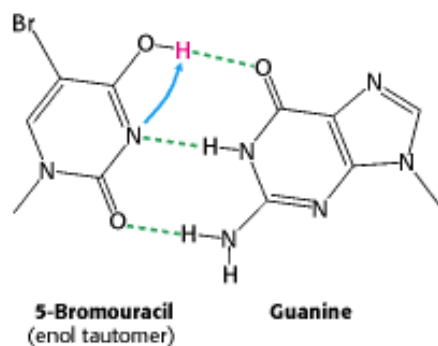


Figure 27.42. Base Pair with 5-Bromouracil. This analog of thymine has a higher tendency to form an enol tautomer than does thymine itself. The pairing of the enol tautomer of 5-bromouracil with guanine will lead to a transition from T-A to C-G.

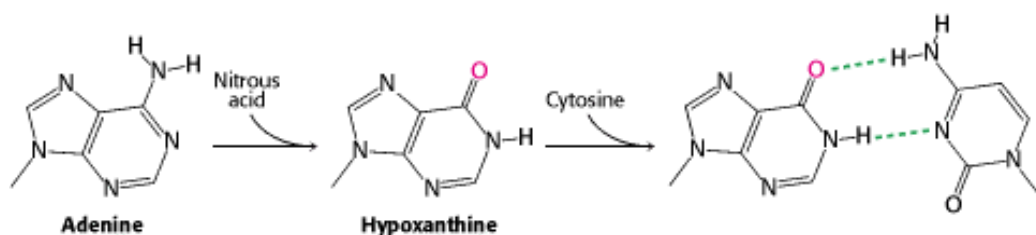
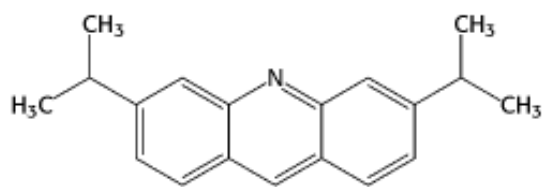
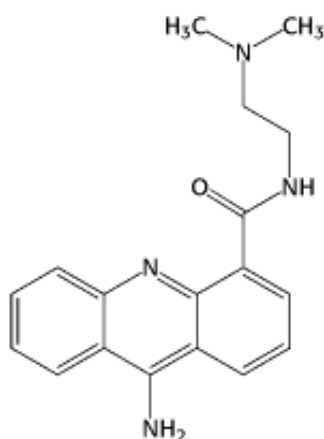


Figure 27.43. Chemical Mutagenesis. Treatment of DNA with nitrous acid results in the conversion of adenine into hypoxanthine. Hypoxanthine pairs with cytosine, inducing a transition from A-T to G-C.

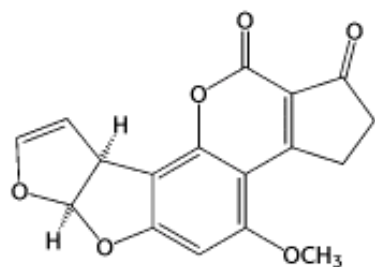


Acridine orange



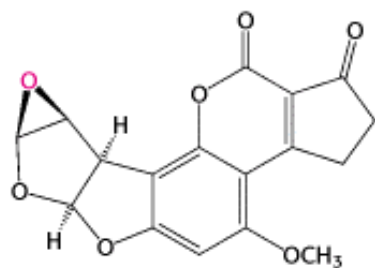
9-Amine-(N-(2-dimethylamino)-ethyl)acridine-4-carboxamide

Figure 27.44. Acridines. Acridine dyes induce frameshift mutations by intercalating into the DNA, leading to the incorporation of an additional base on the opposite strand.



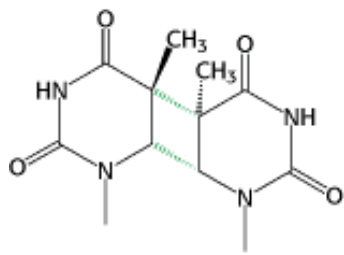
Aflatoxin B₁

↓
Cytochrome P450



Active DNA-modifying agent

Figure 27.45. Aflatoxin Reaction. The compound, produced by molds that grow on peanuts, is activated by cytochrome P450 to form a highly reactive species that modifies bases such as guanine in DNA, leading to mutations.



Thymine dimer

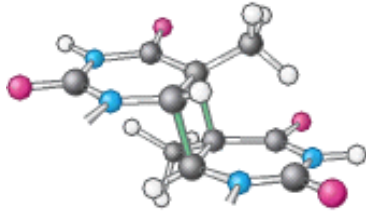


Figure 27.46. Cross-Linked Dimer of Two Thymine Bases. Ultraviolet light induces cross-links between adjacent pyrimidines along one strand of DNA.

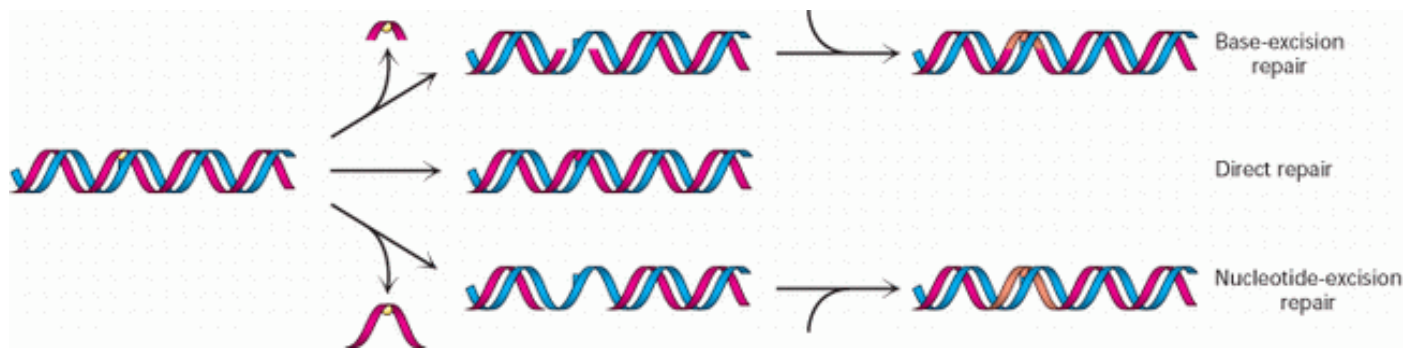


Figure 27.47. Repair Pathways. Three different pathways are used to repair damaged regions in DNA. In base-excision repair, the damaged base is removed and replaced. In direct repair, the damaged region is corrected in place. In nucleotide-excision repair, a stretch of DNA around the site of damage is removed and replaced.

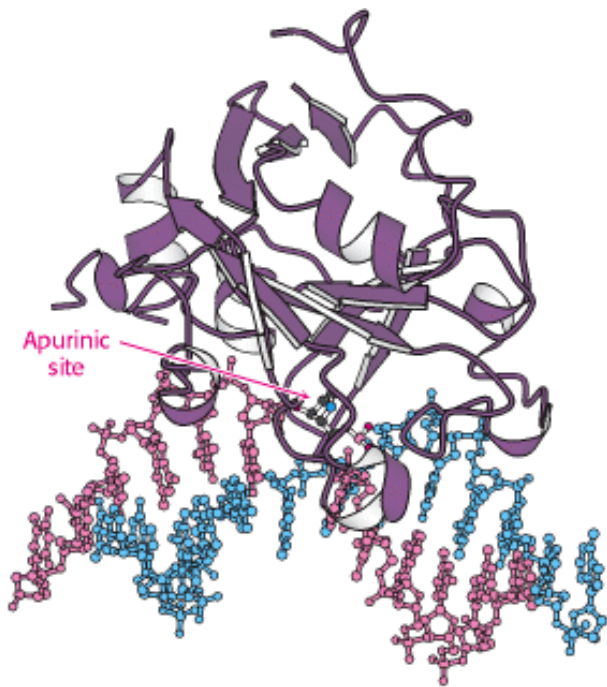


Figure 27.48. Structure of DNA-Repair Enzyme. A complex between the DNA-repair enzyme AlkA and an analog of an apurinic site. Note that the damaged base is flipped out of the DNA double helix into the active site of the enzyme for excision.

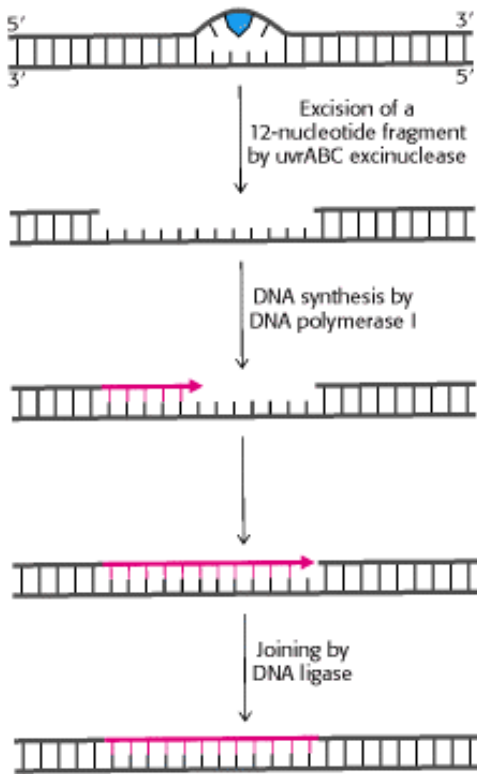


Figure 27.49. Excision Repair. Repair of a region of DNA containing a thymine dimer by the sequential action of a

specific excinuclease, a DNA polymerase, and a DNA ligase. The thymine dimer is shown in blue, and the new region of DNA is in red. [After P. C. Hanawalt. *Endavour* 31(1982):83.]

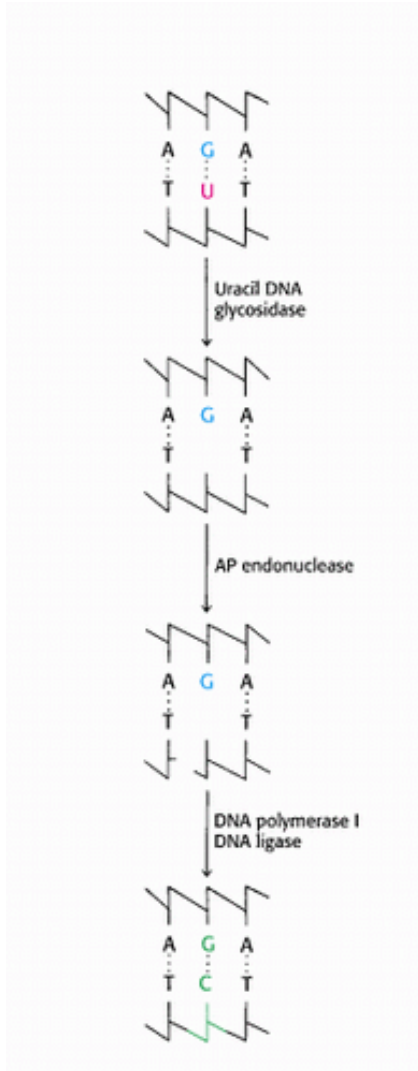


Figure 27.50. Uracil Repair. Uracil bases in DNA, formed by the deamination of cytosine, are excised and replaced by cytosine.

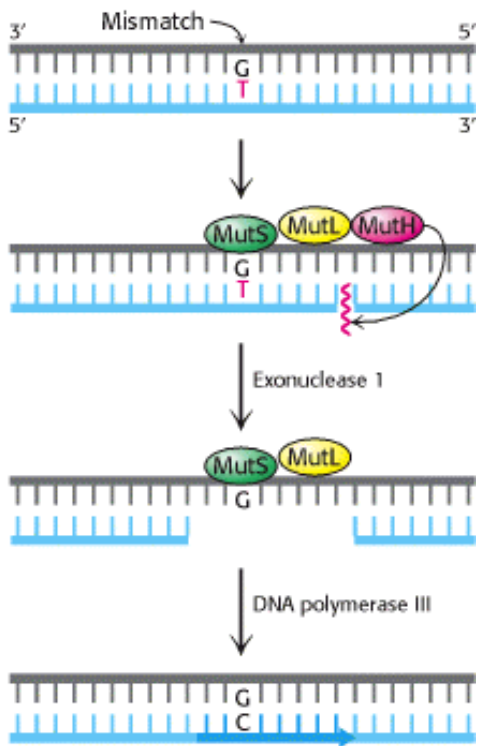


Figure 27.51. Mismatch Repair. DNA mismatch repair in *E. coli* is initiated by the interplay of MutS, MutL, and MutH proteins. A G-T mismatch is recognized by MutS. MutH cleaves the backbone in the vicinity of the mismatch. A segment of the DNA strand containing the erroneous T is removed by exonuclease I and synthesized anew by DNA polymerase III. [After R. F. Service. *Science* 263(1994):1559.]

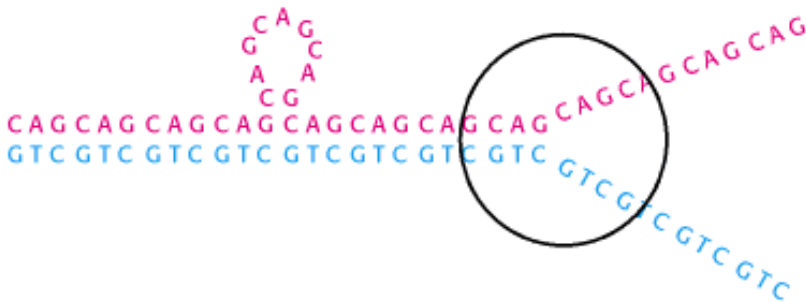


Figure 27.52. Triplet Repeat Expansion. Sequences containing tandem arrays of repeated triplet sequences can be expanded to include more repeats by the looping out of some of the repeats before replication.

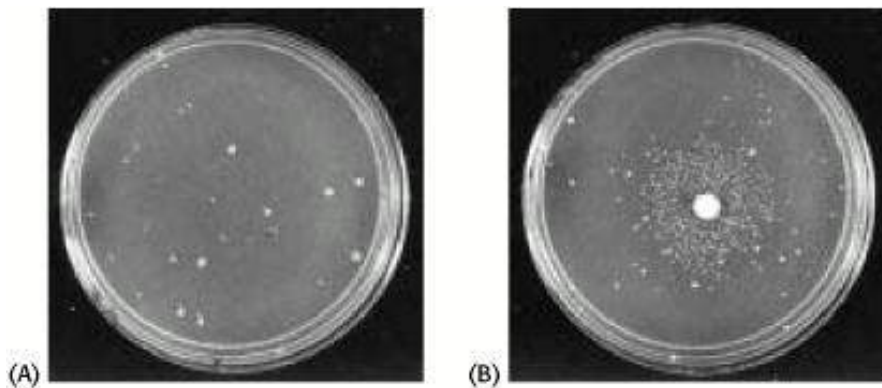


Figure 27.53. Ames Test. (A) A petri dish containing about 10^9 *Salmonella* bacteria that cannot synthesize histidine and (B) a petri dish containing a filter-paper disc with a mutagen, which produces a large number of revertants that can

synthesize histidine. After 2 days, the revertants appear as rings of colonies around the disc. The small number of visible colonies in plate A are spontaneous revertants. [From B. N. Ames, J. McCann, and E. Yamasaki. *Mutat. Res.* 31 (1975):347.]

Summary

DNA Can Assume a Variety of Structural Forms

DNA is a structurally dynamic molecule that can exist in a variety of helical forms: A-DNA, B-DNA (the classic Watson-Crick helix), and Z-DNA. DNA can be bent, kinked, and unwound. In A-, B-, and Z-DNA, two antiparallel chains are held together by Watson-Crick base pairs and stacking interactions between bases in the same strand. The sugar-phosphate backbone is on the outside, and the bases are inside the double helix. A- and B-DNA are right-handed helices. In B-DNA, the base pairs are nearly perpendicular to the helix axis. In A-DNA, the bases are tilted rather than perpendicular. An important structural feature of the B helix is the presence of major and minor grooves, which display different potential hydrogen-bond acceptors and donors according to the base sequence. X-ray analysis of a single crystal of B-DNA reveals that the structure is much more variable than was originally imagined. Dehydration induces the transition from B- to A-DNA. Z-DNA is a left-handed helix. It can be formed in regions of DNA in which purines alternate with pyrimidines, as in CGCG or CACA. Most of the DNA in a cell is in the B-form.

DNA Polymerases Require a Template and a Primer

DNA polymerases are template-directed enzymes that catalyze the formation of phosphodiester bonds by the 3'-hydroxyl group's nucleophilic attack on the innermost phosphorus atom of a deoxyribonucleoside 5'-triphosphate. They cannot start chains de novo; a primer with a free 3'-hydroxyl group is required. DNA polymerases from a variety of sources have important structural features in common as well as a catalytic mechanism requiring the presence of two metal ions. Many DNA polymerases proofread the nascent product; their 3' → 5' exonuclease activity potentially edits the outcome of each polymerization step. A mispaired nucleotide is excised before the next step proceeds. In *E. coli*, DNA polymerase I repairs DNA and participates in replication. Fidelity is further enhanced by an induced fit that results in a catalytically active conformation only when the complex of enzyme, DNA, and correct dNTP is formed. Helicases prepare the way for DNA replication by using ATP hydrolysis to separate the strands of the double helix.

Double-Stranded DNA Can Wrap Around Itself to Form Supercoiled Structures

A key topological property of DNA is its linking number (Lk), which is defined as the number of times one strand of DNA winds around the other in the right-hand direction when the DNA axis is constrained to lie in a plane. Molecules differing in linking number are topoisomers of one another and can be interconverted only by cutting one or both DNA strands; these reactions are catalyzed by topoisomerases. Changes in linking number generally lead to changes in both the number of turns of double helix and the number of turns of superhelix. Topoisomerase II (DNA gyrase) catalyzes the ATP-driven introduction of negative supercoils, which leads to the compaction of DNA and renders it more susceptible to unwinding. Supercoiled DNA is relaxed by topoisomerase I. The unwinding of DNA at the replication fork is catalyzed by an ATP-driven helicase.

DNA Replication of Both Strands Proceeds Rapidly from Specific Start Sites

DNA replication in *E. coli* starts at a unique origin (*oriC*) and proceeds sequentially in opposite directions. More than 20 proteins are required for replication. An ATP-driven helicase unwinds the *oriC* region to create a replication fork. At this fork, both strands of parental DNA serve as templates for the synthesis of new DNA. A short stretch of RNA formed by primase, an RNA polymerase, primes DNA synthesis. One strand of DNA (the leading strand) is synthesized continuously, whereas the other strand (the lagging strand) is synthesized discontinuously, in the form of 1-kb fragments (Okazaki fragments). Both new strands are formed simultaneously by the concerted actions of the highly processive

DNA polymerase III holoenzyme, an asymmetric dimer. The discontinuous assembly of the lagging strand enables $5' \rightarrow 3'$ polymerization at the atomic level to give rise to overall growth of this strand in the $3' \rightarrow 5'$ direction. The RNA primer stretch is hydrolyzed by the $5' \rightarrow 3'$ nuclease activity of DNA polymerase I, which also fills gaps. Finally, nascent DNA fragments are joined by DNA ligase in a reaction driven by ATP or NAD^+ cleavage.

DNA synthesis in eukaryotes is more complex than in prokaryotes. Eukaryotes require thousands of origins of replication to complete replication in a timely fashion. A special RNA-dependent DNA polymerase called telomerase is responsible for the replication of the ends of linear chromosomes.

Double-Stranded DNA Molecules with Similar Sequences Sometimes Recombine

DNA molecules that are similar in nucleotide sequences in a local region can recombine to form DNA duplexes that begin with the sequence of one molecule and continue with the sequence of the other. Recombinases catalyze the formation of these products through the formation and resolution of Holliday junctions. In these structures, four DNA strands come together to form a crosslike structure.

Recombinases cleave DNA strands and form specific adducts in which a tyrosine residue of the enzyme is linked to a $3'$ -phosphoryl group of the DNA. These intermediates then react with $5'$ -hydroxyl groups of other DNA strands to form the new phosphodiester bonds that are present in the recombination products. The reaction mechanisms of recombinases are similar to those of type I topoisomerases.

Mutations Are Produced by Several Types of Changes in the Base Sequence of DNA

Mutations are produced by mistakes in base-pairing, by the covalent modification of bases, and by the deletion and insertion of bases. The $3' \rightarrow 5'$ exonuclease activity of DNA polymerases is critical in lowering the spontaneous mutation rate, which arises in part from mispairing by tautomeric bases. Lesions in DNA are continually being repaired. Multiple repair processes utilize information present in the intact strand to correct the damaged strand. For example, pyrimidine dimers formed by the action of ultraviolet light are excised by uvrABC excinuclease, an enzyme that removes a 12-nucleotide region containing the dimer. Xeroderma pigmentosum, a genetically transmitted disease, is caused by the defective repair of lesions in DNA, such as pyrimidine dimers; patients with this disease usually develop skin cancers. Many cancers of the colon are caused by defective DNA mismatch repair arising from mutations in human genes that have been highly conserved in evolution. Damage to DNA is a fundamental event in both carcinogenesis and mutagenesis. Many potential carcinogens can be detected by their mutagenic action on bacteria.

Key Terms

B-DNA helix

A-DNA helix

major groove

minor groove

Z-DNA helix

DNA polymerase

template

primer

exonuclease

helicase

supercoil

linking number (Lk)

topoisomer

twist (Tw)

writhe (Wr)

topoisomerase

origin of replication

primase

replication fork

Okazaki fragment

lagging strand

leading strand

DNA ligase

processivity

cell cycle

telomere

telomerase

recombinase

Holliday junction

recombination synapse

transition

transversion

mutagen

direct repair

base-excision repair

nucleotide-excision repair

trinucleotide repeat

Ames test

Problems

1. *Activated intermediates.* DNA polymerase I, DNA ligase, and topoisomerase I catalyze the formation of phosphodiester bonds. What is the activated intermediate in the linkage reaction catalyzed by each of these enzymes? What is the leaving group?

See answer

2. *Fuel for a new ligase.* Whether the joining of two DNA chains by known DNA ligases is driven by NAD^+ or ATP depends on the species. Suppose that a new DNA ligase requiring a different energy donor is found. Propose a plausible substitute for NAD^+ or ATP in this reaction.

See answer

3. *Life in a hot tub.* An archaeon (*Sulfolobus acidocaldarius*) found in acidic hot springs contains a topoisomerase that catalyzes the ATP-driven introduction of positive supercoils into DNA. How might this enzyme be advantageous to this unusual organism?

See answer

4. *A cooperative transition.* The transition from B-DNA to Z-DNA occurs over a small change in the superhelix density, which shows that the transition is highly cooperative.

(a) Consider a DNA molecule at the midpoint of this transition. Are B- and Z-DNA regions frequently intermingled or are there long stretches of each?

(b) What does this finding reveal about the energetics of forming a junction between the two kinds of helices?

(c) Would you expect the transition from B- to A-DNA to be more or less cooperative than the one from B- to Z-DNA? Why?

See answer

5. *Molecular motors in replication.* (a) How fast does template DNA spin (expressed in revolutions per second) at an *E. coli* replication fork? (b) What is the velocity of movement (in micrometers per second) of DNA polymerase III holoenzyme relative to the template?

See answer

6. *Wound tighter than a drum.* Why would replication come to a halt in the absence of topoisomerase II?

See answer

7. *Telomeres and cancer.* Telomerase is not active in most human cells. Some cancer biologists have suggested that activation of the telomerase gene would be a requirement for a cell to become cancerous. Explain why this might be the case.

See answer

8. *Nick translation.* Suppose that you wish to make a sample of DNA duplex highly radioactive to use as a DNA probe. You have a DNA endonuclease that cleaves the DNA internally to generate 3'-OH and 5'-phosphate groups, intact DNA polymerase I, and radioactive dNTPs. Suggest a means for making the DNA radioactive.

See answer

9. *Revealing tracks.* Suppose that replication is initiated in a medium containing *moderately* radioactive tritiated thymine. After a few minutes of incubation, the bacteria are transferred to a medium containing *highly* radioactive tritiated thymidine. Sketch the autoradiographic pattern that would be seen for (a) unidirectional replication and (b) bidirectional replication, from a single origin.

See answer

10. *Mutagenic trail.* Suppose that the single-stranded RNA from tobacco mosaic virus was treated with a chemical mutagen, that mutants were obtained having serine or leucine instead of proline at a specific position, and that further treatment of these mutants with the same mutagen yielded phenylalanine at this position.



(a) What are the plausible codon assignments for these four amino acids?

(b) Was the mutagen 5-bromouracil, nitrous acid, or an acridine dye?

See answer

11. *Induced spectrum.* DNA photolyases convert the energy of light in the near ultraviolet or visible region (300–500 nm) into chemical energy to break the cyclobutane ring of pyrimidine dimers. In the absence of substrate, these photoreactivating enzymes do not absorb light of wavelengths longer than 300 nm. Why is the substrate-induced absorption band advantageous?

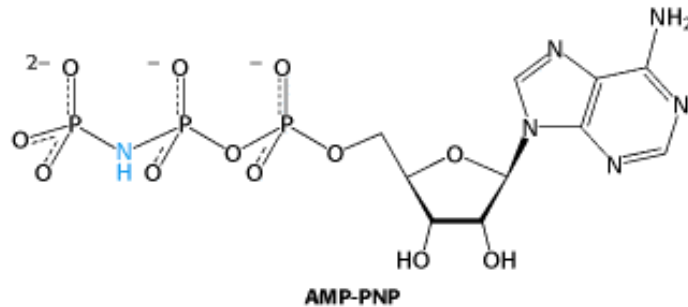
See answer

Mechanism Problems

12. *AMP-induced relaxation.* DNA ligase from *E. coli* relaxes supercoiled circular DNA in the presence of AMP but not in its absence. What is the mechanism of this reaction, and why is it dependent on AMP?

See answer

13. *A revealing analog.* AMP-PNP, the β,γ -imido analog of ATP, is hydrolyzed very slowly by most ATPases.

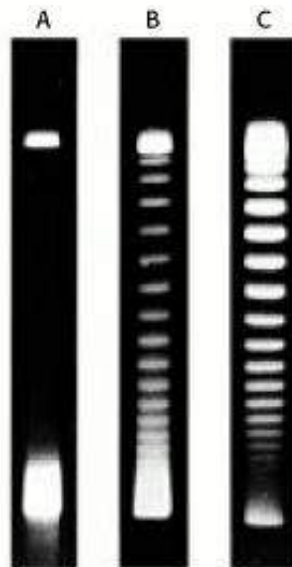


The addition of AMP-PNP to topoisomerase II and circular DNA leads to the negative supercoiling of a single molecule of DNA per enzyme. DNA remains bound to the enzyme in the presence of this analog. What does this finding reveal about the catalytic mechanism?

See answer

Data Interpretation and Chapter Integration Problems

14. *Like a ladder.* Circular DNA from SV40 virus was isolated and subjected to gel electrophoresis. The results are shown in lane A (the control) of the adjoining gel patterns.



- (a) Why does the DNA separate in agarose gel electrophoresis? How does the DNA in each band differ?

The DNA was then incubated with topoisomerase I for 5 minutes and again analyzed by gel electrophoresis with the results shown in lane B.

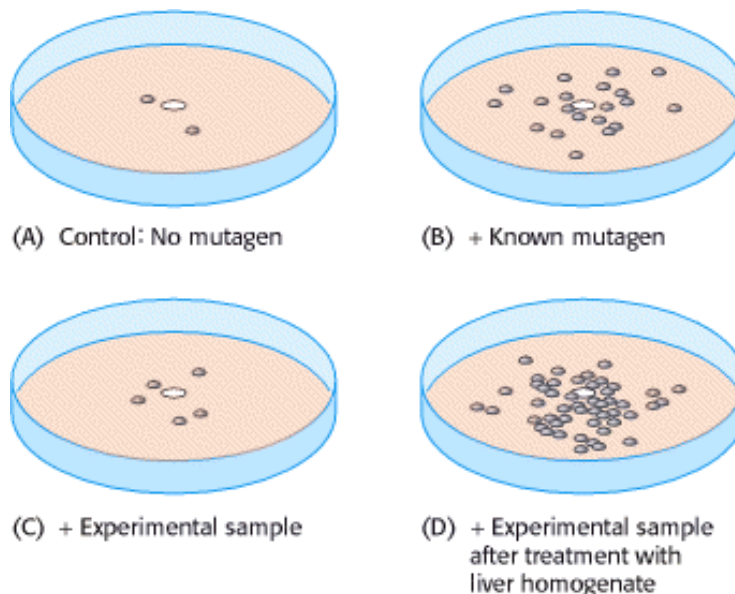
- (b) What types of DNA do the various bands represent?

Another sample of DNA was incubated with topoisomerase I for 30 minutes and again analyzed as shown in lane C.

(c) What is the significance of the fact that more of the DNA is in slower moving forms?

See answer

15. *Ames test*. The adjoining illustration shows four petri plates used for the Ames test. A piece of filter paper (white circle in the center of each plate) was soaked in one of four preparations and then placed on a petri plate. The four preparations contained




(A) purified water (control), (B) a known mutagen, (C) a chemical whose mutagenicity is under investigation, and (D) the same chemical after treatment with liver homogenate. The number of revertants, visible as colonies on the petri plates, was determined in each case.

- (a) What was the purpose of the control plate, which was exposed only to water?
- (b) Why was it wise to use a known mutagen in the experimental system?
- (c) How would you interpret the results obtained with the experimental compound?
- (d) What liver components would you think are responsible for the effects observed part D?

See answer

Media Problem

16.  *Cre-ative duplexes*. Site specific recombinases like Cre require that the sequence between the recombinase binding sites be identical in the two DNA molecules that are to be recombined. Examine the structures of the various Cre-DNA complexes in the **Structural Insights** module on Cre recombinase. How might the recombination mechanism check these sequences to ensure they are identical in the parent DNA molecules?

Selected Readings

Where to begin

- A. Kornberg. 1988. DNA replication *J. Biol. Chem.* 263: 1-4. ([PubMed](#))
- R.E. Dickerson. 1983. The DNA helix and how it is read *Sci. Am.* 249: (6) 94-111.
- J.C. Wang. 1982. DNA topoisomerases *Sci. Am.* 247: (1) 94-109. ([PubMed](#))
- T. Lindahl. 1993. Instability and decay of the primary structure of DNA *Nature* 362: 709-715. ([PubMed](#))
- C.W. Greider and E.H. Blackburn. 1996. Telomeres, telomerase, and cancer *Sci. Am.* 274: (2) 92-97. ([PubMed](#))

Books

- Kornberg, A., and Baker, T. A., 1992. *DNA Replication* (2d ed.). W. H. Freeman and Company.
- Bloomfield, V. A., Crothers, D., Tinoco, I., and Hearst, J., 2000. *Nucleic Acids: Structures, Properties and Functions*. University Science Books.
- Friedberg, E. C., Walker, G. C., Siede, W., 1995. *DNA Repair and Mutagenesis*. American Society for Microbiology.
- Cozzarelli, N. R., and Wang, J. C. (Eds.), 1990. *DNA Topology and Its Biological Effects*. Cold Spring Harbor Laboratory Press.

DNA structure

- T.K. Chiu and R.E. Dickerson. 2000. 1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium *J. Mol. Biol.* 301: 915-945. ([PubMed](#))
- A. Herbert and A. Rich. 1999. Left-handed Z-DNA: Structure and Function *Genetica* 106: 37-47. ([PubMed](#))
- R.E. Dickerson. 1992. DNA Structure from A to Z *Methods Enzymol* 211: 67-111. ([PubMed](#))
- J.R. Quintana, K. Grzeskowiak, K. Yanagi, and R.E. Dickerson. 1992. Structure of a B-DNA decamer with a central T-A step: C-G-A-T-T-A-A-T-C-G *J. Mol. Biol.* 225: 379-395. ([PubMed](#))
- N. Verdaguer, J. Aymami, F.D. Fernandez, I. Fita, M. Coll, D.T. Huynh, J. Igolen, and J.A. Subirana. 1991. Molecular structure of a complete turn of A-DNA *J. Mol. Biol.* 221: 623-635. ([PubMed](#))

DNA topology and topoisomerases

- D. Sikder, S. Unniraman, T. Bhaduri, and V. Nagaraja. 2001. Functional cooperation between topoisomerase I and single strand DNA-binding protein *J. Mol. Biol.* 306: 669-679. ([PubMed](#))
- Z. Yang and J.J. Champoux. 2001. The role of histidine 632 in catalysis by human topoisomerase I *J. Biol. Chem.* 276: 677-685. ([PubMed](#))
- J.M. Fortune and N. Osheroff. 2000. Topoisomerase II as a target for anticancer drugs: When enzymes stop being nice *Prog. Nucleic Acid Res. Mol. Biol.* 64: 221-253. ([PubMed](#))
- R.J. Isaacs, S.L. Davies, M.I. Sandri, C. Redwood, N.J. Wells, and I.D. Hickson. 1998. Physiological regulation of eukaryotic topoisomerase II *Biochim. Biophys. Acta* 1400: 121-137. ([PubMed](#))

J.C. Wang. 1996. DNA topoisomerases *Annu. Rev. Biochem.* 65: 635-692. ([PubMed](#))

J.C. Wang. 1998. Moving one DNA double helix through another by a type II DNA topoisomerase: The story of a simple molecular machine *Q. Rev. Biophys.* 31: 107-144. ([PubMed](#))

C.L. Baird, T.T. Harkins, S.K. Morris, and J.E. Lindsley. 1999. Topoisomerase II drives DNA transport by hydrolyzing one ATP *Proc. Natl. Acad. Sci. USA* 96: 13685-13690. ([PubMed](#)) ([Full Text in PMC](#))

A.V. Vologodskii, S.D. Levene, K.V. Klenin, K.M. Frank, and N.R. Cozzarelli. 1992. Conformational and thermodynamic properties of supercoiled DNA *J. Mol. Biol.* 227: 1224-1243. ([PubMed](#))

L.M. Fisher, C.A. Austin, R. Hopewell, M. Margerrison, M. Oram, S. Patel, D.B. Wigley, G.J. Davies, E.J. Dodson, A. Maxwell, and G. Dodson. 1991. Crystal structure of an N-terminal fragment of the DNA gyrase B protein *Nature* 351: 624-629. ([PubMed](#))

Mechanism of replication

M.J. Davey and M. O'Donnell. 2000. Mechanisms of DNA replication *Curr. Opin. Chem. Biol.* 4: 581-586. ([PubMed](#))

J.L. Keck and J.M. Berger. 2000. DNA replication at high resolution *Chem. Biol.* 7: R63-R71. ([PubMed](#))

T.A. Kunkel and K. Bebenek. 2000. DNA replication fidelity *Annu. Rev. Biochem.* 69: 497-529. ([PubMed](#))

S. Waga and B. Stillman. 1998. The DNA replication fork in eukaryotic cells *Annu. Rev. Biochem.* 67: 721-751. ([PubMed](#))

K.J. Marians. 1992. Prokaryotic DNA replication *Annu. Rev. Biochem.* 61: 673-719. ([PubMed](#))

DNA polymerases and other enzymes of replication

U. Hubscher, H.P. Nasheuer, and J.E. Syvaoja. 2000. Eukaryotic DNA polymerases: A growing family *Trends Biochem. Sci.* 25: 143-147. ([PubMed](#))

S. Doublié, S. Tabor, A.M. Long, C.C. Richardson, and T. Ellenberger. 1998. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution *Nature* 391: 251-258. ([PubMed](#))

B. Arezi and R.D. Kuchta. 2000. Eukaryotic DNA primase *Trends Biochem. Sci.* 25: 572-576. ([PubMed](#))

J. Jager and J.D. Pata. 1999. Getting a grip: Polymerases and their substrate complexes *Curr. Opin. Struct. Biol.* 9: 21-28. ([PubMed](#))

T.A. Steitz. 1999. DNA polymerases: Structural diversity and common mechanisms *J. Biol. Chem.* 274: 17395-17398. ([PubMed](#))

L.S. Beese, V. Derbyshire, and T.A. Steitz. 1993. Structure of DNA polymerase I Klenow fragment bound to duplex DNA *Science* 260: 352-355. ([PubMed](#))

C.S. McHenry. 1991. DNA polymerase III holoenzyme: Components, structure, and mechanism of a true replicative complex *J. Biol. Chem.* 266: 19127-19130. ([PubMed](#))

X.P. Kong, R. Onrust, M. O'Donnell, and J. Kuriyan. 1992. Three-dimensional structure of the beta subunit of *E. coli* DNA polymerase III holoenzyme: A sliding DNA clamp *Cell* 69: 425-437. ([PubMed](#))

A.H. Polesky, T.A. Steitz, N.D. Grindley, and C.M. Joyce. 1990. Identification of residues critical for the polymerase activity of the Klenow fragment of DNA polymerase I from *Escherichia coli* *J. Biol. Chem.* 265: 14579-14591. ([PubMed](#))

J.Y. Lee, C. Chang, H.K. Song, J. Moon, J.K. Yang, H.K. Kim, S.T. Kwon, and S.W. Suh. 2000. Crystal structure of NAD(+)-dependent DNA ligase: Modular architecture and functional implications *EMBO J.* 19: 1119-1129. ([PubMed](#))

D.J. Timson and D.B. Wigley. 1999. Functional domains of an NAD⁺-dependent DNA ligase *J. Mol. Biol.* 285: 73-83. ([PubMed](#))

A.J. Doherty and D.B. Wigley. 1999. Functional domains of an ATP-dependent DNA ligase *J. Mol. Biol.* 285: 63-71. ([PubMed](#))

P.H. von Hippel and E. Delagoutte. 2001. A general model for nucleic acid helicases and their "coupling" within macromolecular machines *Cell* 104: 177-190. ([PubMed](#))

B.K. Tye and S. Sawyer. 2000. The hexameric eukaryotic MCM helicase: Building symmetry from nonidentical parts *J. Biol. Chem.* 275: 34833-34836. ([PubMed](#))

K.J. Marians. 2000. Crawling and wiggling on DNA: Structural insights to the mechanism of DNA unwinding by helicases *Structure Fold Des.* 5: R227-R235.

P. Soultanas and D.B. Wigley. 2000. DNA helicases: 'Inching forward' *Curr. Opin. Struct. Biol.* 10: 124-128. ([PubMed](#))

F. Bachand and C. Autexier. 2001. Functional regions of human telomerase reverse transcriptase and human telomerase RNA required for telomerase activity and RNA-protein interactions *Mol. Cell Biol.* 21: 1888-1897. ([PubMed](#)) ([Full Text in PMC](#))

T.M. Bryan and T.R. Cech. 1999. Telomerase and the maintenance of chromosome ends *Curr. Opin. Cell Biol.* 11: 318-324. ([PubMed](#))

J.D. Griffith, L. Comeau, S. Rosenfield, R.M. Stansel, A. Bianchi, H. Moss, and T. de Lange. 1999. Mammalian telomeres end in a large duplex loop *Cell* 97: 503-514. ([PubMed](#))

M.J. McEachern, A. Krauskopf, and E.H. Blackburn. 2000. Telomeres and their control *Annu. Rev. Genet.* 34: 331-358. ([PubMed](#))

Recombinases

G.D. Van Duyne. 2001. A structural view of cre-loxp site-specific recombination *Annu. Rev. Biophys. Biomol. Struct.* 30: 87-104. ([PubMed](#))

Y. Chen, U. Narendra, L.E. Iype, M.M. Cox, and P.A. Rice. 2000. Crystal structure of a Flp recombinase-Holliday junction complex: Assembly of an active oligomer by helix swapping *Mol. Cell* 6: 885-897. ([PubMed](#))

N.L. Craig. 1997. Target site selection in transposition *Annu. Rev. Biochem.* 66: 437-474. ([PubMed](#))

D.N. Gopaul, F. Guo, and G.D. Van Duyne. 1998. Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination *EMBO J.* 17: 4175-4187. ([PubMed](#))

D.N. Gopaul and G.D. Duyne. 1999. Structure and mechanism in site-specific recombination *Curr. Opin. Struct. Biol.* 9: 14-20. ([PubMed](#))

Mutations and DNA repair

R.J. Michelson and T. Weinert. 2000. Closing the gaps among a web of DNA repair disorders *Bioessays* 22: 966-969. ([PubMed](#))

L. Aravind, D.R. Walker, and E.V. Koonin. 1999. Conserved domains in DNA repair proteins and evolution of repair systems *Nucleic Acids Res.* 27: 1223-1242. ([PubMed](#)) ([Full Text in PMC](#))

C.D. Mol, S.S. Parikh, C.D. Putnam, T.P. Lo, and J.A. Tainer. 1999. DNA repair mechanisms for the recognition and removal of damaged DNA bases *Annu. Rev. Biophys. Biomol. Struct.* 28: 101-128. ([PubMed](#))

S.S. Parikh, C.D. Mol, and J.A. Tainer. 1997. Base excision repair enzyme family portrait: Integrating the structure and chemistry of an entire DNA repair pathway *Structure* 5: 1543-1550. ([PubMed](#))

D.G. Vassilyev and K. Morikawa. 1997. DNA-repair enzymes *Curr. Opin. Struct. Biol.* 7: 103-109. ([PubMed](#))

G.L. Verdine and S.D. Bruner. 1997. How do DNA repair proteins locate damaged bases in the genome? *Chem. Biol.* 4: 329-334. ([PubMed](#))

R.P. Bowater and R.D. Wells. 2000. The intrinsically unstable life of DNA triplet repeats associated with human hereditary disorders *Prog. Nucleic Acid Res. Mol. Biol.* 66: 159-202. ([PubMed](#))

C.J. Cummings and H.Y. Zoghbi. 2000. Fourteen and counting: Unraveling trinucleotide repeat diseases *Hum. Mol. Genet.* 9: 909-916. ([PubMed](#))

Defective DNA repair and cancer

M. Berneburg and A.R. Lehmann. 2001. Xeroderma pigmentosum and related disorders: Defects in DNA repair and transcription *Adv. Genet.* 43: 71-102. ([PubMed](#))

M.W. Lambert and W.C. Lambert. 1999. DNA repair and chromatin structure in genetic diseases *Prog. Nucleic Acid Res. Mol. Biol.* 63: 257-310. ([PubMed](#))

C.H. Buys. 2000. Telomeres, telomerase, and cancer *N. Engl. J. Med.* 342: 1282-1283. ([PubMed](#))

V. Urquidi, D. Tarin, and S. Goodison. 2000. Role of telomerase in cell senescence and oncogenesis *Annu. Rev. Med.* 51: 65-79. ([PubMed](#))

H.T. Lynch, T.C. Smyrk, P. Watson, S.J. Lanspa, J.F. Lynch, P.M. Lynch, R.J. Cavalieri, and C.R. Boland. 1993. Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: An updated review *Gastroenterology* 104: 1535-1549. ([PubMed](#))

R. Fishel, M.K. Lescoe, M.R.S. Rao, N.G. Copeland, N.A. Jenkins, J. Garber, M. Kane, and R. Kolodner. 1993. The human mutator gene homolog *MSH2* and its association with hereditary nonpolyposis colon cancer *Cell* 75: 1027-1038. ([PubMed](#))

B.N. Ames and L.S. Gold. 1991. Endogenous mutagens and the causes of aging and cancer *Mutat. Res.* 250: 3-16. ([PubMed](#))

B.N. Ames. 1979. Identifying environmental chemicals causing mutations and cancer *Science* 204: 587-593. ([PubMed](#))

28. RNA Synthesis and Splicing

DNA stores genetic information in a stable form that can be readily replicated. However, the expression of this genetic information requires its flow from DNA to RNA to protein, as was introduced in [Chapter 5](#). The present chapter deals with how RNA is synthesized and spliced. We begin with transcription in *Escherichia coli* and focus on three questions: What are the properties of promoters (the DNA sites at which RNA transcription is initiated), and how do the promoters function? How do RNA polymerase, the DNA template, and the nascent RNA chain interact with one another? How is transcription terminated?

We then turn to transcription in eukaryotes, beginning with promoter structure and the transcription-factor proteins that regulate promoter activity. A distinctive feature of eukaryotic DNA templates is the presence of enhancer sequences that

can stimulate transcriptional initiation more than a thousand base pairs away from the start site. Primary transcripts in eukaryotes are extensively modified, as exemplified by the capping of the 5' end of an mRNA precursor and the addition of a long poly(A) tail to its 3' end. Most striking is the splicing of mRNA precursors, which is catalyzed by spliceosomes consisting of small nuclear ribonucleoprotein particles (snRNPs). The small nuclear RNA (snRNA) molecules in these complexes play a key role in directing the alignment of splice sites and in mediating catalysis. Indeed, some RNA molecules can splice themselves in the absence of protein. This landmark discovery by Thomas Cech and Sidney Altman revealed that RNA molecules can serve as catalysts and greatly influenced our view of molecular evolution.

RNA splicing is not merely a curiosity. Approximately 15% of all genetic diseases are caused by mutations that affect RNA splicing. Moreover, the same pre-mRNA can be spliced differently in various cell types, at different stages of development, or in response to other biological signals. In addition, individual bases in some pre-mRNA molecules are changed, in a process called *RNA editing*. One of the biggest surprises of the sequencing of the human genome was that only approximately 40,000 genes were identified compared with previous estimates of 100,000 or more. The ability of one gene to encode more than one distinct mRNA and, hence, more than one protein may play a key role in expanding the repertoire of our genomes.

28.0.1. An Overview of RNA Synthesis

RNA synthesis, or *transcription*, is the process of transcribing DNA nucleotide sequence information into RNA sequence information. RNA synthesis is catalyzed by a large enzyme called *RNA polymerase*. The basic biochemistry of RNA synthesis is common to prokaryotes and eukaryotes, although its regulation is more complex in eukaryotes. The close connection between prokaryotic and eukaryotic transcription has been beautifully illustrated by the recently determined three-dimensional structures of representative RNA polymerases from prokaryotes and eukaryotes ([Figure 28.1](#)). Despite substantial differences in size and number of polypeptide subunits, the overall structures of these enzymes are quite similar, revealing a common evolutionary origin.

RNA synthesis, like nearly all biological polymerization reactions, takes place in three stages: *initiation*, *elongation*, and *termination*. RNA polymerase performs multiple functions in this process:

1. It searches DNA for initiation sites, also called *promoter sites* or simply *promoters*. For instance, *E. coli* DNA has about 2000 promoter sites in its 4.8×10^6 bp genome. Because these sequences are on the *same* molecule of DNA as the genes being transcribed, they are called *cis-acting elements*.
2. It unwinds a short stretch of double-helical DNA to produce a single-stranded DNA template from which it takes instructions.
3. It selects the correct ribonucleoside triphosphate and catalyzes the formation of a phosphodiester bond. This process is repeated many times as the enzyme moves unidirectionally along the DNA template. RNA polymerase is completely processive—a transcript is synthesized from start to end by a single RNA polymerase molecule.
4. It detects termination signals that specify where a transcript ends.
5. It interacts with activator and repressor proteins that modulate the rate of transcription initiation over a wide dynamic range. These proteins, which play a more prominent role in eukaryotes than in prokaryotes, are called *transcription factors* or *trans-acting elements*. Gene expression is controlled mainly at the level of transcription, as will be discussed in detail in [Chapter 31](#).

The fundamental reaction of RNA synthesis is the formation of a phosphodiester bond. The 3'-hydroxyl group of the last nucleotide in the chain nucleophilically attacks the α -phosphate group of the incoming nucleoside triphosphate with the concomitant release of a pyrophosphate (see [Figure 5.25](#)). This reaction is thermodynamically favorable, and the subsequent degradation of the pyrophosphate to orthophosphate locks the reaction in the direction of RNA synthesis.

The chemistry of RNA synthesis is identical for all forms of RNA, including messenger RNA, transfer RNA, and ribosomal RNA. The basic steps just outlined also apply to all forms. Their synthetic processes differ mainly in regulation, posttranscriptional processing, and the specific polymerase that participates.

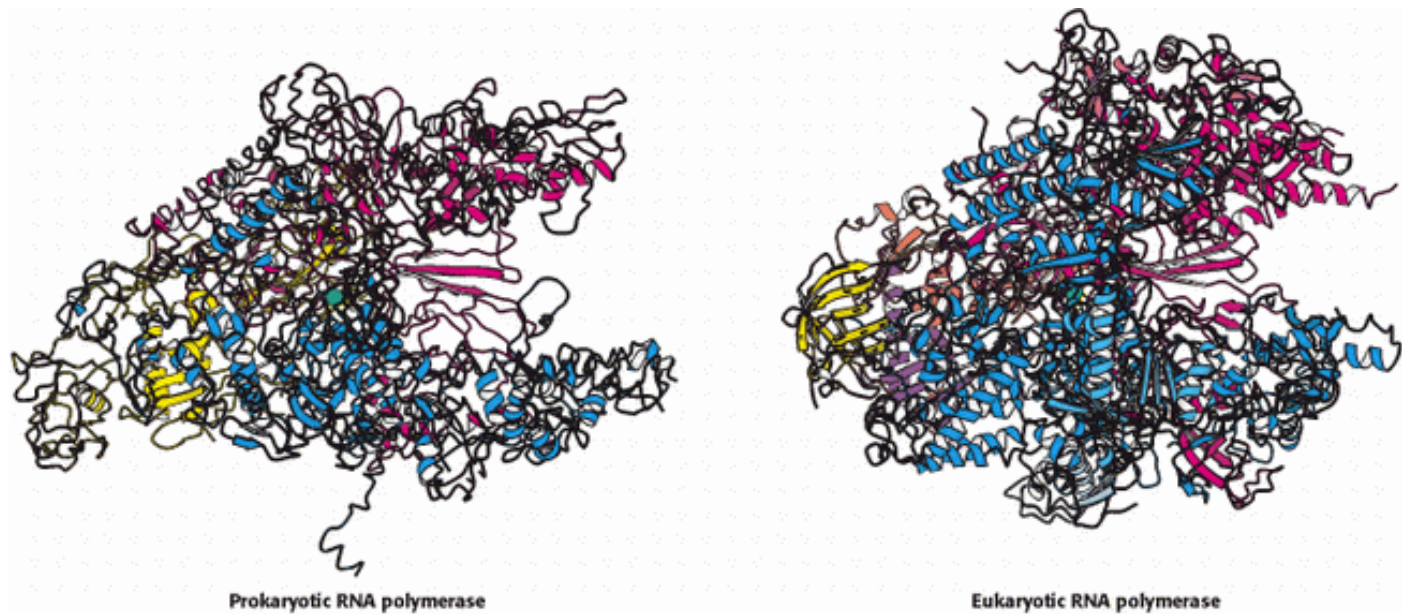
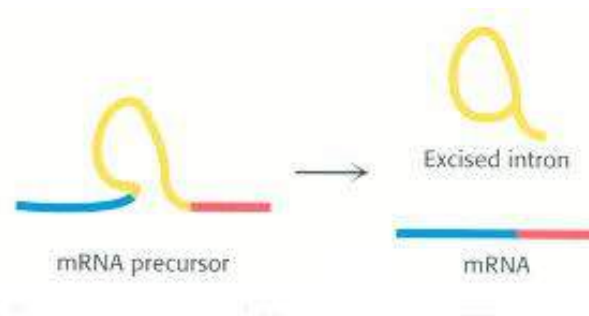
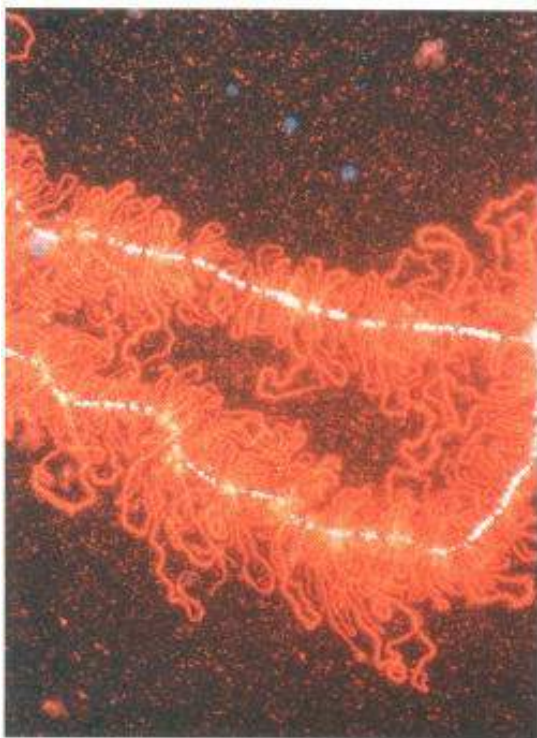


Figure 28.1. RNA Polymerase Structures. The three-dimensional structures of RNA polymerases from a prokaryote (*Thermus aquaticus*) and a eukaryote (*Saccharomyces cerevisiae*). The two largest subunits for each structure are shown in dark red and dark blue. The similarity of these structures reveals that these enzymes have the same evolutionary origin and have many mechanistic features in common.



RNA synthesis is a key step in the expression of genetic information. For eukaryotic cells, the initial RNA transcript (the mRNA precursor) is often spliced, removing introns that do not encode protein sequences. Often, the same pre-mRNA is spliced differently in different cell types or at different developmental stages. In the image at the left, proteins

associated with RNA splicing (stained with a fluorescent antibody) highlight regions of the new genome that are being actively transcribed. [(Left) courtesy of Mark B. Roth and Joseph G. Gall.]

28.1. Transcription Is Catalyzed by RNA Polymerase

We begin our consideration of transcription by examining the process in bacteria such as *E. coli*. RNA polymerase from *E. coli* is a very large (~400 kd) and complex enzyme consisting of four kinds of subunits (Table 28.1). The subunit composition of the entire enzyme, called the *holoenzyme*, is $\alpha_2 \beta \beta' \sigma$. The σ subunit helps find a promoter site where transcription begins, participates in the initiation of RNA synthesis, and then dissociates from the rest of the enzyme. RNA polymerase without this subunit ($\alpha_2 \beta \beta'$) is called the *core enzyme*. The core enzyme contains the catalytic site.

This catalytic site resembles that of DNA polymerase (Section 27.2.2) in that it includes two metal ions in its active form (Figure 28.2). One metal ion remains bound to the enzyme, whereas the other appears to come in with the nucleoside triphosphate and leave with the pyrophosphate. Three conserved aspartate residues of the enzyme participate in binding these metal ions. Note that the overall structures of DNA polymerase and RNA polymerase are quite different; their similar active sites are the products of convergent evolution.

28.1.1. Transcription Is Initiated at Promoter Sites on the DNA Template

Transcription starts at *promoters* on the DNA template. *Promoters are sequences of DNA that direct the RNA polymerase to the proper initiation site for transcription.* Promoter sites can be identified and characterized by a combination of techniques. One powerful technique for characterizing these and other protein-binding sites on DNA is called *footprinting* (Figure 28.3). First, one of the strands of a DNA fragment under investigation is labeled on one end with ^{32}P . RNA polymerase is added to the labeled DNA, and *the complex is digested with DNase just long enough to make an average of one cut in each chain.* A part of the radioactive DNA is treated in the same way but without the addition of RNA polymerase to serve as a control. The resulting DNA fragments are separated according to size by electrophoresis. The gel pattern is highly revealing: a series of bands present in the control sample is absent from the sample containing RNA polymerase. These bands are missing because RNA polymerase shields DNA from cleavages that would give rise to the corresponding fragments.

A striking pattern is evident when the sequences of many prokaryotic promoters are compared. *Two common motifs are present on the 5' (upstream) side of the start site.* They are known as the *-10 sequence* and the *-35 sequence* because they are centered at about 10 and 35 nucleotides upstream of the start site. These sequences are each 6 bp long. Their *consensus (average) sequences*, deduced from analyses of many promoters (Figure 28.4), are



The first nucleotide (the start site) of a transcribed DNA sequence is denoted as +1 and the second one as +2; the nucleotide preceding the start site is denoted as -1. These designations refer to the coding strand of DNA. Recall that the sequence of the *template strand of DNA* is the *complement* of that of the RNA transcript (see Figure 5.26). In contrast, the *coding strand of DNA* has the *same* sequence as that of the RNA transcript except for thymine (T) in place of uracil (U). The coding strand is also known as the *sense (+) strand*, and the template strand as the *antisense (-) strand*.

Promoters differ markedly in their efficacy. Genes with strong promoters are transcribed frequently— as often as every 2 seconds in *E. coli*. In contrast, genes with very weak promoters are transcribed about once in 10 minutes. The -10 and -

35 regions of most strong promoters have sequences that correspond closely to the consensus sequences, whereas weak promoters tend to have multiple substitutions at these sites. Indeed, mutation of a single base in either the -10 sequence or the -35 sequence can diminish promoter activity. The distance between these conserved sequences also is important; a separation of 17 nucleotides is optimal. Thus, *the efficiency or strength of a promoter sequence serves to regulate transcription*. Regulatory proteins that bind to specific sequences near promoter sites and interact with RNA polymerase (Chapter 31) also markedly influence the frequency of transcription of many genes.

28.1.2. Sigma Subunits of RNA Polymerase Recognize Promoter Sites

The $\alpha_2 \beta \beta'$ core of RNA polymerase is unable to start transcription at promoter sites. Rather, the complete $\alpha_2 \beta \beta' \sigma$ holoenzyme is essential for initiation at the correct start site. The σ subunit contributes to specific initiation in two ways. First, *it decreases the affinity of RNA polymerase for general regions of DNA by a factor of 10^4* . In its absence, the core enzyme binds DNA indiscriminately and tightly. Second, the σ subunit enables *RNA polymerase to recognize promoter sites*. A large fragment of a σ subunit was found to have an α helix on its surface; this helix has been implicated in recognizing the 5'-TATAAT sequence of the -10 region (Figure 28.5). The holoenzyme binds to duplex DNA and moves along the double helix in search of a promoter, forming transient hydrogen bonds with exposed hydrogen-donor and -acceptor groups on the base pairs. The search is rapid because RNA polymerase slides along DNA instead of repeatedly binding and dissociating from it. *In other words, the promoter site is encountered by a random walk in one dimension rather than in three dimensions*. The observed rate constant for the binding of RNA polymerase holoenzyme to promoter sequences is $10^{10} \text{ M}^{-1} \text{ s}^{-1}$, more than 100 times as fast as could be accomplished by repeated encounters moving on and off the DNA. The σ subunit is released when the nascent RNA chain reaches nine or ten nucleotides in length. After its release, it can assist initiation by another core enzyme. Thus, *the σ subunit acts catalytically*.

E. coli contains multiple σ factors to recognize several types of promoter sequences contained in *E. coli* DNA. The type that recognizes the consensus sequences described earlier is called σ^{70} because it has a mass of 70 kd. A different σ factor comes into play when the temperature is raised abruptly. *E. coli* responds by synthesizing σ^{32} , which recognizes the promoters of *heat-shock genes*. These promoters exhibit -10 sequences that are somewhat different from the -10 sequence for standard promoters (Figure 28.6). The increased transcription of heat-shock genes leads to the coordinated synthesis of a series of protective proteins. Other σ factors respond to environmental conditions, such as nitrogen starvation. These findings demonstrate that *σ plays a key role in determining where RNA polymerase initiates transcription*.

28.1.3. RNA Polymerase Must Unwind the Template Double Helix for Transcription to Take Place

Although RNA polymerase can search for promoter sites when bound to double-helical DNA, a segment of the helix must be unwound before synthesis can begin. A region of duplex DNA must be unpaired so that nucleotides on one of its strands become accessible for base-pairing with incoming ribonucleoside triphosphates. The DNA template strand selects the correct ribonucleoside triphosphate by forming a Watson-Crick base pair with it (Section 5.2.1), as in DNA synthesis.

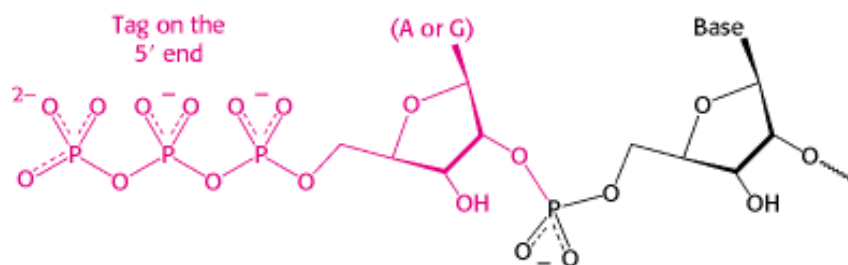
How much of the template DNA is unwound by the polymerase? Because unwinding increases the negative supercoiling of the DNA (Section 27.3.2), this question was answered by analyzing the supercoiling of a circular duplex DNA exposed to varying amounts of RNA polymerase. Topoisomerase I, an enzyme catalyzing the concerted cleavage and resealing of duplex DNA (Section 27.3.3), was then added to relax the part of circular DNA not in contact with polymerase molecules. These DNA samples were analyzed by gel electrophoresis after the removal of bound protein. *The degree of negative supercoiling increased in proportion to the number of RNA polymerase molecules bound per template DNA, showing that the enzyme unwinds DNA. Each bound polymerase molecule unwinds a 17-bp segment of DNA, which corresponds to 1.6 turns of B-DNA helix (Figure 28.7)*.

Negative supercoiling of circular DNA favors the transcription of genes because it facilitates unwinding ([Section 27.3.2](#)). Thus, the introduction of negative supercoils into DNA by topoisomerase II can increase the efficiency of promoters located at distant sites. However, not all promoter sites are stimulated by negative supercoiling. The promoter site for topoisomerase II itself is a noteworthy exception. Negative supercoiling decreases the rate of transcription of this gene, an elegant feedback control ensuring that DNA does not become excessively supercoiled. Negative supercoiling could decrease the efficiency of this promoter by changing the structural relation of the -10 and -35 regions.

The transition from the *closed promoter complex* (in which DNA is double helical) to the *open promoter complex* (in which a DNA segment is unwound) is an essential event in transcription. The stage is now set for the formation of the first phosphodiester bond of the new RNA chain.

28.1.4. RNA Chains Are Formed de Novo and Grow in the 5'-to-3' Direction

In contrast with DNA synthesis, *RNA synthesis can start de novo, without the requirement for a primer*. Most newly synthesized RNA chains carry a highly distinctive tag on the 5' end: the first base at that end is either *pppG* or *pppA*.



The presence of the triphosphate moiety suggests that RNA synthesis starts at the 5' end. The results of labeling experiments with γ - ^{32}P substrates confirmed that RNA chains, like DNA chains, grow in the 5' \rightarrow 3' direction.



28.1.5. Elongation Takes Place at Transcription Bubbles That Move Along the DNA Template

The elongation phase of RNA synthesis begins after the formation of the first phosphodiester bond. An important change is the loss of σ ; recall that the core enzyme without σ binds more strongly to the DNA template. Indeed, RNA polymerase stays bound to its template until a termination signal is reached. The region containing RNA polymerase, DNA, and nascent RNA is called a *transcription bubble* because it contains a locally melted "bubble" of DNA ([Figure 28.8](#)). The newly synthesized RNA forms a hybrid helix with the template DNA strand. This RNA-DNA helix is about 8 bp long, which corresponds to nearly one turn of a double helix ([Section 27.1.3](#)). The 3'-hydroxyl group of the RNA in this hybrid helix is positioned so that it can attack the α -phosphorus atom of an incoming ribonucleoside triphosphate. The core enzyme also contains a binding site for the other DNA strand. About 17 bp of DNA are unwound throughout the elongation phase, as in the initiation phase. The transcription bubble moves a distance of 170 Å (17 nm) in a second, which corresponds to a rate of elongation of about 50 nucleotides per second. Although rapid, it is much slower than the rate of DNA synthesis, which is 800 nucleotides per second.

The lengths of the RNA-DNA hybrid and of the unwound region of DNA stay rather constant as RNA polymerase moves along the DNA template. This finding indicates that DNA is rewound at about the same rate at the rear of RNA polymerase as it is unwound at the front of the enzyme. The RNA-DNA hybrid must also rotate each time a nucleotide is added so that the 3'-OH end of the RNA stays at the catalytic site. The length of the RNA-DNA hybrid is determined by a structure within the enzyme that forces the RNA-DNA hybrid to separate, allowing the RNA chain to exit from the enzyme and the DNA chain to rejoin its DNA partner (Figure 28.9).

It is noteworthy that RNA polymerase lacks nuclease activity. Thus, in contrast with DNA polymerase, *RNA polymerase does not correct the nascent polynucleotide chain. Consequently, the fidelity of transcription is much lower than that of replication.* The error rate of RNA synthesis is of the order of one mistake per 10^4 or 10^5 nucleotides, about 10^5 times as high as that of DNA synthesis. The much lower fidelity of RNA synthesis can be tolerated because mistakes are not transmitted to progeny. For most genes, many RNA transcripts are synthesized; a few defective transcripts are unlikely to be harmful.

28.1.6. An RNA Hairpin Followed by Several Uracil Residues Terminates the Transcription of Some Genes

The termination of transcription is as precisely controlled as its initiation. In the termination phase of transcription, the formation of phosphodiester bonds ceases, the RNA-DNA hybrid dissociates, the melted region of DNA rewinds, and RNA polymerase releases the DNA. What determines where transcription is terminated? *The transcribed regions of DNA templates contain stop signals.* The simplest one is a *palindromic GC-rich region followed by an AT-rich region.* The RNA transcript of this DNA palindrome is self-complementary (Figure 28.10). Hence, its bases can pair to form a hairpin structure with a stem and loop, a structure favored by its high content of G and C residues. Guanine-cytosine base pairs are more stable than adenine-thymine pairs because of the extra hydrogen bond in the base pair. This stable hairpin is followed by a sequence of four or more uracil residues, which also are crucial for termination. The RNA transcript ends within or just after them.

How does this combination hairpin-oligo(U) structure terminate transcription? First, it seems likely that RNA polymerase pauses immediately after it has synthesized a stretch of RNA that folds into a hairpin. Furthermore, the RNA-DNA hybrid helix produced after the hairpin is unstable because its rU-dA base pairs are the weakest of the four kinds. Hence, the pause in transcription caused by the hairpin permits the weakly bound *nascent RNA to dissociate from the DNA template and then from the enzyme.* The solitary DNA template strand rejoins its partner to re-form the DNA duplex, and the transcription bubble closes.

28.1.7. The Rho Protein Helps Terminate the Transcription of Some Genes

RNA polymerase needs no help to terminate transcription at a hairpin followed by several U residues. At other sites, however, termination requires the participation of an additional factor. This discovery was prompted by the observation that some RNA molecules synthesized in vitro by RNA polymerase acting alone are *longer* than those made in vivo. The missing factor, a protein that caused the correct termination, was isolated and named *rho* (ρ). Additional information about the action of ρ was obtained by adding this termination factor to an incubation mixture at various times after the initiation of RNA synthesis (Figure 28.11). RNAs with sedimentation coefficients of 10S, 13S, and 17S were obtained when ρ was added at initiation, a few seconds after initiation, and 2 minutes after initiation, respectively. If no ρ was added, transcription yielded a 23S RNA product. It is evident that the template contains at least three termination sites that respond to ρ (yielding 10S, 13S, and 17S RNA) and one termination site that does not (yielding 23S RNA). Thus, specific termination at a site producing 23S RNA can occur in the absence of ρ . However, ρ detects additional termination signals that are not recognized by RNA polymerase alone.

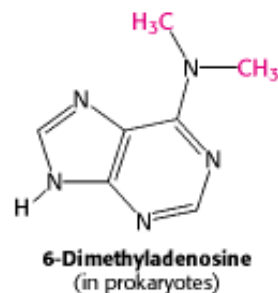
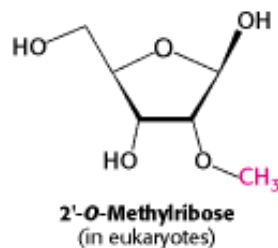
How does ρ provoke the termination of RNA synthesis? *A key clue is the finding that ρ hydrolyzes ATP in the presence of single-stranded RNA but not in the presence of DNA or duplex RNA.* Hexameric ρ , which is structurally similar to and homologous to ATP synthase (Section 18.4.1), specifically binds single-stranded RNA; a stretch of 72 nucleotides is

bound in such a way that the RNA passes through the center of the structure (Figure 28.12). *Rho* is brought into action by sequences located in the nascent RNA that are rich in cytosine and poor in guanine. The ATPase activity of ρ enables the protein to pull the nascent RNA while pursuing RNA polymerase. When ρ catches RNA polymerase at the transcription bubble, it breaks the RNA-DNA hybrid helix by functioning as an RNA-DNA helicase. Given the structural and evolutionary connection, it is possible that the mechanism of action of ρ is similar to that of ATP synthase, with the single-stranded RNA playing the role of the γ subunit.

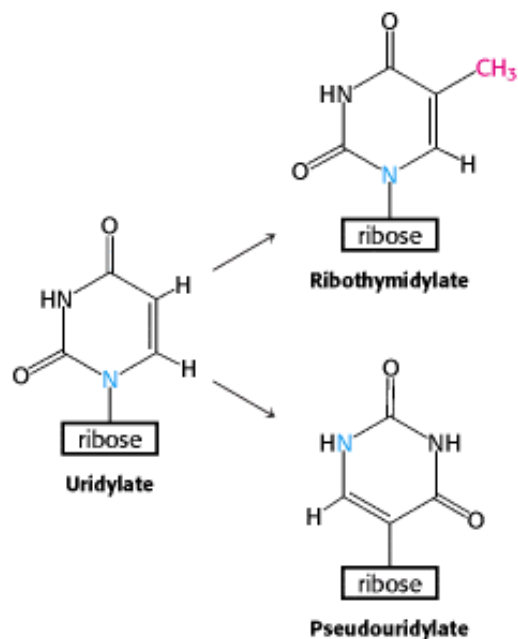
Proteins in addition to ρ mediate and modulate termination. For example, the *nusA* protein enables RNA polymerase in *E. coli* to recognize a characteristic class of termination sites. In *E. coli*, specialized termination signals called *attenuators* are regulated to meet the nutritional needs of the cell (Section 31.4.1). A common feature of protein-independent and protein-dependent termination is that the functioning signals lie in newly synthesized RNA rather than in the DNA template.

28.1.8. Precursors of Transfer and Ribosomal RNA Are Cleaved and Chemically Modified After Transcription

In prokaryotes, messenger RNA molecules undergo little or no modification after synthesis by RNA polymerase. Indeed, many mRNA molecules are translated while they are being transcribed. In contrast, *transfer RNA and ribosomal RNA molecules are generated by cleavage and other modifications of nascent RNA chains*. For example, in *E. coli*, three kinds of rRNA molecules and a tRNA molecule are excised from a single primary RNA transcript that also contains spacer regions (Figure 28.13). Other transcripts contain arrays of several kinds of tRNA or of several copies of the same tRNA. The nucleases that cleave and trim these precursors of rRNA and tRNA are highly precise. *Ribonuclease P*, for example, generates the correct 5' terminus of all tRNA molecules in *E. coli*. This interesting enzyme contains a catalytically active RNA molecule. *Ribonuclease III* excises 5S, 16S, and 23S rRNA precursors from the primary transcript by cleaving double-helical hairpin regions at specific sites.

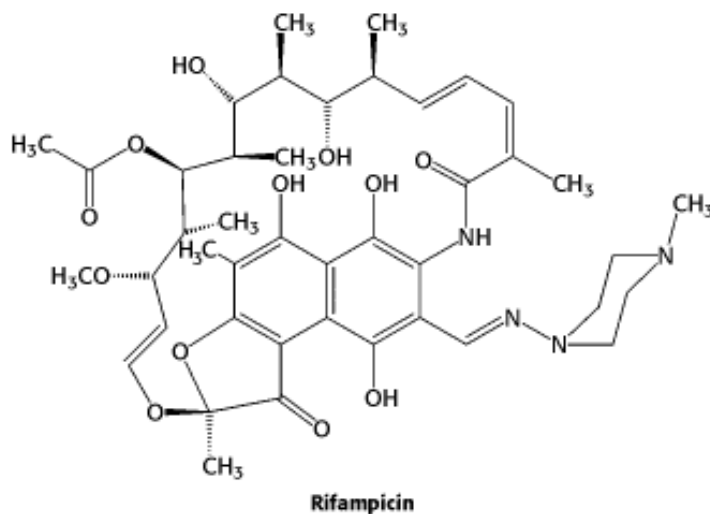


A second type of processing is the *addition of nucleotides to the termini of some RNA chains*. For example, CCA, a terminal sequence required for the function of all tRNAs, is added to the 3' ends of tRNA molecules that do not already possess this terminal sequence. A third type of processing is the *modification of bases and ribose units of ribosomal RNAs*. In prokaryotes, some bases of rRNA are methylated. Unusual bases are found in all tRNA molecules (Section 29.1.2). They are formed by the enzymatic modification of a standard ribonucleotide in a tRNA precursor. For example, uridylylate residues are modified after transcription to form *ribothymidylate* and *pseudouridylate*. These modifications generate diversity, allowing greater structural and functional versatility.



28.1.9. Antibiotic Inhibitors of Transcription

Many antibiotics are highly specific inhibitors of biological processes. Rifampicin and actinomycin are two antibiotics that inhibit transcription, although in quite different ways. *Rifampicin* is a semisynthetic derivative of *rifamycins*, which are derived from a strain of *Streptomyces*.



This antibiotic *specifically inhibits the initiation of RNA synthesis*. Rifampicin does not block the binding of RNA polymerase to the DNA template; rather, it interferes with the formation of the first few phosphodiester bonds in the RNA chain. The structure of a complex between a prokaryotic RNA polymerase and rifampicin reveals that the antibiotic blocks the channel into which the RNA-DNA hybrid generated by the enzyme must pass (Figure 28.14). The binding site is 12 Å from the active site itself. Rifampicin does not hinder chain elongation once initiated, because the RNA-DNA hybrid present in the enzyme prevents the antibiotic from binding.

Actinomycin D, a polypeptide-containing antibiotic from a different strain of *Streptomyces*, inhibits transcription by an entirely different mechanism. *Actinomycin D binds tightly and specifically to double-helical DNA and thereby prevents it from being an effective template for RNA synthesis*. It does not bind to single-stranded DNA or RNA, double-stranded RNA, or RNA-DNA hybrids. The results of spectroscopic and hydrodynamic studies of complexes of actinomycin D and DNA suggested that the phenoxazine ring of actinomycin slips in between neighboring base pairs in DNA. This mode of binding is called *intercalation*. At low concentrations, actinomycin D inhibits transcription without significantly

affecting DNA replication or protein synthesis. Hence, *actinomycin D* is extensively used as a highly specific inhibitor of the formation of new RNA in both prokaryotic and eukaryotic cells. Its ability to inhibit the growth of rapidly dividing cells makes it an effective therapeutic agent in the treatment of some cancers.

Table 28.1. Subunits of RNA polymerase from *E. coli*

Subunit	Gene	Number	Mass(kd)
α	<i>rpoA</i>	2	37
β	<i>rpoB</i>	1	151
β^{\prime}	<i>rpoC</i>	1	155
σ^{70}	<i>rpoD</i>	1	70

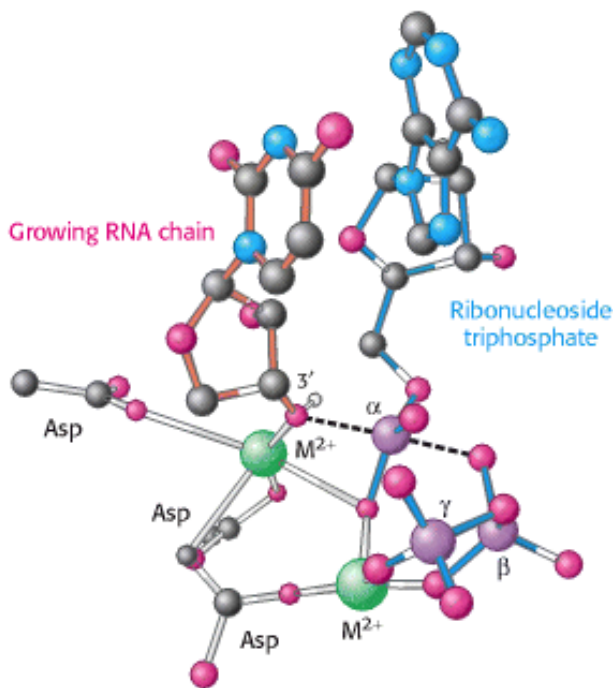


Figure 28.2. RNA Polymerase Active Site. A model of the transition state for phosphodiester-bond formation in the active site of RNA polymerase. The 3'-hydroxyl group of the growing RNA chain attacks the α -phosphate of the incoming nucleoside triphosphate. This transition state is structurally similar to that in DNA polymerase (see [Figure 27.12](#)).

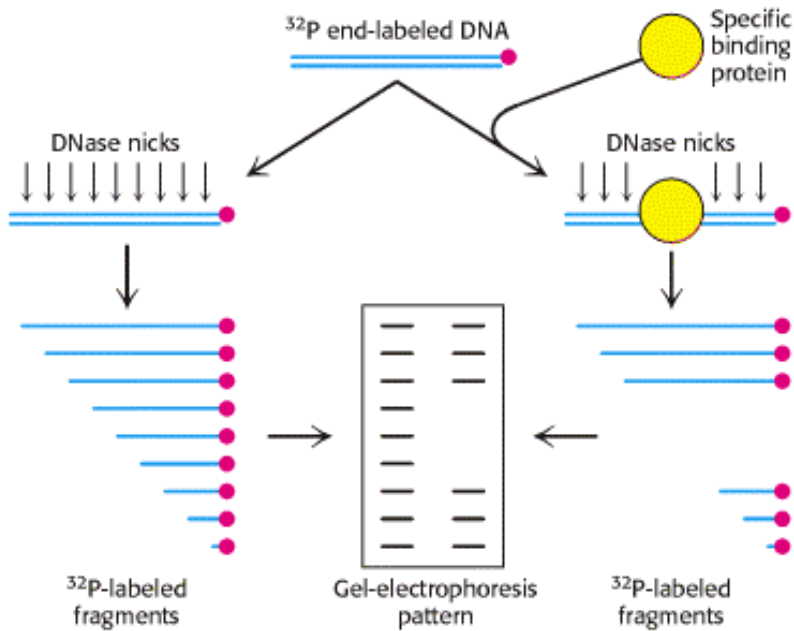


Figure 28.3. Footprinting. One end of a DNA chain is labeled with ^{32}P (shown as a red circle). This labeled DNA is then treated with DNase I such that each fragment is cut only once. The same cleavage is carried out after a protein that binds to specific sites on the DNA has been added. The bound protein protects a segment on the DNA from the action of DNase I. Hence, certain fragments present in the reaction without protein will be missing. These missing bands in the gel pattern identify the binding site on DNA.

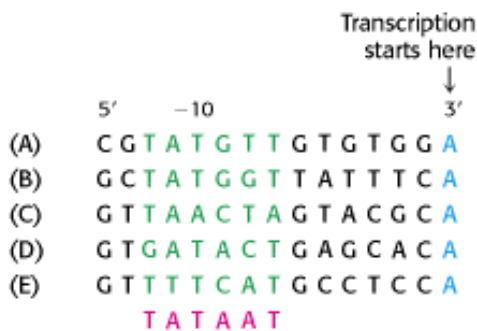


Figure 28.4. Prokaryotic Promoter Sequences. A comparison of five sequences from prokaryotic promoters reveals a recurring sequence of TATAAT centered on position -10. The -10 consensus sequence (in red) was deduced from a large number of promoter sequences. The sequences are from the (A) *lac*, (B) *gal*, and (C) *trp* operons of *E. coli*; (D) from λ phage; and (E) from Φ X174 phage.



Figure 28.5. Structure of the σ Subunit. The structure of a fragment from the *E. coli* subunit σ^{70} reveals the position of an α helix on the protein surface; this helix plays an important role in binding to the -10 TATAAT sequence.



Figure 28.6. Alternative Promoter Sequences. A comparison of the consensus sequences of standard promoters, heat-shock promoters, and nitrogen-starvation promoters of *E. coli*. These promoters are recognized by σ^{70} , σ^{32} , and σ^{54} , respectively.

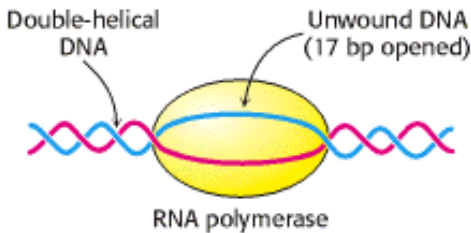


Figure 28.7. DNA Unwinding. RNA polymerase unwinds about 17 base pairs of template DNA.

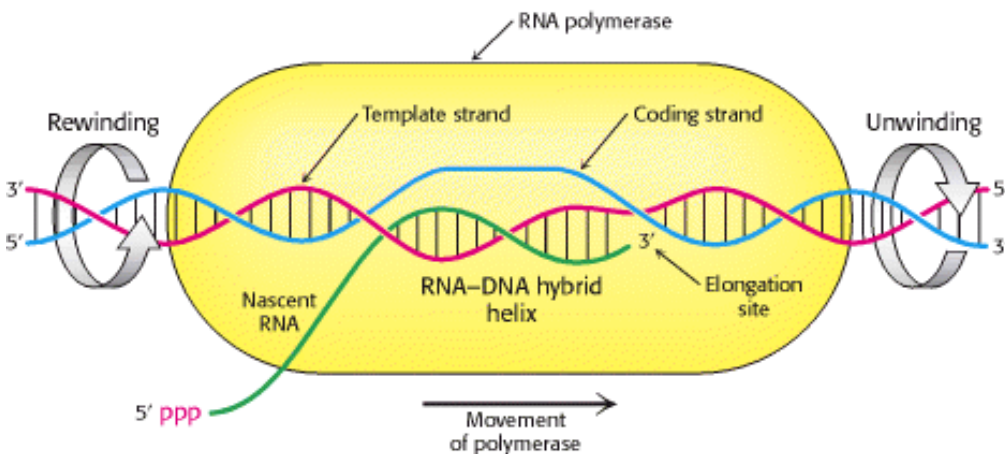


Figure 28.8. Transcription Bubble. A schematic representation of a transcription bubble in the elongation of an RNA transcript. Duplex DNA is unwound at the forward end of RNA polymerase and rewound at its rear end. The RNA-DNA hybrid rotates during elongation.



Figure 28.9. RNA-DNA Hybrid Separation. A structure within RNA polymerase forces the separation of the RNA-DNA hybrid, allowing the DNA strand to exit in one direction and the RNA product to exit in another.

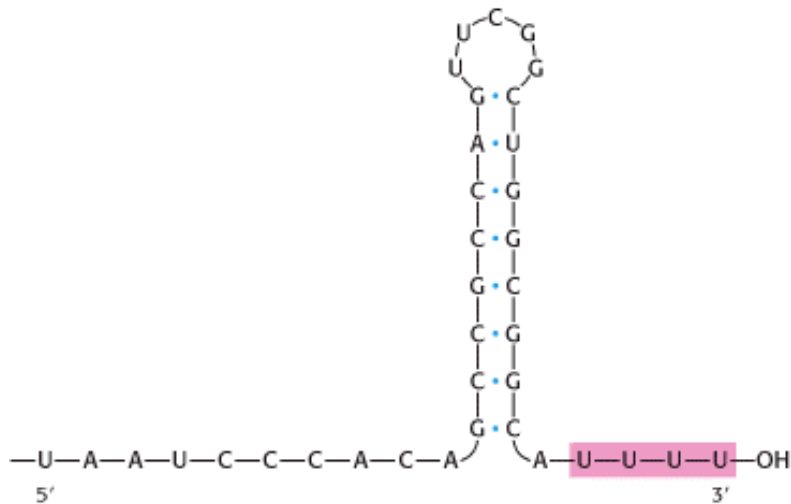


Figure 28.10. Termination Signal. A termination signal found at the 3' end of an mRNA transcript consists of a series of bases that form a stable stem-loop structure and a series of U residues.

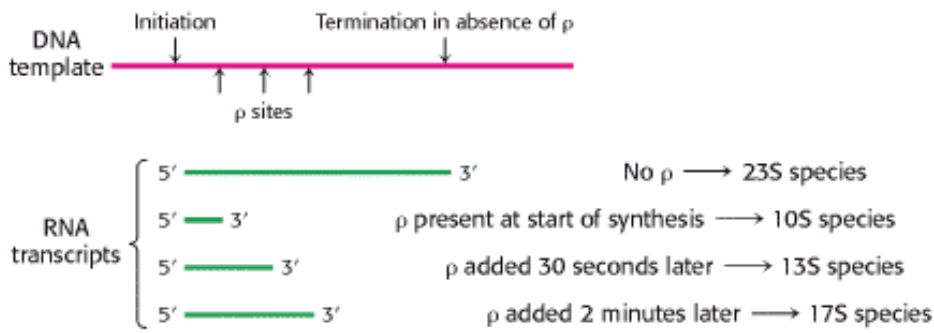


Figure 28.11. Effect of ρ Protein On the Size of RNA Transcripts.

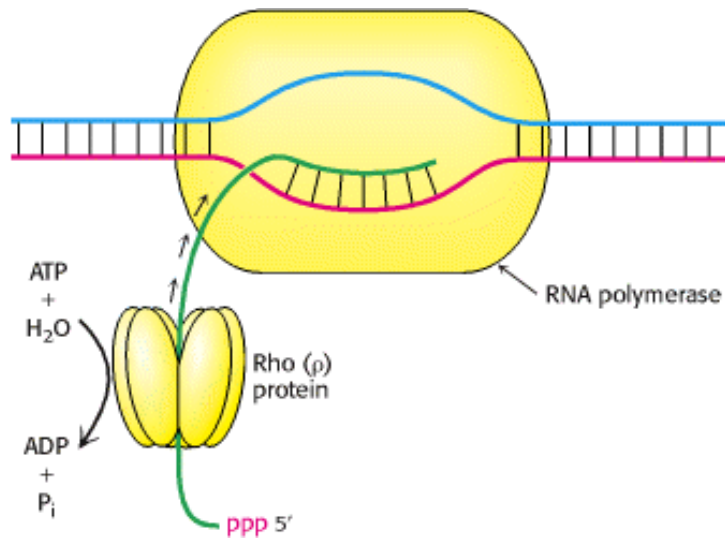


Figure 28.12. Mechanism For the Termination of Transcription by ρ Protein. This protein is an ATP-dependent helicase that binds the nascent RNA chain and pulls it away from RNA polymerase and the DNA template.



Figure 28.13. Primary Transcript. Cleavage of this transcript produces 5S, 16S, and 23S rRNA molecules and a tRNA molecule. Spacer regions are shown in yellow.

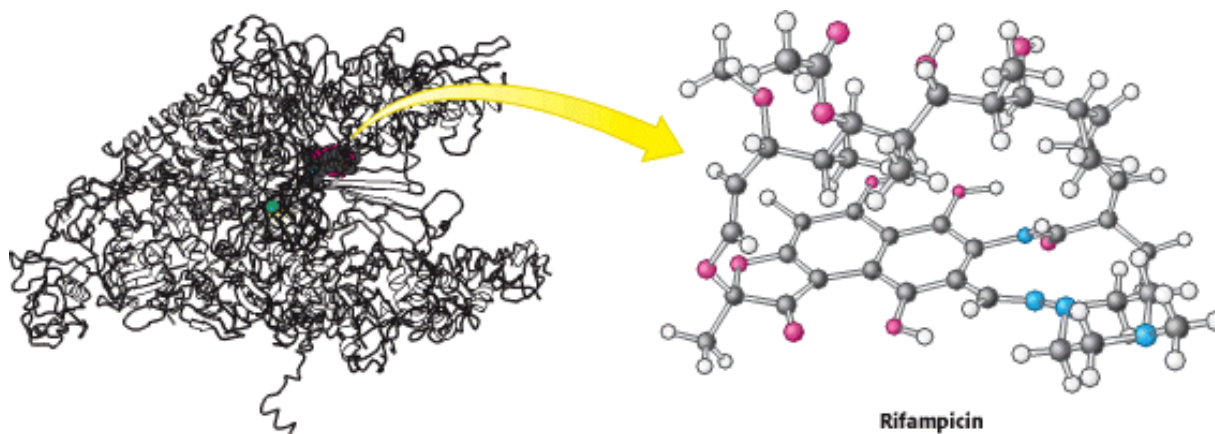


Figure 28.14. Antibiotic Action. Rifampicin binds to a pocket in the channel that is normally occupied by the newly formed RNA-DNA hybrid. Thus the antibiotic blocks elongation after only two or three nucleotides have been added.

28.2. Eukaryotic Transcription and Translation Are Separated in Space and Time

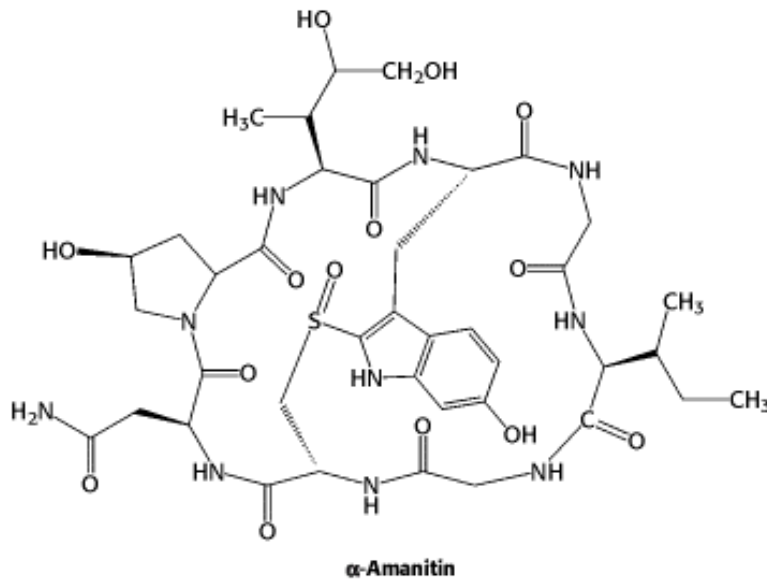
We turn now to transcription in eukaryotes, a much more complex process than in prokaryotes. *In eukaryotes, transcription and translation take place in different cellular compartments:* transcription takes place in the membrane-bounded nucleus, whereas translation takes place outside the nucleus in the cytoplasm. In prokaryotes, the two processes are closely coupled (Figure 28.15). Indeed, the translation of bacterial mRNA begins while the transcript is still being synthesized. *The spatial and temporal separation of transcription and translation enables eukaryotes to regulate gene expression in much more intricate ways, contributing to the richness of eukaryotic form and function.*

A second major difference between prokaryotes and eukaryotes is the extent of RNA processing. Although both prokaryotes and eukaryotes modify tRNA and rRNA, *eukaryotes very extensively process nascent RNA destined to become mRNA.* Primary transcripts (pre-mRNA molecules), the products of RNA polymerase action, acquire a cap at their 5' ends and a poly(A) tail at their 3' ends. Most importantly, *nearly all mRNA precursors in higher eukaryotes are spliced* (Section 5.6.1). Introns are precisely excised from primary transcripts, and exons are joined to form mature mRNAs with continuous messages. Some mRNAs are only a tenth the size of their precursors, which can be as large as 30 kb or more. The pattern of splicing can be regulated in the course of development to generate variations on a theme, such as membrane-bound and secreted forms of antibody molecules. Alternative splicing enlarges the repertoire of proteins in eukaryotes and is a clear illustration of why the proteome is more complex than the genome.

28.2.1. RNA in Eukaryotic Cells Is Synthesized by Three Types of RNA Polymerase

In prokaryotes, RNA is synthesized by a single kind of polymerase. In contrast, the nucleus of a eukaryote contains three types of RNA polymerase differing in template specificity, location in the nucleus, and susceptibility to inhibitors (Table 28.2). All these polymerases are large proteins, containing from 8 to 14 subunits and having a total molecular mass greater than 500 kd. *RNA polymerase I* is located in nucleoli, where it transcribes the tandem array of genes for 18S, 5.8S, and 28S ribosomal RNA (Section 29.3.1). The other ribosomal RNA molecule (5S rRNA, Section 29.3.1) and all the transfer RNA molecules (Section 29.1.2) are synthesized by *RNA polymerase III*, which is located in the nucleoplasm rather than in nucleoli. *RNA polymerase II*, which also is located in the nucleoplasm, synthesizes the precursors of messenger RNA as well as several small RNA molecules, such as those of the splicing apparatus (Section 28.3.5). Although all eukaryotic RNA polymerases are homologous to one another and to prokaryotic RNA polymerase, RNA polymerase II contains a unique carboxyl-terminal domain on the 220-kd subunit; this domain is unusual because it contains multiple repeats of a YSPTSPS consensus sequence. The activities of RNA polymerase II are regulated by phosphorylation on the serine and threonine residues of the carboxyl-terminal domain. Another major distinction among the polymerases lies in their responses to the toxin α -amanitin, a cyclic octapeptide that contains several modified

amino acids.



α -Amanitin is produced by the poisonous mushroom *Amanita phalloides*, which is also called the *death cup* or the *destroying angel* (Figure 28.16). More than a hundred deaths result worldwide each year from the ingestion of poisonous mushrooms. α -Amanitin binds very tightly ($K_d = 10$ nM) to RNA polymerase II and thereby blocks the elongation phase of RNA synthesis. Higher concentrations of α -amanitin (1 μ M) inhibit polymerase III, whereas polymerase I is insensitive to this toxin. This pattern of sensitivity is highly conserved throughout the animal and plant kingdoms.

28.2.2. Cis- And Trans-Acting Elements: Locks and Keys of Transcription

Eukaryotic genes, like their prokaryotic counterparts, require promoters for transcription initiation. Each of the three types of polymerase has distinct promoters. RNA polymerase I transcribes from a single type of promoter, present only in rRNA genes, that encompasses the initiation site. In some genes, RNA polymerase III responds to promoters located in the normal, upstream position; in other genes, it responds to promoters imbedded in the genes, downstream of the initiation site. Promoters for RNA polymerase II can be simple or complex (Section 28.2.3). As is the case for prokaryotes, promoters are always on the same molecule of DNA as the gene they regulate. Consequently, promoters are referred to as *cis-acting elements*.

However, promoters are not the only types of cis-acting DNA sequences. Eukaryotes and their viruses also contain *enhancers*. These DNA sequences, although not promoters themselves, can enormously increase the effectiveness of promoters. Interestingly, the positions of enhancers relative to promoters are not fixed; they can vary substantially. Enhancers play key roles in regulating gene expression in a specific tissue or developmental stage (Section 31.2.4).

The DNA sequences of cis-acting elements are binding sites for proteins called *transcription factors*. Such a protein is sometimes called a *trans-acting factor* because it may be encoded by a gene on a DNA molecule other than that containing the gene being regulated. The binding of a transcription factor to its cognate DNA sequence enables the RNA polymerase to locate the proper initiation site. We will continue our investigation of transcription by examining these cis- and trans-acting elements in turn.

28.2.3. Most Promoters for RNA Polymerase II Contain a TATA Box Near the Transcription Start Site

Promoters for RNA polymerase II, like those for bacterial polymerases, are located on the 5' side of the start site for transcription. The results of mutagenesis experiments, footprinting studies, and comparisons of many higher eukaryotic

genes have demonstrated the importance of several upstream regions. For most genes transcribed by RNA polymerase II, the most important cis-acting element is called the *TATA box* on the basis of its consensus sequence (Figure 28.17). The TATA box is usually centered between positions -30 and -100. Note that the eukaryotic TATA box closely resembles the prokaryotic -10 sequence (TATAAT) but is farther from the start site. The mutation of a single base in the TATA box markedly impairs promoter activity. Thus, the precise sequence, not just a high content of AT pairs, is essential.


The TATA box is necessary but not sufficient for strong promoter activity. Additional elements are located between -40 and -150. Many promoters contain a *CAAT box*, and some contain a *GC box* (Figure 28.18). Constitutive genes (genes that are continuously expressed rather than regulated) tend to have GC boxes in their promoters. The positions of these upstream sequences vary from one promoter to another, in contrast with the quite constant location of the -35 region in prokaryotes. Another difference is that the CAAT box and the GC box can be effective when present on the template (antisense) strand, unlike the -35 region, which must be present on the coding (sense) strand. These differences between prokaryotes and eukaryotes reflect fundamentally different mechanisms for the recognition of cis-acting elements. The -10 and -35 sequences in prokaryotic promoters correspond to binding sites for RNA polymerase and its associated σ factor. In contrast, the TATA, CAAT, and GC boxes and other cis-acting elements in eukaryotic promoters are recognized by proteins other than RNA polymerase itself.

28.2.4. The TATA-Box-Binding Protein Initiates the Assembly of the Active Transcription Complex

Cis-acting elements constitute only part of the puzzle of eukaryotic gene expression. Transcription factors that bind to these elements also are required. For example, RNA polymerase II is guided to the start site by a set of transcription factors known collectively as *TFII* (*TF* stands for transcription factor, and *II* refers to RNA polymerase II). Individual TFII factors are called TFIIA, TFIIB, and so on. Initiation begins with the binding of TFIID to the TATA box (Figure 28.19).

The key initial event is the recognition of the TATA box by the TATA-box-binding protein (TBP), a 30-kd component of the 700-kd TFIID complex. TBP binds 10^5 times as tightly to the TATA box as to noncognate sequences; the dissociation constant of the specific complex is approximately 1 nM. TBP is a saddle-shaped protein consisting of two similar domains (Section 7.3.3; Figure 28.20). The TATA box of DNA binds to the concave surface of TBP. This binding induces large conformational changes in the bound DNA. The double helix is substantially unwound to widen its *minor groove*, enabling it to make extensive contact with the antiparallel β strands on the concave side of TBP. Hydrophobic interactions are prominent at this interface. Four phenylalanine residues, for example, are intercalated between base pairs of the TATA box. The flexibility of AT-rich sequences is generally exploited here in bending the DNA. Immediately outside the TATA box, classical B-DNA resumes. This complex is distinctly asymmetric. The asymmetry is crucial for specifying a unique start site and ensuring that transcription proceeds unidirectionally.

TBP bound to the TATA box is the heart of the initiation complex (see Figure 28.19). The surface of the TBP saddle provides docking sites for the binding of other components (Figure 28.21). Additional transcription factors assemble on this nucleus in a defined sequence. TFIIA is recruited, followed by TFIIB and then TFIIF—an ATP-dependent helicase that initially separates the DNA duplex for the polymerase. Finally, RNA polymerase II and then TFIIE join the other factors to form a complex called the *basal transcription apparatus*. Sometime in the formation of this complex, the carboxyl-terminal domain of the polymerase is phosphorylated on the serine and threonine residues, a process required for successful initiation. The importance of the carboxyl-terminal domain is highlighted by the finding that yeast containing mutant polymerase II with fewer than 10 repeats is not viable. Most of the factors are released before the polymerase leaves the promoter and can then participate in another round of initiation.

 Although bacteria lack TBP, archaea utilize a TBP molecule that is structurally quite similar to the eukaryotic protein. In fact, transcriptional control processes in archaea are, in general, much more similar to those in eukaryotes than are the processes in bacteria. Many components of the eukaryotic transcriptional machinery evolved from an ancestor of archaea.

28.2.5. Multiple Transcription Factors Interact with Eukaryotic Promoters


The basal transcription complex described in [Section 28.2.4](#), initiates transcription at a relatively low frequency. Additional transcription factors that bind to other sites are required to achieve a high rate of mRNA synthesis and to selectively stimulate specific genes. Upstream stimulatory sites in eukaryotic genes are diverse in sequence and variable in position. Their variety suggests that they are recognized by many different specific proteins. Indeed, many transcription factors have been isolated, and their binding sites have been identified by footprinting experiments ([Figure 28.22](#)). For example, *Sp1*, an ~ 100-kd protein from mammalian cells, binds to promoters that contain GC boxes. The duplex DNA of SV40 virus (a cancer-producing virus that infects monkey cells) contains five GC boxes from 50 to 100 bp upstream or downstream of start sites. The *CCAAT-binding transcription factor* (CTF; also called NF1), a 60-kd protein from mammalian cells, binds to the CAAT box. A *heat-shock transcription factor* (HSTF) is expressed in *Drosophila* after an abrupt increase in temperature. This 93-kd DNA-binding protein binds to the consensus sequence

5'-CNNGAANNTCCNNG-3'

Several copies of this sequence, known as the *heat-shock response element*, are present starting at a site 15 bp upstream of the TATA box. HSTF differs from σ^{32} , a heat-shock protein of *E. coli* ([Section 28.1.2](#)), in binding directly to response elements in heat-shock promoters rather than first becoming associated with RNA polymerase.

28.2.6. Enhancer Sequences Can Stimulate Transcription at Start Sites Thousands of Bases Away

The activities of many promoters in higher eukaryotes are greatly increased by another type of cis-acting element called an *enhancer*. Enhancers' sequences have no promoter activity of their own *yet can exert their stimulatory actions over distances of several thousand base pairs. They can be upstream, downstream, or even in the midst of a transcribed gene.* Moreover, enhancers are effective when present on *either DNA strand* (equivalently, in either orientation). Enhancers in yeast are known as *upstream activator sequences* (UASs).

 *A particular enhancer is effective only in certain cells.* For example, the immunoglobulin enhancer functions in B lymphocytes but not elsewhere. Cancer can result if the relation between genes and enhancers is disrupted. In Burkitt lymphoma and B-cell leukemia, a chromosomal translocation brings the proto-oncogene *myc* (a transcription factor itself) under the control of a powerful immunoglobulin enhancer. The consequent dysregulation of the *myc* gene is believed to play a role in the progression of the cancer.

Transcription factors and other proteins that bind to regulatory sites on DNA can be regarded as passwords that cooperatively open multiple locks, giving RNA polymerase access to specific genes. The discovery of promoters and enhancers has opened the door to understanding how genes are selectively expressed in eukaryotic cells. The regulation of gene transcription, discussed in [Chapter 31](#), is the fundamental means of controlling gene expression.

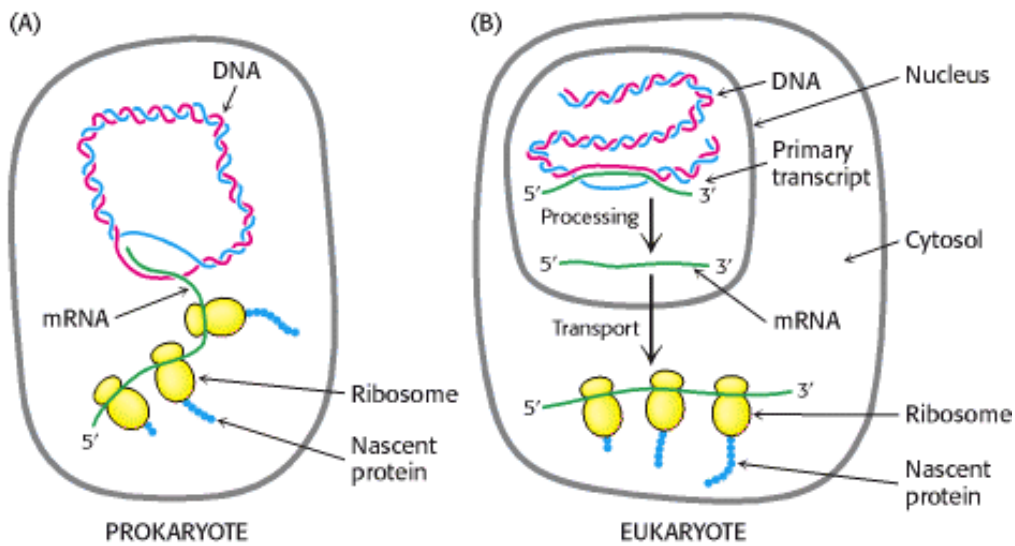


Figure 28.15. Transcription and Translation. These two processes are closely coupled in prokaryotes, whereas they are spatially and temporally separate in eukaryotes. (A) In prokaryotes, the primary transcript serves as mRNA and is used immediately as the template for protein synthesis. (B) In eukaryotes, mRNA precursors are processed and spliced in the nucleus before being transported to the cytosol for translation into protein. [After J. Darnell, H. Lodish, and D. Baltimore. *Molecular Cell Biology*, 2d ed. (Scientific American Books, 1990), p. 230.]

Table 28.2. Eukaryotic RNA polymerases

Type	Location	Cellular transcripts	Effects of α -amanitin
I	Nucleolus	18S, 5.8S, and 28S rRNA	Insensitive
II	Nucleoplasm	mRNA precursors and snRNA	Strongly inhibited
III	Nucleoplasm	tRNA and 5S rRNA	Inhibited by high concentrations

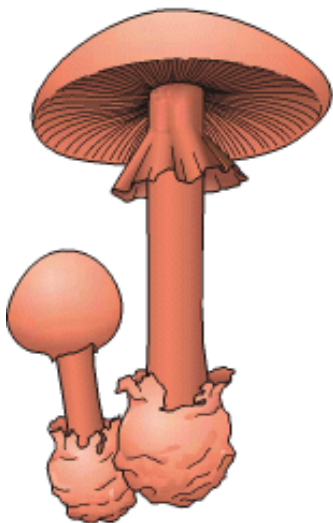


Figure 28.16. RNA Polymerase Poison. *Amanita phalloides*, a poisonous mushroom that produces α -amanitin. [After G. Lincoff and D. H. Mitchel, *Toxic and Hallucinogenic Mushroom Poisoning* (Van Nostrand Reinhold, 1977), p. 30.]

5' T₈₂ A₉₇ T₉₃ A₈₅ A₆₃ A₈₈ A₅₀ 3'
TATA box

Figure 28.17. TATA Box. Comparisons of the sequences of more than 100 eukaryotic promoters led to the consensus sequence shown. The subscripts denote the frequency (%) of the base at that position.

5' GGN**CAAT**CT 3'
CAAT box

5' GG**CGG**G 3'
GC box

Figure 28.18. CAAT Box and GC Box. Consensus sequences for the CAAT and GC boxes of eukaryotic promoters for mRNA precursors.

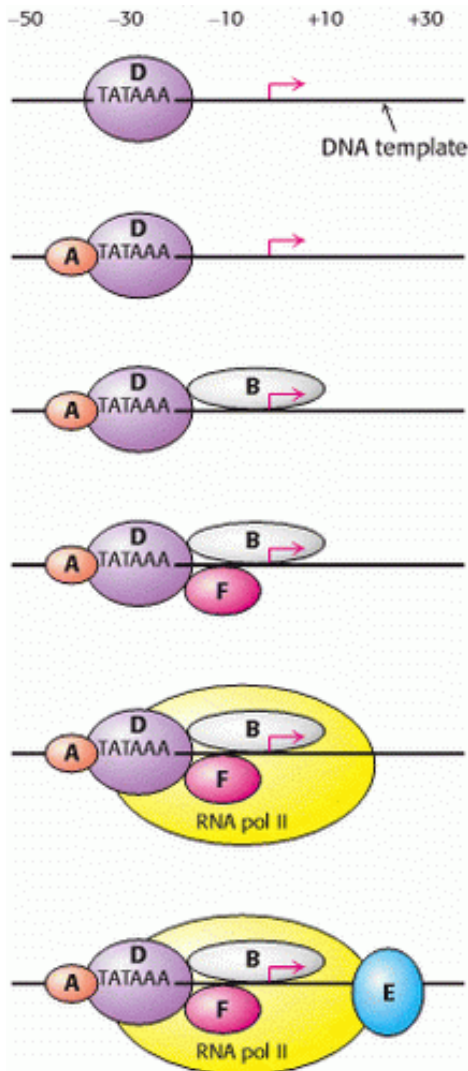


Figure 28.19. Transcription Initiation. Transcription factors TFIIA, B, D, E, and F are essential in initiating transcription by RNA polymerase II. The step-by-step assembly of these general transcription factors begins with the binding of TFIID (purple) to the TATA box. The arrow marks the transcription start site. [After L. Guarente. *Trends Genet.* 8(1992):28.]

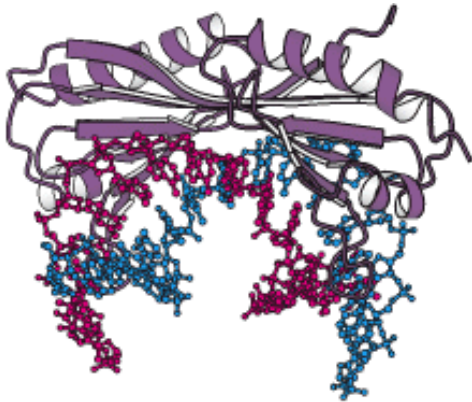


Figure 28.20. Complex Formed by TATA-Box-Binding Protein and DNA. The saddlelike structure of the protein sits atop a DNA fragment that is both significantly unwound and bent. 

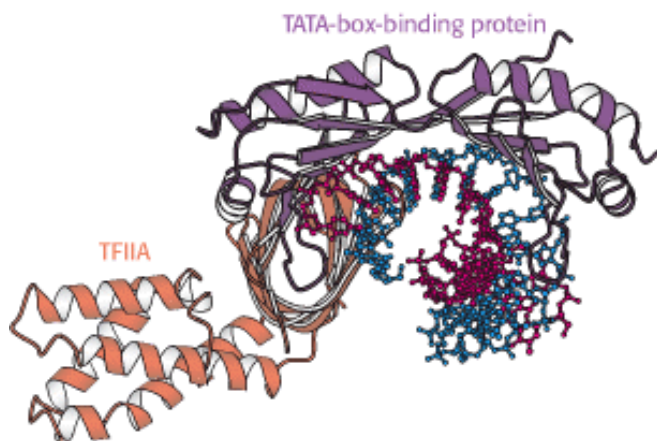



Figure 28.21. Assembly of the Initiation Complex. A ternary complex between the TATA-box-binding protein  (purple), TFIIA (orange), and DNA. TFIIA interacts primarily with the other protein.

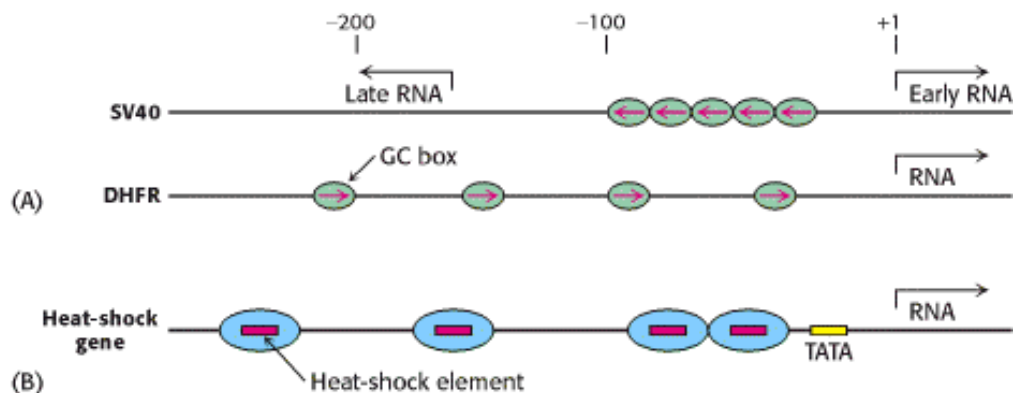


Figure 28.22. Transcription-Factor-Binding Sites. These multiple binding sites for transcription factors were mapped by footprinting. (A) Binding of Sp1 (green) to the SV40 viral promoter and the dihydrofolate reductase (DHFR) promoter. (B) Binding of HSTF (blue) to a *Drosophila* heat-shock promoter. [After W. S. Dynan and R. Tjian. *Nature*

28.3. The Transcription Products of All Three Eukaryotic Polymerases Are Processed

Virtually all the initial products of transcription are further processed in eukaryotes. For example, tRNA precursors are converted into mature tRNAs by a series of alterations: cleavage of a 5' leader sequence, splicing to remove an intron, replacement of the 3'-terminal UU by CCA, and modification of several bases (Figure 28.23). A series of enzymes may act on the ribonucleic acid chain or its constituent bases to achieve the final product.

28.3.1. The Ends of the Pre-mRNA Transcript Acquire a 5' Cap and a 3' Poly(A) Tail

Perhaps the most extensively modified transcription product is the product of RNA polymerase II: the majority of this RNA will be processed to mRNA. The immediate product of an RNA polymerase is sometimes referred to as *pre-mRNA*. Most pre-mRNA molecules are spliced to remove the introns. Moreover, both the 5' and the 3' ends are modified, and both modifications are retained as the pre-mRNA is converted into mRNA (Section 28.3.3). As in prokaryotes, eukaryotic transcription usually begins with A or G. However, the 5' triphosphate end of the nascent RNA chain is immediately modified. First, a phosphate is released by hydrolysis. The diphosphate 5' end then attacks the α -phosphorus atom of GTP to form a very unusual 5'-5' triphosphate linkage. This distinctive terminus is called a *cap* (Figure 28.24). The N-7 nitrogen of the terminal guanine is then methylated by *S*-adenosylmethionine to form *cap 0*. The adjacent riboses may be methylated to form *cap 1* or *cap 2*. Transfer RNA and ribosomal RNA molecules, in contrast with messenger RNAs and small RNAs that participate in splicing, do not have caps. Caps contribute to the stability of mRNAs by protecting their 5' ends from phosphatases and nucleases. In addition, caps enhance the translation of mRNA by eukaryotic proteinsynthesizing systems (Section 29.5).

As mentioned earlier, pre-mRNA is also modified at the 3' end. *Most eukaryotic mRNAs contain a polyadenylate, poly(A), tail at that end*, added after transcription has ended. Thus, DNA does not encode this poly(A) tail. Indeed, the nucleotide preceding poly(A) is not the last nucleotide to be transcribed. Some primary transcripts contain hundreds of nucleotides beyond the 3' end of the mature mRNA.

How is the 3' end of the pre-mRNA given its final form? *Eukaryotic primary transcripts are cleaved by a specific endonuclease that recognizes the sequence AAUAAA* (Figure 28.25). Cleavage does not occur if this sequence or a segment of some 20 nucleotides on its 3' side is deleted. The presence of internal AAUAAA sequences in some mature mRNAs indicates that AAUAAA is only part of the cleavage signal; its context also is important. After cleavage by the endonuclease, a *poly(A) polymerase* adds about 250 adenylate residues to the 3' end of the transcript; ATP is the donor in this reaction.

The role of the poly(A) tail is still not firmly established despite much effort. However, evidence that it enhances translation efficiency and the stability of mRNA is accumulating. Blocking the synthesis of the poly(A) tail by exposure to 3'-*deoxyadenosine* (*cordycepin*) does not interfere with the synthesis of the primary transcript. Messenger RNA devoid of a poly(A) tail can be transported out of the nucleus. However, an mRNA molecule devoid of a poly(A) tail is usually a much less effective template for protein synthesis than is one with a poly(A) tail. Indeed, some mRNAs are stored in an unadenylated form and receive the poly(A) tail only when translation is imminent. The half-life of an mRNA molecule may also be determined in part by the rate of degradation of its poly(A) tail.

28.3.2. RNA Editing Changes the Proteins Encoded by mRNA

The sequence content of some mRNAs is altered after transcription. *RNA editing* is the term for a change in the base sequence of RNA after transcription by processes other than RNA splicing. RNA editing is prominent in some systems

already discussed. *Apolipoprotein B (apo B)* plays an important role in the transport of triacylglycerols and cholesterol by forming an amphipathic spherical shell around the lipids carried in lipoprotein particles (Section 26.3.1). Apo B exists in two forms, a 512-kd *apo B-100* and a 240-kd *apo B-48*. The larger form, synthesized by the liver, participates in the transport of lipids synthesized in the cell. The smaller form, synthesized by the small intestine, carries dietary fat in the form of chylomicrons. Apo B-48 contains the 2152 N-terminal residues of the 4536-residue apo B-100. This truncated molecule can form lipoprotein particles but cannot bind to the low-density-lipoprotein receptor on cell surfaces. What is the biosynthetic relation of these two forms of apo B? One possibility a priori is that apo B-48 is produced by proteolytic cleavage of apo B-100, and another is that the two forms arise from alternative splicing (see Section 28.3.6). The results of experiments show that neither occurs. A totally unexpected and new mechanism for generating diversity is at work: *the changing of the nucleotide sequence of mRNA after its synthesis (Figure 28.26)*. A specific cytidine residue of mRNA is deaminated to uridine, which changes the codon at residue 2153 from CAA (Gln) to UAA (stop). The deaminase that catalyzes this reaction is present in the small intestine, but not in the liver, and is expressed only at certain developmental stages.

RNA editing is not confined to apolipoprotein B. Glutamate opens cation-specific channels in the vertebrate central nervous system by binding to receptors in postsynaptic membranes. RNA editing changes a single glutamine codon (CAG) in the mRNA for the glutamate receptor to the codon for arginine (read as CGG). The substitution of Arg for Gln in the receptor prevents Ca^{2+} , but not Na^{+} , from flowing through the channel. RNA editing is likely much more common than was previously thought. The chemical reactivity of nucleotide bases, including the susceptibility to deamination that necessitates complex DNA-repair mechanisms (Section 27.6.3), has been harnessed as an engine for generating molecular diversity at the RNA and, hence, protein levels.

In trypanosomes (parasitic protozoans), a different kind of RNA editing markedly changes several mitochondrial mRNAs. Nearly half the uridine residues in these mRNAs are *inserted* by RNA editing. A *guide RNA molecule* identifies the sequences to be modified, and a *poly(U) tail* on the guide donates uridine residues to the mRNAs undergoing editing. It is evident that DNA sequences do not always faithfully disclose the sequence of encoded proteins—functionally crucial changes to mRNA can take place.


28.3.3. Splice Sites in mRNA Precursors Are Specified by Sequences at the Ends of Introns

Most genes in higher eukaryotes are composed of exons and introns. The introns must be excised and the exons linked to form the final mRNA in a process called *splicing*. This splicing must be exquisitely sensitive: a one-nucleotide slippage in a splice point would shift the reading frame on the 3' side of the splice to give an entirely different amino acid sequence. Thus, the correct splice site must be clearly marked. Does a particular sequence denote the splice site? The base sequences of thousands of intron-exon junctions within RNA transcripts are known. In eukaryotes from yeast to mammals, these sequences have a common structural motif: *the base sequence of an intron begins with GU and ends with AG*. The consensus sequence at the 5' splice in vertebrates is AGGUAAGU (Figure 28.27). At the 3' end of an intron, the consensus sequence is a stretch of 10 pyrimidines (U or C), followed by any base and then by C, and ending with the invariant AG. Introns also have an important internal site located between 20 and 50 nucleotides upstream of the 3' splice site; it is called the *branch site* for reasons that will be evident shortly. In yeast, the branch site sequence is nearly always UACUAAC, whereas in mammals a variety of sequences are found.

Parts of introns other than the 5' and 3' splice sites and the branch site appear to be less important in determining where splicing takes place. The length of introns ranges from 50 to 10,000 nucleotides. Much of an intron can be deleted without altering the site or efficiency of splicing. Likewise, splicing is unaffected by the insertion of long stretches of DNA into the introns of genes. Moreover, chimeric introns crafted by recombinant DNA methods from the 5' end of one intron and the 3' end of a very different intron are properly spliced, provided that the splice sites and branch site are unaltered. In contrast, mutations in each of these three critical regions lead to aberrant splicing.

Despite our knowledge of splice-site sequences, predicting splicing patterns from genomic DNA sequence information

remains a challenge. Other information that contributes to splice-site selection is present in DNA sequences, but it is more loosely distributed than are the splice-site sequences themselves.

 Aberrant splicing causes some forms of thalassemia, a group of hereditary anemias characterized by the defective synthesis of hemoglobin. In one patient, a mutation of G to A 19 nucleotides away from the normal 3' splice site of the first intron created a new 3' splice site (Figure 28.28). The resulting mRNA contains a series of codons not normally present. The sixth codon after the splice is a stop signal for protein synthesis, and so the aberrant protein ends prematurely. *Mutations affecting splice sites have been estimated to cause 15% of all genetic diseases.*

28.3.4. Splicing Consists of Two Transesterification Reactions

The splicing of nascent mRNA molecules is a complicated process. It requires the cooperation of several small RNAs and proteins that form a large complex called a *spliceosome*. However, the chemistry of the splicing process is simple. Splicing begins with the cleavage of the phosphodiester bond between the upstream exon (exon 1) and the 5' end of the intron (Figure 28.29). The attacking group in this reaction is the 2'-hydroxyl group of an adenylate residue in the branch site. A 2',5'-phosphodiester bond is formed between this A residue and the 5' terminal phosphate of the intron. This reaction is a transesterification.



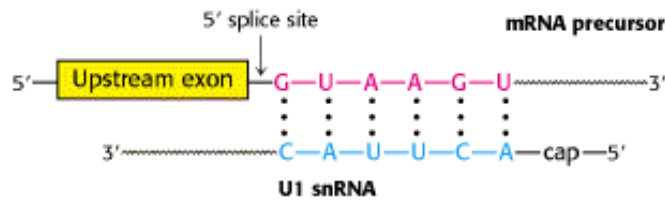
Note that this adenylate residue is also joined to two other nucleotides by normal 3',5'-phosphodiester bonds (Figure 28.30). Hence a *branch* is generated at this site, and a *lariat intermediate* is formed.

The 3'-OH terminus of exon 1 then attacks the phosphodiester bond between the intron and exon 2. Exons 1 and 2 become joined, and the intron is released in lariat form. Again, this reaction is a transesterification. Splicing is thus accomplished by two *transesterification reactions* rather than by hydrolysis followed by ligation. The first reaction generates a free 3'-hydroxyl group at the 3' end of exon 1, and the second reaction links this group to the 5'-phosphate of exon 2. *The number of phosphodiester bonds stays the same during these steps*, which is crucial because it allows the splicing reaction itself to proceed without an energy source such as ATP or GTP.

28.3.5. Small Nuclear RNAs in Spliceosomes Catalyze the Splicing of mRNA Precursors

The nucleus contains many types of small RNA molecules with fewer than 300 nucleotides, referred to as *snRNAs* (*small nuclear RNAs*). A few of them—designated U1, U2, U4, U5, and U6—are essential for splicing mRNA precursors. The secondary structures of these RNAs are highly conserved in organisms ranging from yeast to human beings. These RNA molecules are associated with specific proteins to form complexes termed *snRNPs* (*small nuclear ribonucleoprotein particles*); investigators often speak of them as "snurps." Spliceosomes are large (60S), dynamic assemblies composed of snRNPs, other proteins called *splicing factors*, and the mRNA precursors being processed (Table 28.3).

In mammalian cells, splicing begins with the recognition of the 5' splice site by U1 snRNP (Figure 28.31). In fact, U1 RNA contains a highly conserved six-nucleotide sequence that base pairs to the 5' splice site of the pre-mRNA. This binding initiates spliceosome assembly on the pre-mRNA molecule.



U2 snRNP then binds the branch site in the intron by base-pairing between a highly conserved sequence in U2 snRNA and the pre-mRNA. U2 snRNP binding requires ATP hydrolysis. A preassembled U4-U5-U6 complex joins this complex of U1, U2, and the mRNA precursor to form a complete spliceosome. This association also requires ATP hydrolysis.

A revealing view of the interplay of RNA molecules in this assembly came from examining the pattern of cross-links formed by *psoralen*, a photoactivable reagent that joins neighboring pyrimidines in base-paired regions. These cross-links suggest that splicing takes place in the following way. First, U5 interacts with exon sequences in the 5' splice site and subsequently with the 3' exon. Next, U6 disengages from U4 and undergoes an intramolecular rearrangement that permits base-pairing with U2 and displaces U1 from the spliceosome by interacting with the 5' end of the intron. The U2·U6 helix is indispensable for splicing, suggesting that *U2 and U6 snRNAs probably form the catalytic center of the spliceosome* (Figure 28.32). U4 serves as an inhibitor that masks U6 until the specific splice sites are aligned. These rearrangements result in the first transesterification reaction, generating the lariat intermediate and a cleaved 5' exon.

Further rearrangements of RNA in the spliceosome facilitate the second transesterification. These rearrangements align the free 5' exon with the 3' exon such that the 3'-hydroxyl group of the 5' exon is positioned to nucleophilically attack the 3' splice site to generate the spliced product. U2, U5, and U6 bound to the excised lariat intron are released to complete the splicing reaction.

Many of the steps in the splicing process require ATP hydrolysis. How is the free energy associated with ATP hydrolysis used to power splicing? To achieve the well-ordered rearrangements necessary for splicing, ATP-powered RNA helicases must unwind RNA helices and allow alternative base-pairing arrangements to form. Thus, two features of the splicing process are noteworthy. First, *RNA molecules play key roles in directing the alignment of splice sites and in carrying out catalysis*. Second, *ATP-powered helicases unwind RNA duplex intermediates that facilitate catalysis and induce the release of snRNPs from the mRNA*.

28.3.6. Some Pre-mRNA Molecules Can Be Spliced in Alternative Ways to Yield Different mRNAs

Alternative splicing is a widespread mechanism for generating protein diversity. The differential inclusion of exons into a mature RNA, alternative splicing may be regulated to produce distinct forms of a protein for specific tissues or developmental stages (Figure 28.33). A sample of the growing list of proteins known to result from alternative splicing is presented in Table 28.4, and recent estimates suggest that the RNA products of 30% of human genes are alternatively spliced. *Alternative splicing provides a powerful mechanism for expanding the versatility of genomic sequences*. Suppose, for example, that it is beneficial to have two forms of a protein with somewhat different properties that are expressed in different tissues. The evolution of an alternative splicing pathway provides a route to meeting this need by means other than gene duplication and specialization. Furthermore, alternative splicing provides an opportunity for *combinatorial control*. Consider a gene with five positions at which alternative splicing can take place. With the assumption that these alternative splicing pathways can be regulated independently, a total of $2^5 = 32$ different mRNAs can be generated. Further studies of alternative splicing and the mechanisms of splice-site selection will be crucial to the field of proteomics.

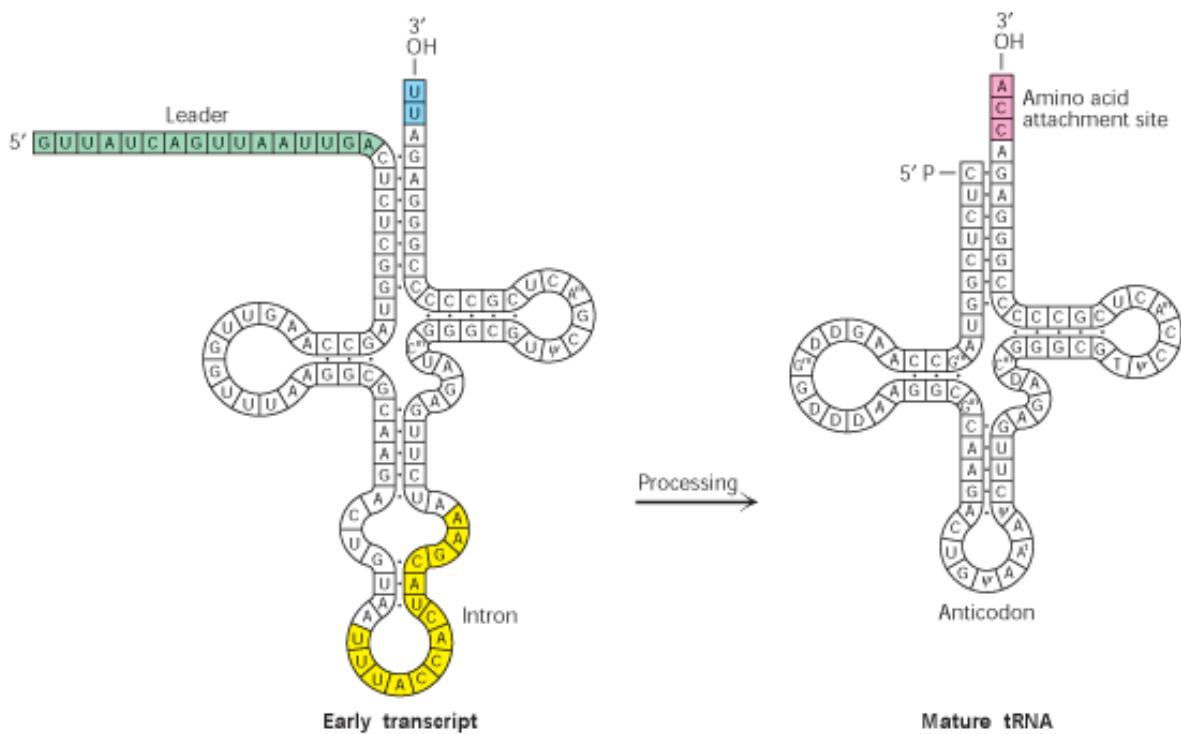


Figure 28.23. Transfer RNA Precursor Processing. The conversion of a yeast tRNA precursor into a mature tRNA requires the removal of a 14-nucleotide intron (yellow), the cleavage of a 5' leader (green), and the removal of UU and the attachment of CCA at the 3' end (red). In addition, several bases are modified.

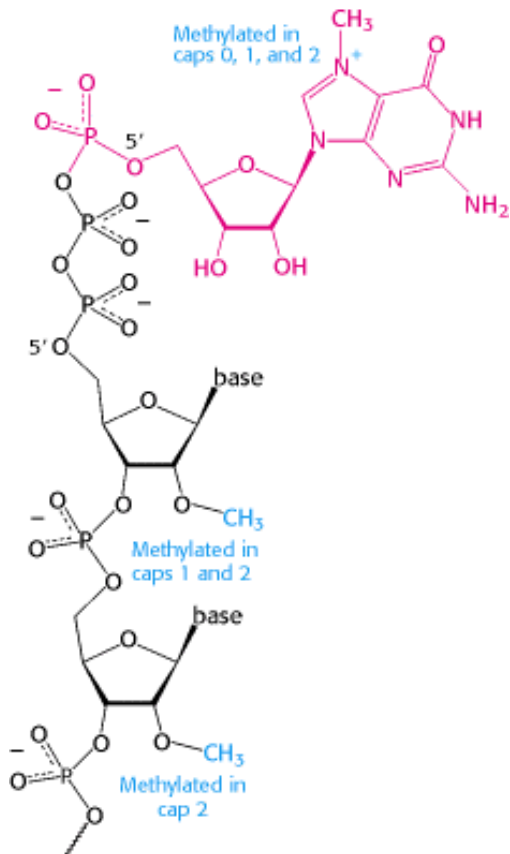


Figure 28.24. Capping the 5' End. Caps at the 5' end of eukaryotic mRNA include 7-methylguanylate (red) attached by

a triphosphate linkage to the ribose at the 5' end. None of the riboses are methylated in cap 0, one is methylated in cap 1, and both are methylated in cap 2.

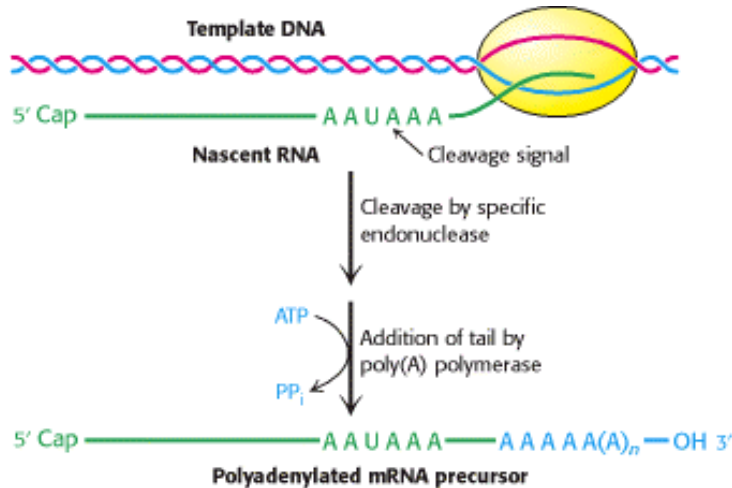


Figure 28.25. Polyadenylation of a Primary Transcript. A specific endonuclease cleaves the RNA downstream of AAUAAA. Poly(A) polymerase then adds about 250 adenylate residues.

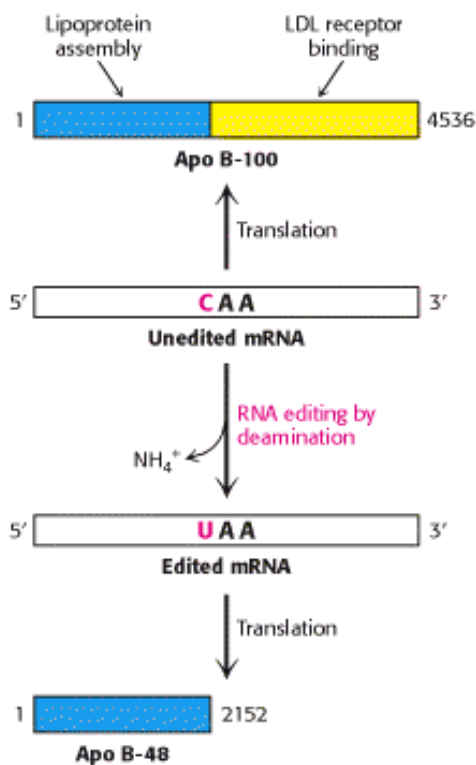


Figure 28.26. RNA Editing. Enzyme-catalyzed deamination of a specific cytosine residue in the mRNA for apolipoprotein B-100 changes a codon for glutamine (CAA) to a stop codon (UAA). Apolipoprotein B-48, a truncated version of the protein lacking the LDL receptor-binding domain, is generated by this posttranscriptional change in the mRNA sequence. [After P. Hodges and J. Scott. *Trends Biochem. Sci.* 17(1992):77.]



Figure 28.27. Splice Sites. Consensus sequences for the 5[′] splice site and the 3[′] splice site are shown. Py stands for pyrimidine.



Figure 28.28. Splicing Defects. Mutation of a single base (G to A) in an intron of the β-globin gene leads to thalassemia. This mutation generates a new 3[′] splice site (blue) akin to the normal one (yellow) but farther upstream.

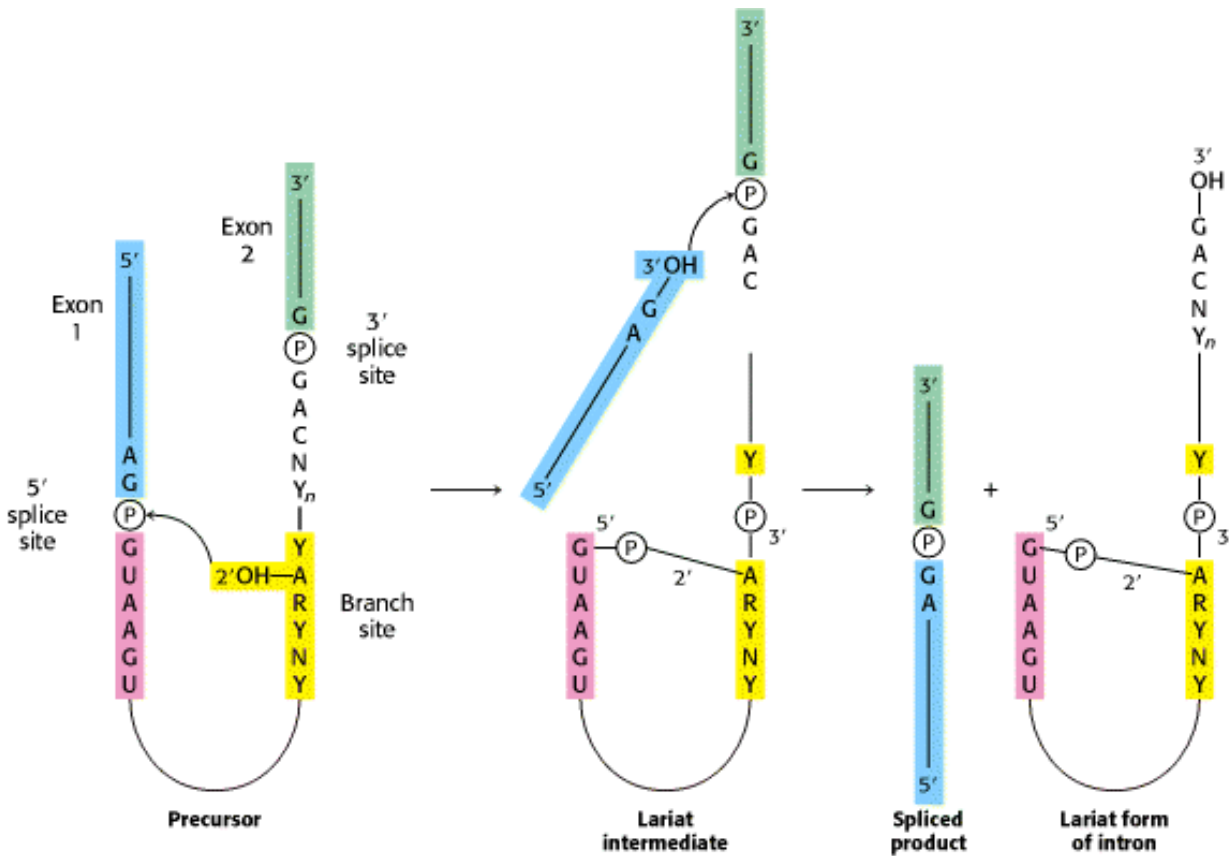


Figure 28.29. Splicing Mechanism Used for mRNA Precursors. The upstream (5[′]) exon is shown in blue, the downstream (3[′]) exon in green, and the branch site in yellow. Y stands for a purine nucleotide, R for a pyrimidine nucleotide, and N for any nucleotide. The 5[′] splice site is attacked by the 2[′]-OH group of the branch-site adenosine residue. The 3[′] splice site is attacked by the newly formed 3[′]-OH group of the upstream exon. The exons are joined, and the intron is released in the form of a lariat. [After P. A. Sharp. *Cell* 2(1985):3980.]

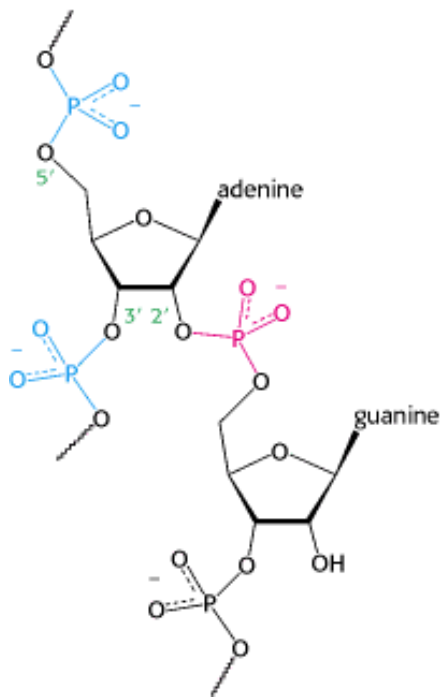


Figure 28.30. Splicing Branch Point. The structure of the branch point in the lariat intermediate in which the adenylate residue is joined to three nucleotides by phosphodiester bonds. The new 2'-to-5' linkage is shown in red, and the usual 3'-to-5' linkages are shown in blue.

Table 28.3. Small nuclear ribonucleoprotein particles (snRNPs) in the splicing of mRNA precursors

snRNP	Size of snRNA(nucleotides)	Role
U1	165	Binds the 5' splice site and then the 3' splice site
U2	185	Binds the branch site and forms part of the catalytic center
U5	116	Binds the 5' splice site
U4	145	Masks the catalytic activity of U6
U6	106	Catalyzes splicing

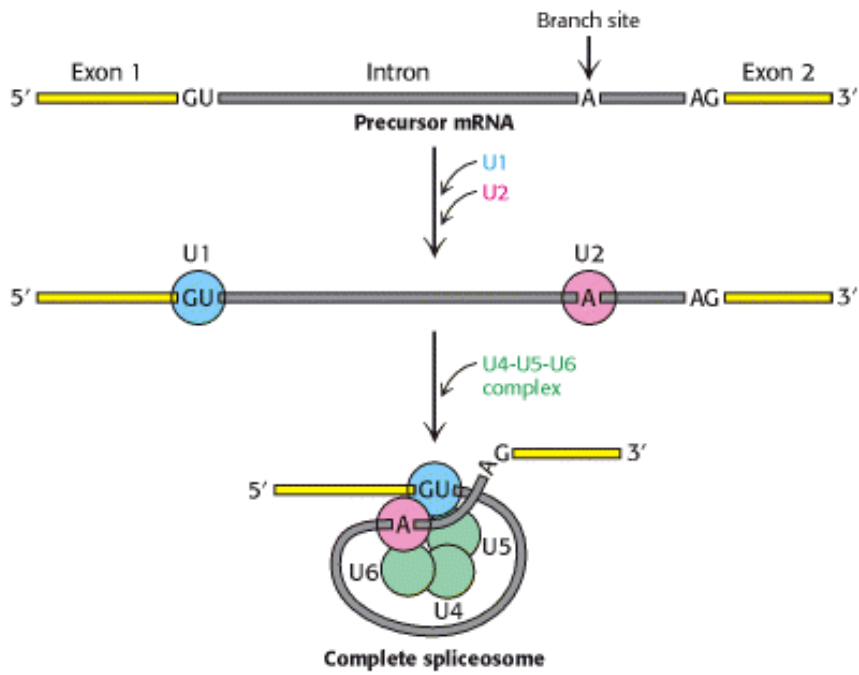


Figure 28.31. Spliceosome Assembly. U1 (blue) binds the 5' splice site and U2 (red) to the branch point. A preformed U4-U5-U6 complex then joins the assembly to form the complete spliceosome.

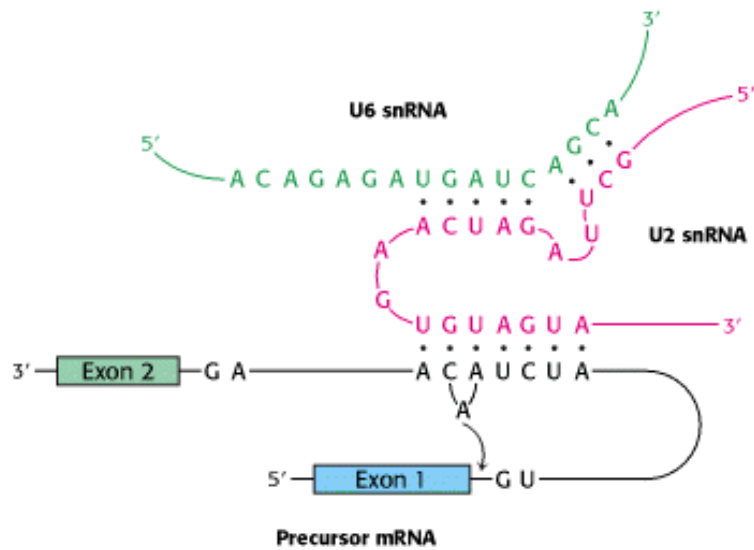


Figure 28.32. Splicing Catalytic Center. The catalytic center of the spliceosome is formed by U2 snRNA (red) and U6 snRNA (green), which are base paired. U2 is also base paired to the branch site of the mRNA precursor. [After H. D. Madhani and C. Guthrie. *Cell* 71(1992):803.]

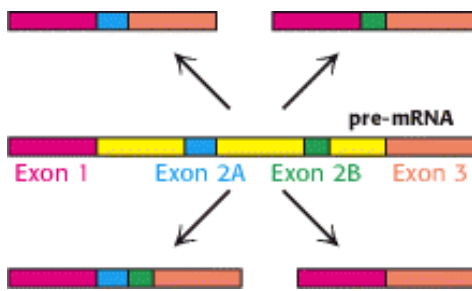


Figure 28.33. Alternative Splicing Patterns. A pre-mRNA with multiple exons is sometimes spliced in different ways. Here, with two alternative exons (exons 2A and 2B) present, the mRNA can be produced with neither, either, or both exons included. More complex alternative splicing patterns also are possible.

Table 28.4. Selected proteins exhibiting alternative RNA splicing

Actin
Alcohol dehydrogenase
Aldolase
K-ras
Calcitonin
Fibrinogen
Fibronectin
Myosin
Nerve growth factor
Tropomyosin
Troponin

Source: R. E. Breitbart, A. Andreadis, and B. Nadal-Ginard. *Annu. Rev. Biochem.* 56(1987):467–495.

28.4. The Discovery of Catalytic RNA Was Revealing in Regard to Both Mechanism and Evolution

RNAs form a surprisingly versatile class of molecules. As we have seen, splicing is catalyzed largely by RNA molecules, with proteins playing a secondary role. Another enzyme that contains a key RNA component is ribonuclease P (RNase P), which catalyzes the maturation of tRNA by removing nucleotides from the 5' end of the precursor molecule (Section 28.1.8). Finally, as we shall see in [Chapter 29](#), the RNA component of ribosomes is the catalyst that carries out protein synthesis.

The versatility of RNA first became clear from observations regarding the processing carried out on ribosomal RNA in a single-cell eukaryote. In *Tetrahymena* (a ciliated protozoan), a 414-nucleotide intron is removed from a 6.4-kb precursor to yield the mature 26S rRNA molecule ([Figure 28.34](#)). In an elegant series of studies of this splicing reaction, Thomas Cech and his coworkers established that the RNA spliced itself to precisely excise the 414-nucleotide intron. These remarkable experiments demonstrated that an RNA molecule can *splice itself* in the absence of protein and, indeed, can have highly specific catalytic activity.

The *self-splicing* reaction requires an added guanosine nucleotide. Nucleotides were originally included in the reaction mixture because it was thought that ATP or GTP might be needed as an energy source. Instead, the nucleotides were found to be necessary as cofactors. The required cofactor proved to be a guanosine unit, in the form of guanosine, GMP, GDP, or GTP. G (denoting any one of these species) serves not as an energy source but as an attacking group that becomes transiently incorporated into the RNA (see [Figure 28.34](#)). G binds to the RNA and then attacks the 5' splice site to form a phosphodiester bond with the 5' end of the intron. This transesterification reaction generates a 3'-OH group at the end of the upstream exon. This newly attached 3'-OH group then attacks the 3' splice. This second transesterification reaction joins the two exons and leads to the release of the 414-nucleotide intron.


Self-splicing depends on the structural integrity of the rRNA precursor. Much of the intron is needed for self-splicing. This molecule, like many RNAs, has a folded structure formed by many double-helical stems and loops ([Figure 28.35](#)). Examination of the three-dimensional structure determined by x-ray crystallography reveals a compact folding structure that is in many ways analogous to the structures of protein enzymes. A welldefined pocket for binding the guanosine is formed within the structure.

Analysis of the base sequence of the rRNA precursor suggested that the 5' splice site is aligned with the catalytic residues by base-pairing between a *pyrimidine-rich region* (CUCUCU) of the upstream exon and a *purine-rich guide sequence* (GGGAGG) within the intron ([Figure 28.36](#)). The intron brings together the guanosine cofactor and the 5' splice site so that the 3'-OH group of G can nucleophilically attack the phosphorus atom at this splice site. Another part of the intron then holds the downstream exon in position for attack by the newly formed 3'-OH group of the upstream exon. A phosphodiester bond is formed between the two exons, and the intron is released as a linear molecule. Like catalysis by protein enzymes, selfcatalysis of bond formation and breakage in this rRNA precursor is highly specific.

The finding of enzymatic activity in the self-splicing intron and in the RNA component of RNase P has opened new areas of inquiry and changed the way in which we think about molecular evolution. The discovery that RNA can be a catalyst as well as an information carrier suggests that an RNA world may have existed early in the evolution of life, before the appearance of DNA and protein ([Section 2.2.2](#)).

Messenger RNA precursors in the mitochondria of yeast and fungi also undergo self-splicing, as do some RNA precursors in the chloroplasts of unicellular organisms such as *Chlamydomonas*. Self-splicing reactions can be classified according to the nature of the unit that attacks the upstream splice site. Group I self-splicing is mediated by a guanosine cofactor, as in *Tetrahymena*. The attacking moiety in group II splicing is the 2'-OH group of a specific adenylate of the intron ([Figure 28.37](#)).

Group I and group II self-splicing resembles spliceosome-catalyzed splicing in two respects. First, in initial step, a ribose hydroxyl group attacks the 5' splice site. The newly formed 3'-OH terminus of the upstream exon then attacks the 3' splice site to form a phosphodiester bond with the downstream exon. Second, both reactions are transesterifications in which the phosphate moieties at each splice site are retained in the products. The number of phosphodiester bonds stays constant. Group II splicing is like the spliceosome-catalyzed splicing of mRNA precursors in several additional ways. The attack at the 5' splice site is carried out by a part of the intron itself (the 2'-OH group of adenosine) rather than by an external cofactor (G). In both cases, the intron is released in the form of a lariat. Moreover, in some instances, the group II intron is transcribed in pieces that assemble through hydrogen bonding to the catalytic intron, in a manner analogous to the assembly of the snRNAs in the spliceosome.

 These similarities have led to the suggestion that the spliceosome-catalyzed splicing of mRNA precursors evolved from RNA-catalyzed self-splicing. Group II splicing may well be an intermediate between group I splicing and the splicing in the nuclei of higher eukaryotes. *A major step in this transition was the transfer of catalytic power from the intron itself to other molecules.* The formation of spliceosomes gave genes a new freedom because introns were no longer constrained to provide the catalytic center for splicing. Another advantage of external catalysts for splicing is that they can be more readily regulated. However, it is important to note that similarities do not establish ancestry. The similarities between group II introns and mRNA splicing may be a result of convergent evolution. Perhaps there are only

a limited number of ways to carry out efficient, specific intron excision. The determination of whether these similarities stem from ancestry or from chemistry will require expanding our understanding of RNA biochemistry.

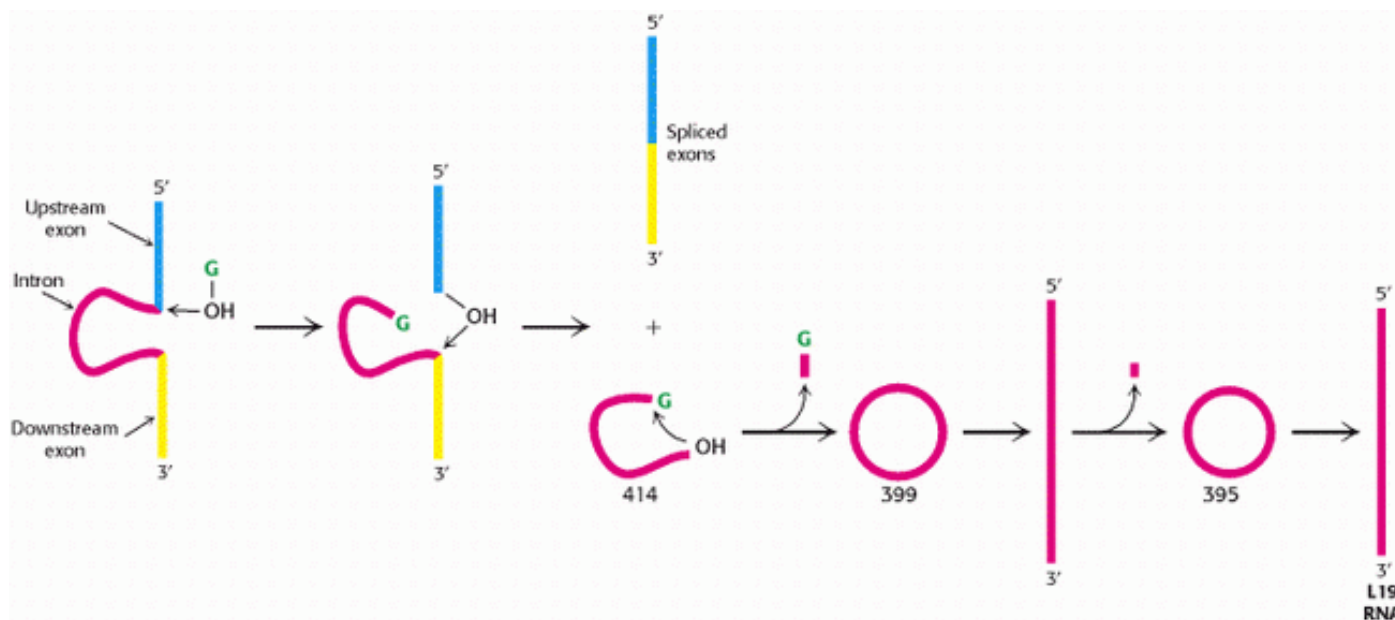


Figure 28.34. Self-Splicing. A ribosomal RNA precursor from *Tetrahymena* splices itself in the presence of a guanosine co-factor (G, shown in green). A 414-nucleotide intron (red) is released in the first splicing reaction. This intron then splices itself twice again to produce a linear RNA that has lost a total of 19 nucleotides. This L19 RNA is catalytically active. [After T. Cech. RNA as an enzyme. Copyright © 1986 by Scientific American, Inc. All rights reserved.]

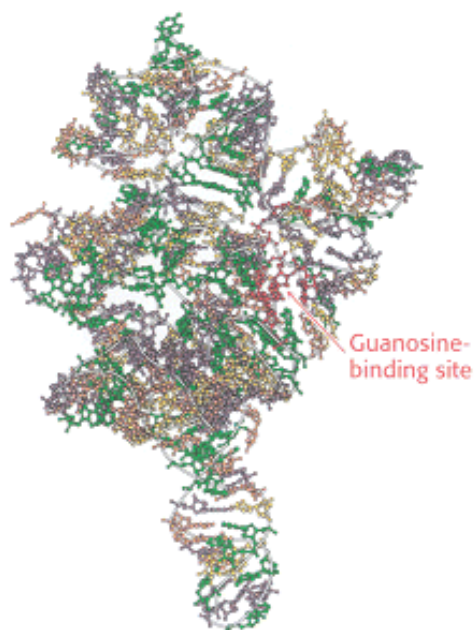


Figure 28.35. Structure of a Self-Splicing Intron. The structure of a large fragment of the self-splicing intron from *Tetrahymena* reveals a complex folding pattern of helices and loops. Bases are shown in green, A; yellow, C; purple, G; and orange, U.

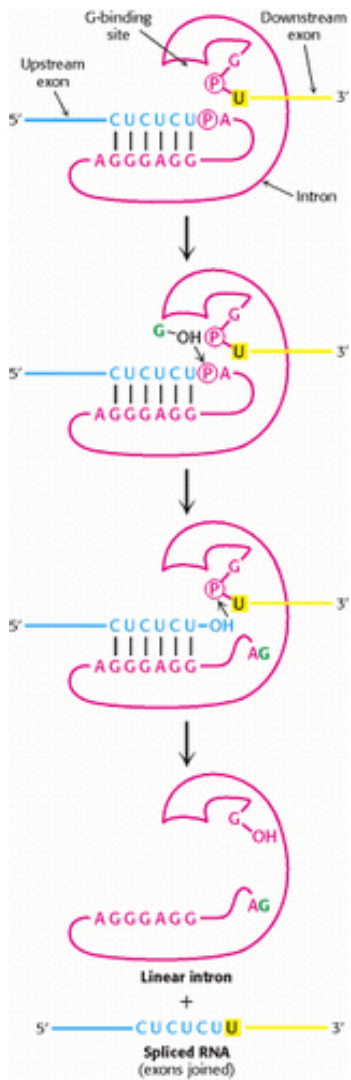
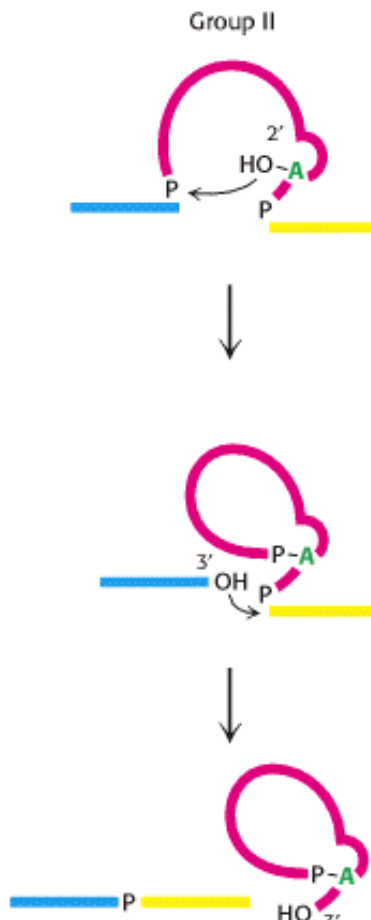
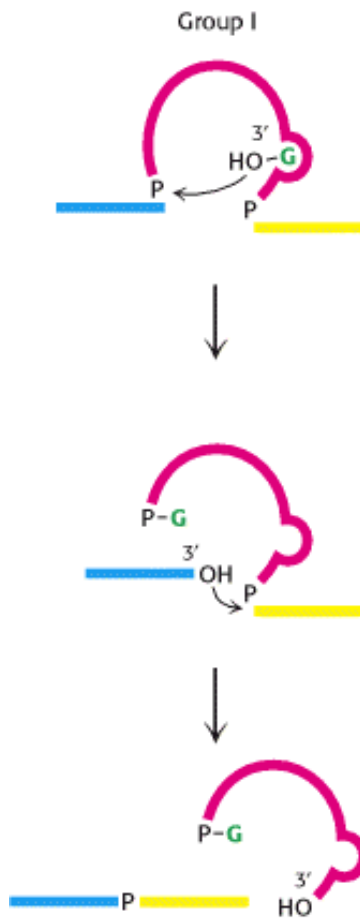


Figure 28.36. Self-Splicing Mechanism. The catalytic mechanism of the selfsplicing intron from *Tetrahymena* includes a series of transesterification reactions. [After T. Cech. RNA as an enzyme. Copyright © 1986 by Scientific American, Inc. All rights reserved.]

SELF-SPICING INTRONS



SPLICEOSOME-CATALYZED SPICING OF NUCLEAR mRNA

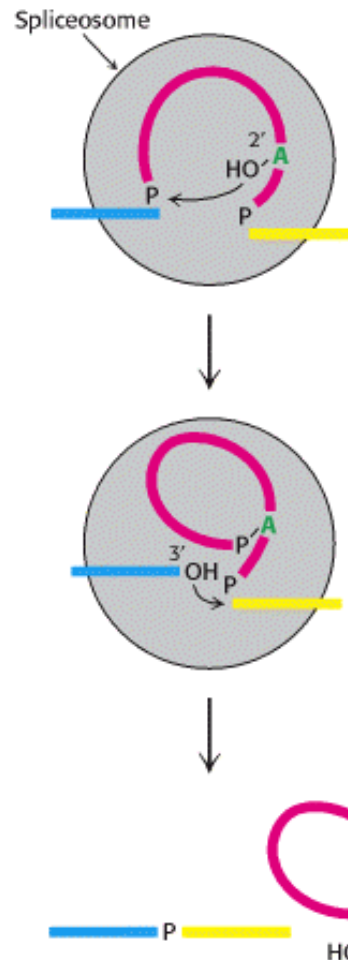


Figure 28.37. Comparison of Splicing Pathways. The exons being joined are shown in blue and yellow and the attacking unit is shown in green. The catalytic site is formed by the intron itself (red) in group I and group II splicing. In contrast, the splicing of nuclear mRNA precursors is catalyzed by snRNAs and their associated proteins in the spliceosome. [After P. A. Sharp. *Science* 235(1987):769.]

Summary

Transcription Is Catalyzed by RNA Polymerase

All cellular RNA molecules are synthesized by RNA polymerases according to instructions given by DNA templates. The activated monomer substrates are ribonucleoside triphosphates. The direction of RNA synthesis is $5' \rightarrow 3'$, as in DNA synthesis. RNA polymerases, unlike DNA polymerases, do not need a primer and do not possess proofreading nuclease activity.

RNA polymerase in *E. coli* is a multisubunit enzyme. The subunit composition of the ~500-kd holoenzyme is $\alpha_2 \beta \beta' \sigma$ and that of the core enzyme is $\alpha_2 \beta \beta'$. Transcription is initiated at promoter sites consisting of two sequences, one centered near -10 and the other near -35; that is, 10 and 35 nucleotides away from the start site in the $5'$ (upstream) direction. The consensus sequence of the -10 region is TATAAT. The σ subunit enables the holoenzyme to recognize promoter sites. When the growth temperature is raised, *E. coli* expresses a special σ that selectively binds the distinctive promoter of heat-shock genes. RNA polymerase must unwind the template double helix for transcription to take place. Unwinding exposes some 17 bases on the template strand and sets the stage for the formation of the first phosphodiester

bond. RNA chains usually start with pppG or pppA. The σ subunit dissociates from the holoenzyme after the initiation of the new chain. Elongation takes place at transcription bubbles that move along the DNA template at a rate of about 50 nucleotides per second. The nascent RNA chain contains stop signals that end transcription. One stop signal is an RNA hairpin, which is followed by several U residues. A different stop signal is read by the *rho* protein, an ATPase. In *E. coli*, precursors of transfer RNA and ribosomal RNA are cleaved and chemically modified after transcription, whereas mRNA is used unchanged as a template for protein synthesis.

Eukaryotic Transcription and Translation Are Separated in Space and Time

RNA synthesis in eukaryotes takes place in the nucleus, whereas protein synthesis takes place in the cytoplasm. There are three types of RNA polymerase in the nucleus: RNA polymerase I makes ribosomal RNA precursors, II makes messenger RNA precursors, and III makes transfer RNA precursors. Eukaryotic promoters are complex, being composed of several different elements. Promoters for RNA polymerase II are located on the 5' side of the start site for transcription. Each consists of a TATA box centered between -30 and -100 and additional upstream sequences. They are recognized by proteins called transcription factors rather than by RNA polymerase II. The saddle-shaped TATA-box-binding protein unwinds and sharply bends DNA at TATA-box sequences and serves as a focal point for the assembly of transcription complexes. The TATA-box-binding protein initiates the assembly of the active transcription complex. The activity of many promoters is greatly increased by enhancer sequences that have no promoter activity of their own. Enhancer sequences can act over distances of several kilobases, and they can be located either upstream or downstream of a gene.

The Transcription Products of All Three Eukaryotic Polymerases Are Processed

The 5' ends of mRNA precursors become capped and methylated in the course of transcription. A 3' poly(A) tail is added to most mRNA precursors after the nascent chain has been cleaved by an endonuclease. RNA editing processes alter the nucleotide sequence of some mRNAs, such as the one for apolipoprotein B.

The splicing of mRNA precursors is carried out by spliceosomes, which consist of small nuclear ribonucleoprotein particles (snRNPs). Splice sites in mRNA precursors are specified by sequences at ends of introns and by branch sites near their 3' ends. The 2'-OH group of an adenosine residue in the branch site attacks the 5' splice site to form a lariat intermediate. The newly generated 3'-OH terminus of the upstream exon then attacks the 3' splice site to become joined to the downstream exon. Splicing thus consists of two transesterification reactions, with the number of phosphodiester bonds remaining constant during reactions. Small nuclear RNAs in spliceosomes catalyze the splicing of mRNA precursors. In particular, U2 and U6 snRNAs form the active centers of spliceosomes.

The Discovery of Catalytic RNA Was Revealing with Regard to Both Mechanism And Evolution

Some RNA molecules, such as the 26S ribosomal RNA precursor from *Tetrahymena*, undergo self-splicing in the absence of protein. A self-modified version of this rRNA intron displays true catalytic activity. Spliceosome-catalyzed splicing may have evolved from self-splicing. The discovery of catalytic RNA has opened new vistas in our exploration of early stages of molecular evolution.

Key Terms

transcription

RNA polymerase

promoter sites

transcription factor

footprinting

consensus sequence

sigma subunit

transcription bubble

rho (ρ) protein

TATA box

enhancer

pre-mRNA

5' cap

poly(A) tail

RNA editing

RNA splicing

spliceosome

small nuclear RNA (snRNA)

alternative splicing

catalytic RNA

self-splicing

Problems

1. *Complements.* The sequence of part of an mRNA is

5'-AUGGGGAACAGCAAGAGUGGGGCCUGUCCAAGGAG-3'

What is the sequence of the DNA coding strand? Of the DNA template strand?

See answer

2. *Checking for errors.* Why is RNA synthesis not as carefully monitored for errors as is DNA synthesis?

See answer

3. *Speed is not of the essence.* Why is it advantageous for DNA synthesis to be more rapid than RNA synthesis?

See answer

4. *Potent inhibitor.* Heparin inhibits transcription by binding to RNA polymerase. What properties of heparin allow it to bind so effectively to RNA polymerase?

See answer

5. *A loose cannon.* Sigma protein by itself does not bind to promoter sites. Predict the effect of a mutation enabling σ to bind to the -10 region in the absence of other subunits of RNA polymerase.

See answer

6. *Stuck sigma.* What would be the likely effect of a mutation that would prevent σ from dissociating from the RNA polymerase core?

See answer

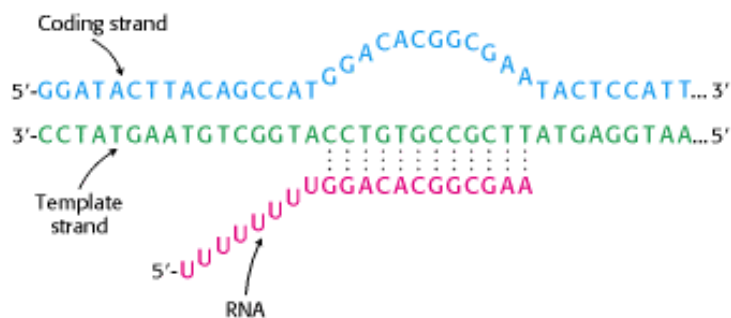
7. *Transcription time.* What is the minimum length of time required for the synthesis by *E. coli* polymerase of an mRNA encoding a 100-kd protein?

See answer

8. *Between bubbles.* How far apart are transcription bubbles on *E. coli* genes that are being transcribed at a maximal rate?

See answer

9. *A revealing bubble.* Consider the synthetic RNA-DNA transcription bubble illustrated here. Let us refer to the coding DNA strand, the template strand, and the RNA strand as strands 1, 2, and 3, respectively.



(a) Suppose that strand 3 is labeled with ^{32}P at its 5' end and that polyacrylamide gel electrophoresis is carried out under nondenaturing conditions. Predict the autoradiographic pattern for (i) strand 3 alone, (ii) strands 1 and 3, (iii) strands 2 and 3, (iv) strands 1, 2, and 3, and (v) strands 1, 2, and 3 and core RNA polymerase.

(b) What is the likely effect of rifampicin on RNA synthesis in this system?

(c) Heparin blocks elongation of the RNA primer if it is added to core RNA polymerase before the onset of transcription but not if added after transcription starts. Account for this difference.

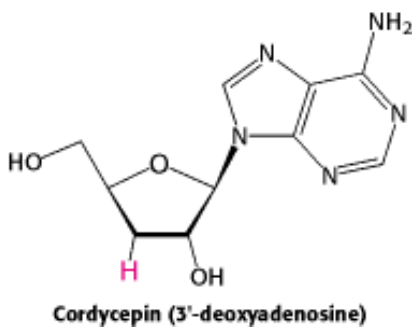
(d) Suppose that synthesis is carried out in the presence of ATP, CTP, and UTP. Compare the length of the longest product obtained with that expected when all four ribonucleoside triphosphates are present.

See answer

10. *Abortive cycling.* Di- and trinucleotides are occasionally released from RNA polymerase at the very start of transcription, a process called abortive cycling. This process requires the restart of transcription. Suggest a plausible explanation for abortive cycling.

See answer

11. *Polymerase inhibition.* Cordycepin inhibits poly(A) synthesis at low concentrations and RNA synthesis at higher concentrations.



- (a) What is the basis of inhibition by cordycepin?
- (b) Why is poly(A) synthesis more sensitive to the presence of cordycepin?
- (c) Does cordycepin need to be modified to exert its effect?

See answer

12. *An extra piece.* What is the amino acid sequence of the extra segment of protein synthesized in a thalassemic patient having a mutation leading to aberrant splicing (see [Figure 28.28](#))? The reading frame after the splice site begins with TCT.

See answer

13. *A long-tailed messenger.* Another thalassemic patient had a mutation leading to the production of an mRNA for the β chain of hemoglobin that was 900 nucleotides longer than the normal one. The poly(A) tail of this mutant mRNA was located a few nucleotides after the only AAUAAA sequence in the additional sequence. Propose a mutation that would lead to the production of this altered mRNA.

See answer

Mechanism Problem

14. *RNA editing.* Many uridine molecules are inserted into some mitochondrial mRNAs in trypanosomes. The uridine residues come from the poly(U) tail of a donor strand. Nucleoside triphosphates do not participate in this reaction. Propose a reaction mechanism that accounts for these findings. (Hint: Relate RNA editing to RNA splicing.)

See answer

Chapter Integration Problems

15. *Proteome complexity.* What processes considered in this chapter make the proteome more complex than the genome? What processes might further enhance this complexity?

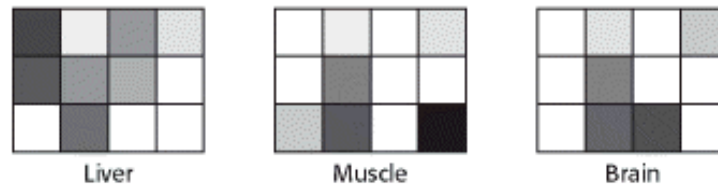
See answer

16. *Separation technique.* Suggest a means by which you could separate mRNA from the other types of RNA in a eukaryotic cell.

See answer

Data Interpretation Problems

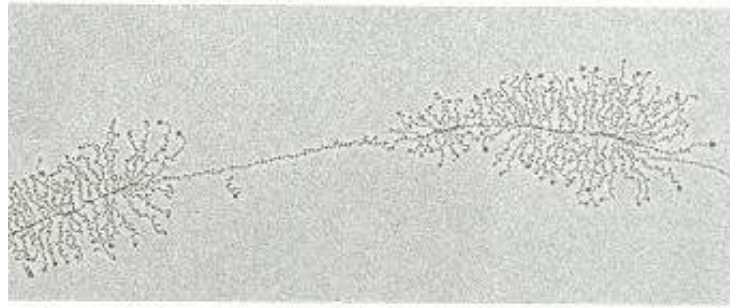
17. *Run-off experiment.* Nuclei were isolated from brain, liver, and muscle. The nuclei were then incubated with α - ^{32}P -UTP under conditions that allow RNA synthesis, except that an inhibitor of RNA initiation was present. The radioactive RNA was isolated and annealed to various DNA sequences that had been attached to a gene chip. In the adjoining graphs, the intensity of the shading indicates roughly how much mRNA was attached to each RNA sequence.



- (a) Why does the intensity of hybridization differ between genes?
- (b) What is the significance of the fact that some of the RNA molecules display different hybridization patterns in different tissues?
- (c) Some genes are expressed in all three tissues. What would you guess is the nature of these genes?
- (d) Suggest a reason why an initiation inhibitor was included in the reaction mixture.

See answer

18. *Christmas trees*. The adjoining autoradiograph depicts several bacterial genes undergoing transcription. Identify the DNA. What are the strands of increasing length? Where is the beginning of transcription? The end of transcription? On the page, what is the direction of RNA synthesis? What can you conclude about the number of enzymes participating in RNA synthesis on a given gene?



See answer

Selected Readings

Where to begin

- N.A. Woychik. 1998. Fractions to functions: RNA polymerase II thirty years later *Cold Spring Harbor Symp. Quant. Biol.* 63: 311-317. ([PubMed](#))
- R. Losick. 1998. Summary: Three decades after sigma *Cold Spring Harbor Symp. Quant. Biol.* 63: 653-666. ([PubMed](#))
- J.E. Darnell Jr.. 1985. RNA *Sci. Am.* 253: (4) 68-78. ([PubMed](#))
- T.R. Cech. 1986. RNA as an enzyme *Sci. Am.* 255: (5) 64-75. ([PubMed](#))
- P.A. Sharp. 1994. Split genes and RNA splicing *Cell* 77: 805- 815. ([PubMed](#))
- T.R. Cech. 1990. Self-splicing and enzymatic activity of an intervening sequence RNA from *Tetrahymena* *Biosci. Rep.* 10: 239-261. ([PubMed](#))
- C. Guthrie. 1991. Messenger RNA splicing in yeast: Clues to why the spliceosome is a ribonucleoprotein *Science* 253: 157-163. ([PubMed](#))

Books

- Lewin, B., 2000. *Genes* (7th ed.). Oxford University Press.
- Kornberg, A., and Baker, T. A., 1992. *DNA Replication* (2d ed.). W. H. Freeman and Company.
- Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J., 2000. *Molecular Cell Biology* (4th ed.). W. H. Freeman and Company.
- Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M., 1987. *Molecular Biology of the Gene* (4th ed.). Benjamin Cummings.
- Gesteland, R. F., Cech, T., and Atkins, J. F. (Eds.), 1999. *The RNA World* (2d ed.). Cold Spring Harbor Laboratory Press.

RNA polymerases

P. Cramer, D.A. Bushnell, and R.D. Kornberg. 2001. Structural basis of transcription: RNA polymerase II at 2.8 Å resolution *Science* 292: 1863-1875. ([PubMed](#))

A.L. Gnatt, P. Cramer, J. Fu, D.A. Bushnell, and R.D. Kornberg. 2001. Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution *Science* 292: 1876-1882. ([PubMed](#))

G. Zhang, E.A. Campbell, L. Minakhin, C. Richter, K. Severinov, and S.A. Darst. 1999. Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution *Cell* 98: 811-824. ([PubMed](#))

E.A. Campbell, N. Korzheva, A. Mustaev, K. Murakami, S. Nair, A. Goldfarb, and S.A. Darst. 2001. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase *Cell* 104: 901-912. ([PubMed](#))

G.M. Cheetham and T.A. Steitz. 1999. Structure of a transcribing T7 RNA polymerase initiation complex *Science* 286: 2305-2309. ([PubMed](#))

R.H. Ebright. 2000. RNA polymerase: Structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II *J. Mol. Biol.* 304: 687-698. ([PubMed](#))

M.R. Paule and R.J. White. 2000. Survey and summary: Transcription by RNA polymerases I and III *Nucleic Acids Res.* 28: 1283-1298. ([PubMed](#)) ([Full Text in PMC](#))

Initiation and elongation

S. Buratowski. 2000. Snapshots of RNA polymerase II transcription initiation *Curr. Opin. Cell Biol.* 12: 320-325. ([PubMed](#))

J.W. Conaway and R.C. Conaway. 1999. Transcription elongation and human disease *Annu. Rev. Biochem.* 68: 301-319. ([PubMed](#))

J.W. Conaway, A. Shilatifard, A. Dvir, and R.C. Conaway. 2000. Control of elongation by RNA polymerase II *Trends Biochem. Sci.* 25: 375-380. ([PubMed](#))

N. Korzheva, A. Mustaev, M. Kozlov, A. Malhotra, V. Nikiforov, A. Goldfarb, and S.A. Darst. 2000. A structural model of transcription elongation *Science* 289: 619-625. ([PubMed](#))

D. Reines, R.C. Conaway, and J.W. Conaway. 1999. Mechanism and regulation of transcriptional elongation by RNA polymerase II *Curr. Opin. Cell Biol.* 11: 342-346. ([PubMed](#))

Promoters, enhancers, and transcription factors

M. Merika and D. Thanos. 2001. Enhanceosomes *Curr. Opin. Genet. Dev.* 11: 205-208. ([PubMed](#))

J.M. Park, B.S. Gim, J.M. Kim, J.H. Yoon, H.S. Kim, J.G. Kang, and Y.J. Kim. 2001. *Drosophila* mediator complex is broadly utilized by diverse gene-specific transcription factors at different types of core promoters *Mol. Cell. Biol.* 21: 2312-2323. ([PubMed](#)) ([Full Text in PMC](#))

S. Fiering, E. Whitelaw, and D.I. Martin. 2000. To be or not to be active: The stochastic nature of enhancer action *Bioessays* 22: 381-387. ([PubMed](#))

M. Hampsey and D. Reinberg. 1999. RNA polymerase II as a control panel for multiple coactivator complexes *Curr. Opin. Genet. Dev.* 9: 132-139. ([PubMed](#))

L. Chen. 1999. Combinatorial gene regulation by eukaryotic transcription factors *Curr. Opin. Struct. Biol.* 9: 48-55. ([PubMed](#))

C.W. Muller. 2001. Transcription factors: Global and detailed views *Curr. Opin. Struct. Biol.* 11: 26-32. ([PubMed](#))

H. Sakurai and T. Fukasawa. 2000. Functional connections between mediator components and general transcription factors of *Saccharomyces cerevisiae* *J. Biol. Chem.* 275: 37251-37256. ([PubMed](#))

P. Droge and B. Muller-Hill. 2001. High local protein concentrations at promoters: Strategies in prokaryotic and eukaryotic cells *Bioessays* 23: 179-183. ([PubMed](#))

J.W. Fickett and A.G. Hatzigeorgiou. 1997. Eukaryotic promoter recognition *Genome Res.* 7: 861-878. ([PubMed](#))

S.T. Smale, A. Jain, J. Kaufmann, K.H. Emami, K. Lo, and I.P. Garraway. 1998. The initiator element: A paradigm for core promoter heterogeneity within metazoan protein-coding genes *Cold Spring Harbor Symp. Quant. Biol.* 63: 21-31. ([PubMed](#))

Y. Kim, J.H. Geiger, S. Hahn, and P.B. Sigler. 1993. Crystal structure of a yeast TBP/TATA-box complex *Nature* 365: 512-520. ([PubMed](#))

J.L. Kim, D.B. Nikolov, and S.K. Burley. 1993. Co-crystal structure of TBP recognizing the minor groove of a TATA element *Nature* 365: 520-527. ([PubMed](#))

R.J. White and S.P. Jackson. 1992. The TATA-binding protein: A central role in transcription by RNA polymerases I, II and III *Trends Genet.* 8: 284-288. ([PubMed](#))

Termination

B.R. Burgess and J.P. Richardson. 2001. RNA passes through the hole of the protein hexamer in the complex with *Escherichia coli* Rho factor *J. Biol. Chem.* 276: 4182-4189. ([PubMed](#))

X. Yu, T. Horiguchi, K. Shigesada, and E.H. Egelman. 2000. Three-dimensional reconstruction of transcription termination factor rho: Orientation of the N-terminal domain and visualization of an RNA-binding site *J. Mol. Biol.* 299: 1279-1287. ([PubMed](#))

B.L. Stitt. 2001. *Escherichia coli* transcription termination factor Rho binds and hydrolyzes ATP using a single class of three sites *Bio-chemistry* 40: 2276-2281. ([PubMed](#))

T.M. Henkin. 2000. Transcription termination control in bacteria *Curr. Opin. Microbiol.* 3: 149-153. ([PubMed](#))

I. Gusarov and E. Nudler. 1999. The mechanism of intrinsic transcription termination *Mol. Cell.* 3: 495-504. ([PubMed](#))

5'-Cap formation and polyadenylation

A.J. Shatkin and J.L. Manley. 2000. The ends of the affair: Capping and polyadenylation *Nat. Struct. Biol.* 7: 838-842. ([PubMed](#))

T.S. Ro-Choi. 1999. Nuclear snRNA and nuclear function (discovery of 5' cap structures in RNA) *Crit. Rev. Eukaryotic Gene Expr.* 9: 107-158. ([PubMed](#))

S. Shuman, Y. Liu, and B. Schwer. 1994. Covalent catalysis in nucleotidyl transfer reactions: Essential motifs in *Saccharomyces cerevisiae* RNA capping enzyme are conserved in *Schizosaccharomyces pombe* and viral capping enzymes and among polynucleotide ligases *Proc. Natl. Acad. Sci. U. S. A.* 91: 12046-12050. ([PubMed](#)) ([Full Text in PMC](#))

J. Bard, A.M. Zhelkovsky, S. Helmling, T.N. Earnest, C.L. Moore, and A. Bohm. 2000. Structure of yeast poly(A) polymerase alone and in complex with 3'-dATP *Science* 289: 1346-1349. ([PubMed](#))

G. Martin, W. Keller, and S. Doublié. 2000. Crystal structure of mammalian poly(A) polymerase in complex with an analog of ATP *EMBO J.* 19: 4193-4203. ([PubMed](#))

J. Zhao, L. Hyman, and C. Moore. 1999. Formation of mRNA 3' ends in eukaryotes: Mechanism, regulation, and interrelationships with other steps in mRNA synthesis *Microbiol. Mol. Biol. Rev.* 63: 405-445. ([PubMed](#)) ([Full Text in PMC](#))

L. Minvielle-Sebastia and W. Keller. 1999. mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription *Curr. Opin. Cell Biol.* 11: 352-357. ([PubMed](#))

E. Wahle and U. Kuhn. 1997. The mechanism of 3' cleavage and polyadenylation of eukaryotic pre-mRNA *Prog. Nucleic Acid Res. Mol. Biol.* 57: 41-71. ([PubMed](#))

RNA editing

J.M. Gott and R.B. Emeson. 2000. Functions and mechanisms of RNA editing *Annu. Rev. Genet.* 34: 499-531. ([PubMed](#))

L. Simpson, O.H. Thiemann, N.J. Savill, J.D. Alfonzo, and D.A. Maslov. 2000. Evolution of RNA editing in trypanosome mitochondria *Proc. Natl. Acad. Sci. USA* 97: 6986-6993. ([PubMed](#)) ([Full Text in PMC](#))

A. Chester, J. Scott, S. Anant, and N. Navaratnam. 2000. RNA editing: Cytidine to uridine conversion in apolipoprotein B mRNA *Biochim. Biophys. Acta* 1494: 1-3. ([PubMed](#))

S. Maas and A. Rich. 2000. Changing genetic information through RNA editing *Bioessays* 22: 790-802. ([PubMed](#))

Splicing of mRNA precursors

H. Stark, P. Dube, R. Luhrmann, and B. Kastner. 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle *Nature* 409: 539-542. ([PubMed](#))

E.E. Strehler and D.A. Zacharias. 2001. Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps *Physiol. Rev.* 81: 21-50. ([PubMed](#))

B.R. Graveley. 2001. Alternative splicing: Increasing diversity in the proteomic world *Trends Genet.* 17: 100-107. ([PubMed](#))

A. Newman. 1998. RNA splicing *Curr. Biol.* 8: R903-R905. ([PubMed](#))

R. Reed. 2000. Mechanisms of fidelity in pre-mRNA splicing *Curr. Opin. Cell Biol.* 12: 340-345. ([PubMed](#))

J.E. Sleeman and A.I. Lamond. 1999. Nuclear organization of pre-mRNA splicing factors *Curr. Opin. Cell Biol.* 11: 372-377. ([PubMed](#))

D.L. Black. 2000. Protein diversity from alternative splicing: A challenge for bioinformatics and post-genome biology *Cell* 103: 367-370. ([PubMed](#))

C.A. Collins and C. Guthrie. 2000. The question remains: Is the spliceosome a ribozyme? *Nat. Struct. Biol.* 7: 850-854. ([PubMed](#))

Self-splicing and RNA catalysis

C. Carola and F. Eckstein. 1999. Nucleic acid enzymes *Curr. Opin. Chem. Biol.* 3: 274-283. ([PubMed](#))

E.A. Doherty and J.A. Doudna. 2000. Ribozyme structures and mechanisms *Annu. Rev. Biochem.* 69: 597-615. ([PubMed](#))

M.J. Fedor. 2000. Structure and function of the hairpin ribozyme *J. Mol. Biol.* 297: 269-291. ([PubMed](#))

R. Hanna and J.A. Doudna. 2000. Metal ions in ribozyme folding and catalysis *Curr. Opin. Chem. Biol.* 4: 166-170. ([PubMed](#))

W.G. Scott. 1998. RNA catalysis *Curr. Opin. Struct. Biol.* 8: 720-726. ([PubMed](#))

W.G. Scott and A. Klug. 1996. Ribozymes: Structure and mechanism in RNA catalysis *Trends Biochem. Sci.* 21: 220-224. ([PubMed](#))

T.R. Cech, D. Herschlag, J.A. Piccirilli, and A.M. Pyle. 1992. RNA catalysis by a group I ribozyme: Developing a model for transition state stabilization *J. Biol. Chem.* 267: 17479-17482. ([PubMed](#))

D. Herschlag and T.R. Cech. 1990. Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme 1: Kinetic description of the reaction of an RNA substrate complementary to the active site *Biochemistry* 29: 10159-10171. ([PubMed](#))

J.A. Piccirilli, J.S. Vyle, M.H. Caruthers, and T.R. Cech. 1993. Metal ion catalysis in the *Tetrahymena* ribozyme reaction *Nature* 361: 85-88. ([PubMed](#))

J.F. Wang, W.D. Downs, and T.R. Cech. 1993. Movement of the guide sequence during RNA catalysis by a group I ribozyme *Science* 260: 504-508. ([PubMed](#))

29. Protein Synthesis

Genetic information is most important because of the proteins that it encodes, in that proteins play most of the functional roles in cells. In [Chapters 27](#) and [28](#), we examined how DNA is replicated and how DNA is transcribed into RNA. We now turn to the mechanism of protein synthesis, a process called *translation* because the four-letter alphabet of nucleic acids is translated into the entirely different twenty-letter alphabet of proteins. Translation is a conceptually more complex process than either replication or transcription, both of which take place within the framework of a common base-pairing language. Befitting its position linking the nucleic acid and protein languages, the process of protein synthesis depends critically on both nucleic acid and protein factors. Protein synthesis takes place on *ribosomes*—enormous complexes containing three large RNA molecules and more than 50 proteins. One of the great triumphs in biochemistry in recent years has been the determination of the structure of the ribosome and its components so that its function can be examined in atomic detail ([Figure 29.1](#)). Perhaps the most significant conclusion from these studies is that *the ribosome is a ribozyme*; that is, the RNA components play the most fundamental roles. These observations strongly support the notion that the ribosome is a surviving inhabitant of the RNA world. As such, the ribosome is rich in information regarding very early steps in evolution.

Transfer RNA molecules (tRNAs), messenger RNA, and many proteins participate in protein synthesis along with ribosomes. The link between amino acids and nucleic acids is first made by enzymes called aminoacyl-tRNA synthetases. By specifically linking a particular amino acid to each tRNA, these enzymes implement the genetic code. This chapter focuses primarily on protein synthesis in prokaryotes because it illustrates many general principles and is relatively well understood. Some distinctive features of protein synthesis in eukaryotes also are presented.

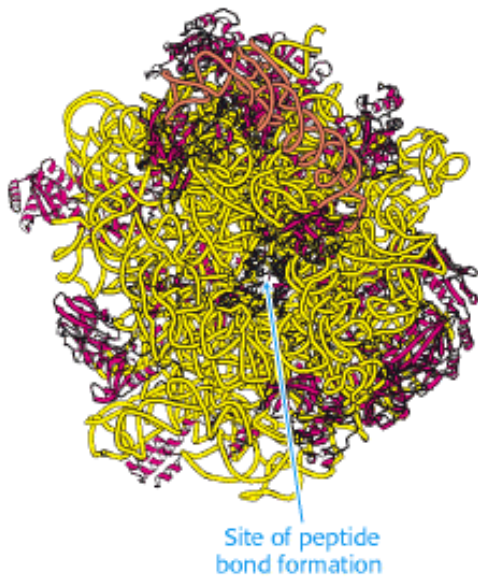
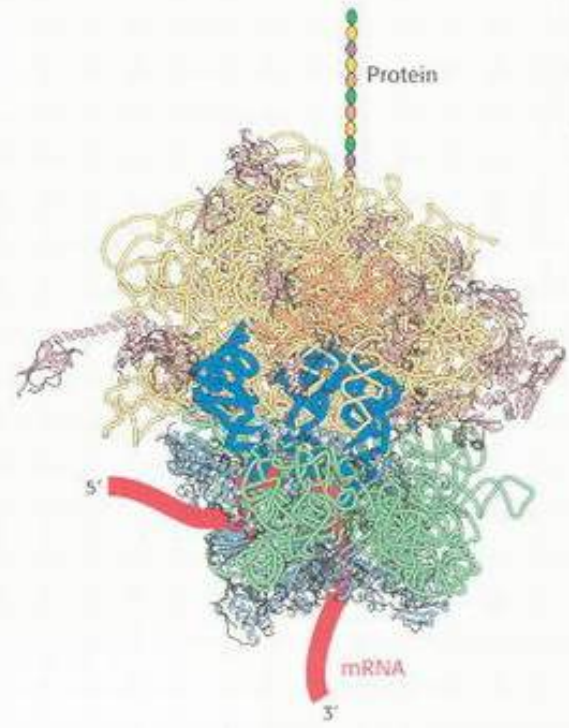


Figure 29.1. Ribosome Structure. The structure of a part of the ribosome showing the site at which peptide-bond formation takes place. This site contains only RNA (shown in yellow), with no protein (red) within 20 Å.



Protein Assembly. The ribosome, shown at the right, is a factory for the manufacture of polypeptides. Amino acids are carried into the ribosome, one at a time, connected to transfer RNA molecules (blue). Each amino acid is joined to the growing polypeptide chain, which detaches from the ribosome only once it is completed. This assembly line approach allows even very long polypeptide chains to be assembled rapidly and with impressive accuracy. [(Left) Doug Martin/Photo Researchers.]

29.1. Protein Synthesis Requires the Translation of Nucleotide Sequences Into Amino Acid Sequences

The basics of protein synthesis are the same across all kingdoms of life, attesting to the fact that the protein-synthesis system arose very early in evolution. A protein is synthesized in the amino-to-carboxyl direction by the sequential addition of amino acids to the carboxyl end of the growing peptide chain (Figure 29.2). The amino acids arrive at the growing chain in activated form as aminoacyl-tRNAs, created by joining the carboxyl group of an amino acid to the 3' end of a transfer RNA molecule. The linking of an amino acid to its corresponding tRNA is catalyzed by an *aminoacyl-tRNA synthetase*. ATP cleavage drives this activation reaction. For each amino acid, there is usually one activating enzyme and at least one kind of tRNA.

29.1.1. The Synthesis of Long Proteins Requires a Low Error Frequency

The process of transcription is analogous to copying, word for word, a page from a book. There is no change of alphabet or vocabulary; so the likelihood of a change in meaning is small. Translating the base sequence of an mRNA molecule into a sequence of amino acids is similar to translating the page of a book into another language. Translation is a complex process, entailing many steps and dozens of molecules. The potential for error exists at each step. The complexity of translation creates a conflict between two requirements: the process must be not only accurate, but also fast enough to meet a cell's needs. How fast is "fast enough"? In *E.coli*, translation takes place at a rate of 40 amino acids per second, a truly impressive speed considering the complexity of the process.

How accurate must protein synthesis be? Let us consider error rates. The probability p of forming a protein with no errors depends on n , the number of amino acid residues, and ϵ , the frequency of insertion of a wrong amino acid:

$$p = (1 - \epsilon)^n$$

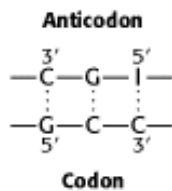
As Table 29.1 shows, an error frequency of 10^{-2} would be intolerable, even for quite small proteins. An ϵ value of 10^{-3} would usually lead to the error-free synthesis of a 300-residue protein (~33 kd) but not of a 1000-residue protein (~110 kd). Thus, the error frequency must not exceed approximately 10^{-4} to produce the larger proteins effectively. Lower error frequencies are conceivable; however, except for the largest proteins, they will not dramatically increase the percentage of proteins with accurate sequences. In addition, such lower error rates are likely to be possible only by a reduction in the rate of protein synthesis because additional time for proofreading will be required. *In fact, the observed values of ϵ are close to 10^{-4} .* An error frequency of about 10^{-4} per amino acid residue was selected in the course of evolution to accurately produce proteins consisting of as many as 1000 amino acids while maintaining a remarkably rapid rate for protein synthesis.

29.1.2. Transfer RNA Molecules Have a Common Design

The fidelity of protein synthesis requires the accurate recognition of three-base *codons* on messenger RNA. Recall that the genetic code relates each amino acid to a three-letter codon (Section 5.5.1). An amino acid cannot itself recognize a codon. Consequently, an amino acid is attached to a specific tRNA molecule that can recognize the codon by Watson-Crick base pairing. *Transfer RNA serves as the adapter molecule that binds to a specific codon and brings with it an amino acid for incorporation into the polypeptide chain.*

Robert Holley first determined the base sequence of a tRNA molecule in 1965, as the culmination of 7 years of effort. Indeed, his study of yeast alanyl-tRNA provided the first complete sequence of any nucleic acid. This adapter molecule is a single chain of 76 ribonucleotides (Figure 29.3). The 5' terminus is phosphorylated (pG), whereas the 3' terminus has a free hydroxyl group. The *amino acid attachment site* is the 3'-hydroxyl group of the adenosine residue at the 3'

terminus of the molecule. The sequence IGC in the middle of the molecule is the *anticodon*. It is complementary to GCC, one of the codons for alanine.



The sequences of several other tRNA molecules were determined a short time later. Hundreds of sequences are now known. The striking finding is that all of them can be arranged in a cloverleaf pattern in which about half the residues are base-paired (Figure 29.4). Hence, *tRNA molecules have many common structural features*. This finding is not unexpected, because all tRNA molecules must be able to interact in nearly the same way with the ribosomes, mRNAs, and protein factors that participate in translation.

All known transfer RNA molecules have the following features:

1. Each is a single chain containing between 73 and 93 ribonucleotides (~25 kd).
2. They contain *many unusual bases*, typically between 7 and 15 per molecule. Some are methylated or dimethylated derivatives of A, U, C, and G formed by enzymatic modification of a precursor tRNA (Section 28.1.8). Methylation prevents the formation of certain base pairs, thereby rendering some of the bases accessible for other interactions. In addition, methylation imparts a hydrophobic character to some regions of tRNAs, which may be important for their interaction with synthetases and ribosomal proteins. Other modifications alter codon recognition, as will be discussed shortly.



3. About half the nucleotides in tRNAs are base-paired to form double helices. Five groups of bases are not base paired in this way: the 3' *CCA terminal region*, which is part of a region called the *acceptor stem*; the *T ψ C loop*, which acquired its name from the sequence ribothymine-pseudouracil-cytosine; the "*extra arm*," which contains a variable number of residues; the *DHU loop*, which contains several dihydrouracil residues; and the *anticodon loop*. The structural diversity generated by this combination of helices and loops containing modified bases ensures that the tRNAs can be uniquely distinguished, though structurally similar overall.

4. The 5' end of a tRNA is phosphorylated. The 5' terminal residue is usually pG.

5. The activated amino acid is attached to a hydroxyl group of the adenosine residue located at the end of the 3' CCA component of the acceptor stem. This region is single stranded at the 3' end of mature rRNAs.

6. The anticodon is present in a loop near the center of the sequence.

29.1.3. The Activated Amino Acid and the Anticodon of tRNA Are at Opposite Ends of the L-Shaped Molecule

The three-dimensional structure of a tRNA molecule was first determined in 1974 through x-ray crystallographic studies carried out in the laboratories of Alexander Rich and Aaron Klug. The structure determined, that of yeast phenylalanyl-tRNA, is highly similar to all structures subsequently determined for other tRNA molecules. The most important properties of the tRNA structure are:

1. The molecule is *L-shaped* (Figure 29.5).
2. There are *two* apparently continuous segments of double helix. These segments are like A-form DNA, as expected for an RNA helix (Section 27.1.1). The base-pairing predicted from the sequence analysis is correct. The helix containing the 5' and 3' ends stacks on top of the helix that ends in the T ψ C loop to form one arm of the L; the remaining two helices stack to form the other (Figure 29.6).
3. Most of the bases in the nonhelical regions participate in hydrogenbonding interactions, even if the interactions are not like those in Watson-Crick base pairs.
4. The CCA terminus containing the *amino acid attachment site* extends from one end of the L. This single-stranded region can change conformation during amino acid activation and protein synthesis.
5. The anticodon loop is at the other end of the L, making accessible the three bases that make up the anticodon.

Thus, the architecture of the tRNA molecule is well suited to its role as adaptor; the anticodon is available to interact with an appropriate codon on mRNA while the end that is linked to an activated amino acid is well positioned to participate in peptide-bond formation.

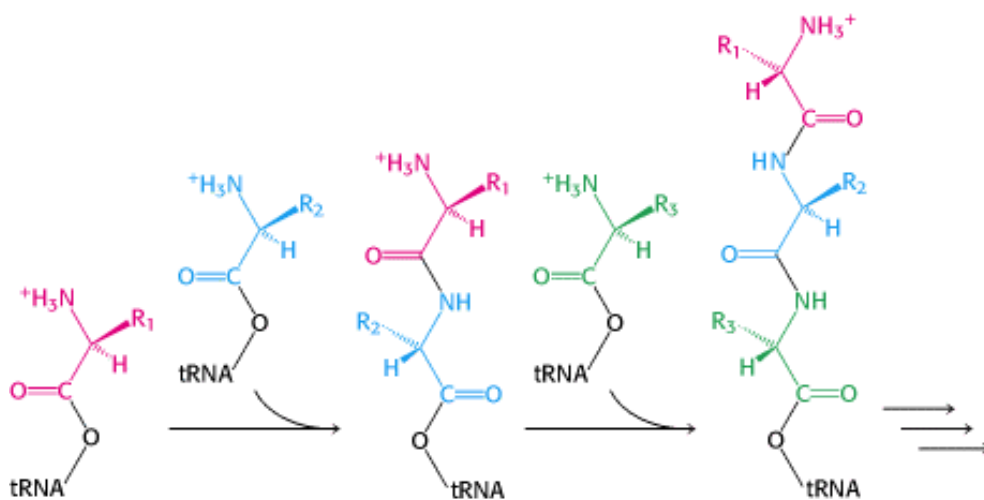


Figure 29.2. Polypeptide-Chain Growth. Proteins are synthesized by the successive addition of amino acids to the carboxyl terminus.

Table 29.1. Accuracy of protein synthesis

Frequency of inserting an incorrect amino acid	Probability of synthesizing an error-free protein		
	Number of amino acid residues		
	100	300	1000
10^{-2}	0.366	0.049	0.000
10^{-3}	0.905	0.741	0.368
10^{-4}	0.990	0.970	0.905
10^{-5}	0.999	0.997	0.990

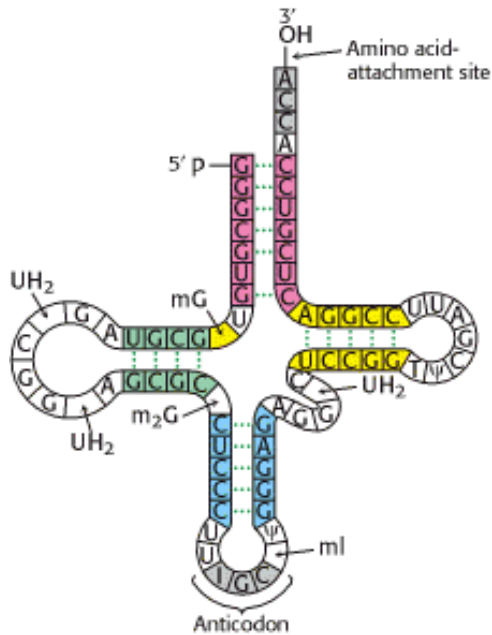


Figure 29.3. Alanine-tRNA Sequence. The base sequence of yeast alanyl-tRNA and the deduced cloverleaf secondary structure are shown. Modified nucleosides are abbreviated as follows: methylinosine (mI), dihydrouridine (UH₂), ribothymidine (T), pseudouridine (Ψ), methylguanosine (mG), and dimethylguanosine (m₂G). Inosine (I), another modified nucleoside, is part of the anticodon.

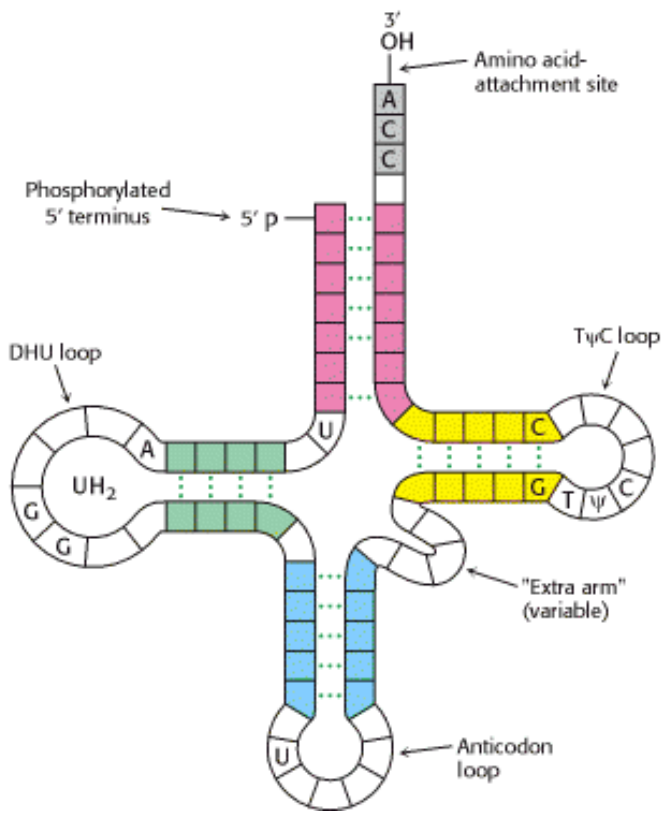


Figure 29.4. General Structure of tRNA Molecules. Comparison of the base sequences of many tRNAs reveals a number of conserved features.

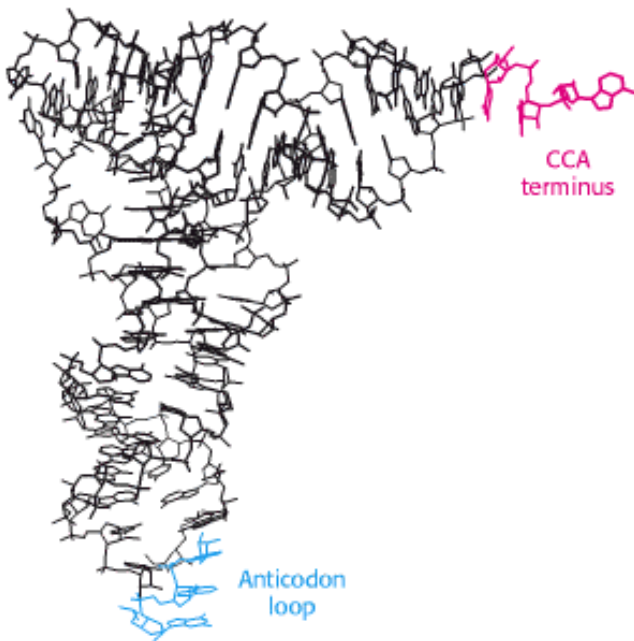


Figure 29.5. L-Shaped tRNA Structure. A skeletal model of yeast phenylalanyl-tRNA reveals the L-shaped structure. The CCA region is at the end of one arm, and the anticodon loop is at the end of the other.

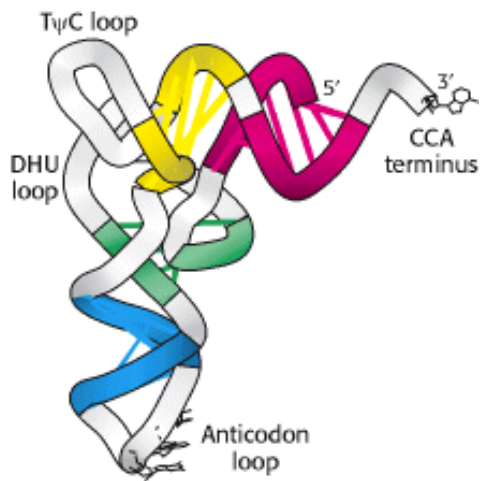


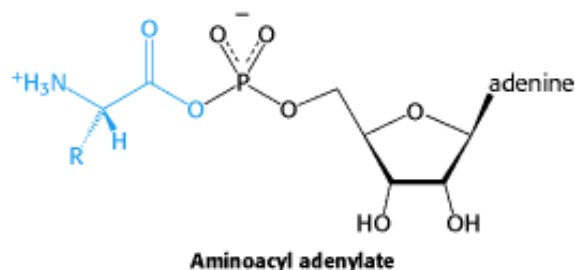
Figure 29.6. Helix Stacking in tRNA. The four helices of the secondary structure of tRNA (see [Figure 29.4](#)) stack to form an L-shaped structure.

29.2. Aminoacyl-Transfer RNA Synthetases Read the Genetic Code

The linkage of an amino acid to a tRNA is crucial for two reasons. *First, the attachment of a given amino acid to a particular tRNA establishes the genetic code.* When an amino acid has been linked to a tRNA, it will be incorporated into a growing polypeptide chain at a position dictated by the anticodon of the tRNA. Second, the formation of a peptide bond between free amino acids is not thermodynamically favorable. The amino acid must first be activated for protein synthesis to proceed. *The activated intermediates in protein synthesis are amino acid esters,* in which the carboxyl group of an amino acid is linked to either the 2'- or the 3'-hydroxyl group of the ribose unit at the 3' end of tRNA. An amino acid ester of tRNA is called an *aminoacyl-tRNA* or sometimes a *charged tRNA* ([Figure 29.7](#)).

29.2.1. Amino Acids Are First Activated by Adenylation

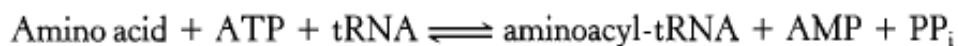
The activation reaction is catalyzed by specific *aminoacyl-tRNA synthetases*, which are also called *activating enzymes*. The first step is the formation of an *aminoacyl adenylate* from an amino acid and ATP. This activated species is a mixed anhydride in which the carboxyl group of the amino acid is linked to the phosphoryl group of AMP; hence, it is also known as *aminoacyl-AMP*.



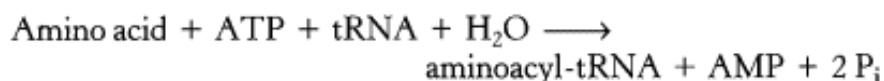
The next step is the transfer of the aminoacyl group of aminoacyl-AMP to a particular tRNA molecule to form *aminoacyl-tRNA*.



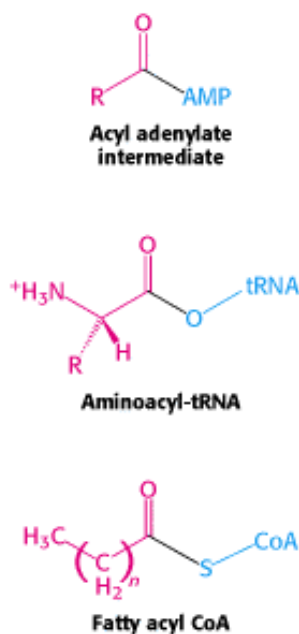
The sum of these activation and transfer steps is



The ΔG° of this reaction is close to 0, because the free energy of hydrolysis of the ester bond of aminoacyl-tRNA is similar to that for the hydrolysis of ATP to AMP and PP_i . As we have seen many times, the reaction is driven by the hydrolysis of pyrophosphate. The sum of these three reactions is highly exergonic:



Thus, *the equivalent of two molecules of ATP are consumed in the synthesis of each aminoacyl-tRNA*. One of them is consumed in forming the ester linkage of aminoacyl-tRNA, whereas the other is consumed in driving the reaction forward.



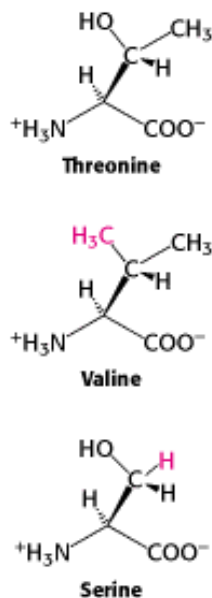
The activation and transfer steps for a particular amino acid are catalyzed by the same aminoacyl-tRNA synthetase. Indeed, *the aminoacyl-AMP intermediate does not dissociate from the synthetase*. Rather, it is tightly bound to the active site of the enzyme by noncovalent interactions. Aminoacyl-AMP is normally a transient intermediate in the synthesis of aminoacyl-tRNA, but it is relatively stable and readily isolated if tRNA is absent from the reaction mixture.

We have already encountered an acyl adenylate intermediate in fatty acid activation ([Section 22.2.2](#)). The major difference between these reactions is that the acceptor of the acyl group is CoA in fatty acid activation and tRNA in amino acid activation. The energetics of these biosyntheses are very similar: both are made irreversible by the hydrolysis of pyrophosphate.

29.2.2. Aminoacyl-tRNA Synthetases Have Highly Discriminating Amino Acid Activation Sites

Each aminoacyl-tRNA synthetase is highly specific for a given amino acid. Indeed, a synthetase will incorporate the incorrect amino acid only once in 10^4 or 10^5 catalytic reactions. How is this level of specificity achieved? Each aminoacyl-tRNA synthetase takes advantage of the properties of its amino acid substrate. Let us consider the challenge

faced by threonyl-tRNA synthetase. Threonine is particularly similar to two other amino acids—namely, valine and serine. Valine has almost exactly the same shape as threonine, except that it has a methyl group in place of a hydroxyl group. Like threonine, serine has a hydroxyl group but lacks the methyl group. How can the threonyl-tRNA synthetase avoid coupling these incorrect amino acids to threonyl-tRNA?



The structure of the amino acid-binding site of threonyl-tRNA synthetase reveals how valine is avoided (Figure 29.8). The enzyme contains a zinc ion, bound to the enzyme by two histidine residues and one cysteine residue. Like carbonic anhydrase (Section 9.2.1), the remaining coordination sites are available for substrate binding. Threonine coordinates to the zinc ion through its amino group and its side-chain hydroxyl group. The side-chain hydroxyl group is further recognized by an aspartate residue that hydrogen bonds to it. The methyl group present in valine in place of this hydroxyl group cannot participate in these interactions; it is excluded from this active site and, hence, does not become adenylated and transferred to threonyl-tRNA (abbreviated tRNA^{Thr}). Note that the carboxylate group of the amino acid is available to attack the α -phosphate group of ATP to form the aminoacyl adenylate. Other aminoacyl-tRNA synthetases have different strategies for recognizing their cognate amino acids; the use of a zinc ion appears to be unique to threonyl-tRNA synthetase.

The zinc site is less well suited to discrimination against serine because this amino acid does have a hydroxyl group that can bind to the zinc. Indeed, with only this mechanism available, threonyl-tRNA synthetase does mistakenly couple serine to threonyl-tRNA at a rate 10^{-2} to 10^{-3} times that for threonine. As noted in Section 29.1.1, this error rate is likely to lead to many translation errors. How is a higher level of specificity achieved?

29.2.3. Proofreading by Aminoacyl-tRNA Synthetases Increases the Fidelity of Protein Synthesis

Threonyl-tRNA synthetase can be incubated with tRNA^{Thr} that has been covalently linked with serine (Ser-tRNA^{Thr}); the tRNA has been "mischarged." The reaction is immediate: a rapid hydrolysis of the aminoacyl-tRNA forms serine and free tRNA. In contrast, incubation with correctly charged Thr-tRNA^{Thr} results in no reaction. Thus, threonyl-tRNA synthetase contains an additional functional site that hydrolyzes Ser-tRNA^{Thr} but not Thr-tRNA^{Thr}. This editing site provides an opportunity for the synthetase to correct its mistakes and improve its fidelity to less than one mistake in 10^4 . The results of structural and mutagenesis studies revealed that the editing site is more than 20 Å from the activation site (Figure 29.9). This site readily accepts and cleaves Ser-tRNA^{Thr} but does not cleave Thr-tRNA^{Thr}. The discrimination of serine from threonine is relatively easy because threonine contains an *extra* methyl group; a site that conforms to the structure of serine will sterically exclude threonine.

Most aminoacyl-tRNA synthetases contain editing sites in addition to acylation sites. These complementary pairs of sites function as a *double sieve* to ensure very high fidelity. In general, the acylation site rejects amino acids that are *larger* than the correct one because there is insufficient room for them, whereas the hydrolytic site cleaves activated species that are *smaller* than the correct one.

The structure of the complex between threonyl-tRNA synthetase and its substrate reveals that the aminoacylated-CCA can swing out of the activation site and into the editing site (Figure 29.10). Thus, the aminoacyl-tRNA can be edited without dissociating from the synthetase. This proofreading, which depends on the conformational flexibility of a short stretch of polynucleotide sequence, is entirely analogous to that of DNA polymerase (Section 27.2.4). In both cases, editing without dissociation significantly improves fidelity with only modest costs in time and energy.

A few synthetases achieve high accuracy without editing. For example, tyrosyl-tRNA synthetase has no difficulty discriminating between tyrosine and phenylalanine; the hydroxyl group on the tyrosine ring enables tyrosine to bind to the enzyme 10^4 times as strongly as phenylalanine. *Proof-reading has been selected in evolution only when fidelity must be enhanced beyond what can be obtained through an initial binding interaction.*

29.2.4. Synthetases Recognize the Anticodon Loops and Acceptor Stems of Transfer RNA Molecules

How do synthetases choose their tRNA partners? This enormously important step is the point at which "translation" takes place—at which the correlation between the amino acid and the nucleic acid worlds is made. In a sense, aminoacyl-tRNA synthetases are the only molecules in biology that "know" the genetic code. Their precise recognition of tRNAs is as important for high-fidelity protein synthesis as is the accurate selection of amino acids.

A priori, the anticodon of tRNA would seem to be a good identifier because each type of tRNA has a different one. Indeed, *some synthetases recognize their tRNA partners primarily on the basis of their anticodons*, although they may also recognize other aspects of tRNA structure. The most direct evidence comes from the results of crystallographic studies of complexes formed between synthetases and their cognate tRNAs. Consider, for example, the structure of the complex between threonyl-tRNA synthetase and tRNA^{Thr} (Figure 29.11). As expected, the CCA arm extends into the zinc-containing activation site, where it is well positioned to accept threonine from threonyl adenylate. The enzyme interacts extensively not only with the acceptor stem of the tRNA, but also with the anticodon loop. The interactions with the anticodon loop are particularly revealing. The bases within the sequence CGU of the anticodon each participate in hydrogen bonds with the enzyme; those in which G and U take part appear to be more important because the C can be replaced by G or U with no loss of acylation efficiency. The importance of the anticodon bases is further underscored by studies of tRNA^{Met}. Changing the anticodon sequence of this tRNA from CAU to GGU allows tRNA^{Met} to be aminoacylated by threonyl-tRNA synthetase nearly as well as tRNA^{Thr}, despite considerable differences in sequence elsewhere in the structure.

The structure of another complex between a tRNA and an aminoacyl-tRNA synthetase, that of glutamyl-tRNA synthetase, again reveals extensive interactions with both the anticodon loop and the acceptor stem (Figure 29.12). In addition, contacts are made near the "elbow" of the tRNA molecule, particularly with the base pair formed by G in position 10 and C in position 25 (denoted position 10:25). Reversal of this base pair from G · C to C · G results in a fourfold decrease in the rate of aminoacylation as well as a fourfold increase in the K_M value for glutamine. The results of mutagenesis studies supply further evidence regarding tRNA specificity, even for aminoacyl-tRNA synthetases for which structures have not yet been determined. For example, *E. coli* tRNA^{Cys} differs from tRNA^{Ala} at 40 positions and contains a C · G base pair at the 3:70 position. When this C · G base pair is changed to the non-Watson-Crick G · U base pair, tRNA^{Cys} is recognized by alanyl-tRNA synthetase as though it were tRNA^{Ala}. This finding raised the question whether a fragment of tRNA suffices for aminoacylation by alanyl-tRNA synthetase. Indeed, a "microhelix" containing just 24 of the 76 nucleotides of the native tRNA is specifically aminoacylated by the alanyl-tRNA synthetase. This microhelix contains only the acceptor stem and a hairpin loop (Figure 29.13). Thus, specific aminoacylation is possible for some synthetases even if the anticodon loop is completely lacking.

29.2.5. Aminoacyl-tRNA Synthetases Can Be Divided into Two Classes



Structural Insights, Aminoacyl-tRNA Synthetases. The first parts of the tutorial focus on the structural differences that distinguish class I and class II aminoacyl-tRNA synthetases. The final section of the tutorial looks at the editing process that most tRNA synthetases use to correct tRNA acylation errors.

At least one aminoacyl-tRNA synthetase exists for each amino acid. The diverse sizes, subunit composition, and sequences of these enzymes were bewildering for many years. Could it be that essentially all synthetases evolved independently? The determination of the three-dimensional structures of several synthetases followed by more-refined sequence comparisons revealed that different synthetases are, in fact, related. Specifically, synthetases fall into two classes, termed *class I* and *class II*, each of which includes enzymes specific for 10 of the 20 amino acids (Table 29.2). Glutamyl-tRNA synthetase is a representative of class I. The activation domain for class I has a Rossmann fold (Section 16.1.10). Threonyl-tRNA synthetase (see Figure 29.11) is a representative of class II. The activation domain for class II consists largely of β strands. Intriguingly, synthetases from the two classes bind to different faces of the tRNA molecule (Figure 29.14). The CCA arm of tRNA adopts different conformations to accommodate these interactions; the arm is in the helical conformation observed for free tRNA (see Figures 29.5 and 29.6) for class II enzymes and in a hairpin conformation for class I enzymes. These two classes also differ in other ways.

1. Class I enzymes acylate the 2'-hydroxyl group of the terminal adenosine of tRNA, whereas class II enzymes (except the enzyme for Phe-tRNA) acylate the 3'-hydroxyl group.
2. These two classes bind ATP in different conformations.
3. Most class I enzymes are monomeric, whereas most class II enzymes are dimeric.

Why did two distinct classes of aminoacyl-tRNA synthetases evolve? The observation that the two classes bind to distinct faces of tRNA suggests at least two possibilities. First, recognition sites on both faces of tRNA may have been required to allow the recognition of 20 different tRNAs. Second, it appears possible that, in some cases, a class I enzyme and a class II enzyme can bind to a tRNA molecule simultaneously without colliding with each other. In this way, enzymes from the two classes could work together to modify specific tRNA molecules.

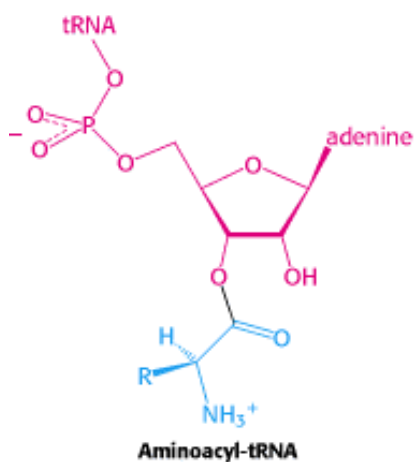


Figure 29.7. Aminoacyl-tRNA. Amino acids are coupled to tRNAs through ester linkages to either the 2'- or the 3'-hydroxyl group of the 3'-adenosine residue. A linkage to the 3'-hydroxyl group is shown.

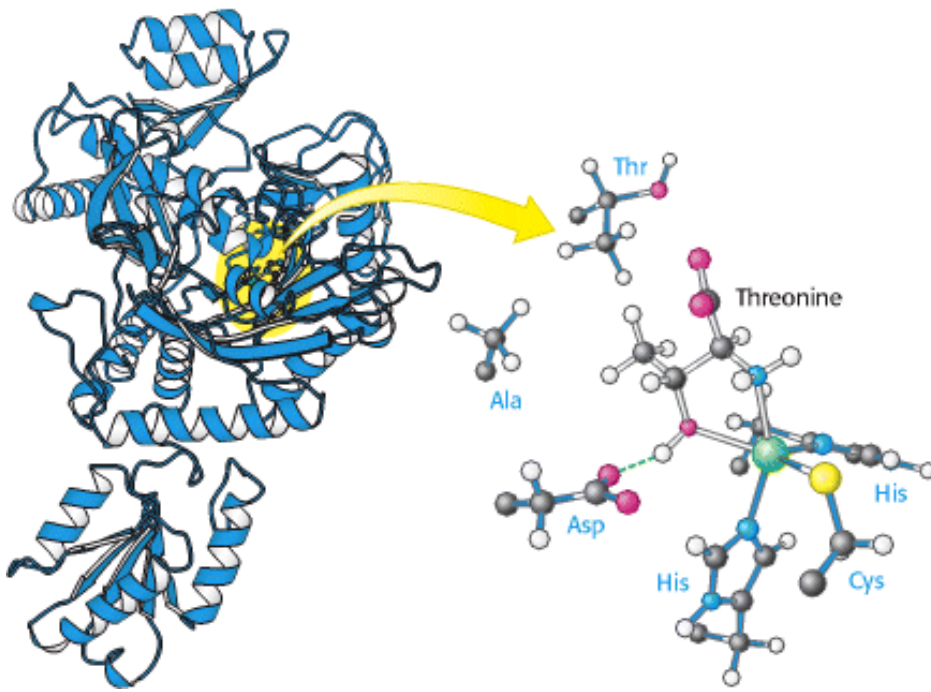


Figure 29.8. Structure of Threonyl-tRNA Synthetase. The structure of a large fragment of threonyl-tRNA synthetase reveals that the amino acid-binding site includes a zinc ion that coordinates threonine through its amino and hydroxyl groups. Only one subunit of the dimeric enzyme is shown in this and subsequent figures.

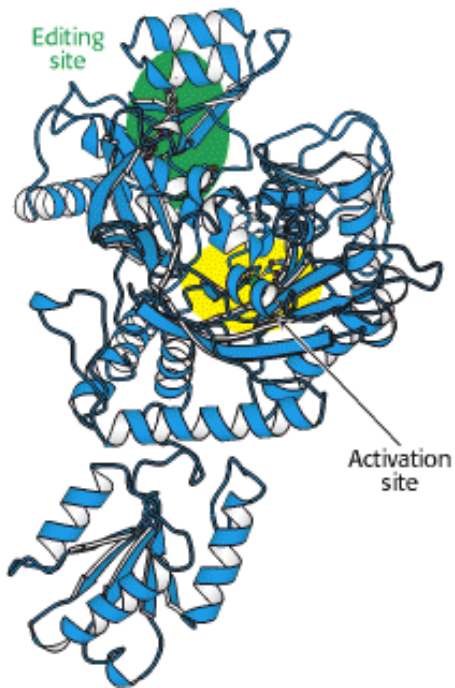


Figure 29.9. Editing Site. The results of mutagenesis studies revealed the position of the editing site (shown in green) in threonyl-tRNA synthetase.

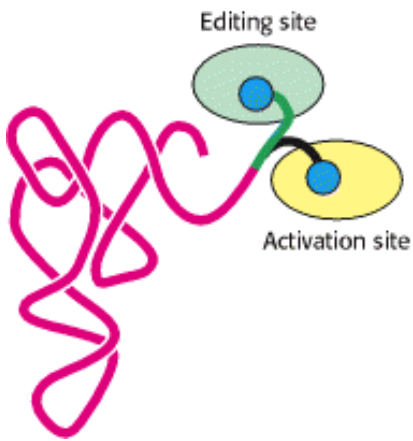


Figure 29.10. Editing of Aminoacyl-tRNA. The flexible CCA arm of an aminoacyl-tRNA can move the amino acid between the activation site and the editing site. If the amino acid fits well into the editing site, the amino acid is removed by hydrolysis.

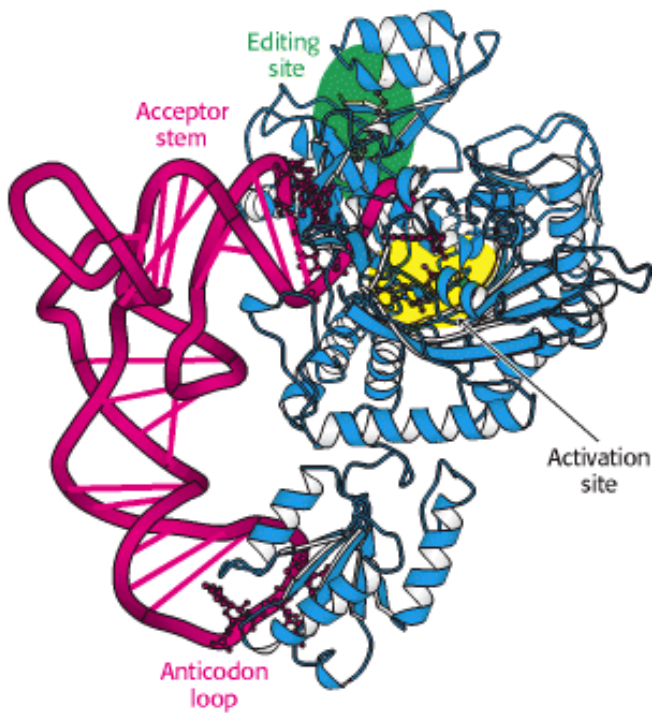


Figure 29.11. Threonyl-tRNA Synthetase Complex. The structure of the complex between threonyl-tRNA synthetase and tRNA^{Thr} reveals that the synthetase binds to both the acceptor stem and the anticodon loop.

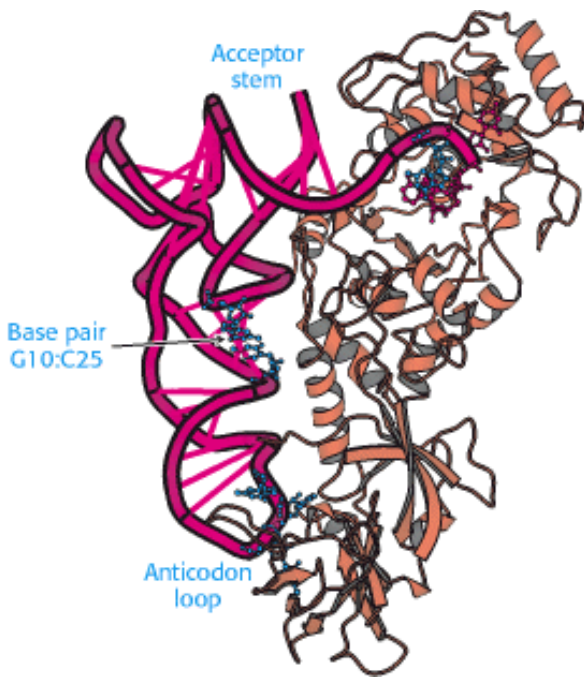


Figure 29.12. Glutaminyl-tRNA Synthetase Complex. The structure of this complex reveals that the synthetase interacts with base pair G10:C25 in addition to the acceptor stem and anticodon loop.

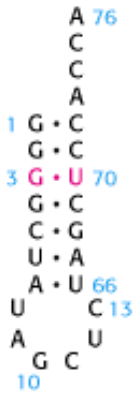


Figure 29.13. Microhelix Recognized by Alanyl-tRNA Synthetase. A stem-loop containing just 24 nucleotides corresponding to the acceptor stem is aminoacylated by alanyl-tRNA synthetase.

Table 29.2. Classification and subunit structure of aminoacyl-tRNA synthetases in *E. coli*

Class I	Class II
Arg (α)	Ala (α_4)
Cys (α)	Asn (α_2)
Gln (α)	Asp (α_2)
Glu (α)	Gly ($\alpha_2 \beta_2$)

Ile (α) His (α_2)
 Leu (α) Lys (α_2)
 Met (α) Phe (α_2 β_2)
 Trp (α_2) Ser (α_2)
 Tyr (α_2) Pro (α_2)
 Val (α) Thr (α_2)

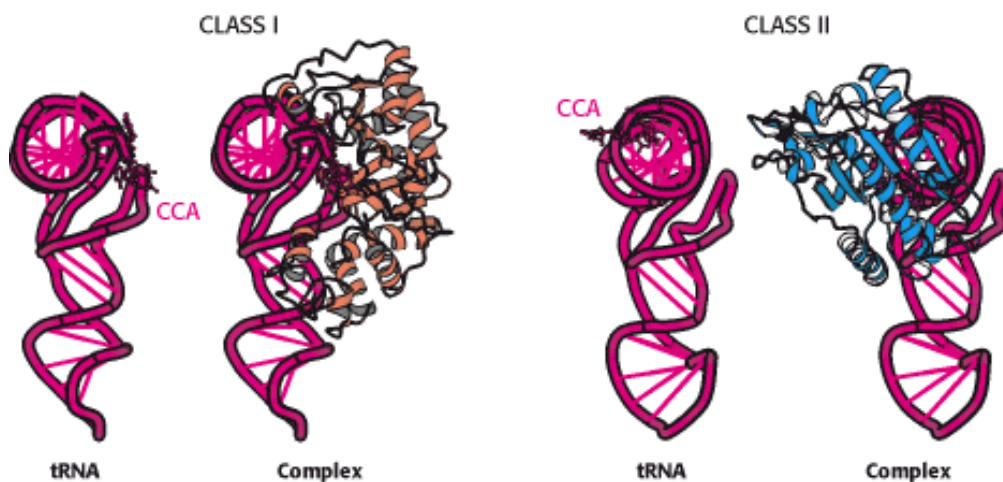


Figure 29.14. Classes of Aminoacyl-tRNA Synthetases. Class I and class II synthetases recognize different faces of the tRNA molecule. The CCA arm of tRNA adopts different conformations in complexes with the two classes of synthetase.

29.3. A Ribosome Is a Ribonucleoprotein Particle (70S) Made of a Small (30S) and a Large (50S) Subunit

We turn now to ribosomes, the molecular machines that coordinate the interplay of charged tRNAs, mRNA, and proteins that leads to protein synthesis. An *E. coli* ribosome is a ribonucleoprotein assembly with a mass of about 2700 kd, a diameter of approximately 200 Å, and a sedimentation coefficient of 70S. The 20,000 ribosomes in a bacterial cell constitute nearly a fourth of its mass.


A ribosome can be dissociated into a *large subunit* (50S) and a *small subunit* (30S) (Figure 29.15). These subunits can be further split into their constituent proteins and RNAs. The 30S subunit contains 21 different proteins (referred to as S1 through S21) and a 16S RNA molecule. The 50S subunit contains 34 different proteins (L1 through L34) and two RNA molecules, a 23S and a 5S species. A ribosome contains one copy of each RNA molecule, two copies of the L7 and L12 proteins, and one copy of each of the other proteins. The L7 protein is identical with L12 except that its amino terminus is acetylated. Only one protein is common to both subunits: S20 is identical with L26. Both the 30S and the 50S subunits can be reconstituted in vitro from their constituent proteins and RNA, as was first achieved by Masayasu Nomura in 1968. *This reconstitution is an outstanding example of the principle that supramolecular complexes can form spontaneously from their macromolecular constituents.*

Electron microscopic studies of the ribosome at increasingly high resolution provided views of the overall structure and revealed the positions of tRNA-binding sites. Astounding progress on the structure of the ribosome has been made by x-ray crystallographic methods, after the pioneering work by Ada Yonath. The structures of both the 30S and the 50S subunits have been determined at or close to atomic resolution, and the elucidation of the structure of intact 70S

ribosomes at a similar resolution is following rapidly (Figure 29.16). The determination of this structure requires the positioning of more than 100,000 atoms. The features of these structures are in remarkable agreement with interpretations of less-direct experimental probes. These structures provide an invaluable framework for examining the mechanism of protein synthesis.

29.3.1. Ribosomal RNAs (5S, 16S, and 23S rRNA) Play a Central Role in Protein Synthesis

The prefix *ribo* in the name *ribosome* is apt, because RNA constitutes nearly two-thirds of the mass of these large molecular assemblies. The three RNAs present—5S, 16S, and 23S—are critical for ribosomal architecture and function. They are formed by cleavage of primary 30S transcripts and further processing. The base-pairing patterns of these molecules were deduced by comparing the nucleotide sequences of many species to detect conserved features, in combination with chemical modification and digestion experiments (Figure 29.17). The striking finding is that *ribosomal RNAs* (rRNAs) *are folded into defined structures with many short duplex regions*. This conclusion and essentially all features of the secondary structure have been confirmed by the x-ray crystallographically determined structures.

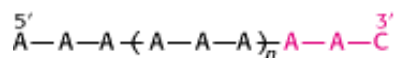
 For many years, ribosomal proteins were presumed to orchestrate protein synthesis and ribosomal RNAs were presumed to serve primarily as structural scaffolding. The current view is almost the reverse. The discovery of catalytic RNA made biochemists receptive to the possibility that RNA plays a much more active role in ribosomal function. The detailed structures make it clear that the key sites in the ribosome are composed almost entirely of RNA. Contributions from the proteins are minor. Many of the proteins have elongated structures that "snake" their way into the RNA matrix (Figure 29.18). The almost inescapable conclusion is that the ribosome initially consisted only of RNA and that the proteins were added later to fine tune its functional properties. This conclusion has the pleasing consequence of dodging a "chicken and egg" question—namely, How can complex proteins be synthesized if complex proteins are required for protein synthesis?

29.3.2. Proteins Are Synthesized in the Amino-to-Carboxyl Direction

Before the mechanism of protein synthesis could be examined, several key facts had to be established. The results of pulse-labeling studies by Howard Dintzis established that protein synthesis proceeds sequentially from the amino terminus. Reticulocytes (young red blood cells) that were actively synthesizing hemoglobin were treated with [³H] leucine. In a period of time shorter than that required to synthesize a complete chain, samples of hemoglobin were taken, separated into α and β chains, and analyzed for the distribution of ³H within their sequences. In the earliest samples, only regions near the carboxyl ends contained radioactivity. In later samples, radioactivity was present closer to the amino terminus as well. This distribution is the one expected if the amino-terminal regions of some chains had already been partly synthesized before the addition of the radioactive amino acid. Thus, *protein synthesis begins at the amino terminus and extends toward the carboxyl terminus*.

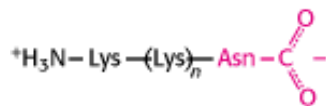
29.3.3. Messenger RNA Is Translated in the 5'-to-3' Direction

The sequence of amino acids in a protein is translated from the nucleotide sequence in mRNA. In which direction is the message read? The answer was established by using the synthetic polynucleotide



as the template in a cell-free protein-synthesizing system. AAA encodes lysine, whereas AAC encodes asparagine. The

polypeptide product was



Because asparagine was the carboxyl-terminal residue, we can conclude that the codon AAC was the last to be read. Hence, *the direction of translation is 5' → 3'.*

The direction of translation has important consequences. Recall that transcription also occurs in the 5' → 3' direction (Section 28.1.4). If the direction of translation were opposite that of transcription, only fully synthesized mRNA could be translated. In contrast, because the directions are the same, mRNA can be translated while it is being synthesized. In prokaryotes, almost no time is lost between transcription and translation. The 5' end of mRNA interacts with ribosomes very soon after it is made, much before the 3' end of the mRNA molecule is finished. *An important feature of prokaryotic gene expression is that translation and transcription are closely coupled in space and time.* Many ribosomes can be translating an mRNA molecule simultaneously. This parallel synthesis markedly increases the efficiency of mRNA translation. The group of ribosomes bound to an mRNA molecule is called a *polyribosome* or a *polysome* (Figure 29.19).

29.3.4. The Start Signal Is AUG (or GUG) Preceded by Several Bases That Pair with 16S rRNA

How does protein synthesis start? The simplest possibility would be for the first three nucleotides of each mRNA to serve as the first codon; no special start signal would then be needed. However, the experimental fact is that translation does not begin immediately at the 5' terminus of mRNA. Indeed, the first translated codon is nearly always more than 25 nucleotides away from the 5' end. Furthermore, in prokaryotes, many mRNA molecules are *polycistronic*, or polygenic—that is, they encode two or more polypeptide chains. For example, a single mRNA molecule about 7000 nucleotides long specifies five enzymes in the biosynthetic pathway for tryptophan in *E. coli*. Each of these five proteins has its own start and stop signals on the mRNA. In fact, *all known mRNA molecules contain signals that define the beginning and end of each encoded polypeptide chain.*

A clue to the mechanism of initiation was the finding that nearly half the amino-terminal residues of proteins in *E. coli* are methionine. In fact, the initiating codon in mRNA is AUG (methionine) or, much less frequently, GUG (valine). What additional signals are necessary to specify a translation start site? The first step toward answering this question was the isolation of initiator regions from a number of mRNAs. This isolation was accomplished by using pancreatic ribonuclease to digest mRNA-ribosome complexes (formed under conditions of chain initiation but not elongation). In each case, a sequence of about 30 nucleotides was protected from digestion. As expected, each initiator region displays an AUG (or GUG) codon (Figure 29.20). In addition, each initiator region contains a purine-rich sequence centered about 10 nucleotides on the 5' side of the initiator codon.

The role of this purine-rich region, called the *Shine-Dalgarno sequence*, became evident when the sequence of 16S rRNA was elucidated. The 3' end of this rRNA component of the 30S subunit contains a sequence of several bases that is complementary to the purine-rich region in the initiator sites of mRNA. Mutagenesis of the CCUCC sequence near the 3' end of 16S rRNA to ACACA markedly interferes with the recognition of start sites in mRNA. This and other evidence shows that the initiator region of mRNA binds to the 16S rRNA very near its 3' end. The number of base pairs linking mRNA and 16S rRNA ranges from three to nine. Thus, *two kinds of interactions determine where protein synthesis starts: (1) the pairing of mRNA bases with the 3' end of 16S rRNA and (2) the pairing of the initiator codon on mRNA*

with the anticodon of an initiator tRNA molecule.

29.3.5. Bacterial Protein Synthesis Is Initiated by Formylmethionyl Transfer RNA

The methionine residue found at the amino-terminal end of *E. coli* proteins is usually modified. In fact, *protein synthesis in bacteria starts with N-formylmethionine (fMet)*. A special tRNA brings formylmethionine to the ribosome to initiate protein synthesis. This *initiator tRNA* (abbreviated as tRNA_f) differs from the one that inserts methionine in internal positions (abbreviated as tRNA_m). The subscript "f" indicates that methionine attached to the initiator tRNA can be formylated, whereas it cannot be formyl-ated when attached to tRNA_m. In approximately one-half of *E. coli* proteins, *N*-formylmethionine is removed when the nascent chain is 10 amino acids long.

Methionine is linked to these two kinds of tRNAs by the same amino-acyl-tRNA synthetase. A specific enzyme then formylates the amino group of methionine that is attached to tRNA_f (Figure 29.21). The activated formyl donor in this reaction is *N*¹⁰-formyltetrahydrofolate (Section 24.2.6). It is significant that free methionine and methionyl-tRNA_m are not substrates for this transformylase.

29.3.6. Ribosomes Have Three tRNA-Binding Sites That Bridge the 30S and 50S Subunits

A snapshot of a significant moment in protein synthesis was obtained by determining the structure of the 70S ribosome bound to three tRNA molecules and a fragment of mRNA (Figure 29.22). As expected, the mRNA fragment is bound within the 30S subunit. Each of the tRNA molecules bridges between the 30S and 50S subunits. At the 30S end, two of the three tRNA molecules are bound to the mRNA fragment through anticodon-codon base pairs. These binding sites are called the A site (for *aminoacyl*) and the P site (for *peptidyl*). The third tRNA molecule is bound to an adjacent site called the E site (for *exit*).

The other end of each tRNA molecule interacts with the 50S subunit. The acceptor stems of the tRNA molecules occupying the A site and the P site converge at a site where a peptide bond is formed. Further examination of this site reveals that a tunnel connects this site to the back of the ribosome (Figure 29.23). *The polypeptide chain passes through this tunnel during synthesis.*

29.3.7. The Growing Polypeptide Chain Is Transferred Between tRNAs on Peptide-Bond Formation

Protein synthesis begins with the interaction of the 30S subunit and mRNA through the Shine-Delgarno sequence. On formation of this complex, the initiator tRNA charged with formylmethionine binds to the initiator AUG codon, and the 50S subunit binds to the 30S subunit to form the complete 70S ribosome. How does the polypeptide chain increase in length (Figure 29.24)? The three sites in our snapshot of protein synthesis provide a clue. The initiator tRNA is bound in the P site on the ribosome. A charged tRNA with an anticodon complementary to the codon in the A site then binds. The stage is set for the formation of a peptide bond: the formylmethionine molecule linked to the initiator tRNA will be transferred to the amino group of the amino acid in the A site. The transfer takes place in a ribosome site called the *peptidyl transferase center*.

The amino group of the aminoacyl-tRNA in the A site is well positioned to attack the ester linkage between the initiator tRNA and the formylmethionine molecule (Figure 29.25). The peptidyl transferase center includes bases that promote this reaction by helping to form an -NH₂ group on the A site aminoacyl-tRNA and by helping to stabilize the tetrahedral intermediate that forms. This reaction is, in many ways, analogous to the reverse of the reaction catalyzed by serine proteases such as chymotrypsin (Section 9.1.2). The peptidyl-tRNA is analogous to the acyl-enzyme form of a serine protease. In a serine protease, the acyl-enzyme is generated with the use of the free energy associated with cleaving an amide bond. In the ribosome, the free energy necessary to form the analogous species, an amino-acyl-tRNA, comes from

the ATP that is cleaved by the aminoacyl-tRNA synthetase before the arrival of the tRNA at the ribosome.

With the peptide bond formed, the peptide chain is now attached to the tRNA in the A site on the 30S subunit while a change in the interaction with the 50S subunit has placed that tRNA and its peptide in the P site of the large subunit. The tRNA in the P site of the 30S subunit is now uncharged. For translation to proceed, the mRNA must be moved (or *translocated*) so that the codon for the next amino acid to be added is in the A site. This translocation takes place through the action of a protein enzyme called *elongation factor G* (Section 29.4.3), driven by the hydrolysis of GTP. On completion of this step, the peptidyl-tRNA is now fully in the P site, and the uncharged initiator tRNA is in the E site and has been disengaged from the mRNA. On dissociation of the initiator tRNA, the ribosome has returned to its initial state except that the peptide chain is attached to a different tRNA, the one corresponding to the first codon past the initiating AUG. Note that *the peptide chain remains in the P site on the 50S subunit throughout this cycle*, presumably growing into the tunnel. This cycle is repeated as new aminoacyl-tRNAs move into the A site, allowing the polypeptide to be elongated indefinitely.

We can now understand why the amino terminus of the initial methionine molecule is modified by the attachment of a formyl group. Chemical reactivity may have dictated this modification (Figure 29.26). Suppose that the amino-terminus is not blocked. After the first peptidyl-transfer reaction, a dipeptide is linked to the tRNA in the P site. If a free amino group is present in the terminal amino acid, this amino group can attack the carbonyl group of the ester linkage to the tRNA, forming a very stable six-membered ring and terminating translation.

29.3.8. Only the Codon-Anticodon Interactions Determine the Amino Acid That Is Incorporated

On the basis of the mechanism described in Section 29.3.7, the base-pairing interaction between the anticodon on the incoming tRNA and the codon in the A site on mRNA determines which amino acid is added to the polypeptide chain. Does the amino acid attached to the tRNA play any role in this process? This question was answered in the following way. First, cysteine was attached to its cognate tRNA. The attached cysteine unit was then converted into alanine by adding Raney nickel to Cys-tRNA^{Cys}; the reaction removed the sulfur atom from the cysteine residue without affecting its linkage to tRNA. Thus, a *mischarged aminoacyl-tRNA* was produced in which alanine was covalently attached to a tRNA specific for cysteine.



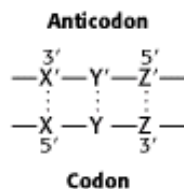
Does this mischarged tRNA recognize the codon for cysteine or for alanine? The answer came when the tRNA was added to a cell-free protein-synthesizing system. The template was a random copolymer of U and G in the ratio of 5:1, which normally incorporates cysteine (encoded by UGU) but not alanine (encoded by GCN). However, alanine was incorporated into a polypeptide when Ala-tRNA^{Cys} was added to the incubation mixture. The same result was obtained when mRNA for hemoglobin served as the template and [¹⁴C]alanyl-tRNA^{Cys} was used as the mischarged aminoacyl-tRNA. The only radioactive tryptic peptide produced was one that normally contained cysteine but not alanine. Thus, *the amino acid in aminoacyl-tRNA does not play a role in selecting a codon*.

In recent years, the ability of mischarged tRNAs to transfer their amino acid cargo to a growing polypeptide chain has been used to synthesize peptides with amino acids not found in proteins incorporated into specific sites in a protein. Aminoacyl-tRNAs are first linked to these unnatural amino acids by chemical methods. These mischarged aminoacyl-tRNAs are added to a cell-free protein-synthesizing system along with specially engineered mRNA that contains codons corresponding to the anticodons of the mischarged aminoacyl-tRNAs in the desired positions. The proteins produced

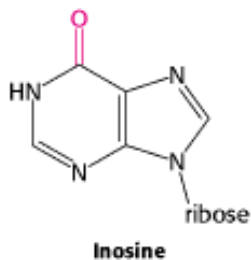
have unnatural amino acids in the expected positions. More than 100 different unnatural amino acids have been incorporated in this way. However, only L-amino acids can be used; apparently this stereochemistry is required for peptide-bond formation to take place.

29.3.9. Some Transfer RNA Molecules Recognize More Than One Codon Because of Wobble in Base-Pairing

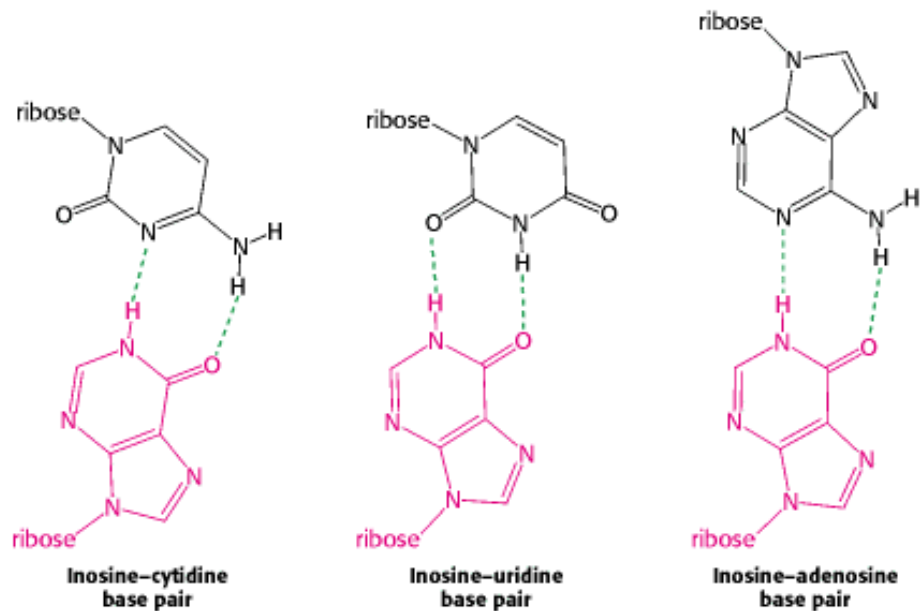
What are the rules that govern the recognition of a codon by the anticodon of a tRNA? A simple hypothesis is that each of the bases of the codon forms a Watson-Crick type of base pair with a complementary base on the anticodon. The codon and anticodon would then be lined up in an antiparallel fashion. In the diagram in the margin, the prime denotes the complementary base. Thus X and X' would be either A and U (or U and A) or G and C (or C and G). According to this model, a particular anticodon can recognize only one codon.



The facts are otherwise. As found experimentally, *some pure tRNA molecules can recognize more than one codon*. For example, the yeast alanyl-tRNA binds to *three* codons: GCU, GCC, and GCA. The first two bases of these codons are the same, whereas the third is different. Could it be that recognition of the third base of a codon is sometimes less discriminating than recognition of the other two? The pattern of degeneracy of the genetic code indicates that this might be so. XYU and XYC always encode the same amino acid; XYA and XYG usually do. Francis Crick surmised from these data that the steric criteria might be less stringent for pairing of the third base than for the other two. Models of various base pairs were built to determine which ones are similar to the standard A · U and G · C base pairs with regard to the distance and angle between the glycosidic bonds. Inosine was included in this study because it appeared in several anticodons. With the assumption of some steric freedom ("wobble") in the pairing of the third base of the codon, the combinations shown in [Table 29.3](#) seemed plausible.



The *wobble hypothesis* is now firmly established. The anticodons of tRNAs of known sequence bind to the codons predicted by this hypothesis. For example, the anticodon of yeast alanyl-tRNA is IGC. This tRNA recognizes the codons GCU, GCC, and GCA. Recall that, by convention, nucleotide sequences are written in the 5' → 3' direction unless otherwise noted. Hence, I (the 5' base of this anticodon) pairs with U, C, or A (the 3' base of the codon), as predicted.



Two generalizations concerning the codon-anticodon interaction can be made:

1. The first two bases of a codon pair in the standard way. Recognition is precise. Hence, *codons that differ in either of their first two bases must be recognized by different tRNAs*. For example, both UUA and CUA encode leucine but are read by different tRNAs.
2. The first base of an anticodon determines whether a particular tRNA molecule reads one, two, or three kinds of codons: C or A (one codon), U or G (two codons), or I (three codons). Thus, *part of the degeneracy of the genetic code arises from imprecision (wobble) in the pairing of the third base of the codon with the first base of the anticodon*. We see here a strong reason for the frequent appearance of inosine, one of the unusual nucleosides, in anticodons. *Inosine maximizes the number of codons that can be read by a particular tRNA molecule*. The inosines in tRNA are formed by deamination of adenosine after synthesis of the primary transcript.

Why is wobble tolerated in the third position of the codon but not in the first two? The 30S subunit has two adenine bases (A1492 and A1493 in the 16S RNA) that form hydrogen bonds on the minor-groove side of the codon-anticodon duplex. These interactions serve to check whether Watson-Crick base pairs are present in the first two positions of the codon- anticodon duplex. No such inspection device is present for the third position so more-varied base pairs are tolerated. This mechanism for ensuring fidelity is analogous to the minor-groove interactions utilized by DNA polymerase for a similar purpose (Section 27.2.3). *Thus, the ribosome plays an active role in decoding the codon-anticodon interactions.*

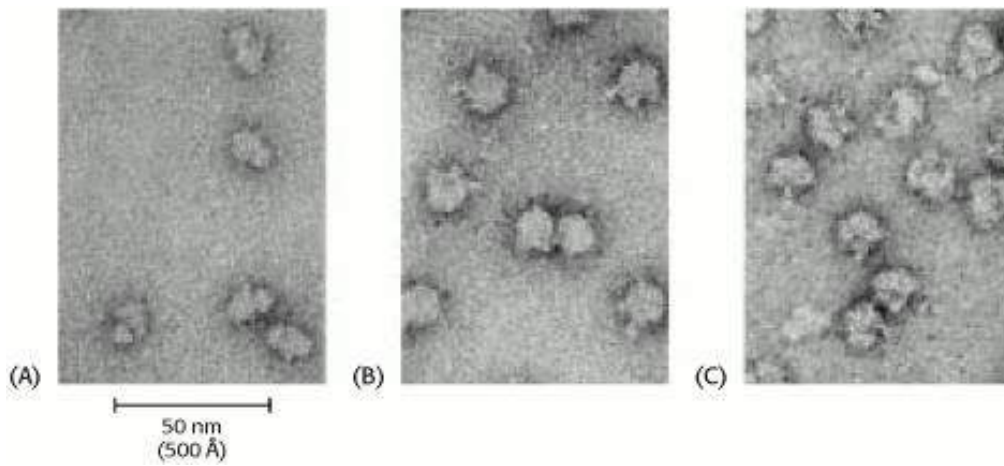


Figure 29.15. Ribosomes at Low Resolution. Electron micrographs of (A) 30S subunits, (B) 50S subunits, and (C) 70S ribosomes. [Courtesy of Dr. James Lake.]

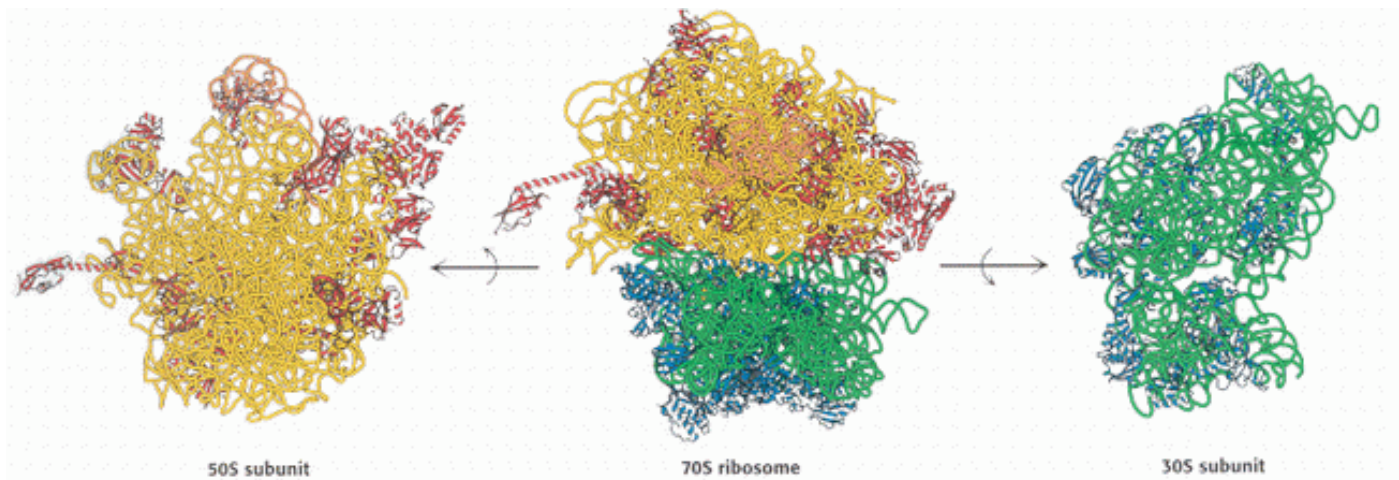


Figure 29.16. The Ribosome at High Resolution. Detailed models of the ribosome based on the results of x-ray crystallographic studies of the 70S ribosome and the 30S and 50S subunits. 23S RNA is shown in yellow, 5S RNA in orange, 16S RNA in green, proteins of the 50S subunit in red, and proteins of the 30S subunit in blue.

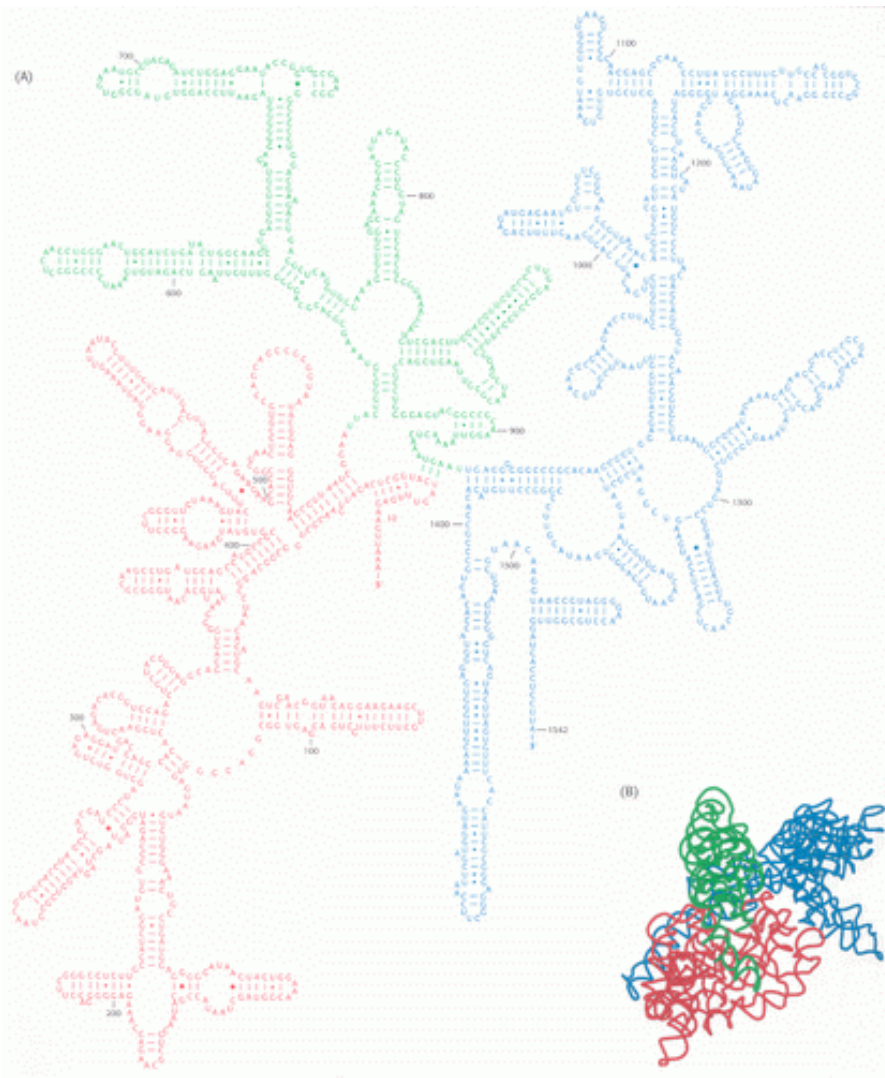


Figure 29.17. Ribosomal RNA Folding Pattern. (A) The secondary structure of 16S ribosomal RNA deduced from sequence comparison and the results of chemical studies. (B) The tertiary structure of 16S RNA determined by x-ray crystallography. [Part A courtesy of Dr. Bryn Weiser and Dr. Harry Noller.]

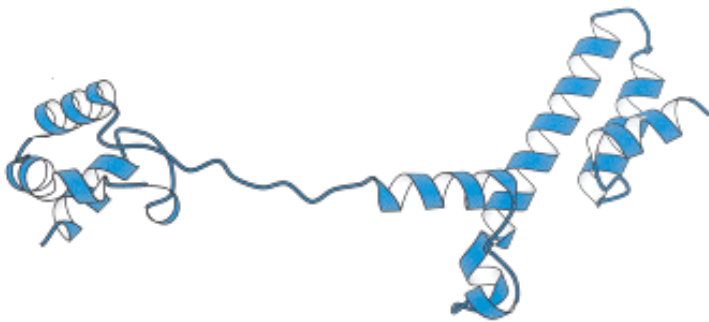


Figure 29.18. Ribosomal Protein Structure. The structure of ribosomal protein L19 of the 50S ribosomal subunit reveals a long segment of extended structure that fits through some of the cavities within the 23S RNA molecule.

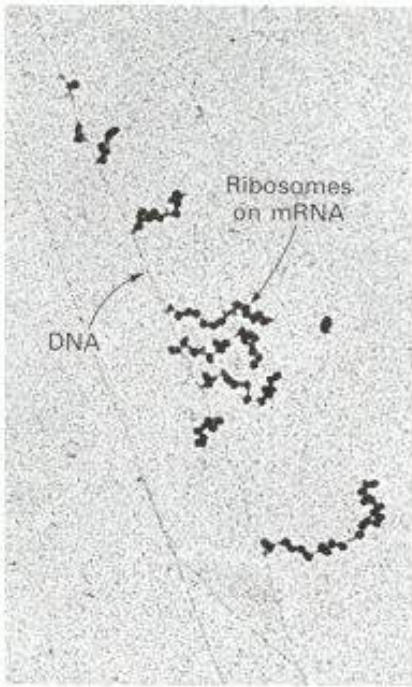


Figure 29.19. Polysomes. Transcription of a segment of DNA from *E. coli* generates mRNA molecules that are immediately translated by multiple ribosomes. [From O. L. Miller, Jr., B. A. Hamkalo, and C. A. Thomas, Jr. *Science* 169 (1970):392.]

5'	3'		
AGCAC	GAGGG	AAAUCUGAUGGAACGCUAC	<i>E. coli trpA</i>
UUUGGAU	GGAGUGA	AAACGAUGGCGAUUGCA	<i>E. coli araB</i>
GGUAAC	CAGGUAAC	AAACAUGCGAGUGUUG	<i>E. coli thrA</i>
CAAUUC	AGGGUGG	UGA AUGGAAACCAAGUA	<i>E. coli lacI</i>
AAUCU	UGGAGG	CUUUUUUAUGGUUCGUUCU	ϕ X174 phage A protein
UAACUA	AAGGAUGAA	AUGCAUGUCUAAGACA	Q β phage replicase
UCCUA	AGGAGGU	UUGACCUAUGCGAGCUUUU	R17 phage A protein
AUGUAC	UAAGGAGGU	UGAUGGAACAA CGC	λ phage <i>cro</i>
		<div style="display: flex; justify-content: space-around; width: 100%;"> Pairs with 16S rRNA Pairs with initiator tRNA </div>	

Figure 29.20. Initiation Sites. Sequences of mRNA initiation sites for protein synthesis in some bacterial and viral mRNA molecules. Comparison of these sequences reveals some recurring features.

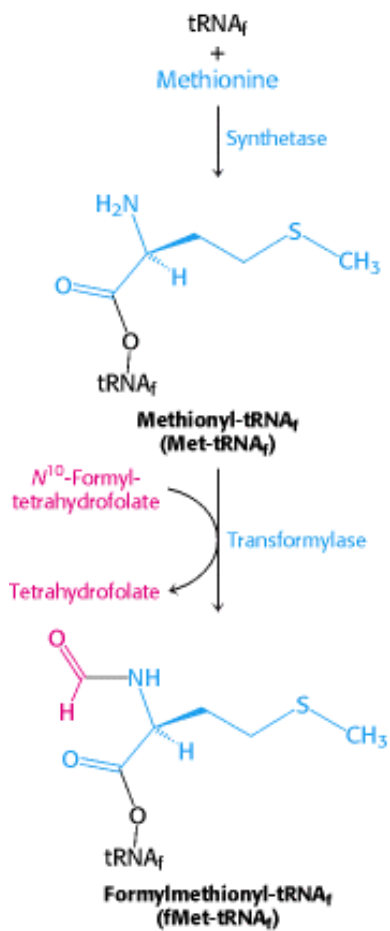


Figure 29.21. Formylation of Methionyl-tRNA. Initiator tRNA (tRNA_f) is first charged with methionine, and then a formyl group is transferred to the methionyl- tRNA_f from N^{10} -formyltetrahydrofolate.

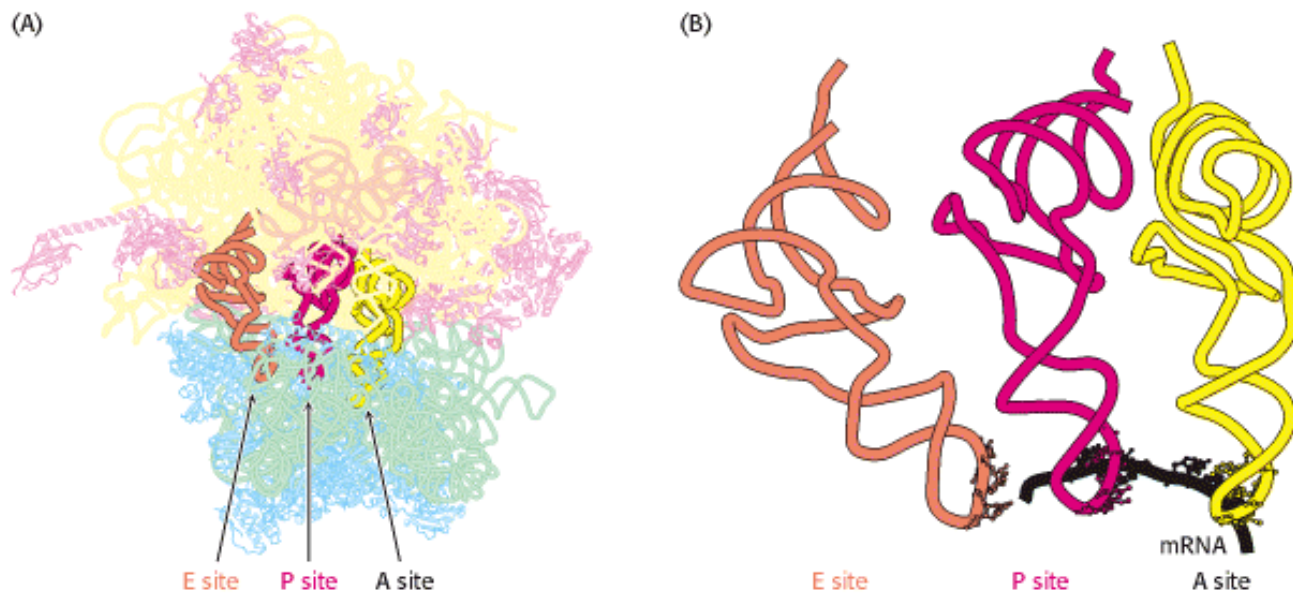


Figure 29.22. Transfer RNA-Binding Sites. (A) Three tRNA-binding sites are present on the 70S ribosome. They are called the A (for aminoacyl), P (for peptidyl), and E (for exit) sites. Each tRNA molecule contacts both the 30S and the 50S subunit. (B) The tRNA molecules in sites A and P are base paired with mRNA.

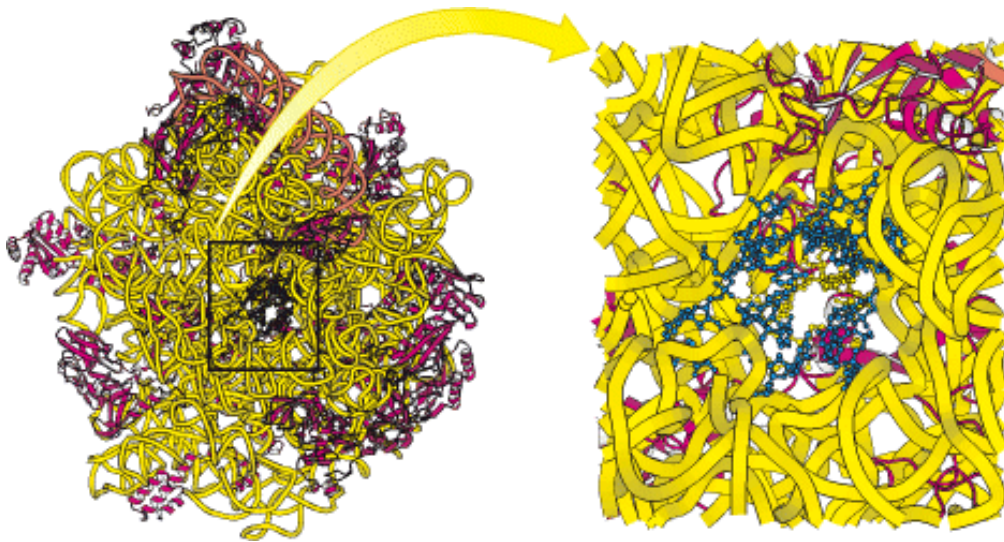


Figure 29.23. Polypeptide Escape Path. A tunnel passes through the 50S subunit beginning at the site of peptide-bond formation (shown in blue). The growing polypeptide chain passes through this tunnel.

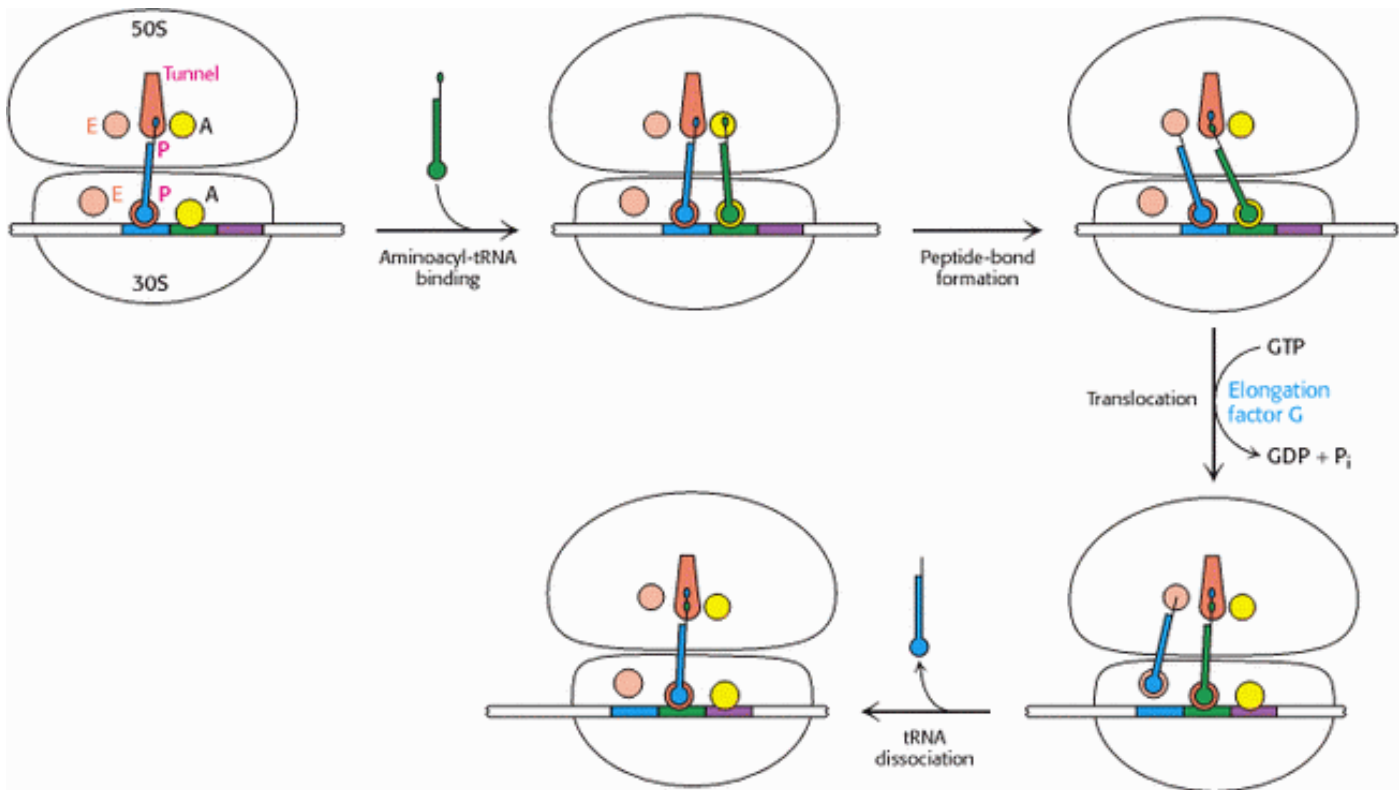


Figure 29.24. Mechanism of Protein Synthesis. The cycle begins with peptidyl-tRNA in the P site. An aminoacyl-tRNA binds in the A site. With both sites occupied, a new peptide bond is formed. The tRNAs and the mRNA are translocated through the action of elongation factor G, which moves the deacylated tRNA to the E site. Once there, it is free to dissociate to complete the cycle.

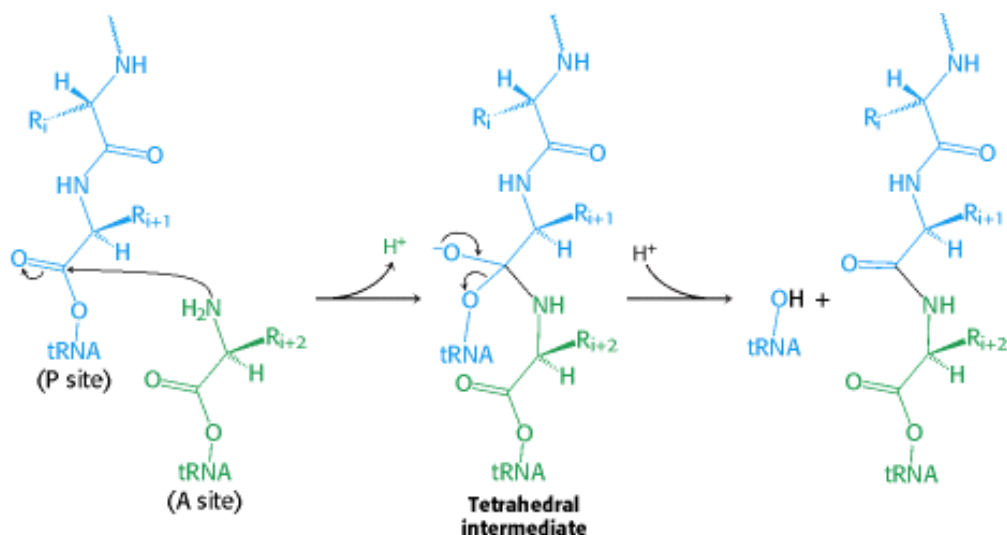


Figure 29.25. Peptide-Bond Formation. The amino group of the aminoacyl-tRNA attacks the carbonyl group of the ester linkage of the peptidyl-tRNA to form a tetrahedral intermediate. This intermediate collapses to form the peptide bond and release the deacylated tRNA.

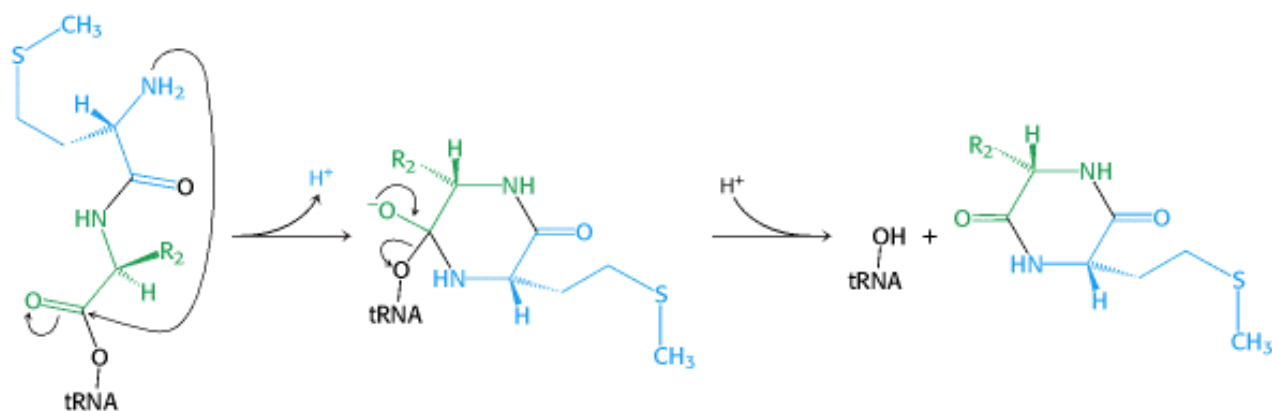


Figure 29.26. A Role for Formylation. With a free terminal amino group, dipeptidyl-tRNA can cyclize to cleave itself from tRNA. Formylation of the amino terminus blocks this reaction.

Table 29.3. Allowed pairings at the third base of the codon according to the wobble hypothesis

First base of anticodon	Third base of codon
C	G
A	U
U	A or G
G	U or C
I	U, C, or A

29.4. Protein Factors Play Key Roles in Protein Synthesis

Although rRNA is paramount in the process of translation, protein factors also are required for the efficient synthesis of a protein. Protein factors participate in the initiation, elongation, and termination of protein synthesis. P-loop NTPases of the G-protein family play particularly important roles. Recall that these proteins serve as molecular switches as they cycle between a GTP-bound form and a GDP-bound form (Section 15.1.2).

29.4.1. Formylmethionyl-tRNA_f Is Placed in the P Site of the Ribosome During Formation of the 70S Initiation Complex

Messenger RNA and formylmethionyl-tRNA_f must be brought to the ribosome for protein synthesis to begin. How is this accomplished? Three protein *initiation factors* (IF1, IF2, and IF3) are essential. The 30S ribosomal subunit first forms a complex with IF1 and IF3 (Figure 29.27). The binding of these factors to the 30S subunit prevents it from prematurely joining the 50S subunit to form a dead-end 70S complex, devoid of mRNA and fMet-tRNA_f. Initiation factor 2, a member of the G-protein family, binds GTP, and the concomitant conformational change enables IF₂ to associate with formylmethionyl-tRNA_f. The IF2-GTP-initiator tRNA complex binds with mRNA (correctly positioned by the Shine-Dalgarno sequence interaction with the 16S rRNA) and the 30S subunit to form the *30S initiation complex*. The hydrolysis of GTP bound to IF2 on entry of the 50S subunit leads to the release of the initiation factors. The result is a *70S initiation complex*.

When the 70S initiation complex has been formed, the ribosome is ready for the elongation phase of protein synthesis. The fMet-tRNA_f molecule occupies the P site on the ribosome. The other two sites for tRNA molecules, the A site and the E site, are empty. Formylmethionyl-tRNA_f is positioned so that its anticodon pairs with the initiating AUG (or GUG) codon on mRNA. This interaction sets the reading frame for the translation of the entire mRNA.

29.4.2. Elongation Factors Deliver Aminoacyl-tRNA to the Ribosome

The second phase of protein synthesis is the elongation cycle. This phase begins with the insertion of an aminoacyl-tRNA into the empty A site on the ribosome. The particular species inserted depends on the mRNA codon in the A site. The cognate aminoacyl-tRNA does not simply leave the synthetase and diffuse to the A site. Rather, it is delivered to the A site in association with a 43-kd protein called *elongation factor Tu* (EF-Tu). Elongation factor Tu, another member of the G-protein family, binds aminoacyl-tRNA only in the GTP form (Figure 29.28). The binding of EF-Tu to aminoacyl-tRNA serves two functions. First, EF-Tu protects the delicate ester linkage in aminoacyl-tRNA from hydrolysis. Second, the GTP in EF-Tu is hydrolyzed to GDP when an appropriate complex between the EF-Tu-aminoacyl-tRNA complex and the ribosome has formed. If the anticodon is not properly paired with the codon, hydrolysis does not take place and the aminoacyl-tRNA is not transferred to the ribosome. This mechanism allows the free energy of GTP hydrolysis to contribute to the fidelity of protein synthesis.

How is EF-Tu in the GDP form reset to bind another aminoacyl-tRNA? *Elongation Factor Ts*, a second elongation factor, joins the EF-Tu complex and induces the dissociation of GDP. Finally, GTP binds to EF-Tu, and EF-Ts is concomitantly released. It is noteworthy that *EF-Tu does not interact with fMet-tRNA_f*. Hence, this initiator tRNA is not delivered to the A site. In contrast, Met-tRNA_m, like all other aminoacyl-tRNAs, does bind to EF-Tu. These findings account for the fact that *internal AUG codons are not read by the initiator tRNA*. Conversely, initiation factor 2 recognizes fMet-tRNA_f but no other tRNA.

This GTP-GDP cycle of EF-Tu is reminiscent of those of the heterotrimeric G proteins in signal transduction (Section 15.1.2) and the Ras proteins in growth control (Section 15.4.2). This similarity is due to their evolutionary heritage, inasmuch as the amino-terminal domain of EF-Tu is homologous to the P-loop NTPase domains in the other G proteins.

The other two domains of the tripartite EF-Tu are distinctive; they mediate interactions with aminoacyl-tRNA and the ribosome. In all these related enzymes, the change in conformation between the GTP and the GDP forms leads to a change in interaction partners. A further similarity is the requirement that an additional protein catalyze the exchange of GTP for GDP; an activated receptor plays the role of EF-Ts for a heterotrimeric G protein, as does Sos for Ras.

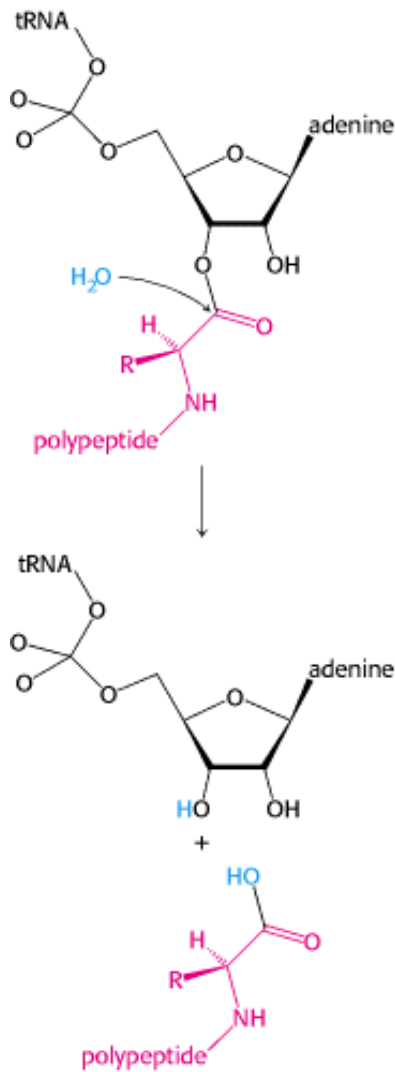
29.4.3. The Formation of a Peptide Bond Is Followed by the GTP-Driven Translocation of tRNAs and mRNA

After the correct aminoacyl-tRNA has been placed in the A site, the transfer of the polypeptide chain from the tRNA in the P site is a spontaneous process, driven by the formation of the stronger peptide bond in place of the ester linkage. However, protein synthesis cannot continue without the translocation of the mRNA and the tRNAs within the ribosome. The mRNA must move by a distance of three nucleotides as the deacylated tRNA moves out of the P site into the E site on the 30S subunit and the peptidyl-tRNA moves out of the A site into the P site on the 30S subunit. The result is that the next codon is positioned in the A site for interaction with the incoming aminoacyl-tRNA.

Translocation is mediated by *elongation factor G* (EF-G, also called *translocase*). The structure of EF-G is exceptional in revealing some aspects of its mode of action (Figure 29.29). The structure of EF-G closely resembles that of the complex between EF-Tu and tRNA. This is an example of *molecular mimicry*; a protein domain evolved so that it mimics the shape of a tRNA molecule. This structural similarity, as well as other experimental data, suggests a mechanism for the translocation process (Figure 29.30). First, EF-G in the GTP form binds to the ribosome, primarily through the interaction of its EF-Tu-like domain with the 50S subunit. The binding site includes proteins L11 and the L7-L12 dimer. The tRNA-like domain of EF-G interacts with the 30S subunit. The binding of EF-G to the ribosome in this manner stimulates the GTPase activity of EF-G. On GTP hydrolysis, EF-G undergoes a conformational change that forces its arm deeper into the A site on the 30S subunit. To accommodate this domain, the peptidyl-tRNA in the A site moves to the P site, carrying the mRNA and the deacylated tRNA with it. The ribosome may be prepared for these rearrangements by the initial binding of EF-G as well. The dissociation of EF-G leaves the ribosome ready to accept the next aminoacyl-tRNA into the A site.


29.4.4. Protein Synthesis Is Terminated by Release Factors That Read Stop Codons

The final phase of translation is termination. How does the synthesis of a polypeptide chain come to an end when a stop codon is encountered? Aminoacyl-tRNA does not normally bind to the A site of a ribosome if the codon is UAA, UGA, or UAG, because normal cells do not contain tRNAs with anticodons complementary to these stop signals. Instead, these *stop codons are recognized by release factors* (RFs), which are proteins. One of these release factors, RF1, recognizes UAA or UAG. A second factor, RF2, recognizes UAA or UGA. A third factor, RF3, another G protein homologous to EF-Tu, mediates interactions between RF1 or RF2 and the ribosome.



Release factors use a Trojan horse strategy to free the polypeptide chain. One of the most impressive properties of the ribosome is *not* that it catalyzes peptide-bond formation; the formation of a peptide bond by the reaction between an amino group and an ester is a facile chemical reaction. Instead, a more impressive feature crucial to ribosome function is that the peptidyl-tRNA ester linkage is not broken by premature hydrolysis. The exclusion of water from the peptidyl transferase center is crucial in preventing such hydrolysis, which would lead to release of the polypeptide chain. The structure of a prokaryotic release factor has not yet been determined. However, the structure of a eukaryotic release factor, though probably not truly homologous to its prokaryotic counterpart, reveals the strategy ([Figure 29.31](#)).

The structure resembles that of a tRNA by molecular mimicry. The sequence Gly-Gly-Gln, present in both eukaryotes and prokaryotes, occurs at the end of the structure corresponding to the acceptor stem of a tRNA. This region binds a water molecule. Disguised as an aminoacyl-tRNA, the release factor may carry this water molecule into the peptidyl transferase center and, assisted by the catalytic apparatus of the ribosome, promote this water molecule's attack on the ester linkage, freeing the polypeptide chain. The detached polypeptide leaves the ribosome. Transfer RNA and messenger RNA remain briefly attached to the 70S ribosome until the entire complex is dissociated in a GTP-dependent fashion by ribosome release factor (RRF) and EF-G. Ribosome release factor is an essential factor for prokaryotic translation.

 The structure of RRF, too, resembles tRNA ([Figure 29.32](#)). However, the known tRNA-mimicking structures of RRF, EF-G, and the release factors are distinct; they do not appear to have been generated from a common ancestor. Thus, convergent evolution has provided a similar solution—looking sufficiently like a tRNA to interact with the tRNA-binding sites on the ribosome—to several problems. The effects of *divergent* evolution are evident in the protein factors that participate in translation, most notably in the form of the homologous G proteins, EF-Tu, EF-G, IF2, and RF3.

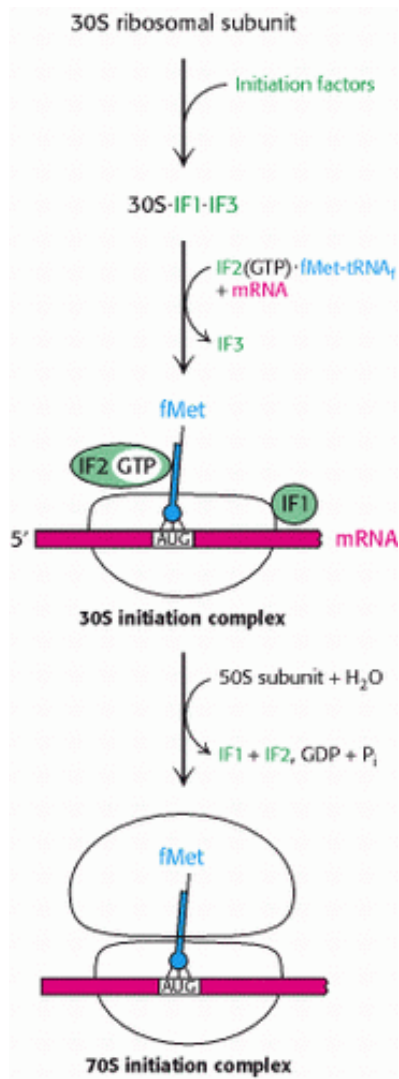


Figure 29.27. Translation Initiation in Prokaryotes. Initiation factors aid the assembly first of the 30S initiation complex and then of the 70S initiation complex.

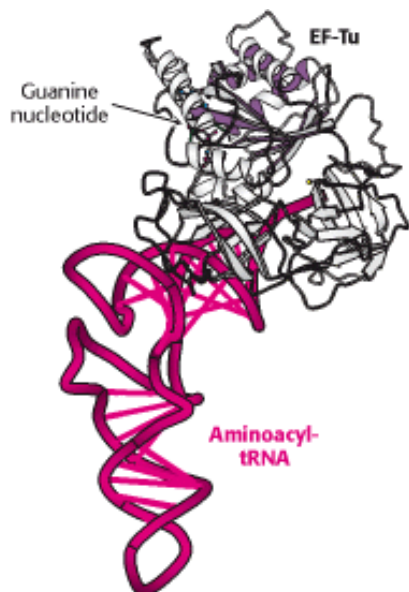


Figure 29.28. Structure of Elongation Factor Tu. The structure of a complex between elongation factor Tu (EF-Tu) and an aminoacyl-tRNA. The amino-terminal domain of EF-Tu is a P-loop NTPase domain similar to those in other G proteins.

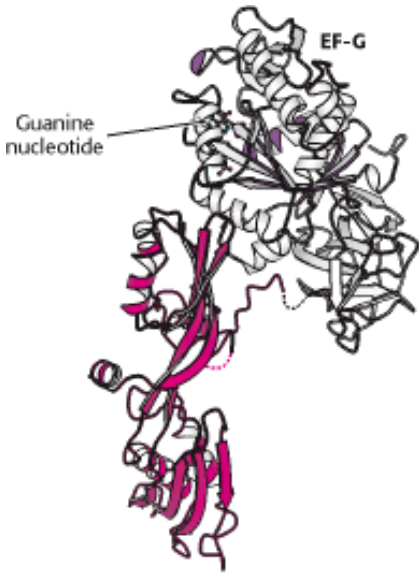


Figure 29.29. Molecular Mimicry. The structure of elongation factor G (EF-G) is remarkably similar in shape to that of the EF-Tu-tRNA complex (see Figure 29.28). The amino-terminal region of EF-G is homologous to EF-Tu, and the carboxyl-terminal region (shown in red) comprises a set of protein domains that adopted the shape of a tRNA molecule over the course of evolution.

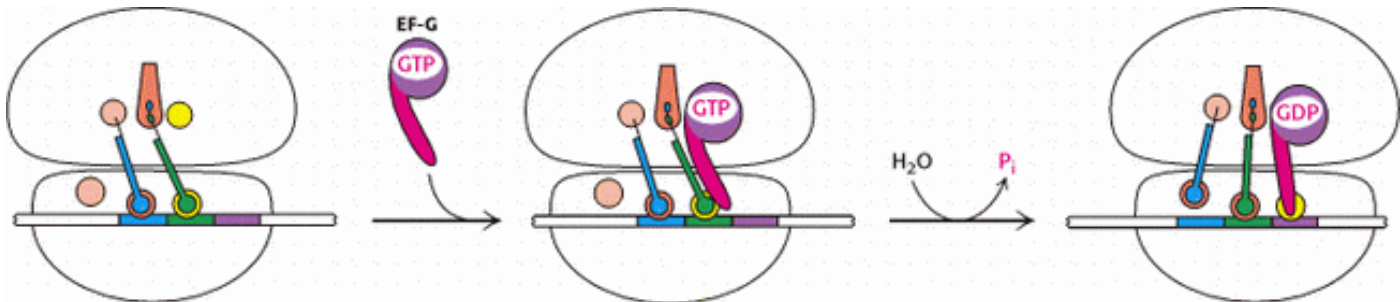


Figure 29.30. Translocation Mechanism. In the GTP form, EF-G binds to the EF-Tu-binding site on the 50S subunit. This stimulates GTP hydrolysis, inducing a conformational change in EF-G, and driving the stem of EF-G into the A site on the 30S subunit. To accommodate this domain, the tRNAs and mRNA move through the ribosome by a distance corresponding to one codon.

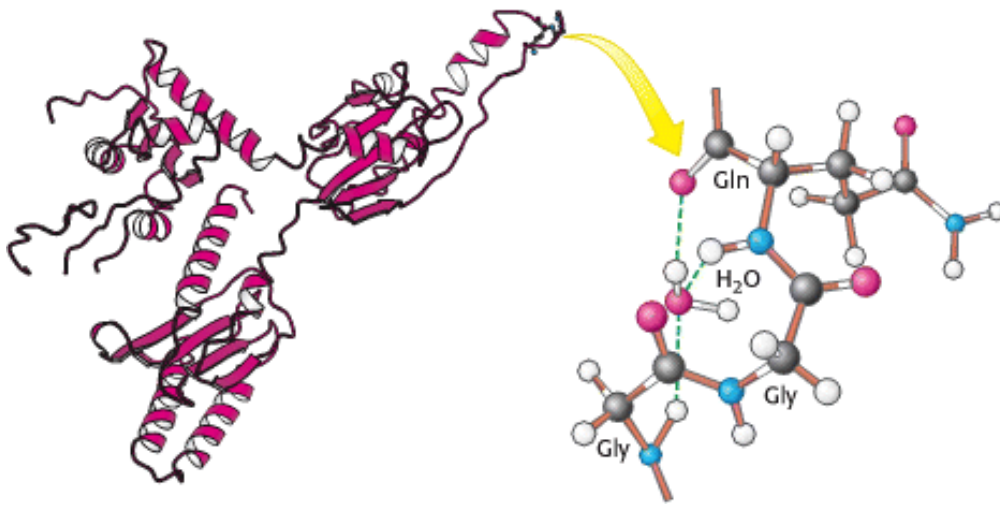


Figure 29.31. Structure of a Release Factor. The structure of a eukaryotic release factor reveals a tRNA-like fold. The α acceptor-stem mimic includes the sequence Gly-Gly-Gln at its tip. This region appears to bind a water molecule, which may be brought into the peptidyl transferase center. There it can participate in the cleavage of the peptidyl-tRNA ester bond, with the aid of the glutamine residue and the ribosomal catalytic apparatus.

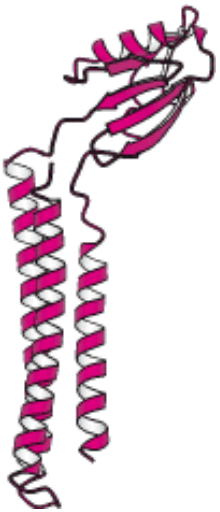


Figure 29.32. Structure of Ribosome Release Factor (RRF). RRF is another protein that resembles tRNA. The α helices of this protein mimic the tRNA structure. In contrast, in EF-G, β strands are the mimics, revealing an independent evolutionary origin.

29.5. Eukaryotic Protein Synthesis Differs from Prokaryotic Protein Synthesis Primarily in Translation Initiation

The basic plan of protein synthesis in eukaryotes and archaea is similar to that in bacteria. The major structural and mechanistic themes recur in all domains of life. However, eukaryotic protein synthesis entails more protein components than does prokaryotic protein synthesis, and some steps are more intricate. Some noteworthy similarities and differences are as follows:

- 1. Ribosomes.** Eukaryotic ribosomes are larger. They consist of a 60S large subunit and a 40S small subunit, which come together to form an 80S particle having a mass of 4200 kd, compared with 2700 kd for the prokaryotic 70S ribosome. The 40S subunit contains an 18S RNA that is homologous to the prokaryotic 16S RNA. The 60S subunit contains three RNAs: the 5S and 28S RNAs are the counterparts of the prokaryotic 5S and 23S molecules; its 5.8S RNA is unique to


eukaryotes.

2. Initiator tRNA. In eukaryotes, the initiating amino acid is methionine rather than *N*-formylmethionine. However, as in prokaryotes, a special tRNA participates in initiation. This aminoacyl-tRNA is called Met-tRNA_i or Met-tRNA_f (the subscript "i" stands for initiation, and "f" indicates that it can be formylated in vitro).

3. Initiation. The initiating codon in eukaryotes is always AUG. Eukaryotes, in contrast with prokaryotes, do not use a specific purine-rich sequence on the 5' side to distinguish initiator AUGs from internal ones. Instead, the AUG nearest the 5' end of mRNA is usually selected as the start site. A 40S ribosome attaches to the cap at the 5' end of eukaryotic mRNA (Section 28.3.1) and searches for an AUG codon by moving step-by-step in the 3' direction (Figure 29.33). This scanning process in eukaryotic protein synthesis is powered by helicases that hydrolyze ATP. Pairing of the anticodon of Met-tRNA_i with the AUG codon of mRNA signals that the target has been found. In almost all cases, eukaryotic mRNA has only one start site and hence is the template for a single protein. In contrast, a prokaryotic mRNA can have multiple Shine-Dalgarno sequences and, hence, start sites, and it can serve as a template for the synthesis of several proteins. Eukaryotes utilize many more initiation factors than do prokaryotes, and their interplay is much more intricate. The prefix *eIF* denotes a eukaryotic initiation factor. For example, eIF-4E is a protein that binds directly to the 7-methylguanosine cap (Section 28.3.1), whereas eIF-4A is a helicase. The difference in initiation mechanism between prokaryotes and eukaryotes is, in part, a consequence of the difference in RNA processing. The 5' end of mRNA is readily available to ribosomes immediately after transcription in prokaryotes. In contrast, pre-mRNA must be processed and transported to the cytoplasm in eukaryotes before translation is initiated. Thus, there is ample opportunity for the formation of complex secondary structures that must be removed to expose signals in the mature mRNA. The 5' cap provides an easily recognizable starting point. In addition, the complexity of eukaryotic translation initiation provides another mechanism for gene expression that we shall explore further in Chapter 31.

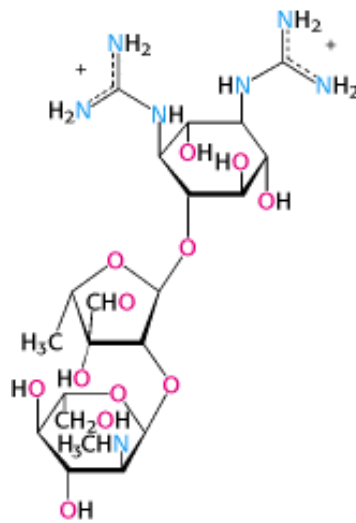
4. Elongation and termination. Eukaryotic elongation factors EF1 α and EF1 β γ are the counterparts of prokaryotic EF-Tu and EF-Ts. The GTP form of EF1 α delivers aminoacyl-tRNA to the A site of the ribosome, and EF1 β γ catalyzes the exchange of GTP for bound GDP. Eukaryotic EF2 mediates GTP-driven translocation in much the same way as does prokaryotic EF-G. Termination in eukaryotes is carried out by a single release factor, eRF1, compared with two in prokaryotes. Finally, eIF3, like its prokaryotic counterpart IF3, prevents the reassociation of ribosomal subunits in the absence of an initiation complex.

29.5.1. Many Antibiotics Work by Inhibiting Protein Synthesis

 The differences between eukaryotic and prokaryotic ribosomes can be exploited for the development of antibiotics (Table 29.4). For example, the antibiotic *puromycin* inhibits protein synthesis by causing nascent prokaryotic polypeptide chains to be released before their synthesis is completed. Puromycin is an analog of the terminal aminoacyl-adenosine part of aminoacyl-tRNA (Figure 29.34).


It binds to the A site on the ribosome and inhibits the entry of aminoacyl-tRNA. Furthermore, puromycin contains an α -amino group. This amino group, like the one on aminoacyl-tRNA, forms a peptide bond with the carboxyl group of the growing peptide chain. The product, a peptide having a covalently attached puromycin residue at its carboxyl end, dissociates from the ribosome.

Streptomycin, a highly basic trisaccharide, interferes with the binding of formylmethionyl-tRNA to ribosomes and thereby prevents the correct initiation of protein synthesis. Other *aminoglycoside antibiotics* such as neomycin, kanamycin, and gentamycin interfere with the *decoding site* located near nucleotide 1492 in 16S rRNA of the 30S subunit (Section 29.3.9). *Chloramphenicol* acts by inhibiting peptidyl transferase activity. *Erythromycin* binds to the 50S subunit and blocks translocation. Finally, *cyclohexamide* blocks peptidyl transferase activity in eukaryotic ribosomes, making a useful laboratory tool for blocking protein synthesis in eukaryotic cells.



Streptomycin

29.5.2. Diphtheria Toxin Blocks Protein Synthesis in Eukaryotes by Inhibiting Translocation

 Diphtheria was a major cause of death in childhood before the advent of effective immunization. The lethal effects of this disease are due mainly to a protein toxin produced by *Corynebacterium diphtheriae*, a bacterium that grows in the upper respiratory tract of an infected person. The gene that encodes the toxin comes from a lysogenic phage that is harbored by some strains of *C. diphtheriae*. A few micrograms of diphtheria toxin is usually lethal in an unimmunized person because it inhibits protein synthesis. The toxin is cleaved shortly after entering a target cell into a 21-kd A fragment and a 40-kd B fragment. *The A fragment of the toxin catalyzes the covalent modification of an important component of the protein-synthesizing machinery, whereas the B fragment enables the A fragment to enter the cytosol of its target cell.*

A single A fragment of the toxin in the cytosol can kill a cell. Why is it so lethal? The target of the A fragment is EF2, the elongation factor catalyzing translocation in eukaryotic protein synthesis. EF2 contains *diphthamide*, an unusual amino acid residue of unknown function that is formed by posttranslational modification of histidine. The A fragment catalyzes the transfer of the adenosine diphosphate ribose unit of NAD^+ to a nitrogen atom of the diphthamide ring (Figure 29.35). *This ADP-ribosylation of a single side chain of EF2 blocks its capacity to carry out translocation of the growing polypeptide chain.* Protein synthesis ceases, accounting for the remarkable toxicity of diphtheria toxin.

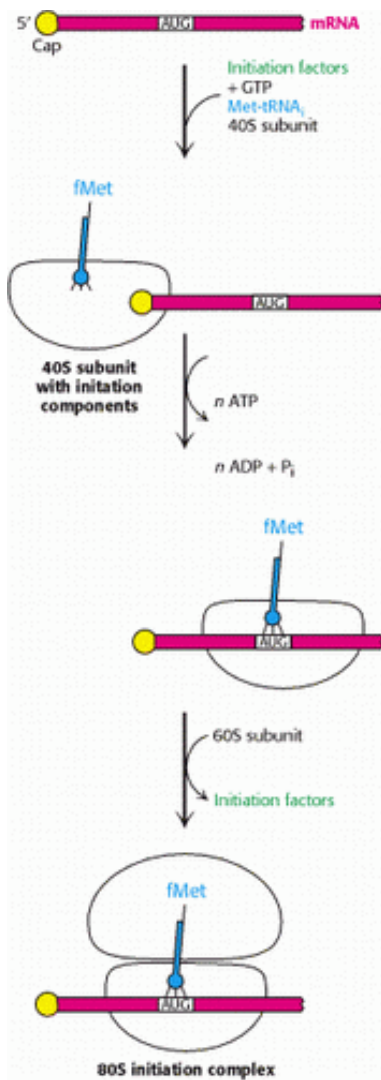


Figure 29.33. Eukaryotic Translation Initiation. In eukaryotes, translation initiation starts with the assembly of a complex on the 5' cap that includes the 40S subunit and Met-tRNA₁. Driven by ATP hydrolysis, this complex scans the mRNA until the first AUG is reached. The 60S subunit is then added to form the 80S initiation complex.

Table 29.4. Antibiotic inhibitors of protein synthesis

Antibiotic	Action
Streptomycin and other aminoglycosides	Inhibit initiation and cause misreading of mRNA (prokaryotes)
Tetracycline	Binds to the 30S subunit and inhibits binding of aminoacyl-tRNAs (prokaryotes)
Chloramphenicol	Inhibits the peptidyl transferase activity of the 50S ribosomal subunit (prokaryotes)
Cycloheximide	Inhibits the peptidyl transferase activity of the 60S ribosomal subunit (eukaryotes)
Erythromycin	Binds to the 50S subunit and inhibits translocation (prokaryotes)
Puromycin	Causes premature chain termination by acting as an analog of aminoacyl-tRNA (prokaryotes and eukaryotes)

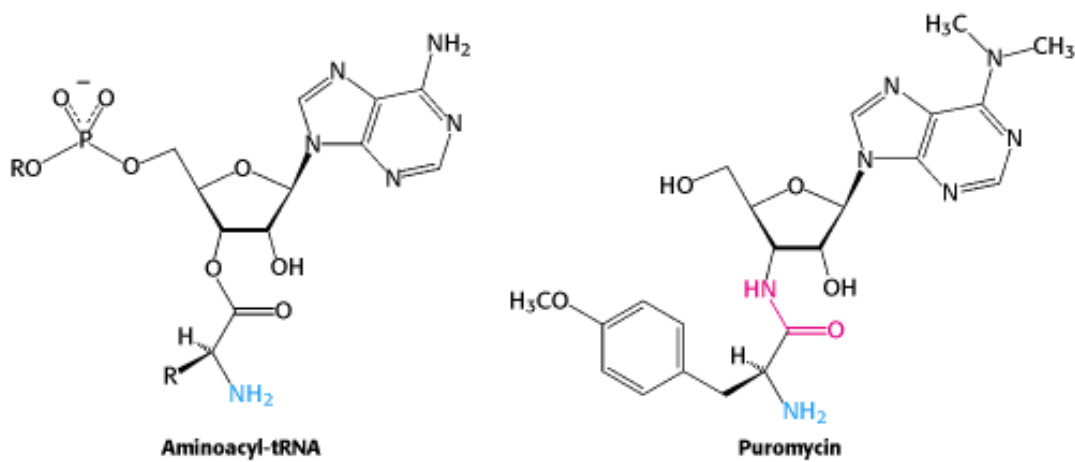


Figure 29.34. Antibiotic Action of Puromycin. Puromycin resembles the aminoacyl terminus of an aminoacyl-tRNA. Its amino group joins the carbonyl group of the growing polypeptide chain to form an adduct that dissociates from the ribosome. This adduct is stable because puromycin has an amide (shown in red) rather than an ester linkage.

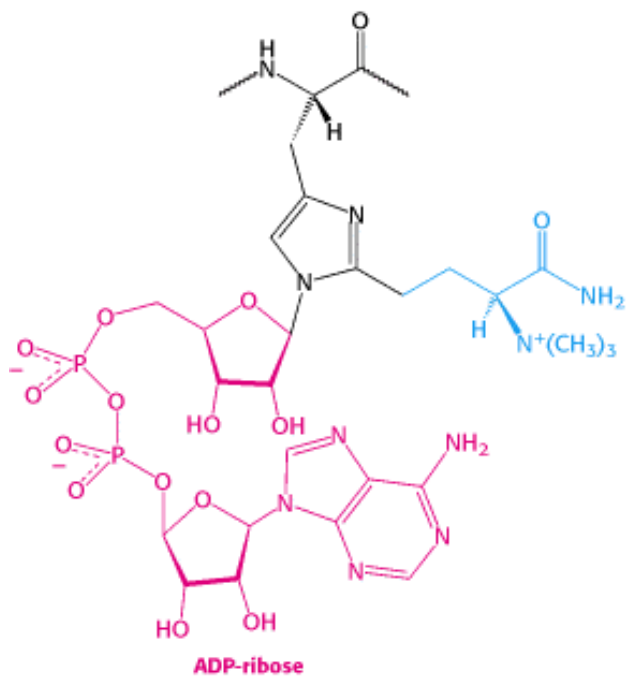


Figure 29.35. Blocking of Translocation by Diphtheria Toxin. Diphtheria toxin blocks protein synthesis in eukaryotes by catalyzing the transfer of an ADP-ribose unit from NAD^+ to diphthamide, a modified amino acid residue in elongation factor 2 (translocase). Diphthamide is formed by a posttranslational modification (blue) of a histidine residue.

Summary

Protein Synthesis Requires the Translation of Nucleotide Sequences into Amino Acid Sequences

Protein synthesis is called translation because information present as a nucleic acid sequence is translated into a different language, the sequence of amino acids in a protein. This complex process is mediated by the coordinated interplay of more than a hundred macromolecules, including mRNA, rRNAs, tRNAs, aminoacyl-tRNA synthetases, and protein factors. Given that proteins typically comprise from 100 to 1000 amino acids, the frequency at which an incorrect amino acid is incorporated in the course of protein synthesis must be less than 10^{-4} . Transfer RNAs are the adaptors that make the link between a nucleic acid and an amino acid. These molecules, single chains of about 80 nucleotides, have an L-shaped structure.

Aminoacyl-Transfer-RNA Synthetases Read the Genetic Code

Each amino acid is activated and linked to a specific transfer RNA by an enzyme called an aminoacyl-tRNA synthetase. Such an enzyme links the carboxyl group of an amino acid to the 2'- or 3'-hydroxyl group of the adenosine unit of a CCA sequence at the 3' end of the tRNA by an ester linkage. There is at least one specific aminoacyl-tRNA synthetase and at least one specific tRNA for each amino acid. A synthetase utilizes both the functional groups and the shape of its cognate amino acid to prevent the attachment of an incorrect amino acid to a tRNA. Some synthetases have a separate active site at which incorrectly linked amino acids are removed by hydrolysis. A synthetase recognizes the anticodon, the acceptor stem, and sometimes other parts of its tRNA substrate. By specifically recognizing both amino acids and tRNAs, aminoacyl-tRNA synthetases implement the instruction of the genetic code. There exist two evolutionary distinct classes of synthetases, each recognizing 10 amino acids. The two classes recognize opposite faces of tRNA molecules.

A Ribosome Is a Ribonucleoprotein Particle (70S) Made of a Small (30S) and a Large (50S) Subunit

Protein synthesis takes place on ribosomes—ribonucleoprotein particles (about two-thirds RNA and one-third protein) consisting of large and small subunits. In *E. coli*, the 70S ribosome (2700 kd) is made up of 30S and 50S subunits. The 30S subunit consists of 16S ribosomal RNA and 21 different proteins; the 50S subunit consists of 23S and 5S rRNA and 34 different proteins. The structure of almost all components of the ribosome have now been determined at or near atomic resolution.

Proteins are synthesized in the amino-to-carboxyl direction, and mRNA is translated in the $5' \rightarrow 3'$ direction. The start signal on prokaryotic mRNA is AUG (or GUG) preceded by a purine-rich sequence that can base-pair with 16S rRNA. In prokaryotes, transcription and translation are closely coupled. Several ribosomes can simultaneously translate an mRNA, forming a polysome.

The ribosome includes three sites for tRNA binding called the A (aminoacyl) site, the P (peptidyl) site, and the E (exit) site. With a tRNA attached to the growing peptide chain in the P site, an aminoacyl-tRNA binds to the A site. A peptide bond is formed when the amino group of the aminoacyl-tRNA nucleophilically attacks the ester carbonyl group of the peptidyl-tRNA. On peptide-bond formation, the tRNAs and mRNA must be translocated for the next cycle to begin. The deacylated tRNA moves to the E site and then leaves the ribosome, and the peptidyl-tRNA moves from the A site into the P site.

The codons of messenger RNA recognize the anticodons of transfer RNAs rather than the amino acids attached to the tRNAs. A codon on mRNA forms base pairs with the anticodon of the tRNA. Some tRNAs are recognized by more than one codon because pairing of the third base of a codon is less crucial than that of the other two (the wobble mechanism).

Protein Factors Play Key Roles in Protein Synthesis

Protein synthesis takes place in three phases: initiation, elongation, and termination. In prokaryotes, mRNA, formylmethionyl-tRNA_f (the special initiator tRNA that recognizes AUG), and a 30S ribosomal subunit come together with the assistance of initiation factors to form a 30S initiation complex. A 50S ribosomal subunit then joins this complex to form a 70S initiation complex, in which fMet-tRNA_f occupies the P site of the ribosome.

Elongation factor Tu delivers the appropriate aminoacyl-tRNA to the ribosome A (aminoacyl) site as an EF-Tu · aminoacyl-tRNA · GTP ternary complex. EF-Tu serves both to protect the aminoacyl-tRNA from premature cleavage and to increase the fidelity of protein synthesis by ensuring that the correct codon-anticodon pairing has taken place before hydrolyzing GTP and releasing aminoacyl-tRNA into the A site. Elongation factor G uses the free energy of GTP hydrolysis to drive translocation. Protein synthesis is terminated by release factors, which recognize the termination codons UAA, UGA, and UAG and cause the hydrolysis of the ester bond between the polypeptide and tRNA.

Eukaryotic Protein Synthesis Differs from Prokaryotic Protein Synthesis Primarily in Translation Initiation

The basic plan of protein synthesis in eukaryotes is similar to that of prokaryotes, but there are some significant differences between them. Eukaryotic ribosomes (80S) consist of a 40S small subunit and a 60S large subunit. The initiating amino acid is again methionine, but it is not formylated. The initiation of protein synthesis is more complex in eukaryotes than in prokaryotes. The AUG closest to the 5' end of mRNA is nearly always the start site. The 40S ribosome finds this site by binding to the 5' cap and then scanning the RNA until AUG is reached. The regulation of translation in eukaryotes provides a means for regulating gene expression. Many antibiotics act by blocking prokaryotic gene expression.

Key Terms

translation

ribosome

transfer RNA (tRNA)

codon

anticodon

aminoacyl-tRNA synthetase

50S subunit

30S subunit

polysome

Shine-Dalgarno sequence

peptidyl transferase center

wobble hypothesis

initiation factor

elongation factor

elongation factor Tu (EF-Tu)

elongation factor Ts (EF-Ts)

elongation factor G (EF-G)

molecular mimicry

release factor

Problems

1. *Synthetase mechanism.* The formation of isoleucyl-tRNA proceeds through the reversible formation of an enzyme-bound Ile-AMP intermediate. Predict whether ^{32}P -labeled ATP is formed from $^{32}\text{PP}_i$ when each of the following sets of components is incubated with the specific activating enzyme:

(a) ATP and $^{32}\text{PP}_i$

(b) tRNA, ATP, and $^{32}\text{PP}_i$

(c) Isoleucine, ATP, and $^{32}\text{PP}_i$

See answer

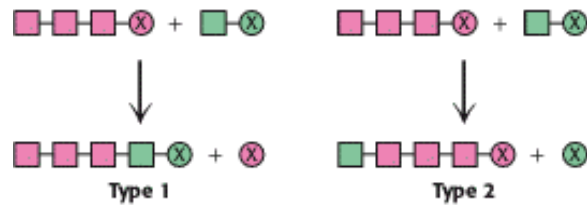
2. *Light and heavy ribosomes.* Ribosomes were isolated from bacteria grown in a "heavy" medium (^{13}C and ^{15}N) and from bacteria grown in a "light" medium (^{12}C and ^{14}N). These 60S ribosomes were added to an in vitro system actively engaged in protein synthesis. An aliquot removed several hours later was analyzed by density-gradient centrifugation. How many bands of 70S ribosomes would you expect to see in the density gradient?

See answer

3. *The price of protein synthesis.* What is the smallest number of molecules of ATP and GTP consumed in the synthesis of a 200-residue protein, starting from amino acids? Assume that the hydrolysis of PP_i is equivalent to the hydrolysis of ATP for this calculation.

See answer

4. *Contrasting modes of elongation.* The two basic mechanisms for the elongation of biomolecules are represented in the adjoining illustration. In type 1, the activating group (X) is released from the growing chain. In type 2, the activating group is released from the incoming unit as it is added to the growing chain. Indicate whether each of the following biosyntheses is by means of a type 1 or a type 2 mechanism:



- (a) Glycogen synthesis
 (b) Fatty acid synthesis
 (c) $C_5 \rightarrow C_{10} \rightarrow C_{15}$ in cholesterol synthesis
 (d) DNA synthesis
 (e) RNA synthesis
 (f) Protein synthesis

See answer

5. *Suppressing frameshifts.* The insertion of a base in a coding sequence leads to a shift in the reading frame, which in most cases produces a nonfunctional protein. Propose a mutation in a tRNA that might suppress frameshifting.

See answer

6. *Tagging a ribosomal site.* Design an affinity-labeling reagent for one of the tRNA binding sites in *E. coli* ribosomes.

See answer

7. *Viral mutation.* An mRNA transcript of a T7 phage gene contains the base sequence



Predict the effect of a mutation that changes the G marked by an arrow to A.

See answer

8. *Two synthetic modes.* Compare and contrast protein synthesis by ribosomes with protein synthesis by the solid-phase method (see [Section 4.4](#)).

See answer

9. *Enhancing fidelity.* Compare the accuracy of (a) DNA replication, (b) RNA synthesis, and (c) protein synthesis. Which mechanisms are used to ensure the fidelity of each of these processes?

See answer

10. *Triggered GTP hydrolysis.* Ribosomes markedly accelerate the hydrolysis of GTP bound to the complex of EF-Tu and aminoacyl-tRNA. What is the biological significance of this enhancement of GTPase activity by ribosomes?

See answer

11. *Blocking translation.* Devise an experimental strategy for switching off the expression of a specific mRNA without changing the gene encoding the protein or the gene's control elements.

See answer

12. *Directional problem.* Suppose that you have a protein synthesis system that is actively synthesizing a protein designated A. Furthermore, you know that protein A has four trypsin-sensitive sites, equally spaced in the protein, that, on digestion with trypsin, yield the peptides A₁, A₂, A₃, A₄, and A₅. Peptide A₁ is the amino-terminal peptide, and A₅ is the carboxyl peptide. Finally, you know that your system requires 4 minutes to synthesize a complete protein A. At $t = 0$, you add all 20 amino acids, each carrying a ¹⁴C label.

(a) At $t = 1$ minute, you isolate intact protein A from the system, cleave it with trypsin, and isolate the five peptides. Which peptide is most heavily labeled?

(b) At $t = 3$ minutes, what will be the order of labeling of peptides from greatest to least?

(c) What does this experiment tell you about the direction of protein synthesis?

See answer

13. *Translator.* Aminoacyl-tRNA synthetases are the only component of gene expression that decodes the genetic code. Explain.

See answer

14. *A timing device.* EF-Tu, a member of the G-protein family, plays a crucial role in the elongation process of translation. Suppose that a slowly hydrolyzable analog of GTP were added to an elongating system. What would be the effect on rate of protein synthesis?

See answer

Mechanism Problems

15. *Molecular attack.* What is the nucleophile in the reaction catalyzed by peptidyl transferase? Write out a plausible mechanism for this reaction.

See answer

16. *Evolutionary amino acid choice.* Ornithine is structurally similar to lysine except ornithine's side chain is one methylene group shorter than that of lysine. Attempts to chemically synthesize and isolate ornithinyl-tRNA proved unsuccessful. Propose a mechanistic explanation. (Hint: Six-membered rings are more stable than seven-membered rings).

See answer

Chapter Integration Problems

17. *Déjà vu.* Which protein in G-protein cascades plays a role similar to that of elongation factor Ts?

See answer

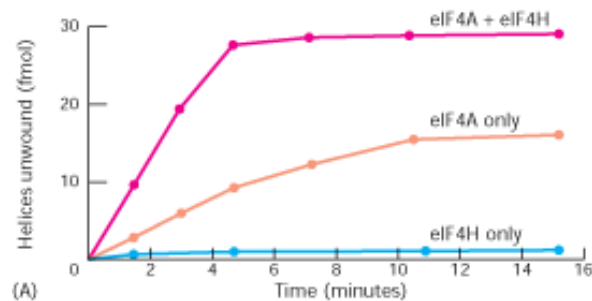
18. *Family resemblance.* Eukaryotic elongation factor 2 is inhibited by ADP ribosylation catalyzed by diphtheria toxin. What other G proteins are sensitive to this mode of inhibition?

See answer

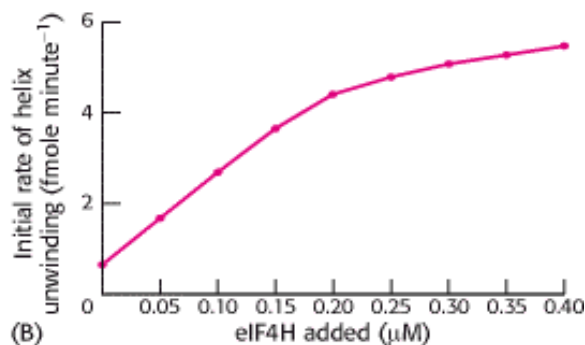
Data Interpretation Problem

19. *Helicase helper.* The initiation factor eIF4A displays ATP-dependent RNA helicase activity. Another initiation factor, eIF4H, has been proposed to assist the action of eIF4A. Graph A shows some of the experimental results from an assay that can measure the activity of eIF4A helicase in the presence of eIF4H.

- (a) What are the effects on eIF4A helicase activity in the presence of eIF4H?
- (b) Why did measuring the helicase activity of eIF4H alone serve as an important control?

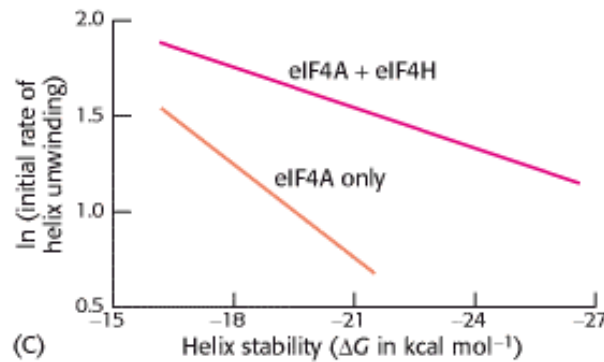


- (c) The initial rate of helicase activity of 0.2 μM of eIF4A was then measured with varying amounts of eIF4H (graph B). What ratio of eIF4H to eIF4A yielded optimal activity?



- (d) Next, the effect of RNA-RNA helix stability on the initial rate of unwinding in the presence and absence of

eIF4H was tested (graph C). How does the effect of eIF4H vary with helix stability?




(e) How might eIF4H affect the helicase activity of eIF4A?

[Data after N. J. Richter, G. W. Rodgers, Jr., J. O. Hensold, and W. C. Merrick, 1999. Further biochemical and kinetic characterization of human eukaryotic initiation factor 4H. *J. Biol. Chem.* 274:35415–35424.]

See answer

Media Problem

20.  *Same difference?* Thr-tRNA synthetase (class II) contains an editing site that recognizes and hydrolyzes misacylated Ser-tRNA^{Thr}. This proofreading ability is needed because serine can fit into and make almost as many favorable interactions with the aminoacylation site as threonine, allowing Thr-tRNA synthetase to mistakenly couple serine to threonyl-tRNA. Which aminoacyl-tRNA synthetase is most likely to have a comparable need to proofread and edit tRNAs misacylated with valine? Given the class of aminoacyl-tRNA synthetases to which this enzyme belongs, how similar do you expect its editing site to be to that of Thr-tRNA synthetase?

Selected Readings

Where to start

- A.E. Dahlberg. 2001. Ribosome structure: The ribosome in action *Science* 292: 868-869. ([PubMed](#))
- M. Ibba, A.W. Curnow, and D. Söll. 1997. Aminoacyl-tRNA synthesis: Divergent routes to a common goal *Trends Biochem. Sci.* 22: 39-42. ([PubMed](#))
- B.K. Davis. 1999. Evolution of the genetic code *Prog. Biophys. Mol. Biol.* 72: 157-243. ([PubMed](#))
- P. Schimmel and L. Ribas de Pouplana. 2000. Footprints of aminoacyl-tRNA synthetases are everywhere *Trends Biochem. Sci.* 25: 207-209. ([PubMed](#))

Books

- Gesteland, R. F., Cech, T., and Atkins, J. F. (Eds.), 1999. *The RNA World*. Cold Spring Harbor Laboratory Press.
- Garret, R., Douthwaite, S. R., Liljas, A., Matheson, A. T, Moore, P. B., and Noller, H. F., 2000. *The Ribosome: Structure, Function, Antibiotics and Cellular Interactions*. The American Society for Microbiology.

Aminoacyl-tRNA synthetases

- M. Ibba and D. Söll. 2000. Aminoacyl-tRNA synthesis *Annu. Rev. Biochem.* 69: 617-650. ([PubMed](#))
- R. Sankaranarayanan, A.C. Dock-Bregeon, B. Rees, M. Bovee, J. Caillet, P. Romby, C.S. Francklyn, and D. Moras. 2000. Zinc ion mediated amino acid discrimination by threonyl-tRNA synthetase *Nat. Struct. Biol.* 7: 461-465. ([PubMed](#))
- R. Sankaranarayanan, A.C. Dock-Bregeon, P. Romby, J. Caillet, M. Springer, B. Rees, C. Ehresmann, B. Ehresmann, and D. Moras. 1999. The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site *Cell* 97: 371-381. ([PubMed](#))
- A. Dock-Bregeon, R. Sankaranarayanan, P. Romby, J. Caillet, M. Springer, B. Rees, C.S. Francklyn, C. Ehresmann, and D. Moras. 2000. Transfer RNA-mediated editing in threonyl-tRNA synthetase: The class II solution to the double discrimination problem *Cell* 103: 877-884. ([PubMed](#))
- L. Serre, G. Verdon, T. Choinowski, N. Hervouet, J.L. Risler, and C. Zelwer. 2001. How methionyl-tRNA synthetase creates its amino acid recognition pocket upon l-methionine binding *J. Mol. Biol.* 306: 863-876. ([PubMed](#))
- P.J. Beuning and K. Musier-Forsyth. 2000. Hydrolytic editing by a class II aminoacyl-tRNA synthetase *Proc. Natl. Acad. Sci. USA* 97: 8916-8920. ([PubMed](#)) ([Full Text in PMC](#))
- M.L. Bovee, W. Yan, B.S. Sproat, and C.S. Francklyn. 1999. tRNA discrimination at the binding step by a class II aminoacyl-tRNA synthetase *Biochemistry* 38: 13725-13735. ([PubMed](#))
- S. Fukai, O. Nureki, S. Sekine, A. Shimada, J. Tao, D.G. Vassylyev, and S. Yokoyama. 2000. Structural basis for double-sieve discrimination of l-valine from l-isoleucine and l-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase *Cell* 103: 793-803. ([PubMed](#))
- L.R. de Pouplana and P. Schimmel. 2000. A view into the origin of life: Aminoacyl-tRNA synthetases *Cell. Mol. Life Sci.* 57: 865-870. ([PubMed](#))
- C.W. Carter Jr. 1993. Cognition, mechanism, and evolutionary relationships in aminoacyl-tRNA synthetases *Annu. Rev. Biochem.* 62: 715-748. ([PubMed](#))

Transfer RNA

- M. Ibba, H.D. Becker, C. Stathopoulos, D.L. Tumbula, and D. Söll. 2000. The adaptor hypothesis revisited *Trends Biochem. Sci.* 25: 311-316. ([PubMed](#))
- B. Weisblum. 1999. Back to Camelot: Defining the specific role of tRNA in protein synthesis *Trends Biochem. Sci.* 24: 247-250. ([PubMed](#))
- J. Normanly and J. Abelson. 1989. Transfer RNA identity *Annu. Rev. Biochem.* 58: 1029-1049. ([PubMed](#))
- R. Basavappa and P.B. Sigler. 1991. The 3 Å crystal structure of yeast initiator tRNA: Functional implications in initiator/elongator discrimination *EMBO J.* 10: 3105-3111. ([PubMed](#))

Ribosomes and ribosomal RNAs

- P.B. Moore. 2001. The ribosome at atomic resolution *Biochemistry* 40: 3243-3250. ([PubMed](#))
- A. Yonath and F. Franceschi. 1998. Functional universality and evolutionary diversity: Insights from the structure of the ribosome *Structure* 6: 679-684. ([PubMed](#))
- M.M. Yusupov, G.Z. Yusupova, A. Baucom, K. Lieberman, T.N. Earnest, J.H. Cate, and H.F. Noller. 2001. Crystal structure of the ribosome at 5.5 Å resolution *Science* 292: 883-896. ([PubMed](#))

N. Ban, P. Nissen, J. Hansen, P.B. Moore, and T.A. Steitz. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution *Science* 289: 905-920. ([PubMed](#))

A.P. Carter, W.M. Clemons, D.E. Brodersen, R.J. Morgan-Warren, B.T. Wimberly, and V. Ramakrishnan. 2000. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics *Nature* 407: 340-348. ([PubMed](#))

B.T. Wimberly, D.E. Brodersen, W.M. Clemons, R.J. Morgan-Warren, A.P. Carter, C. Vornheim, T. Hartsch, and V. Ramakrishnan. 2000. Structure of the 30S ribosomal subunit *Nature* 407: 327-339. ([PubMed](#))

S.C. Agalarov, G. Sridhar Prasad, P.M. Funke, C.D. Stout, and J.R. Williamson. 2000. Structure of the S15,S6,S18-rRNA complex: Assembly of the 30S ribosome central domain *Science* 288: 107-113. ([PubMed](#))

J. Frank. 2000. The ribosome: A macromolecular machine par excellence *Chem. Biol.* 7: R133-R141. ([PubMed](#))

S.A. Woodson and N.B. Leontis. 1998. Structure and dynamics of ribosomal RNA *Curr. Opin. Struct. Biol.* 8: 294-300. ([PubMed](#))

A. Yonath and H.G. Wittmann. 1988. Approaching the molecular structure of ribosomes *Biophys. Chem.* 29: 17-29. ([PubMed](#))

Initiation factors

A.P. Carter, W.M. Clemons Jr, D.E. Brodersen, R.J. Morgan-Warren, T. Hartsch, B.T. Wimberly, and V. Ramakrishnan. 2001. Crystal structure of an initiation factor bound to the 30S ribosomal subunit *Science* 291: 498-501. ([PubMed](#))

M. Guennegues, E. Caserta, L. Brandi, R. Spurio, S. Meunier, C.L. Pon, R. Boelens, and C.O. Gualerzi. 2000. Mapping the fMet-tRNA(f)(Met) binding site of initiation factor IF2 *EMBO J.* 19: 5233-5240. ([PubMed](#))

J.H. Lee, S.K. Choi, A. Roll-Mecak, S.K. Burley, and T.E. Dever. 1999. Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2 *Proc Natl. Acad. Sci. USA* 96: 4342-4347. ([PubMed](#)) ([Full Text in PMC](#))

S. Meunier, R. Spurio, M. Czisch, R. Wechselberger, M. Guennegues, C.O. Gualerzi, and R. Boelens. 2000. Structure of the fMet-tRNA(fMet)-binding domain of *B. stearothermophilus* initiation factor IF2 *EMBO J.* 19: 1918-1926. ([PubMed](#))

Elongation factors

H. Stark, M.V. Rodnina, H.J. Wieden, M. van Heel, and W. Wintermeyer. 2000. Large-scale movement of elongation factor G and extensive conformational change of the ribosome during translocation *Cell* 100: 301-309. ([PubMed](#))

M. Baensch, R. Frank, and J. Kohl. 1998. Conservation of the amino-terminal epitope of elongation factor Tu in eubacteria and Archaea *Microbiology* 144: 2241-2246. ([PubMed](#))

L. Krasny, J.R. Mesters, L.N. Tieleman, B. Kraal, V. Fucik, R. Hilgenfeld, and J. Jonak. 1998. Structure and expression of elongation factor Tu from *Bacillus stearothermophilus* *J. Mol. Biol.* 283: 371-381. ([PubMed](#))

T. Pape, W. Wintermeyer, and M.V. Rodnina. 1998. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the *E. coli* ribosome *EMBO J.* 17: 7490-7497. ([PubMed](#))

O. Piepenburg, T. Pape, J.A. Pleiss, W. Wintermeyer, O.C. Uhlenbeck, and M.V. Rodnina. 2000. Intact aminoacyl-tRNA is required to trigger GTP hydrolysis by elongation factor Tu on the ribosome *Biochemistry* 39: 1734-1738. ([PubMed](#))

Peptide-bond formation and translocation

M. Yarus and M. Welch. 2000. Peptidyl transferase: Ancient and exiguous *Chem. Biol.* 7: R187-R190. ([PubMed](#))

C. Rodriguez-Fonseca, H. Phan, K.S. Long, B.T. Porse, S.V. Kirillov, R. Amils, and R.A. Garrett. 2000. Puromycin-rRNA interaction sites at the peptidyl transferase center *RNA* 6: 744-754. ([PubMed](#))

S.N. Vladimirov, Z. Druzina, R. Wang, and B.S. Cooperman. 2000. Identification of 50S components neighboring 23S rRNA nucleotides A2448 and U2604 within the peptidyl transferase center of *Escherichia coli* ribosomes *Biochemistry* 39: 183-193. ([PubMed](#))

J. Frank and R.K. Agrawal. 2000. A ratchet-like inter-subunit reorganization of the ribosome during translocation *Nature* 406: 318-322. ([PubMed](#))

Termination

T. Fujiwara, K. Ito, and Y. Nakamura. 2001. Functional mapping of ribosome-contact sites in the ribosome recycling factor: A structural view from a tRNA mimic *RNA* 7: 64-70. ([PubMed](#))

K.K. Kim, K. Min, and S.W. Suh. 2000. Crystal structure of the ribosome recycling factor from *Escherichia coli* *EMBO J.* 19: 2362-2370. ([PubMed](#))

D.V. Freistroffer, M. Kwiatkowski, R.H. Buckingham, and M. Ehrenberg. 2000. The accuracy of codon recognition by polypeptide release factors *Proc. Natl. Acad. Sci. USA* 97: 2046-2051. ([PubMed](#)) ([Full Text in PMC](#))

V. Heurgue-Hamard, R. Karimi, L. Mora, J. MacDougall, C. Leboeuf, G. Grentzmann, M. Ehrenberg, and R.H. Buckingham. 1998. Ribosome release factor RF4 and termination factor RF3 are involved in dissociation of peptidyl-tRNA from the ribosome *EMBO J.* 17: 808-816. ([PubMed](#))

L.L. Kisselev and R.H. Buckingham. 2000. Translational termination comes of age *Trends Biochem. Sci.* 25: 561-566. ([PubMed](#))

Fidelity and proofreading

M. Ibba and D. Söll. 1999. Quality control mechanisms during translation *Science* 286: 1893-1897. ([PubMed](#))

M.V. Rodnina and W. Wintermeyer. 2001. Ribosome fidelity: tRNA discrimination, proofreading and induced fit *Trends Biochem. Sci.* 26: 124-130. ([PubMed](#))

C.G. Kurland. 1992. Translational accuracy and the fitness of bacteria *Annu. Rev. Genet.* 26: 29-50. ([PubMed](#))

Fersht, A., 1999. *Structure and Mechanism in Protein Science : A Guide to Enzyme Catalysis and Protein Folding* . W. H. Freeman and Company.

Eukaryotic protein synthesis

M. Kozak. 1999. Initiation of translation in prokaryotes and eukaryotes *Gene* 234: 187-208. ([PubMed](#))

B.S. Negrutskii and A.V. El'skaya. 1998. Eukaryotic translation elongation factor 1 alpha: Structure, expression, functions, and possible role in aminoacyl-tRNA channeling *Prog. Nucleic Acid Res. Mol. Biol.* 60: 47-78. ([PubMed](#))

T. Preiss and M.W. Hentze. 1999. From factors to mechanisms: Translation and translational control in eukaryotes *Curr. Opin. Genet. Dev.* 9: 515-521. ([PubMed](#))

M. Bushell, W. Wood, M.J. Clemens, and S.J. Morley. 2000. Changes in integrity and association of eukaryotic protein synthesis initiation factors during apoptosis *Eur. J. Biochem.* 267: 1083-1091. ([PubMed](#))

S. Das, R. Ghosh, and U. Maitra. 2001. Eukaryotic translation initiation factor 5 functions as a GTPase-activating protein *J. Biol. Chem.* 276: 6720-6726. ([PubMed](#))

J.H. Lee, S.K. Choi, A. Roll-Mecak, S.K. Burley, and T.E. Dever. 1999. Universal conservation in translation initiation revealed by human and archaeal homologs of bacterial translation initiation factor IF2 *Proc. Natl. Acad. Sci. USA* 96: 4342-4347. ([PubMed](#)) ([Full Text in PMC](#))

T.V. Pestova and C.U. Hellen. 2000. The structure and function of initiation factors in eukaryotic protein synthesis *Cell Mol. Life Sci.* 57: 651-674. ([PubMed](#))

Antibiotics and toxins

L. Belova, T. Tenson, L. Xiong, P.M. McNicholas, and A.S. Mankin. 2001. A novel site of antibiotic action in the ribosome: Interaction of evernimicin with the large ribosomal subunit *Proc. Natl. Acad. Sci. USA* 98: 3726-3731. ([PubMed](#)) ([Full Text in PMC](#))

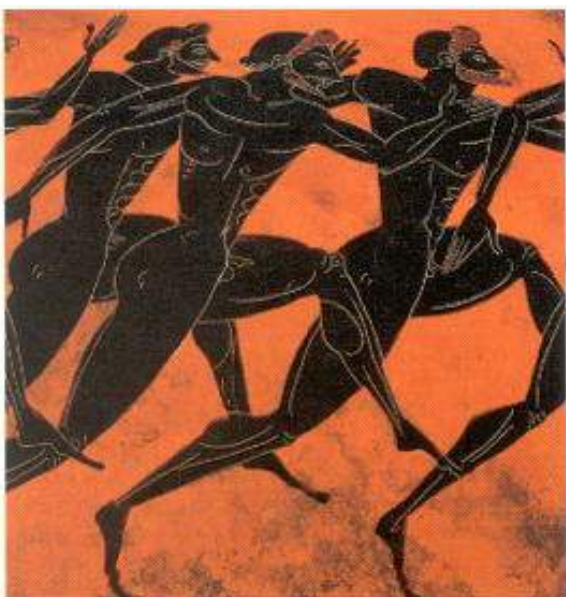
D.E. Brodersen, W.M. Clemons Jr, A.P. Carter, R.J. Morgan-Warren, B.T. Wimberly, and V. Ramakrishnan. 2000. The structural basis for the action of the antibiotics tetracycline, pactamycin, and hygromycin B on the 30S ribosomal subunit *Cell* 103: 1143-1154. ([PubMed](#))

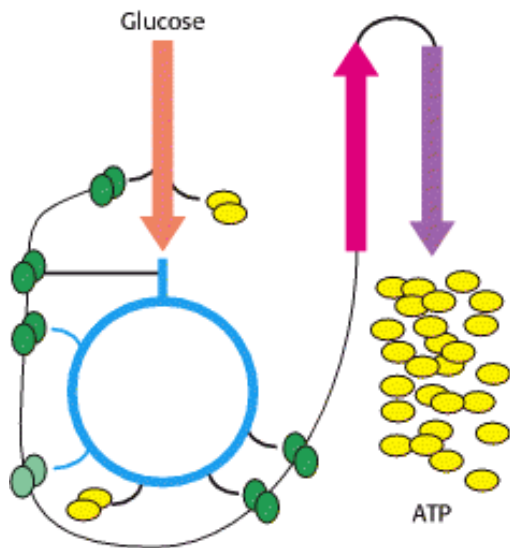
B.T. Porse and R.A. Garrett. 1999. Ribosomal mechanics, antibiotics, and GTP hydrolysis *Cell* 97: 423-426. ([PubMed](#))

D. Eisenberg. 1992. The crystal structure of diphtheria toxin *Nature* 357: 216-222. ([PubMed](#))

30. The Integration of Metabolism

We have been examining the biochemistry of metabolism one pathway at a time, but in living systems many pathways are operating simultaneously. Each pathway must be able to sense the status of the others to function optimally to meet the needs of an organism. How is the intricate network of reactions in metabolism coordinated? This chapter presents some of the principles underlying the *integration of metabolism* in mammals. We begin with a recapitulation of the strategy of metabolism and of recurring motifs in its regulation. We then turn to the interplay of different pathways in regard to the flow of molecules at three key crossroads: glucose 6-phosphate, pyruvate, and acetyl CoA. We consider the differences in the metabolic patterns of the brain, muscle, adipose tissue, kidney, and liver. Finally, we examine how the interplay between these tissues is altered in a variety of metabolic perturbations. These discussions will illustrate how biochemical knowledge illuminates the functioning of the organism.





Interplay of metabolic pathways for energy production. At left, the image shows a detail of runners on a Greek amphora painted in the sixth century, B.C. Athletic feats, and others as seemingly simple as maintenance of blood glucose levels, require elaborate metabolic integration. The schematic above represents the oxidation of glucose to yield ATP in a process requiring interplay among glycolysis, the citric acid cycle, and oxidative phosphorylation. These are a few of the many metabolic pathways that must be coordinated to meet the demands of living. [(Left) Metropolitan Museum of Art, Rogers Fund, 1914 (14.130.12). Copyright © 1977 by the Metropolitan Museum of Art.]

30.1. Metabolism Consist of Highly Interconnected Pathways

The basic strategy of catabolic metabolism is to form ATP, reducing power, and building blocks for biosyntheses. Let us briefly review these central themes:

1. *ATP is the universal currency of energy.* The high phosphoryl transfer potential of ATP enables it to serve as the energy source in muscle contraction, active transport, signal amplification, and biosyntheses. The hydrolysis of an ATP molecule changes the equilibrium ratio of products to reactants in a coupled reaction by a factor of about 10^8 . Hence, *a thermodynamically unfavorable reaction sequence can be made highly favorable by coupling it to the hydrolysis of a sufficient number of ATP molecules.*
2. *ATP is generated by the oxidation of fuel molecules such as glucose, fatty acids, and amino acids.* The common intermediate in most of these oxidations is acetyl CoA. The carbon atoms of the acetyl unit are completely oxidized to CO_2 by the citric acid cycle with the concomitant formation of NADH and FADH_2 . These electron carriers then transfer their high-potential electrons to the respiratory chain. The subsequent flow of electrons to O_2 leads to the pumping of protons across the inner mitochondrial membrane (Figure 30.1). This proton gradient is then used to synthesize ATP. Glycolysis also generates ATP, but the amount formed is much smaller than that in oxidative phosphorylation. The oxidation of glucose to pyruvate yields only 2 molecules of ATP, whereas the complete oxidation of glucose to CO_2 yields 30 molecules of ATP.
3. *NADPH is the major electron donor in reductive biosyntheses.* In most biosyntheses, the products are more reduced than the precursors, and so reductive power is needed as well as ATP. The high-potential electrons required to drive these reactions are usually provided by NADPH. The pentose phosphate pathway supplies much of the required NADPH.
4. *Biomolecules are constructed from a small set of building blocks.* The highly diverse molecules of life are synthesized from a much smaller number of precursors. The metabolic pathways that generate ATP and NADPH also provide building blocks for the biosynthesis of more-complex molecules. For example, acetyl CoA, the common intermediate in the breakdown of most fuels, supplies a two-carbon unit in a wide variety of biosyntheses, such as those leading to fatty acids, prostaglandins, and cholesterol. Thus, *the central metabolic pathways have anabolic as well as catabolic roles.*

5. *Biosynthetic and degradative pathways are almost always distinct.* For example, the pathway for the synthesis of fatty acids is different from that of their degradation. This separation enables both biosynthetic and degradative pathways to be thermodynamically favorable at all times. A biosynthetic pathway is made exergonic by coupling it to the hydrolysis of a sufficient number of ATP molecules. The separation of biosynthetic and degradative pathways contributes greatly to the effectiveness of metabolic control.

"To every thing there is a season, and a time to every purpose under the heaven:

A time to be born, and a time to die; a time to plant, and a time to pluck up that which is planted;

A time to kill, and a time to heal; a time to break down, and a time to build up."

Ecclesiastes 3:1-3

30.1.1. Recurring Motifs in Metabolic Regulation

Anabolism and catabolism must be precisely coordinated. Metabolic networks sense and respond to information on the status of their component pathways. The information is received and metabolism is controlled in several ways:

Pasteur effect-

The inhibition of glycolysis by respiration, discovered by Louis Pasteur in studying fermentation by yeast. The consumption of carbohydrate is about sevenfold lower under aerobic conditions than under anaerobic ones. The inhibition of phosphofructokinase by citrate and ATP accounts for much of the Pasteur effect.

1. *Allosteric interactions.* The flow of molecules in most metabolic pathways is determined primarily by the activities of certain enzymes rather than by the amount of substrate available. Enzymes that catalyze essentially irreversible reactions are likely control sites, and the first irreversible reaction in a pathway (the committed step) is nearly always tightly controlled. Enzymes catalyzing committed steps are allosterically regulated, as exemplified by phosphofructokinase in glycolysis and acetyl CoA carboxylase in fatty acid synthesis. Allosteric interactions enable such enzymes to rapidly detect diverse signals and to adjust their activity accordingly.

2. *Covalent modification.* Some regulatory enzymes are controlled by covalent modification in addition to allosteric interactions. For example, the catalytic activity of glycogen phosphorylase is enhanced by phosphorylation, whereas that of glycogen synthase is diminished. Specific enzymes catalyze the addition and removal of these modifying groups (Figure 30.2). Why is covalent modification used in addition to noncovalent allosteric control? The covalent modification of an essential enzyme in a pathway is often the final step in an amplifying cascade and allows metabolic pathways to be rapidly switched on or off by very low concentrations of triggering signals. In addition, covalent modifications usually last longer (from seconds to minutes) than do reversible allosteric interactions (from milliseconds to seconds).

3. *Enzyme levels.* The amounts of enzymes, as well as their activities, are controlled. The rates of synthesis and degradation of many regulatory enzymes are altered by hormones. The basics of this control were considered in [Chapter 28](#); we will return to the topic in Chapter 31.

4. *Compartmentation.* The metabolic patterns of eukaryotic cells are markedly affected by the presence of compartments (Figure 30.3). The fates of certain molecules depend on whether they are in the cytosol or in mitochondria, and so their flow across the inner mitochondrial membrane is often regulated. For example, fatty acids are transported into mitochondria for degradation only when energy is required, whereas fatty acids in the cytosol are esterified or exported.

5. *Metabolic specializations of organs.* Regulation in higher eukaryotes is enhanced by the existence of organs with different metabolic roles. Metabolic specialization is the result of differential gene expression.

30.1.2. Major Metabolic Pathways and Control Sites

Let us now review the roles of the major pathways of metabolism and the principal sites for their control:

1. *Glycolysis.* This sequence of reactions in the cytosol converts one molecule of glucose into two molecules of pyruvate with the concomitant generation of two molecules each of ATP and NADH. The NAD^+ consumed in the reaction catalyzed by glyceraldehyde 3-phosphate dehydrogenase must be regenerated for glycolysis to proceed. Under anaerobic conditions, as in highly active skeletal muscle, this regeneration is accomplished by the reduction of pyruvate to lactate. Alternatively, under aerobic conditions, NAD^+ is regenerated by the transfer of electrons from NADH to O_2 through the electron-transport chain. Glycolysis serves two main purposes: it degrades glucose to generate ATP, and it provides carbon skeletons for biosyntheses.

Phosphofructokinase, which catalyzes the committed step in glycolysis, is the most important control site. ATP is both a substrate in the phosphoryl transfer reaction and a regulatory molecule. A high level of ATP inhibits phosphofructokinase—the regulatory sites are distinct from the substrate-binding sites and have a lower affinity for the nucleotide. This inhibitory effect is enhanced by citrate and reversed by AMP (Figure 30.4). Thus, the rate of glycolysis depends on the need for ATP, as signaled by the ATP/AMP ratio, and on the availability of building blocks, as signaled by the level of citrate. *In liver, the most important regulator of phosphofructokinase activity is fructose 2,6-bisphosphate (F-2,6-BP).* Recall that the level of F-2,6-BP is determined by the activity of the kinase that forms it from fructose 6-phosphate and of the phosphatase that hydrolyzes the 2-phosphoryl group (Section 16.2.2). When the blood-glucose level is low, a glucagon-triggered cascade leads to activation of the phosphatase and inhibition of the kinase in the liver. The level of F-2,6-BP declines and, consequently, so does phosphofructokinase activity. Hence, glycolysis is slowed, and the spared glucose is released into the blood for use by other tissues.

2. *Citric acid cycle and oxidative phosphorylation.* The reactions of this common pathway for the oxidation of fuel molecules—carbohydrates, amino acids, and fatty acids—take place inside mitochondria. Most fuels enter the cycle as acetyl CoA. The complete oxidation of an acetyl unit by the citric acid cycle generates one molecule of GTP and four pairs of electrons in the form of three molecules of NADH and one molecule FADH_2 . These electrons are transferred to O_2 through the electron-transport chain, which results in the formation of a proton gradient that drives the synthesis of nine molecules of ATP. The electron donors are oxidized and recycled back to the citric acid cycle only if ADP is simultaneously phosphorylated to ATP. *This tight coupling, called respiratory control, ensures that the rate of the citric acid cycle matches the need for ATP.* An abundance of ATP also diminishes the activities of two enzymes in the cycle—*isocitrate dehydrogenase* and *α -ketoglutarate dehydrogenase*. The citric acid cycle has an anabolic role as well. In concert with *pyruvate carboxylase*, the citric acid cycle provides intermediates for biosyntheses, such as succinyl CoA for the formation of porphyrins and citrate for the formation of fatty acids.

3. *Pentose phosphate pathway.* This series of reactions, which takes place in the cytosol, consists of two stages. The first stage is the oxidative decarboxylation of glucose 6-phosphate. Its purpose is the production of NADPH for reductive biosyntheses and the formation of ribose 5-phosphate for the synthesis of nucleotides. Two molecules of NADPH are generated in the conversion of glucose 6-phosphate into ribose 5-phosphate. The dehydrogenation of glucose 6-phosphate is the committed step in this pathway. This reaction is controlled by the level of NADP^+ , the electron acceptor

(Figure 30.5).

The second stage of the pentose phosphate pathway is the nonoxidative, reversible metabolism of five-carbon phosphosugars into phosphorylated three-carbon and six-carbon glycolytic intermediates. Thus, the nonoxidative branch can either introduce riboses into glycolysis for catabolism or generate riboses from glycolytic intermediates for biosyntheses.

4. Gluconeogenesis. Glucose can be synthesized by the liver and kidneys from noncarbohydrate precursors such as lactate, glycerol, and amino acids. The major entry point of this pathway is pyruvate, which is carboxylated to oxaloacetate in mitochondria. Oxaloacetate is then metabolized in the cytosol to form phosphoenolpyruvate. The other distinctive means of gluconeogenesis are two hydrolytic steps that bypass the irreversible reactions of glycolysis. *Gluconeogenesis and glycolysis are usually reciprocally regulated so that one pathway is minimally active while the other is highly active.* For example, AMP inhibits and citrate activates fructose 1,6-bisphosphatase, an essential enzyme in gluconeogenesis, whereas these molecules have opposite effects on phosphofructokinase, the pacemaker of glycolysis (Figure 30.6). Fructose-2,6-bisphosphate also coordinates these processes by inhibiting fructose 1,6-bisphosphatase. Hence, when glucose is abundant, the high level of F-2,6-BP inhibits gluconeogenesis and activates glycolysis.

5. Glycogen synthesis and degradation. Glycogen, a readily mobilizable fuel store, is a branched polymer of glucose residues (Figure 30.7). In glycogen degradation, a phosphorylase catalyzes the cleavage of glycogen by orthophosphate to yield glucose 1-phosphate, which is rapidly converted into glucose 6-phosphate for further metabolism. In glycogen synthesis, the activated intermediate is UDP-glucose, which is formed from glucose 1-phosphate and UTP. Glycogen synthase catalyzes the transfer of glucose from UDP-glucose to the terminal glucose residue of a growing strand. *Glycogen degradation and synthesis are coordinately controlled by a hormone-triggered amplifying cascade so that the phosphorylase is active when synthase is inactive and vice versa.* Phosphorylation and noncovalent allosteric interactions (Section 21.5) regulate these enzymes.

6. Fatty acid synthesis and degradation. Fatty acids are synthesized in the cytosol by the addition of two-carbon units to a growing chain on an acyl carrier protein. Malonyl CoA, the activated intermediate, is formed by the carboxylation of acetyl CoA. Acetyl groups are carried from mitochondria to the cytosol as citrate by the citrate-malate shuttle. In the cytosol, citrate is cleaved to yield acetyl CoA. In addition to transporting acetyl CoA, *citrate in the cytosol stimulates acetyl CoA carboxylase, the enzyme catalyzing the committed step.* When ATP and acetyl CoA are abundant, the level of citrate increases, which accelerates the rate of fatty acid synthesis (Figure 30.8).

A different pathway in a different compartment degrades fatty acids. Carnitine transports fatty acids into mitochondria, where they are degraded to acetyl CoA in the mitochondrial matrix by β -oxidation. The acetyl CoA then enters the citric acid cycle if the supply of oxaloacetate is sufficient. Alternatively, acetyl CoA can give rise to ketone bodies. The FADH₂ and NADH formed in the β -oxidation pathway transfer their electrons to O₂ through the electron-transport chain. Like the citric acid cycle, β -oxidation can continue only if NAD⁺ and FAD are regenerated. Hence, *the rate of fatty acid degradation also is coupled to the need for ATP.* Malonyl CoA, the precursor for fatty acid synthesis, inhibits fatty acid degradation by inhibiting the formation of acyl carnitine by carnitine acyl transferase 1, thus preventing the translocation of fatty acids into mitochondria (Figure 30.9).

30.1.3. Key Junctions: Glucose 6-phosphate, Pyruvate, and Acetyl CoA

The factors governing the flow of molecules in metabolism can be further understood by examining three important molecules: glucose 6-phosphate, pyruvate, and acetyl CoA. Each of these molecules has several contrasting fates:

1. Glucose 6-phosphate. Glucose entering a cell is rapidly phosphorylated to glucose 6-phosphate and is subsequently stored as glycogen, degraded to pyruvate, or converted into ribose 5-phosphate (Figure 30.10). Glycogen is formed when

glucose 6-phosphate and ATP are abundant. In contrast, glucose 6-phosphate flows into the glycolytic pathway when ATP or carbon skeletons for biosyntheses are required. Thus, the conversion of glucose 6-phosphate into pyruvate can be anabolic as well as catabolic. The third major fate of glucose 6-phosphate, to flow through the pentose phosphate pathway, provides NADPH for reductive biosyntheses and ribose 5-phosphate for the synthesis of nucleotides. Glucose 6-phosphate can be formed by the mobilization of glycogen or it can be synthesized from pyruvate and glucogenic amino acids by the gluconeogenic pathway.

2. Pyruvate. This three-carbon α -ketoacid is another major metabolic junction (Figure 30.11). Pyruvate is derived primarily from glucose 6-phosphate, alanine, and lactate. Pyruvate can be reduced to lactate by lactate dehydrogenase to regenerate NAD^+ . This reaction enables glycolysis to proceed transiently under anaerobic conditions in active tissues such as contracting muscle. The lactate formed in active tissue is subsequently oxidized back to pyruvate, in other tissues. The essence of this interconversion buys time and shifts part of the metabolic burden of active muscle to other tissues. Another readily reversible reaction in the cytosol is the transamination of pyruvate, an α -ketoacid, to alanine, the corresponding amino acid. Conversely, several amino acids can be converted into pyruvate. Thus, *transamination is a major link between amino acid and carbohydrate metabolism.*

A third fate of pyruvate is its carboxylation to oxaloacetate inside mitochondria, the first step in gluconeogenesis. This reaction and the subsequent conversion of oxaloacetate into phosphoenolpyruvate bypass an irreversible step of glycolysis and hence enable glucose to be synthesized from pyruvate. The carboxylation of pyruvate is also important for replenishing intermediates of the citric acid cycle. Acetyl CoA activates pyruvate carboxylase, enhancing the synthesis of oxaloacetate, when the citric acid cycle is slowed by a paucity of this intermediate.

A fourth fate of pyruvate is its oxidative decarboxylation to acetyl CoA. *This irreversible reaction inside mitochondria is a decisive reaction in metabolism: it commits the carbon atoms of carbohydrates and amino acids to oxidation by the citric acid cycle or to the synthesis of lipids.* The pyruvate dehydrogenase complex, which catalyzes this irreversible funneling, is stringently regulated by multiple allosteric interactions and covalent modifications. Pyruvate is rapidly converted into acetyl CoA only if ATP is needed or if two-carbon fragments are required for the synthesis of lipids.

3. Acetyl CoA. The major sources of this activated two-carbon unit are the oxidative decarboxylation of pyruvate and the β -oxidation of fatty acids (see Figure 30.11). Acetyl CoA is also derived from ketogenic amino acids. The fate of acetyl CoA, in contrast with that of many molecules in metabolism, is quite restricted. The acetyl unit can be completely oxidized to CO_2 by the citric acid cycle. Alternatively, 3-hydroxy-3-methylglutaryl CoA can be formed from three molecules of acetyl CoA. This six-carbon unit is a precursor of cholesterol and of *ketone bodies*, which are transport forms of acetyl units released from the liver for use by some peripheral tissues. A third major fate of acetyl CoA is its export to the cytosol in the form of citrate for the synthesis of fatty acids.

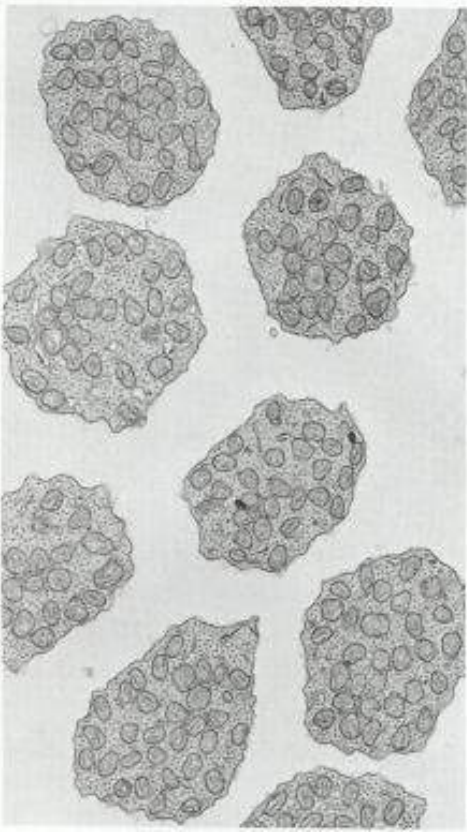


Figure 30.1. Electron Micrograph of Mitochondria. Numerous mitochondria occupy the inner segment of retinal rod cells. These photoreceptor cells generate large amounts of ATP and are highly dependent on a continuous supply of O_2 . [Courtesy of Dr. Michael Hogan.]

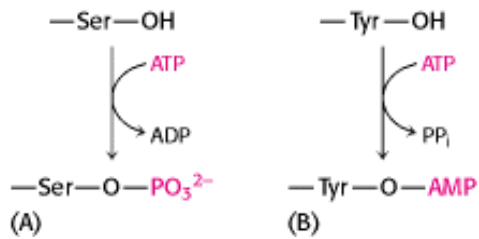


Figure 30.2. Covalent Modifications. Covalent modifications. Examples of reversible covalent modifications of proteins: (A) phosphorylation, (B) adenylation.

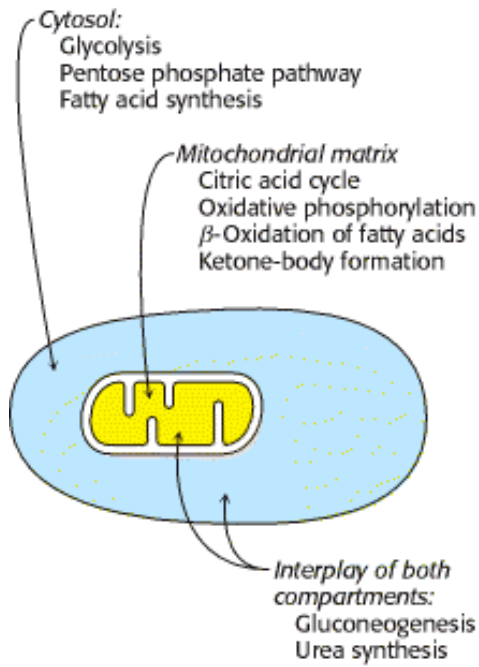


Figure 30.3. Compartmentation of the Major Pathways of Metabolism.

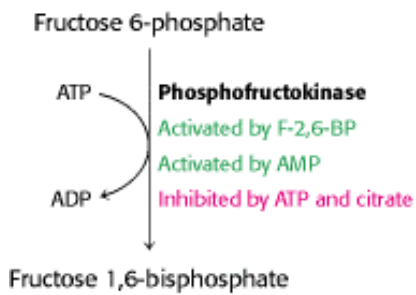


Figure 30.4. Regulation of Glycolysis. Phosphofructokinase is the key enzyme in the regulation of glycolysis.

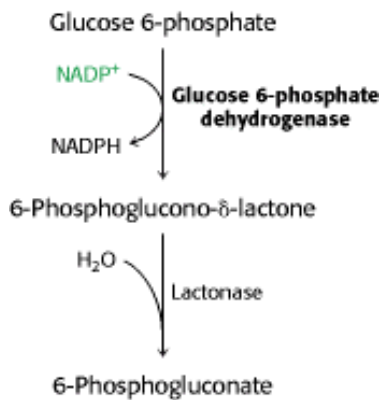


Figure 30.5. Regulation of the Pentose Phosphate Pathway. The dehydrogenation of glucose 6-phosphate is the committed step in the pentose phosphate pathway.

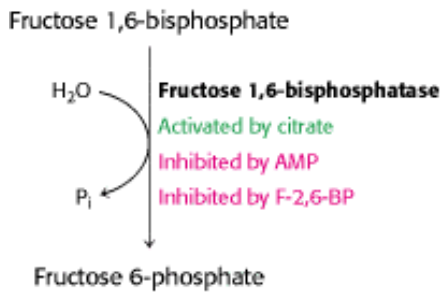


Figure 30.6. Regulation of Gluconeogenesis. Fructose 1,6-bisphosphatase is the principal enzyme controlling the rate of gluconeogenesis.

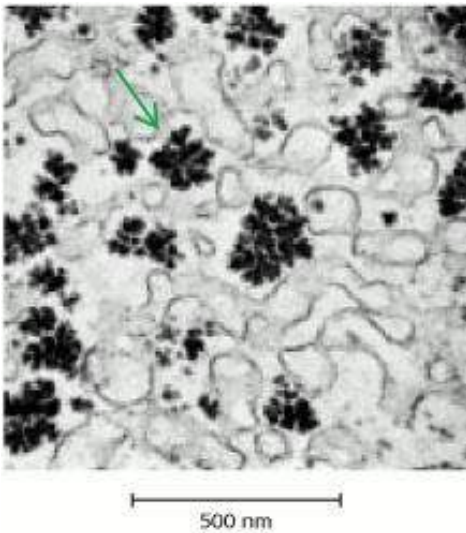


Figure 30.7. Glycogen Granules. The electron micrograph shows part of a liver cell containing glycogen particles. [Courtesy of Dr. George Palade.]

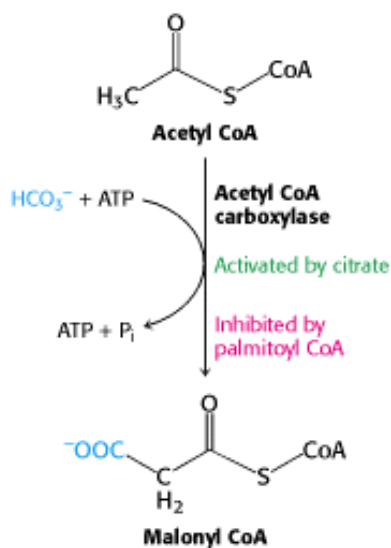


Figure 30.8. Regulation of Fatty Acid Synthesis. Acetyl CoA carboxylase is the key control site in fatty acid synthesis.

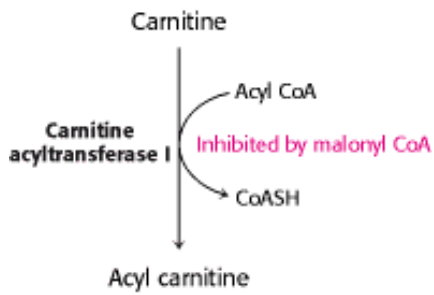


Figure 30.9. Control of Fatty Acid Degradation. Malonyl CoA inhibits fatty acid degradation by inhibiting the formation of acyl carnitine.

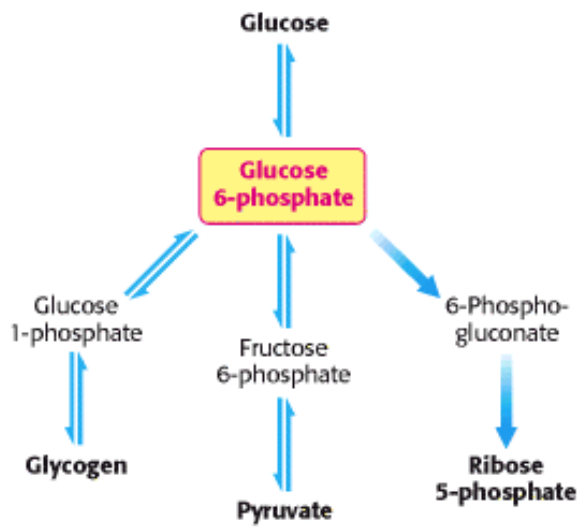


Figure 30.10. Metabolic Fates of Glucose 6-Phosphate.

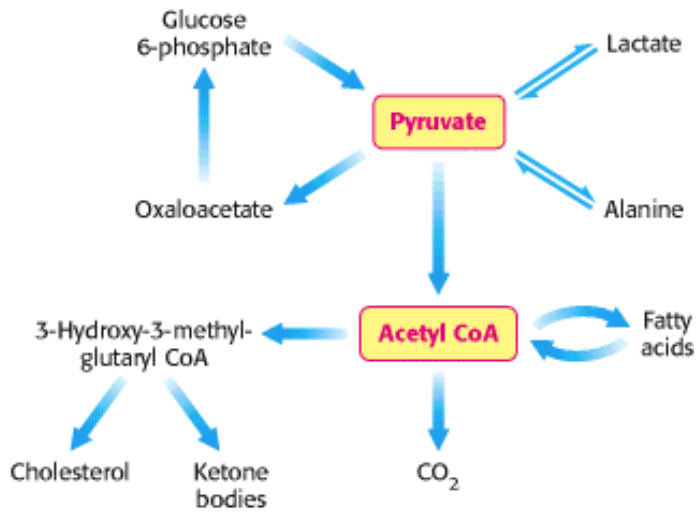


Figure 30.11. Major Metabolic Fates of Pyruvate and Acetyl CoA in Mammals.

30.2. Each Organ Has a Unique Metabolic Profile

The metabolic patterns of the brain, muscle, adipose tissue, kidney, and liver are strikingly different. Let us consider how these organs differ in their use of fuels to meet their energy needs:

1. Brain. *Glucose is virtually the sole fuel for the human brain, except during prolonged starvation.* The brain lacks fuel stores and hence requires a continuous supply of glucose. It consumes about 120 g daily, which corresponds to an energy input of about 420 kcal (1760 kJ), accounting for some 60% of the utilization of glucose by the whole body in the resting state. Much of the energy, estimates suggest from 60% to 70%, is used to power transport mechanisms that maintain the $\text{Na}^+\text{-K}^+$ membrane potential required for the transmission of the nerve impulses. The brain must also synthesize neurotransmitters and their receptors to propagate nerve impulses. Overall, glucose metabolism remains unchanged during mental activity, although local increases are detected when a subject performs certain tasks.

Glucose is transported into brain cells by the glucose transporter GLUT3. This transporter has a low value of K_M for glucose (1.6 mM), which means that it is saturated under most conditions. Thus, the brain is usually provided with a constant supply of glucose. Noninvasive ^{13}C nuclear magnetic resonance measurements have shown that the concentration of glucose in the brain is about 1 mM when the plasma level is 4.7 mM (84.7 mg/dl), a normal value. Glycolysis slows down when the glucose level approaches the K_M value of hexokinase ($\sim 50\ \mu\text{M}$), the enzyme that traps glucose in the cell (Section 16.1.1). This danger point is reached when the plasma-glucose level drops below about 2.2 mM (39.6 mg/dl) and thus approaches the K_M value of GLUT3.

Fatty acids do not serve as fuel for the brain, because they are bound to albumin in plasma and so do not traverse the blood-brain barrier. In starvation, *ketone bodies generated by the liver partly replace glucose as fuel for the brain.*

2. Muscle. *The major fuels for muscle are glucose, fatty acids, and ketone bodies.* Muscle differs from the brain in having a large store of glycogen (1200 kcal, or 5000 kJ). In fact, about three-fourths of all the glycogen in the body is stored in muscle (Table 30.1). This glycogen is readily converted into glucose 6-phosphate for use within muscle cells. Muscle, like the brain, lacks glucose 6-phosphatase, and so it does not export glucose. Rather, *muscle retains glucose, its preferred fuel for bursts of activity.*

In actively contracting skeletal muscle, the rate of glycolysis far exceeds that of the citric acid cycle, and much of the pyruvate formed is reduced to lactate, some of which flows to the liver, where it is converted into glucose (Figure 30.12).

These interchanges, known as the Cori cycle (Section 16.4.2), shift part of the metabolic burden of muscle to the liver. In addition, a large amount of alanine is formed in active muscle by the transamination of pyruvate. Alanine, like lactate, can be converted into glucose by the liver. Why does the muscle release alanine? Muscle can absorb and transaminate branched-chain amino acids; however, it cannot form urea. Consequently, the nitrogen is released into the blood as alanine. The liver absorbs the alanine, removes the nitrogen for disposal as urea, and processes the pyruvate to glucose or fatty acids. The metabolic pattern of resting muscle is quite different. *In resting muscle, fatty acids are the major fuel, meeting 85% of the energy needs.*

Unlike skeletal muscle, heart muscle functions almost exclusively aerobically, as evidenced by the density of mitochondria in heart muscle. Moreover, the heart has virtually no glycogen reserves. Fatty acids are the heart's main source of fuel, although ketone bodies as well as lactate can serve as fuel for heart muscle. In fact, heart muscle consumes acetoacetate in preference to glucose.

3. Adipose tissue. *The triacylglycerols stored in adipose tissue are an enormous reservoir of metabolic fuel* (see [Table 30.1](#)). In a typical 70-kg man, the 15 kg of triacylglycerols have an energy content of 135,000 kcal (565,000 kJ). Adipose tissue is specialized for the esterification of fatty acids and for their release from triacylglycerols. In human beings, the liver is the major site of fatty acid synthesis. Recall that these fatty acids are esterified in the liver to glycerol phosphate to form triacylglycerol and are transported to the adipose tissue in lipoprotein particles, such as very low density lipoproteins ([Section 26.3.1](#)). Triacylglycerols are not taken up by adipocytes; rather, they are first hydrolyzed by an extracellular lipoprotein lipase for uptake. This lipase is stimulated by processes initiated by insulin. After the fatty acids enter the cell, the principal task of adipose tissue is to activate these fatty acids and transfer the resulting CoA derivatives to glycerol in the form of glycerol 3-phosphate. This essential intermediate in lipid biosynthesis comes from the reduction of the glycolytic intermediate dihydroxyacetone phosphate. Thus, *adipose cells need glucose for the synthesis of triacylglycerols* ([Figure 30.13](#)).

Triacylglycerols are hydrolyzed to fatty acids and glycerol by intracellular lipases. The release of the first fatty acid from a triacylglycerol, the rate-limiting step, is catalyzed by a hormone-sensitive lipase that is reversibly phosphorylated. The hormone epinephrine stimulates the formation of cyclic AMP, the intracellular messenger in the amplifying cascade, which activates a protein kinase—a recurring theme in hormone action. Triacylglycerols in adipose cells are continually being hydrolyzed and resynthesized. Glycerol derived from their hydrolysis is exported to the liver. Most of the fatty acids formed on hydrolysis are reesterified if glycerol 3-phosphate is abundant. In contrast, they are released into the plasma if glycerol 3-phosphate is scarce because of a paucity of glucose. Thus, *the glucose level inside adipose cells is a major factor in determining whether fatty acids are released into the blood*.

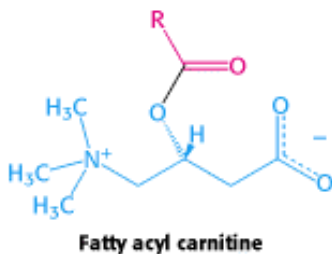
4. The kidney. *The major purpose of the kidney is to produce urine*, which serves as a vehicle for excreting metabolic waste products and for maintaining the osmolarity of the body fluids. The blood plasma is filtered nearly 60 times each day in the renal tubules. Most of the material filtered out of the blood is reabsorbed; so only 1 to 2 liters of urine is produced. Water-soluble materials in the plasma, such as glucose, and water itself are reabsorbed to prevent wasteful loss. The kidneys require large amounts of energy to accomplish the reabsorption. Although constituting only 0.5% of body mass, kidneys consume 10% of the oxygen used in cellular respiration. Much of the glucose that is reabsorbed is carried into the kidney cells by the sodium-glucose cotransporter. Recall that this transporter is powered by the $\text{Na}^+\text{-K}^+$ gradient, which is itself maintained by the $\text{Na}^+\text{-K}^+$ ATPase ([Section 13.4](#)). During starvation, the kidney becomes an important site of gluconeogenesis and may contribute as much as half of the blood glucose.

5. Liver. *The metabolic activities of the liver are essential for providing fuel to the brain, muscle, and other peripheral organs*. Indeed, the liver, which can be from 2% to 4% of body weight, is an organism's metabolic hub ([Figure 30.14](#)). Most compounds absorbed by the intestine first pass through the liver, which is thus able to regulate the level of many metabolites in the blood.

Let us first consider how the liver metabolizes carbohydrates. The liver removes two-thirds of the glucose from the blood and all of the remaining monosaccharides. Some glucose is left in the blood for use by other tissues. The absorbed glucose is converted into glucose 6-phosphate by hexokinase and the liver-specific glucokinase. Glucose 6-phosphate, as already stated, has a variety of fates, although the liver uses little of it to meet its own energy needs. Much of the glucose 6-phosphate is converted into glycogen. As much as 400 kcal (1700 kJ) can be stored in this form in the liver. Excess glucose 6-phosphate is metabolized to acetyl CoA, which is used to form fatty acids, cholesterol, and bile salts. The pentose phosphate pathway, another means of processing glucose 6-phosphate, supplies the NADPH for these reductive biosyntheses. The liver can produce glucose for release into the blood by breaking down its store of glycogen and by carrying out gluconeogenesis. The main precursors for gluconeogenesis are lactate and alanine from muscle, glycerol from adipose tissue, and glucogenic amino acids from the diet.

The liver also plays a central role in the regulation of lipid metabolism. When fuels are abundant, fatty acids derived from the diet or synthesized by the liver are esterified and secreted into the blood in the form of very low density lipoprotein (see [Figure 30.15](#)). However, in the fasting state, the liver converts fatty acids into ketone bodies. How is the fate of liver fatty acids determined? The selection is made according to whether the fatty acids enter the mitochondrial

matrix. Recall that long-chain fatty acids traverse the inner mitochondrial membrane only if they are esterified to carnitine. Carnitine acyltransferase I (also known as carnitine palmitoyl transferase I), which catalyzes the formation of acyl carnitine, is inhibited by malonyl CoA, the committed intermediate in the synthesis of fatty acids (see [Figure 30.9](#)). Thus, *when malonyl CoA is abundant, long-chain fatty acids are prevented from entering the mitochondrial matrix, the compartment of β -oxidation and ketone-body formation. Instead, fatty acids are exported to adipose tissue for incorporation into triacylglycerols.* In contrast, the level of malonyl CoA is low when fuels are scarce. Under these conditions, fatty acids liberated from adipose tissues enter the mitochondrial matrix for conversion into ketone bodies.



The liver also plays an essential role in dietary amino acid metabolism. The liver absorbs the majority of amino acids, leaving some in the blood for peripheral tissues. The priority use of amino acids is for protein synthesis rather than catabolism. By what means are amino acids directed to protein synthesis in preference to use as a fuel? The K_M value for the aminoacyl-tRNA synthetases is lower than that of the enzymes taking part in amino acid catabolism. Thus, amino acids are used to synthesize aminoacyl-tRNAs before they are catabolized. When catabolism does take place, the first step is the removal of nitrogen, which is subsequently processed to urea. The liver secretes from 20 to 30 g of urea a day. The α -ketoacids are then used for gluconeogenesis or fatty acid synthesis. Interestingly, the liver cannot remove nitrogen from the branch-chain amino acids (leucine, isoleucine, and valine). Transamination takes place in the muscle.

How does the liver meet its own energy needs? α -Ketoacids derived from the degradation of amino acids are the liver's own fuel. In fact, the main role of glycolysis in the liver is to form building blocks for biosyntheses. Furthermore, the liver cannot use acetoacetate as a fuel, because it has little of the transferase needed for acetoacetate's activation to acetyl CoA. Thus, the liver eschews the fuels that it exports to muscle and the brain.

Table 30.1. Fuel reserves in a typical 70-kg man

Organ	Available energy in kcal (kJ)		
	Glucose or glycogen	Triacylglycerols	Mobilizable proteins
Blood	60 (250)	45 (200)	0 (0)
Liver	400 (1700)	450 (2000)	400 (1700)
Brain	8 (30)	0 (0)	0 (0)
Muscle	1,200 (5000)	450 (2000)	24,000 (100,000)
Adipose tissue	80 (330)	135,000 (560,000)	40 (170)

Source: After G. F. Cahill, Jr. *Clin. Endocrinol. Metab.* 5(1976):398.

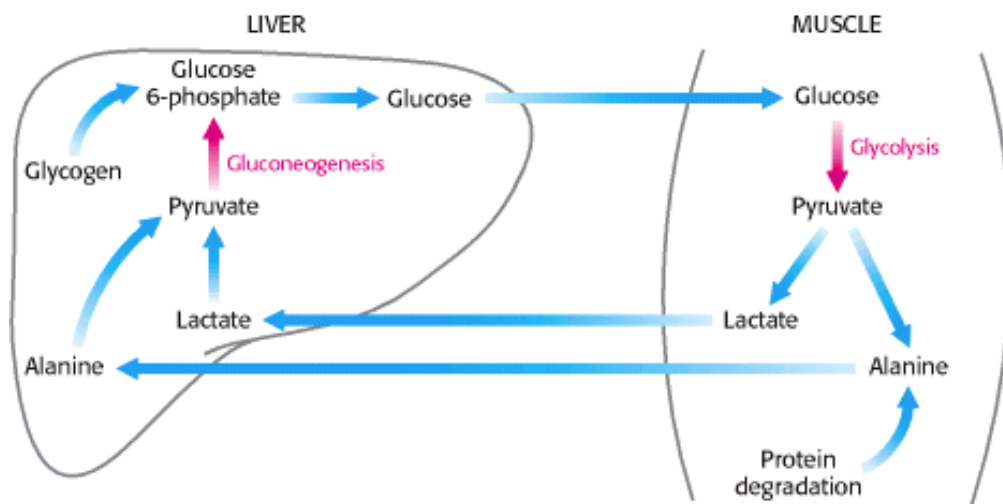


Figure 30.12. Metabolic Interchanges between Muscle and Liver.

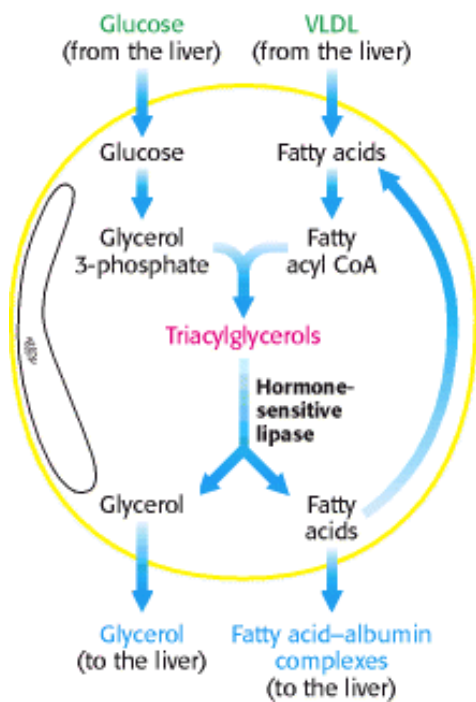


Figure 30.13. Synthesis and Degradation of Triacylglycerols by Adipose Tissue. Fatty acids are delivered to adipose cells in the form of triacylglycerols contained in very low density lipoproteins (VLDLs).

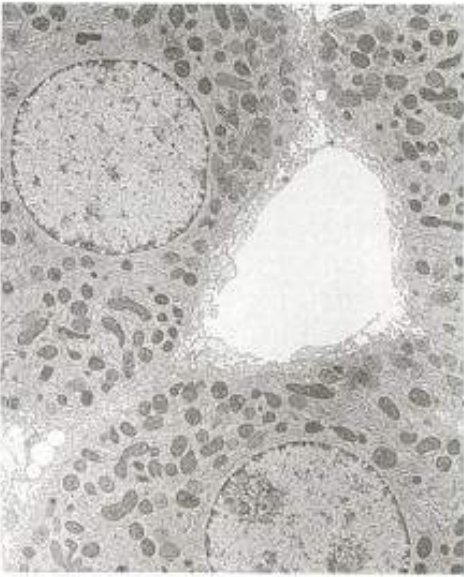


Figure 30.14. Electron Micrograph of Liver Cells. The liver plays an essential role in the integration of metabolism. [Courtesy of Dr. Ann Hubbard.]

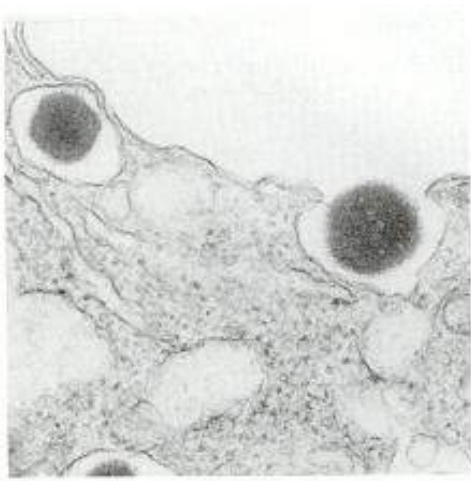


Figure 30.15. Insulin Secretion. The electron micrograph shows the release of insulin from a pancreatic β cell. One secretory granule is on the verge of fusing with the plasma membrane and releasing insulin into the extracellular space, and the other has already released the hormone. [Courtesy of Dr. Lelio Orci. L. Orci, J.-D. Vassalli, and A. Perrelet. *Sci. Am.* 259 (September 1988):85–94.]

30.3. Food Intake and Starvation Induce Metabolic Changes

We shall now consider the biochemical responses to a series of physiological conditions. Our first example is the *starved-fed cycle*, which we all experience in the hours after an evening meal and through the night's fast. This nightly starved-fed cycle has three stages: the postabsorptive state after a meal, the early fasting during the night, and the refeed state after breakfast. A major goal of the many biochemical alterations in this period is to maintain *glucose homeostasis* —that is, a constant blood-glucose level.

1. The well-fed, or postabsorptive, state. After we consume and digest an evening meal, glucose and amino acids are transported from the intestine to the blood. The dietary lipids are packaged into chylomicrons and transported to the blood by the lymphatic system. This fed condition leads to the secretion of insulin, which is one of the two most

important regulators of fuel metabolism, the other regulator being glucagon. The secretion of the hormone *insulin* by the β cells of the pancreas is stimulated by glucose and the parasympathetic nervous system (Figure 30.15). *In essence, insulin signals the fed state—it stimulates the storage of fuels and the synthesis of proteins in a variety of ways.* For instance, insulin initiates protein kinase cascades—it stimulates glycogen synthesis in both muscle and the liver and suppresses gluconeogenesis by the liver. Insulin also accelerates glycolysis in the liver, which in turn increases the synthesis of fatty acids.

The liver helps to limit the amount of glucose in the blood during times of plenty by storing it as glycogen so as to be able to release glucose in times of scarcity. How is the excess blood glucose present after a meal removed? Insulin accelerates the uptake of blood glucose into the liver by GLUT2. The level of glucose 6-phosphate in the liver rises because only then do the catalytic sites of glucokinase become filled with glucose. Recall that glucokinase is active only when blood-glucose levels are high. Consequently, *the liver forms glucose 6-phosphate more rapidly as the blood-glucose level rises. The increase in glucose 6-phosphate coupled with insulin action leads to a buildup of glycogen stores.* The hormonal effects on glycogen synthesis and storage are reinforced by a direct action of glucose itself. *Phosphorylase a is a glucose sensor in addition to being the enzyme that cleaves glycogen.* When the glucose level is high, the binding of glucose to phosphorylase *a* renders the enzyme susceptible to the action of a phosphatase that converts it into phosphorylase *b*, which does not readily degrade glycogen. Thus, *glucose allosterically shifts the glycogen system from a degradative to a synthetic mode.*

The high insulin level in the fed state also promotes *the entry of glucose into muscle and adipose tissue.* Insulin stimulates the synthesis of glycogen by muscle as well as by the liver. The entry of glucose into adipose tissue provides glycerol 3-phosphate for the synthesis of triacylglycerols. The action of insulin also extends to amino acid and protein metabolism. Insulin promotes the uptake of branched-chain amino acids (valine, leucine, and isoleucine) by muscle. Indeed, insulin has a general stimulating effect on protein synthesis, which favors a building up of muscle protein. In addition, it inhibits the intracellular degradation of proteins.

2. The early fasting state. The blood-glucose level begins to drop several hours after a meal, leading to a decrease in insulin secretion and a rise in *glucagon* secretion; glucagon is secreted by the α cells of the pancreas in response to a *low blood-sugar level in the fasting state.* Just as insulin signals the fed state, glucagon signals the starved state. It serves to mobilize glycogen stores when there is no dietary intake of glucose. *The main target organ of glucagon is the liver.* Glucagon stimulates glycogen breakdown and inhibits glycogen synthesis by triggering the cyclic AMP cascade leading to the phosphorylation and activation of phosphorylase and the inhibition of glycogen synthase (Section 21.5). Glucagon also inhibits fatty acid synthesis by diminishing the production of pyruvate and by lowering the activity of acetyl CoA carboxylase by maintaining it in an unphosphorylated state. In addition, glucagon stimulates gluconeogenesis in the liver and blocks glycolysis by lowering the level of F-2,6-BP.

All known actions of glucagon are mediated by protein kinases that are activated by cyclic AMP. The activation of the cyclic AMP cascade results in a higher level of phosphorylase *a* activity and a lower level of glycogen synthase *a* activity. Glucagon's effect on this cascade is reinforced by the diminished binding of glucose to phosphorylase *a*, which makes the enzyme less susceptible to the hydrolytic action of the phosphatase. Instead, the phosphatase remains bound to phosphorylase *a*, and so the synthase stays in the in-active phosphorylated form. Consequently, there is a rapid mobilization of glycogen.

The large amount of glucose formed by the hydrolysis of glucose 6-phosphate derived from glycogen is then released from the liver into the blood. The entry of glucose into muscle and adipose tissue decreases in response to a low insulin level. The diminished utilization of glucose by muscle and adipose tissue also contributes to the maintenance of the bloodglucose level. The net result of these actions of glucagon is to *markedly increase the release of glucose by the liver.*


Both muscle and liver use fatty acids as fuel when the blood-glucose level drops. Thus, *the blood-glucose level is kept at or above 80 mg/dl by three major factors: (1) the mobilization of glycogen and the release of glucose by the liver, (2) the release of fatty acids by adipose tissue, and (3) the shift in the fuel used from glucose to fatty acids by muscle and the*

liver.

What is the result of depletion of the liver's glycogen stores? Gluconeogenesis from lactate and alanine continues, but this process merely replaces glucose that had already been converted into lactate and alanine by the peripheral tissues. Moreover, the brain oxidizes glucose completely to CO_2 and H_2O . Thus, for the net synthesis of glucose to occur, another source of carbons is required. Glycerol released from adipose tissue on lipolysis provides some of the carbons, with the remaining carbons coming from the hydrolysis of muscle proteins.

3. *The refed state.* What are the biochemical responses to a hearty breakfast? Fat is processed exactly as it is processed in the normal fed state. However, this is not the case for glucose. The liver does not initially absorb glucose from the blood, but rather leaves it for the peripheral tissues. Moreover, the liver remains in a gluconeogenic mode. Now, however, the newly synthesized glucose is used to replenish the liver's glycogen stores. As the blood-glucose levels continue to rise, the liver completes the replenishment of its glycogen stores and begins to process the remaining excess glucose for fatty acid synthesis.

30.3.1. Metabolic Adaptations in Prolonged Starvation Minimize Protein Degradation

 What are the adaptations if fasting is prolonged to the point of starvation? A typical well-nourished 70-kg man has fuel reserves totaling about 161,000 kcal (670,000 kJ; see [Table 30.1](#)). The energy need for a 24-hour period ranges from about 1600 kcal (6700 kJ) to 6000 kcal (25,000 kJ), depending on the extent of activity. Thus, stored fuels suffice to meet caloric needs in starvation for 1 to 3 months. However, the carbohydrate reserves are exhausted in only a day.

Even under starvation conditions, the blood-glucose level must be maintained above 2.2 mM (40 mg/dl). *The first priority of metabolism in starvation is to provide sufficient glucose to the brain and other tissues (such as red blood cells) that are absolutely dependent on this fuel.* However, precursors of glucose are not abundant. Most energy is stored in the fatty acyl moieties of triacylglycerols. Recall that fatty acids cannot be converted into glucose, because acetyl CoA cannot be transformed into pyruvate ([Section 22.3.7](#)). The glycerol moiety of triacylglycerol can be converted into glucose, but only a limited amount is available. The only other potential source of glucose is amino acids derived from the breakdown of proteins. However, proteins are not stored, and so any breakdown will necessitate a loss of function. Thus, *the second priority of metabolism in starvation is to preserve protein, which is accomplished by shifting the fuel being used from glucose to fatty acids and ketone bodies* ([Figure 30.16](#)).

The metabolic changes on the first day of starvation are like those after an overnight fast. The low blood-sugar level leads to decreased secretion of insulin and increased secretion of glucagon. *The dominant metabolic processes are the mobilization of triacylglycerols in adipose tissue and gluconeogenesis by the liver. The liver obtains energy for its own needs by oxidizing fatty acids released from adipose tissue.* The concentrations of acetyl CoA and citrate consequently increase, which switches off glycolysis. The uptake of glucose by muscle is markedly diminished because of the low insulin level, whereas fatty acids enter freely. Consequently, *muscle shifts almost entirely from glucose to fatty acids for fuel.* The β -oxidation of fatty acids by muscle halts the conversion of pyruvate into acetyl CoA, because acetyl CoA stimulates the phosphorylation of the pyruvate dehydrogenase complex, which renders it inactive ([Section 17.2.1](#)). Hence, pyruvate, lactate, and alanine are exported to the liver for conversion into glucose. Glycerol derived from the cleavage of triacylglycerols is another raw material for the synthesis of glucose by the liver.

Proteolysis also provides carbon skeletons for gluconeogenesis. During starvation, degraded proteins are not replenished and serve as carbon sources for glucose synthesis. Initial sources of protein are those that turn over rapidly, such as proteins of the intestinal epithelium and the secretions of the pancreas. Proteolysis of muscle protein provides some of three-carbon precursors of glucose. However, survival for most animals depends on being able to move rapidly, which requires a large muscle mass, and so muscle loss must be minimized.


How is the loss of muscle curtailed? After about 3 days of starvation, the liver forms large amounts of acetoacetate and d-

3-hydroxybutyrate (ketone bodies; [Figure 30.17](#)). Their synthesis from acetyl CoA increases markedly because the citric acid cycle is unable to oxidize all the acetyl units generated by the degradation of fatty acids. Gluconeogenesis depletes the supply of oxaloacetate, which is essential for the entry of acetyl CoA into the citric acid cycle. Consequently, the liver produces large quantities of ketone bodies, which are released into the blood. At this time, *the brain begins to consume appreciable amounts of acetoacetate in place of glucose*. After 3 days of starvation, about a third of the energy needs of the brain are met by ketone bodies ([Table 30.2](#)). The heart also uses ketone bodies as fuel.

After several weeks of starvation, ketone bodies become the major fuel of the brain. Acetoacetate is activated by the transfer of CoA from succinyl CoA to give acetoacetyl CoA ([Figure 30.18](#)). Cleavage by thiolase then yields two molecules of acetyl CoA, which enter the citric acid cycle. In essence, *ketone bodies are equivalents of fatty acids that can pass through the blood-brain barrier*. Only 40 g of glucose is then needed per day for the brain, compared with about 120 g in the first day of starvation. *The effective conversion of fatty acids into ketone bodies by the liver and their use by the brain markedly diminishes the need for glucose. Hence, less muscle is degraded than in the first days of starvation.* The breakdown of 20 g of muscle daily compared with 75 g early in starvation is most important for survival. A person's survival time is mainly determined by the size of the triacylglycerol depot.

What happens after depletion of the triacylglycerol stores? The only source of fuel that remains is proteins. Protein degradation accelerates, and death inevitably results from a loss of heart, liver, or kidney function.

30.3.2. Metabolic Derangements in Diabetes Result from Relative Insulin Insufficiency and Glucagon Excess

 We now consider *diabetes mellitus*, a complex disease characterized by grossly abnormal fuel usage: *glucose is overproduced by the liver and underutilized by other organs*. The incidence of diabetes mellitus (usually referred to simply as *diabetes*) is about 5% of the population. Indeed, diabetes is the most common serious metabolic disease in the world; it affects hundreds of millions. *Type I diabetes, or insulin-dependent diabetes mellitus (IDDM)*, is caused by autoimmune destruction of the insulin-secreting β cells in the pancreas and usually begins before age 20. The term insulin-dependent means that the individual requires insulin to live. Most diabetics, in contrast, have a normal or even higher level of insulin in their blood, but they are quite unresponsive to the hormone. This form of the disease—known as *type II, or non-insulin-dependent, diabetes mellitus (NIDDM)*—typically arises later in life than does the insulin-dependent form.

Diabetes-

Named for the excessive urination in the disease. Aretaeus, a Cappadocian physician of the second century a.d., wrote: "The epithet diabetes has been assigned to the disorder, being something like passing of water by a siphon." He perceptively characterized diabetes as "being a melting-down of the flesh and limbs into urine."

From Latin, meaning "sweetened with honey." Refers to the presence of sugar in the urine of patients having the disease.

Mellitus distinguishes this disease from diabetes *insipidus*, which is caused by impaired renal reabsorption of water.


In type I diabetes, insulin is absent and consequently glucagon is present at higher-than-normal levels. In essence, the diabetic person is in biochemical starvation mode despite a high concentration of blood glucose. Because insulin is deficient, *the entry of glucose into cells is impaired*. The liver becomes stuck in a gluconeogenic and ketogenic state. The excessive level of glucagon relative to insulin leads to a decrease in the amount of F-2,6-BP in the liver. Hence,

glycolysis is inhibited and gluconeogenesis is stimulated because of the opposite effects of F-2,6-BP on phosphofruktokinase and fructose-1,6-bisphosphatase (Section 16.4; see also Figures 30.4 and 30.6). The high glucagon/insulin ratio in diabetes also promotes glycogen breakdown. Hence, *an excessive amount of glucose is produced by the liver and released into the blood*. Glucose is excreted in the urine (hence the name *mellitus*) when its concentration in the blood exceeds the reabsorptive capacity of the renal tubules. Water accompanies the excreted glucose, and so an untreated diabetic in the acute phase of the disease is hungry and thirsty.

Because carbohydrate utilization is impaired, a lack of insulin leads to the uncontrolled breakdown of lipids and proteins. Large amounts of acetyl CoA are then produced by β -oxidation. However, much of the acetyl CoA cannot enter the citric acid cycle, because there is insufficient oxaloacetate for the condensation step. Recall that mammals can synthesize oxaloacetate from pyruvate, a product of glycolysis, but not from acetyl CoA; instead, they generate ketone bodies. *A striking feature of diabetes is the shift in fuel usage from carbohydrates to fats; glucose, more abundant than ever, is spurned*. In high concentrations, ketone bodies overwhelm the kidney's capacity to maintain acid-base balance. The untreated diabetic can go into a coma because of a lowered blood pH level and dehydration.

Type II, or non-insulin-dependent, diabetes accounts for more than 90% of the cases and usually develops in middle-aged, obese people. The exact cause of type II diabetes remains to be elucidated, although a genetic basis seems likely.

30.3.3. Caloric Homeostasis: A Means of Regulating Body Weight

 In the United States, obesity has become an epidemic, with nearly 20% of adults classified as obese. Obesity is identified as a risk factor in a host of pathological conditions including diabetes mellitus, hypertension, and cardiovascular disease. The cause of obesity is quite simple in the vast majority of cases—more food is consumed than is needed, and the excess calories are stored as fat.

Although the proximal cause of obesity is simple, the biochemical means by which caloric homeostasis and appetite control are usually maintained is enormously complex, but two important signal molecules are insulin and leptin. A protein consisting of 146 amino acids, *leptin* is a hormone secreted by adipocytes in direct proportion to fat mass. Leptin acts through a membrane receptor (related in structure and mechanism of action to the growth-hormone receptor; Section 15.4) in the hypothalamus to generate satiation signals. During periods when more energy is expended than ingested (the starved state), adipose tissue loses mass. Under these conditions, the secretion of both leptin and insulin declines, fuel utilization is increased, and energy stores are used. The converse is true when calories are consumed in excess.

The importance of leptin to obesity is dramatically illustrated in mice. Mice lacking leptin are obese and will lose weight if given leptin. Mice that lack the leptin receptor are insensitive to leptin administration. Preliminary evidence indicates that leptin and its receptor play a role in human obesity, but the results are not as clear-cut as in the mouse. The interplay of genes and their products to control caloric homeostasis will be an exciting area of research for some time to come.

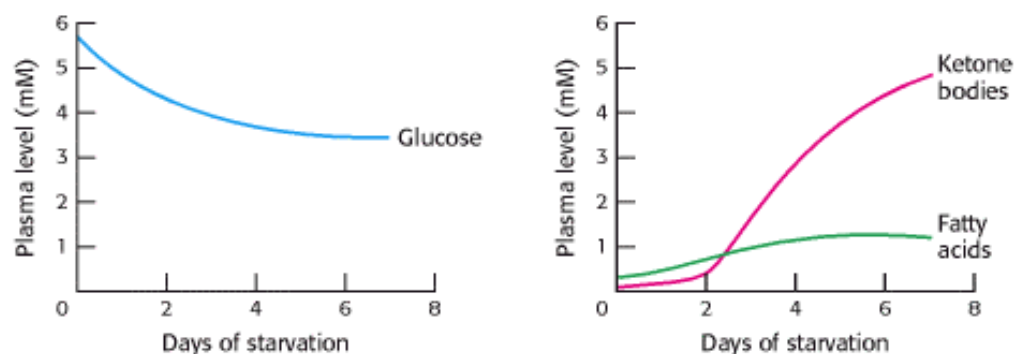


Figure 30.16. Fuel Choice During Starvation. The plasma levels of fatty acids and ketone bodies increase in starvation, whereas that of glucose decreases.

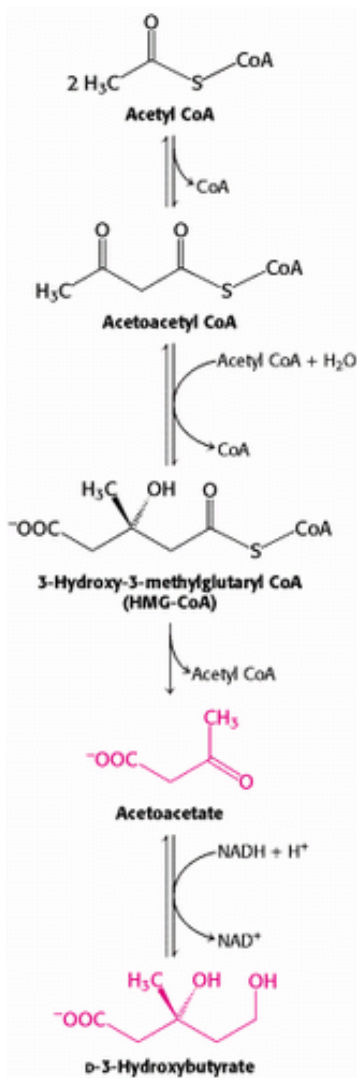


Figure 30.17. Synthesis of Ketone Bodies by the Liver.

Table 30.2. Fuel metabolism in starvation

Fuel exchanges and consumption	Amount formed or consumed in 24 hours (grams)	
	3d day	40th day
Fuel use by the brain		
Glucose	100	40
Ketone bodies	50	100
All other use of glucose	50	40
Fuel mobilization		
Adipose-tissue lipolysis	180	180
Muscle-protein degradation	75	20
Fuel output of the liver		

Glucose	150	80
Ketone bodies	150	150

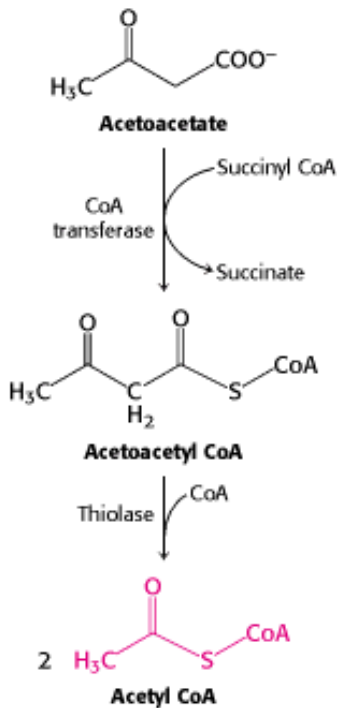


Figure 30.18. Entry of Ketone Bodies Into the Citric Acid Cycle.

30.4. Fuel Choice During Exercise Is Determined by Intensity and Duration of Activity

The fuels used in anaerobic exercises—sprinting, for example—differ from those used in aerobic exercises—such as distance running. The selection of fuels during these different forms of exercise illustrates many important facets of energy transduction and metabolic integration. ATP directly powers myosin, the protein immediately responsible for converting chemical energy into movement ([Chapter 34](#)). However, the amount of ATP in muscle is small. Hence, the power output and, in turn, the velocity of running depend on the rate of ATP production from other fuels. As shown in [Table 30.3](#), *creatine phosphate* (phosphocreatine) can swiftly transfer its high-potential phosphoryl group to ADP to generate ATP ([Section 14.1.5](#)). However, the amount of creatine phosphate, like that of ATP itself, is limited. Creatine phosphate and ATP can power intense muscle contraction for 5 to 6 s. Maximum speed in a sprint can thus be maintained for only 5 to 6 s (see [Figure 14.7](#)). Thus, the winner in a 100-meter sprint is the runner who slows down the least.

A 100-meter sprint is powered by stored ATP, creatine phosphate, and anaerobic glycolysis of muscle glycogen. The conversion of muscle glycogen into lactate can generate a good deal more ATP, but the rate is slower than that of phosphoryl-group transfer from creatine phosphate. During a ~10-second sprint, the ATP level in muscle drops from 5.2 to 3.7 mM, and that of creatine phosphate decreases from 9.1 to 2.6 mM. The essential role of anaerobic glycolysis is manifested in the elevation of the blood-lactate level from 1.6 to 8.3 mM. The release of H⁺ from the intensely active muscle concomitantly lowers the blood pH from 7.42 to 7.24. This pace cannot be sustained in a 1000-meter run (~132 s) for two reasons. First, creatine phosphate is consumed within a few seconds. Second, the lactate produced would cause acidosis. Thus, alternative fuel sources are needed.

The complete oxidation of muscle glycogen to CO₂ substantially increases the energy yield, but this aerobic process is a good deal slower than anaerobic glycolysis. However, as the distance of a run increases, aerobic respiration, or oxidative phosphorylation, becomes increasingly important. For instance, *part of the ATP consumed in a 1000-meter run must come from oxidative phosphorylation*. Because ATP is produced more slowly by oxidative phosphorylation than by glycolysis (see [Table 30.3](#)), the pace is necessarily slower than in a 100-meter sprint. The championship velocity for the 1000-meter run is about 7.6 m/s, compared with approximately 10.2 m/s for the 100-meter event ([Figure 30.19](#)).

The running of a marathon (26 miles 385 yards, or 42,200 meters), requires a different selection of fuels and is characterized by cooperation between muscle, liver, and adipose tissue. Liver glycogen complements muscle glycogen as an energy store that can be tapped. However, the total body glycogen stores (103 mol of ATP at best) are insufficient to provide the 150 mol of ATP needed for this grueling ~2-hour event. Much larger quantities of ATP can be obtained by the oxidation of fatty acids derived from the breakdown of *fat in adipose tissue*, but the maximal rate of ATP generation is slower yet than that of glycogen oxidation and is more than tenfold slower than that with creatine phosphate. Thus, *ATP is generated much more slowly from high-capacity stores than from limited ones*, accounting for the different velocities of anaerobic and aerobic events.

ATP generation from fatty acids is essential for distance running. However, a marathon would take about 6 hours to run if all the ATP came from fatty acid oxidation, because it is much slower than glycogen oxidation. Elite runners consume about equal amounts of glycogen and fatty acids during a marathon to achieve a mean velocity of 5.5 m/s, about half that of a 100-meter sprint. How is an optimal mix of these fuels achieved? *A low blood-sugar level leads to a high glucagon/insulin ratio, which in turn mobilizes fatty acids from adipose tissue*. Fatty acids readily enter muscle, where they are degraded by β oxidation to acetyl CoA and then to CO₂. The elevated acetyl CoA level decreases the activity of the pyruvate dehydrogenase complex to block the conversion of pyruvate into acetyl CoA. Hence, fatty acid oxidation decreases the funneling of sugar into the citric acid cycle and oxidative phosphorylation. Glucose is spared so that just enough remains available at the end of the marathon. The simultaneous use of both fuels gives a higher mean velocity than would be attained if glycogen were totally consumed before the start of fatty acid oxidation.

Table 30.3. Fuel sources for muscle contraction

Fuel source	Maximal rate of ATP production (mmol/s)	Total ~P available (mmol)
Muscle ATP		223
Creatine phosphate	73.3	446
Conversion of muscle glycogen into lactate	39.1	6,700
Conversion of muscle glycogen into CO ₂	16.7	84,000
Conversion of liver glycogen into CO ₂	6.2	19,000
Conversion of adipose-tissue fatty acids into CO ₂	6.7	4,000,000

Note: Fuels stored are estimated for a 70-kg person having a muscle mass of 28 kg.

Source: After E. Hultman and R. C. Harris. In *Principles of Exercise Biochemistry*, J. R. Poortmans (Ed.). (Karger, 1988), pp. 78–119.

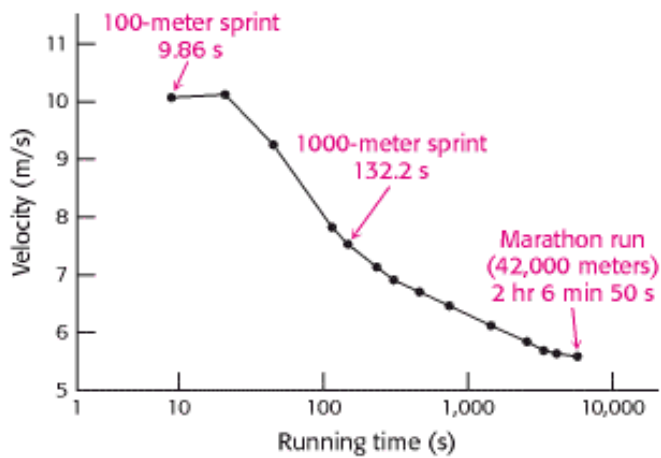

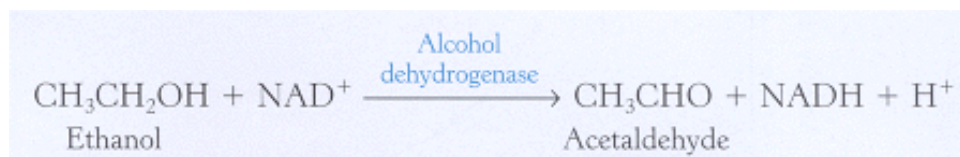


Figure 30.19. Dependence of the Velocity of Running on the Duration of the Race. The values shown are world track records .

30.5. Ethanol Alters Energy Metabolism in the Liver

 Ethanol has been a part of the human diet for centuries. However, its consumption in excess can result in a number of health problems, most notably liver damage. What is the biochemical basis of these health problems?

Ethanol cannot be excreted and must be metabolized, primarily by the liver. This metabolism occurs by two pathways. The first pathway comprises two steps. The first step, catalyzed by the enzyme *alcohol dehydrogenase*, takes place in the cytoplasm:



The second step, catalyzed by *aldehyde dehydrogenase*, takes place in mitochondria:



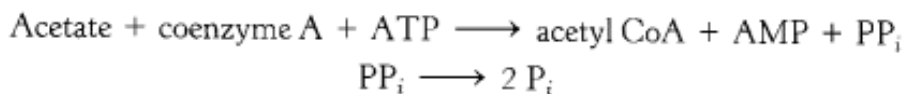
Note that *ethanol consumption leads to an accumulation of NADH*. This high concentration of NADH inhibits gluconeogenesis by preventing the oxidation of lactate to pyruvate. In fact, the high concentration of NADH will cause the reverse reaction to predominate, and lactate will accumulate. The consequences may be hypoglycemia and lactic acidosis.

The NADH glut also inhibits fatty acid oxidation. The metabolic purpose of fatty acid oxidation is to generate NADH for ATP generation by oxidative phosphorylation, but an alcohol consumer's NADH needs are met by ethanol metabolism. In fact, the excess NADH signals that conditions are right for fatty acid synthesis. Hence, triacylglycerols accumulate in the liver, leading to a condition known as "fatty liver."

The second pathway for ethanol metabolism is called the ethanolinducible *microsomal ethanol-oxidizing system* (MEOS). This cytochrome P450-dependent pathway (Section 26.4.2) generates acetaldehyde and subsequently acetate

while oxidizing biosynthetic reducing power, NADPH, to NADP⁺. Because it uses oxygen, this pathway generates free radicals that damage tissues. Moreover, because the system consumes NADPH, the antioxidant glutathione cannot be regenerated (Section 20.5), exacerbating the oxidative stress.

What are the effects of the other metabolites of ethanol? Liver mitochondria can convert acetate into acetyl CoA in a reaction requiring ATP. The enzyme is the thiokinase that normally activates short-chain fatty acids.



However, further processing of the acetyl CoA by the citric acid cycle is blocked, because NADH inhibits two important regulatory enzymes— isocitrate dehydrogenase and α -ketoglutarate dehydrogenase. The accumulation of acetyl CoA has several consequences. First, ketone bodies will form and be released into the blood, exacerbating the acidic condition already resulting from the high lactate concentration. The processing of the acetate in the liver becomes inefficient, leading to a buildup of acetaldehyde. This very reactive compound forms covalent bonds with many important functional groups in proteins, impairing protein function. If ethanol is consistently consumed at high levels, the acetaldehyde can significantly damage the liver, eventually leading to cell death.

Liver damage from excessive ethanol consumption occurs in three stages. The first stage is the aforementioned development of fatty liver. In the second stage—alcoholic hepatitis—groups of cells die and inflammation results. This stage can itself be fatal. In stage three—cirrhosis—fibrous structure and scar tissue are produced around the dead cells. Cirrhosis impairs many of the liver's biochemical functions. The cirrhotic liver is unable to convert ammonia into urea, and blood levels of ammonia rise. Ammonia is toxic to the nervous system and can cause coma and death. Cirrhosis of the liver arises in about 25% of alcoholics, and about 75% of all cases of liver cirrhosis are the result of alcoholism. Viral hepatitis is a nonalcoholic cause of liver cirrhosis.

Summary

Metabolism Consists of Highly Interconnected Pathways

The basic strategy of metabolism is simple: form ATP, reducing power, and building blocks for biosyntheses. This complex network of reactions is controlled by allosteric interactions and reversible covalent modifications of enzymes and changes in their amounts, by compartmentation, and by interactions between metabolically distinct organs. The enzyme catalyzing the committed step in a pathway is usually the most important control site. Opposing pathways such as gluconeogenesis and glycolysis are reciprocally regulated so that one pathway is usually quiescent when the other is highly active.

Each Organ Has a Unique Metabolic Profile

The metabolic patterns of the brain, muscle, adipose tissue, kidney, and liver are very different. Glucose is essentially the sole fuel for the brain in a well-fed person. During starvation, ketone bodies (acetoacetate and 3-hydroxybutyrate) become the predominant fuel of the brain. Adipose tissue is specialized for the synthesis, storage, and mobilization of triacylglycerols. The kidney produces urine and reabsorbs glucose. The diverse metabolic activities of the liver support the other organs. The liver can rapidly mobilize glycogen and carry out gluconeogenesis to meet the glucose needs of other organs. It plays a central role in the regulation of lipid metabolism. When fuels are abundant, fatty acids are synthesized, esterified, and sent from the liver to adipose tissue. In the fasting state, however, fatty acids are converted into ketone bodies by the liver.

Food Intake and Starvation Induce Metabolic Changes

Insulin signals the fed state: it stimulates the formation of glycogen and triacylglycerols and the synthesis of proteins. In contrast, glucagon signals a low blood-glucose level: it stimulates glycogen breakdown and gluconeogenesis by the liver and triacylglycerol hydrolysis by adipose tissue. After a meal, the rise in the blood-glucose level leads to increased secretion of insulin and decreased secretion of glucagon. Consequently, glycogen is synthesized in muscle and the liver. When the blood-glucose level drops several hours later, glucose is then formed by the degradation of glycogen and by the gluconeogenic pathway, and fatty acids are released by the hydrolysis of triacylglycerols. The liver and muscle then use fatty acids instead of glucose to meet their own energy needs so that glucose is conserved for use by the brain.

The metabolic adaptations in starvation serve to minimize protein degradation. Large amounts of ketone bodies are formed by the liver from fatty acids and released into the blood within a few days after the onset of starvation. After several weeks of starvation, ketone bodies become the major fuel of the brain. The diminished need for glucose decreases the rate of muscle breakdown, and so the likelihood of survival is enhanced.

Diabetes mellitus, the most common serious metabolic disease, is due to metabolic derangements resulting in an insufficiency of insulin and an excess of glucagon relative to the needs of the individual. The result is an elevated blood-glucose level, the mobilization of triacylglycerols, and excessive ketone-body formation. Accelerated ketone-body formation can lead to acidosis, coma, and death in untreated insulin-dependent diabetics.

Fuel Choice During Exercise Is Determined by Intensity and Duration of Activity

Sprinting and marathon running are powered by different fuels to maximize power output. The 100-meter sprint is powered by stored ATP, creatine phosphate, and anaerobic glycolysis. In contrast, the oxidation of both muscle glycogen and fatty acids derived from adipose tissue is essential in the running of a marathon, a highly aerobic process.

Ethanol Alters Energy Metabolism in the Liver

The oxidation of ethanol results in an unregulated overproduction of NADH, which has several consequences. A rise in the blood levels of lactic acid and ketone bodies causes a fall in blood pH, or acidosis. The liver is damaged because the excess NADH causes excessive fat formation as well as the generation of acetaldehyde, a reactive molecule. Severe liver damage can result.

Key Terms

allosteric interaction

covalent modification

glycolysis

phosphofructokinase

citric acid cycle

oxidative phosphorylation

pentose phosphate pathway

gluconeogenesis

glycogen synthesis

glycogen degradation

glucose 6-phosphate

pyruvate

acetyl CoA

ketone body

starved-fed cycle

glucose homeostasis

insulin

glucagon

caloric homeostasis

leptin

creatine phosphate

Problems

1. *Distinctive organs.* What are the key enzymatic differences between liver, kidney, muscle, and brain that account for their differing utilization of metabolic fuels?

See answer

2. *Missing enzymes.* Predict the major consequence of each of the following enzymatic deficiencies:

- (a) Hexokinase in adipose tissue
- (b) Glucose 6-phosphatase in liver
- (c) Carnitine acyltransferase I in skeletal muscle
- (d) Glucokinase in liver
- (e) Thiolase in brain
- (f) Kinase in liver that synthesizes fructose 2,6-bisphosphate

See answer

3. *Contrasting milieux.* Cerebrospinal fluid has a low content of albumin and other proteins compared with plasma.

(a) What effect does this lower content have on the concentration of fatty acids in the extracellular medium of the brain?

(b) Propose a plausible reason for the selection by the brain of glucose rather than fatty acids as the prime fuel.

(c) How does the fuel preference of muscle complement that of the brain?

See answer

4. *Metabolic energy and power.* The rate of energy expenditure of a typical 70-kg person at rest is about 70 watts (W), like a light bulb.

(a) Express this rate in kilojoules per second and in kilocalories per second.

(b) How many electrons flow through the mitochondrial electron-transport chain per second under these conditions?

(c) Estimate the corresponding rate of ATP production.

(d) The total ATP content of the body is about 50 g. Estimate how often an ATP molecule turns over in a person at rest.

See answer

5. *Respiratory quotient (RQ).* This classic metabolic index is defined as the volume of CO₂ released divided by the volume of O₂ consumed.

(a) Calculate the RQ values for the complete oxidation of glucose and of tripalmitoylglycerol.

(b) What do RQ measurements reveal about the contributions of different energy sources during intense exercise? (Assume that protein degradation is negligible.)

See answer

6. *Camel's hump.* Compare the H₂O yield from the complete oxidation of 1 g of glucose with that of 1 g of tripalmitoylglycerol. Relate these values to the evolutionary selection of the contents of a camel's hump.

See answer

7. *The wages of sin.* How long does one have to jog to offset the calories obtained from eating 10 macadamia nuts (18 kcal/nut)? [Assume an incremental power consumption of 400 W.]

See answer

8. *Sweet hazard.* Ingesting large amounts of glucose before a marathon might seem to be a good way of increasing the fuel stores. However, experienced runners do not ingest glucose before a race. What is the biochemical reason for their avoidance of this potential fuel? (Hint: Consider the effect of glucose ingestion on the level of insulin.)

See answer

9. *An effect of diabetes.* Insulin-dependent diabetes is often accompanied by hypertriglyceridemia, which is an excess blood level of triacylglycerides in the form of very low density lipoproteins. Suggest a biochemical explanation.

See answer

10. *Sharing the wealth.* The hormone glucagon signifies the starved state, yet it inhibits glycolysis in the liver. How does this inhibition of an energy-production pathway benefit the organism?

See answer

11. *Compartmentation.* Glycolysis takes place in the cytoplasm, whereas fatty acid degradation takes place in mitochondria. What metabolic pathways depend on the interplay of reactions that take place in both compartments?

See answer

12. *Kwashiorkor.* The most common form of malnutrition in children in the world, kwashiorkor is caused by a diet having ample calories but little protein. The high levels of carbohydrate result in high levels of insulin. What is the effect of high levels of insulin on

(a) lipid utilization?

(b) protein metabolism?

(c) Children suffering from kwashiorkor often have large distended bellies caused by water from the blood leaking into extracellular spaces. Suggest a biochemical basis for this condition.

See answer

13. *Oxygen deficit.* After light exercise, the oxygen consumed in recovery is approximately equal to the oxygen deficit, which is the amount of additional oxygen that would have been consumed had oxygen consumption reached steady state immediately. How is the oxygen consumed in recovery used?

See answer

14. *Excess post-exercise oxygen consumption.* The oxygen consumed after strenuous exercise stops is significantly greater than the oxygen deficit and is termed *excess post-exercise oxygen consumption* (EPOC). Why is so much more oxygen required after intense exercise?

See answer

15. *Psychotropic effects.* Ethanol is unusual in that it is freely soluble in both water and lipids. Thus, it has access to all regions of the highly vascularized brain. Although the molecular basis of ethanol action in the brain is not clear, it is evident that ethanol influences a number of neurotransmitter receptors and ion channels. Suggest a biochemical explanation for the diverse effects of ethanol.

See answer

16. *Fiber type*. Skeletal muscle has several distinct fiber types. Type I is used primarily for aerobic activity, whereas type II is specialized for short, intense bursts of activity. How could you distinguish between these types of muscle fiber if you viewed them with an electron microscope?

See answer

17. *Tour de France*. Cyclists in the Tour de France (more than 2000 miles in 3 weeks) require about 200,000 kcal of energy, or 10,000 kcal day⁻¹ (a resting male requires \approx 2000 kcal day⁻¹).

(a) With the assumptions that the energy yield of ATP is about 12 kcal mol⁻¹ and that ATP has a molecular weight of 503 gmol⁻¹, how much ATP would be expended by a Tour de France cyclist?

(b) Pure ATP can be purchased at the cost of approximately \$150 per gram. How much would it cost to power a cyclist through the Tour de France if the ATP had to be purchased?

See answer

Selected Readings

Where to start

G.J. Kemp. 2000. Studying metabolic regulation in human muscle *Biochem. Soc. Trans.* 28: 100-103. ([PubMed](#))

G.E. Lienhard, J.W. Slot, D.E. James, and M.M. Mueckler. 1992. How cells absorb glucose *Sci. Am.* 266: (1) 86-91. ([PubMed](#))

Books

Fell, D., 1997. *Understanding the Control of Metabolism*. Portland Press.

Frayn, K. N., 1996. *Metabolic Regulation: A Human Perspective*. Portland Press.

Hargreaves, M., and Thompson, M. (Eds.) 1999. *Biochemistry of Exercise X. Human Kinetics*.

Harris, R. A., and Crabb, D. W. 1997. Metabolic interrelationships. In *Textbook of Biochemistry with Clinical Correlations* (pp. 525 – 562), edited by T. M. Devlin. Wiley-Liss.

Fuel metabolism

F. Rolland, J. Winderickx, and J.M. Thevelein. 2001. Glucosensing mechanism in eukaryotic cells *Trends Biochem. Sci.* 26: 310-317. ([PubMed](#))

B.B. Rasmussen and R.R. Wolfe. 1999. Regulation of fatty acid oxidation in skeletal muscle *Annu. Rev. Nutr.* 19: 463-484. ([PubMed](#))

P.W. Hochachka. 2000. Oxygen, homeostasis, and metabolic regulation *Adv. Exp. Med. Biol.* 475: 311-335. ([PubMed](#))

E. Holm, O. Sedlaczek, and E. Grips. 1999. Amino acid metabolism in liver disease *Curr. Opin. Clin. Nutr. Metab. Care* 2: 47-53. ([PubMed](#))

A.J. Wagenmakers. 1998. Protein and amino acid metabolism in human muscle *Adv. Exp. Med. Biol.* 441: 307-319. ([PubMed](#))

Metabolic adaptations in starvation

G. Baverel, B. Ferrier, and M. Martin. 1995. Fuel selection by the kidney: Adaptation to starvation *Proc. Nutr. Soc.* 54: 197-212. ([PubMed](#))

I.A. MacDonald and J. Webber. 1995. Feeding, fasting and starvation: Factors affecting fuel utilization *Proc. Nutr. Soc.* 54: 267-274. ([PubMed](#))

G.F. Cahill Jr. 1976. Starvation in man *Clin. Endocrinol. Metab.* 5: 397-415. ([PubMed](#))

M.C. Sugden, M.J. Holness, and T.N. Palmer. 1989. Fuel selection and carbon flux during the starved-to-fed transition *Biochem. J.* 263: 313-323. ([PubMed](#))

Diabetes mellitus

G.A. Rutter. 2000. Diabetes: The importance of the liver *Curr. Biol.* 10: R736-R738. ([PubMed](#))

A.R. Saltiel. 2001. New perspectives into the molecular pathogenesis and treatment of type 2 diabetes *Cell* 104: 517-529. ([PubMed](#))

G.I. Bell, S.J. Pilikis, I.T. Weber, and K.S. Polonsky. 1996. Glucokinase mutations, insulin secretion, and diabetes mellitus *Annu. Rev. Physiol.* 58: 171-186. ([PubMed](#))

D.J. Withers and M. White. 2000. Perspective: The insulin signaling system — a common link in the pathogenesis of type 2 diabetes *Endocrinology* 141: 1917-1921. ([PubMed](#))

Taylor, S. I., 1995. Diabetes mellitus. In *The Metabolic Basis of Inherited Diseases* (7th ed.; pp. 935 – 936), edited by C. R. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, J. B. Stanbury, J. B. Wyngaarden, and D. S. Fredrickson. McGraw-Hill.

Exercise metabolism

R.G. Shulman and D.L. Rothman. 2001. The "glycogen shunt" in exercising muscle: A role for glycogen in muscle energetics and fatigue *Proc. Natl. Acad. Sci. USA* 98: 457-461. ([PubMed](#)) ([Full Text in PMC](#))

T. Gleason. 1996. Post-exercise lactate metabolism: A comparative review of sites, pathways, and regulation *Annu. Rev. Physiol.* 58: 556-581.

J.O. Holloszy and W.M. Kohrt. 1996. Regulation of carbohydrate and fat metabolism during and after exercise *Annu. Rev. Nutr.* 16: 121-138. ([PubMed](#))

P.W. Hochachka and G.B. McClelland. 1997. Cellular metabolic homeostasis during large-scale change in ATP turnover rates in muscles *J. Exp. Biol.* 200: 381-386. ([PubMed](#))

J.F. Horowitz and S. Klein. 2000. Lipid metabolism during endurance exercise *Am. J. Clin. Nutr.* 72: 558S-563S. ([PubMed](#))

A.J. Wagenmakers. 1999. Muscle amino acid metabolism at rest and during exercise *Diabetes Nutr. Metab.* 12: 316-322. ([PubMed](#))

Ethanol metabolism

S. Stewart, D. Jones, and C.P. Day. 2001. Alcoholic liver disease: New insights into mechanisms and preventive strategies *Trends Mol. Med.* 7: 408-413. ([PubMed](#))

C.S. Lieber. 2000. Alcohol: Its metabolism and interaction with nutrients *Annu. Rev. Nutr.* 20: 395-430. ([PubMed](#))

O. Niemela. 1999. Aldehyde-protein adducts in the liver as a result of ethanol-induced oxidative stress *Front. Biosci.* 1:

H. Riveros-Rosas, A. Julian-Sanchez, and E. Pina. 1997. Enzymology of ethanol and acetaldehyde metabolism in mammals *Arch. Med. Res.* 28: 453-471. ([PubMed](#))

I. Diamond and A.S. Gordon. 1997. Cellular and molecular neuroscience of alcoholism *Physiol. Rev.* 77: 1-20. ([PubMed](#))

31. The Control of Gene Expression

Bacteria are highly versatile and responsive organisms: the rate of synthesis of some proteins in bacteria may vary more than a 1000-fold in response to the supply of nutrients or to environmental challenges. Cells of multicellular organisms also respond to varying conditions. Such cells exposed to hormones and growth factors will change substantially in shape, growth rate, and other characteristics. Moreover, many different *cell types* are present in multicellular organisms. For example, cells from muscle and nerve tissue show strikingly different morphologies and other properties, yet they contain exactly the same DNA. These diverse properties are the result of differences in gene expression.

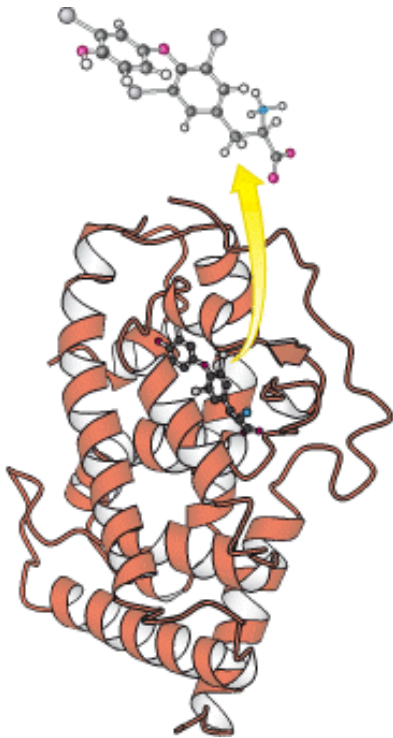
Gene expression is the combined process of the transcription of a gene into mRNA, the processing of that mRNA, and its translation into protein (for protein-encoding genes). A comparison of the gene-expression patterns of cells from the pancreas, which secretes digestive enzymes, and the liver, the site of lipid transport and energy transduction, reveals marked differences in the genes that are highly expressed ([Table 31.1](#)), a difference consistent with the physiological roles of these tissues.

How is gene expression controlled? Gene activity is controlled first and foremost at the level of transcription. Much of this control is achieved through the interplay between proteins that bind to specific DNA sequences and their DNA-binding sites. In this chapter, we shall see how signals from the environment of a cell can alter this interplay to induce changes in gene expression. We first consider gene-regulation mechanisms in prokaryotes and particularly in *E. coli*, because these processes have been extensively investigated in this organism. We then turn to eukaryotic gene regulation. In the chapter's final section, we explore mechanisms for regulating gene expression past the level of transcription.

Table 31.1. Highly expressed protein-encoding genes of the pancreas and liver (as percentage of total mRNA pool)

Rank	Pancreas	% Liver	%
1	Procarboxypeptidase A1	7.6 Albumin	3.5
2	Pancreatic trypsinogen 2	5.5 Apolipoprotein A-I	2.8
3	Chymotrypsinogen	4.4 Apolipoprotein C-I	2.5
4	Pancreatic trypsin 1	3.7 Apolipoprotein C-III	2.1
5	Elastase IIIB	2.4 ATPase 6/8	1.5
6	Protease E	1.9 Cytochrome oxidase 3	1.1
7	Pancreatic lipase	1.9 Cytochrome oxidase 2	1.1
8	Procarboxypeptidase B	1.7 α -1-Antitrypsin	1.0
9	Pancreatic amylase	1.7 Cytochrome oxidase 1	0.9
10	Bile salt-stimulated lipase	1.4 Apolipoprotein E	0.9

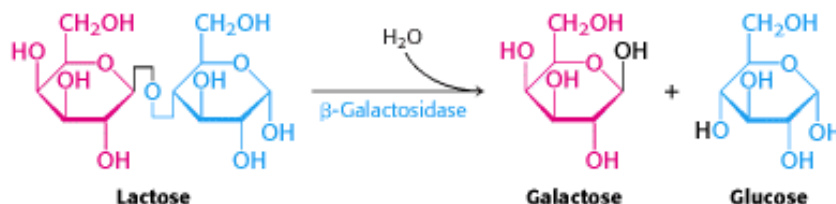
Sources: Data for pancreas from V. E. Velculescu, L. Zhang, B. Vogelstein, and K. W. Kinzler, Serial analysis of gene expression, *Science* 270(1995):484–487. Data for liver from T. Yamashita, S. Hashimoto, S. Kaneko, S. Nagai, N. Toyoda, T. Suzuki, K. Kobayashi, and K. Matsushima, Comprehensive gene expression profile of a normal human liver, *Biochem. Biophys. Res. Commun.* 269(2000):110–116.



Programming Gene Expression. Complex biological processes often involve coordinated control of the expression of many genes. The maturation of a tadpole into a frog is largely controlled by thyroid hormone. This hormone regulates gene expression by binding to a protein, the thyroid hormone receptor, shown at the right. In response to hormone binding, this protein binds to specific DNA sites in the genome and modulates the expression of nearby genes. [(Left) Shanon Cummings/Dembinsky Photo Associates.]

31.1. Prokaryotic DNA-Binding Proteins Bind Specifically to Regulatory Sites in Operons

Bacteria such as *E. coli* usually rely on glucose as their source of carbon and energy. However, when glucose is scarce, *E. coli* can use lactose as their carbon source even though this disaccharide does not lie on any major metabolic pathways. An essential enzyme in the metabolism of lactose is β -galactosidase, which hydrolyzes lactose into galactose and glucose. These products are then metabolized by pathways discussed in [Chapter 16](#).



This reaction can be conveniently followed in the laboratory through the use of alternative galactoside substrates that form colored products such as X-Gal ([Figure 31.1](#)). An *E. coli* cell growing on a carbon source such as glucose or glycerol contains fewer than 10 molecules of β -galactosidase. In contrast, the same cell will contain several thousand molecules of the enzyme when grown on lactose ([Figure 31.2](#)). The presence of lactose in the culture medium induces a large increase in the amount of β -galactosidase by eliciting the synthesis of new enzyme molecules rather than by activating a preexisting but inactive precursor.

A crucial clue to the mechanism of gene regulation was the observation that two other proteins are synthesized in concert with β -galactosidase—namely, *galactoside permease* and *thiogalactoside transacetylase*. The permease is required for the transport of lactose across the bacterial cell membrane. The transacetylase is not essential for lactose metabolism but appears to play a role in the detoxification of compounds that also may be transported by the permease. Thus, *the expression levels of a set of enzymes that all contribute to the adaptation to a given change in the environment change together*. Such a coordinated unit of gene expression is called an *operon*.

31.1.1. An Operon Consists of Regulatory Elements and Protein-Encoding Genes

The parallel regulation of β -galactosidase, the permease, and the transacetylase suggested that the expression of genes encoding these enzymes is controlled by a common mechanism. Francois Jacob and Jacques Monod proposed the *operon model* to account for this parallel regulation as well as the results of other genetic experiments ([Figure 31.3](#)). The genetic elements of the model are a *regulator gene*, a regulatory DNA sequence called an *operator site*, and a *set of structural genes*.

The regulator gene encodes a *repressor* protein that binds to the operator site. The binding of the repressor to the operator prevents transcription of the structural genes. The operator and its associated structural genes constitute the operon. For the *lactose (lac) operon*, the *i* gene encodes the repressor, *o* is the operator site, and the *z*, *y*, and *a* genes are the structural genes for β -galactosidase, the permease, and the transacetylase, respectively. The operon also contains a promoter site (denoted by *p*), which directs the RNA polymerase to the correct transcription initiation site. The *z*, *y*, and *a* genes are transcribed to give a single mRNA molecule that encodes all three proteins. An mRNA molecule encoding more than one protein is known as a *polygenic* or *polycistronic* transcript.

31.1.2. The *lac* Operator Has a Symmetric Base Sequence

The operator site of the *lac* operon has been extensively studied ([Figure 31.4](#)). The nucleotide sequence of the operator site shows a nearly perfect inverted repeat, indicating that the DNA in this region has an approximate twofold axis of

symmetry. Recall that cleavage sites for restriction enzymes such as *EcoRV* have similar symmetry properties (Section 9.3.3). Symmetry in the operator site usually corresponds to symmetry in the repressor protein that binds the operator site. *Symmetry matching is a recurring theme in protein-DNA interactions.*

31.1.3. The *lac* Repressor Protein in the Absence of Lactose Binds to the Operator and Blocks Transcription

How does the *lac* repressor inhibit the expression of the *lac* operon? The *lac* repressor can exist as a dimer of 37-kd subunits, and two dimers often come together to form a tetramer. In the absence of lactose, the repressor binds very tightly and rapidly to the operator. When the *lac* repressor is bound to DNA, it prevents bound RNA polymerase from locally unwinding the DNA to expose the bases that will act as the template for the synthesis of the RNA strand (Section 28.1.3).

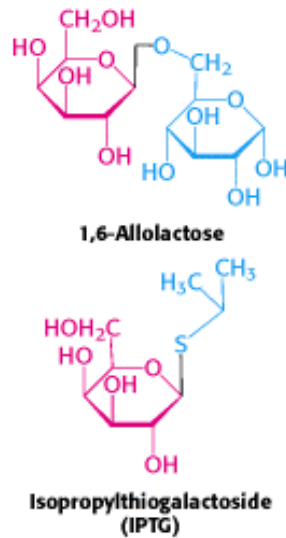
How does the *lac* repressor locate the operator site in the *E. coli* chromosome? The *lac* repressor binds 4×10^6 times as strongly to operator DNA as it does to random sites in the genome. This high degree of selectivity allows the repressor to find the operator efficiently even with a large excess (4.6×10^6) of other sites within the *E. coli* genome. The dissociation constant for the repressor-operator complex is approximately 0.1 pM (10^{-13} M). The rate constant for association ($\approx 10^{10}$ $\text{M}^{-1} \text{s}^{-1}$) is strikingly high, indicating that the repressor finds the operator by diffusing along a DNA molecule (a one-dimensional search) rather than encountering it from the aqueous medium (a three-dimensional search).

Inspection of the complete *E. coli* genome sequence reveals two sites within 500 bp of the primary operator site that approximate the sequence of the operator. Other *lac* repressor dimers can bind to these sites, particularly when aided by cooperative interactions with the *lac* repressor dimer at the primary operator site. No other sites that closely match sequence of the *lac* operator site are present in the rest of the *E. coli* genome sequence. Thus, *the DNA-binding specificity of the lac repressor is sufficient to specify a nearly unique site within the E. coli genome.*

The three-dimensional structure of the *lac* repressor has been determined in various forms. Each monomer consists of a small amino-terminal domain that binds DNA and a larger domain that mediates the formation of the dimer and the tetramer (Figure 31.5). A pair of the amino-terminal domains come together to form the functional DNA-binding unit. Complexes between the *lac* repressor and oligonucleotides that contain the *lac* operator sequence have been structurally characterized. The *lac* repressor binds DNA by inserting an α helix into the major groove of DNA and making a series of contacts with the edges of the base pairs as well as with the phosphodiester backbone (Figure 31.6). For example, an arginine residue in the α helix forms a pair of hydrogen bonds with a guanine residue within the operator. Other bases are not directly contacted but may still be important for binding by virtue of their effects on local DNA structure. As expected, the twofold axis of the operator coincides with a twofold axis that relates the two DNA-binding domains.


31.1.4. Ligand Binding Can Induce Structural Changes in Regulatory Proteins

How does the presence of lactose trigger expression from the *lac* operon? Interestingly, lactose itself does not have this effect; rather, *allolactose*, a combination of galactose and glucose with an α -1,6 rather than an α -1,4 linkage, does. Allolactose is thus referred to as the *inducer* of the *lac* operon. Allolactose is a side product of the β -galactosidase reaction produced at low levels by the few molecules of β -galactosidase that are present before induction. Some other β -galactosides such as *isopropylthiogalactoside (IPTG)* are potent inducers of β -galactosidase expression, although they are not substrates of the enzyme. IPTG is useful in the laboratory as a tool for inducing gene expression.




How does the presence of the inducer modulate gene expression? *When the lac repressor is bound to the inducer, the repressor's affinity for operator DNA is greatly reduced.* The inducer binds in the center of the large domain within each monomer. This binding leads to local conformational changes that are transmitted to the interface with the DNA-binding domains (Figure 31.7). The relation between the two small DNA-binding domains is modified so that they cannot easily contact DNA simultaneously, leading to a dramatic reduction in DNA-binding affinity.

Let us recapitulate the processes that regulate gene expression in the lactose operon (Figure 31.8). In the absence of inducer, the *lac* repressor is bound to DNA in a manner that blocks RNA polymerase from transcribing the *z*, *y*, and *a* genes. Thus, very little β -galactosidase, permease, or transacetylase are produced. The addition of lactose to the environment leads to the formation of allolactose. This inducer binds to the *lac* repressor, leading to conformational changes and the release of DNA by the *lac* repressor. With the operator site unoccupied, RNA polymerase can then transcribe the other *lac* genes and the bacterium will produce the proteins necessary for the efficient utilization of lactose.

 The structure of the large domain of the *lac* repressor is similar to those of a large class of proteins that are present in *E. coli* and other bacteria. This family of homologous proteins binds ligands such as sugars and amino acids at their centers. Remarkably, domains of this family are utilized by eukaryotes in taste proteins and in neurotransmitter receptors, as will be discussed in Chapter 32.

31.1.5. The Operon Is a Common Regulatory Unit in Prokaryotes

Many other gene-regulatory networks function in ways analogous to those of the *lac* operon. For example, genes taking part in purine and, to a lesser degree, pyrimidine biosynthesis are repressed by the *pur* repressor. This dimeric protein is 31% identical in sequence with the *lac* repressor and has a similar three-dimensional structure. However, the behavior of the *pur* repressor is opposite that of the *lac* repressor: whereas the *lac* repressor is released from DNA by binding to a small molecule, the *pur* repressor binds DNA specifically only when bound to a small molecule. Such a small molecule is called a corepressor. For the *pur* repressor, the corepressor can be either guanine or hypoxanthine. The dimeric *pur* repressor binds to inverted repeat DNA sites of the form 5'-ANGCAANCGNTTNCNT-3', in which the bases shown in boldface type are particularly important. Examination of the *E. coli* genome sequence reveals the presence of more than 20 such sites, regulating 19 operons including more than 25 genes (Figure 31.9).

 Because the DNA binding sites for these regulatory proteins are relatively short, it is likely that they evolved independently from one another and are not related by divergence from an ancestral regulatory site. Once a ligand-regulated DNA-binding protein is present in a cell, binding sites may evolve adjacent to additional genes, allowing them to become regulated in a physiologically appropriate manner. Binding sites for the *pur* repressor have evolved in the regulatory regions of a wide range of genes taking part in nucleotide biosynthesis. All such genes can then be regulated in a concerted manner.

31.1.6. Transcription Can Be Stimulated by Proteins That Contact RNA Polymerase

All the DNA-binding proteins discussed thus far function by inhibiting transcription until some environmental condition, such as the presence of lactose, is met. There are also DNA-binding proteins that stimulate transcription. One particularly well studied example is the *catabolite activator protein (CAP)*, which is also known as the cAMP response protein (CRP). When bound to cAMP, CAP, which also is a sequence-specific DNA-binding protein, stimulates the transcription of lactose- and arabinose-catabolizing genes. Within the *lac* operon, CAP binds to an inverted repeat that is centered near position -61 relative to the start site for transcription ([Figure 31.10](#)). CAP functions as a dimer of identical subunits.

The CAP-cAMP complex stimulates the initiation of transcription by approximately a factor of 50. A major factor in this stimulation is the recruitment of RNA polymerase to promoters to which CAP is bound. Studies have been undertaken to localize the surfaces on CAP and on the α subunit of RNA polymerase that participate in these interactions ([Figure 31.11](#)). These energetically favorable protein-protein contacts increase the likelihood that transcription will be initiated at sites to which the CAP-cAMP complex is bound. Thus, in regard to the *lac* operon, gene expression is maximal when the binding of allolactose relieves the inhibition by the *lac* repressor, and the CAP-cAMP complex stimulates the binding of RNA polymerase.

The *E. coli* genome contains many CAP-binding sites in positions appropriate for interactions with RNA polymerase. Thus, an increase in the cAMP level inside an *E. coli* bacterium results in the formation of CAP-cAMP complexes that bind to many promoters and stimulate the transcription of genes encoding a variety of catabolic enzymes. When grown on glucose, *E. coli* have a very low level of catabolic enzymes such as β -galactosidase. Clearly, it would be wasteful to synthesize these enzymes when glucose is abundant. The inhibitory effect of glucose, called *catabolite repression*, is due to the ability of glucose to lower the intracellular concentration of cyclic AMP.

31.1.7. The Helix-Turn-Helix Motif Is Common to Many Prokaryotic DNA-Binding Proteins

The structures of many prokaryotic DNA-binding proteins have now been determined, and amino acid sequences are known for many more. Strikingly, the DNA-binding surfaces of many (but not all) of these proteins consist of a pair of α helices separated by a tight turn ([Figure 31.12](#)). This *helix-turn-helix motif* is present in the *lac* repressor family, CAP, and many other gene-regulatory proteins. In complexes with DNA, the second of these two helices (often called the *recognition helix*) lies in the major groove, where amino acid side chains make contact with the edges of base pairs, whereas residues of the first helix participate primarily in contacts with the DNA backbone. Helix-turn-helix motifs are very frequently present on proteins that bind DNA as dimers, and thus two of the units will be present, one on each monomer. In this case, the two helix-turn-helix units are related by twofold symmetry along the DNA double helix.

Although the helix-turn-helix motif is the most commonly observed DNA-binding unit in prokaryotes, not all regulatory proteins bind DNA through such units. A striking example is provided by the *E. coli* methionine repressor ([Figure 31.13](#)). This protein binds DNA through the insertion of a pair of β strands into the major groove. We shall shortly encounter a variety of other DNA-binding motifs found in eukaryotic cells.

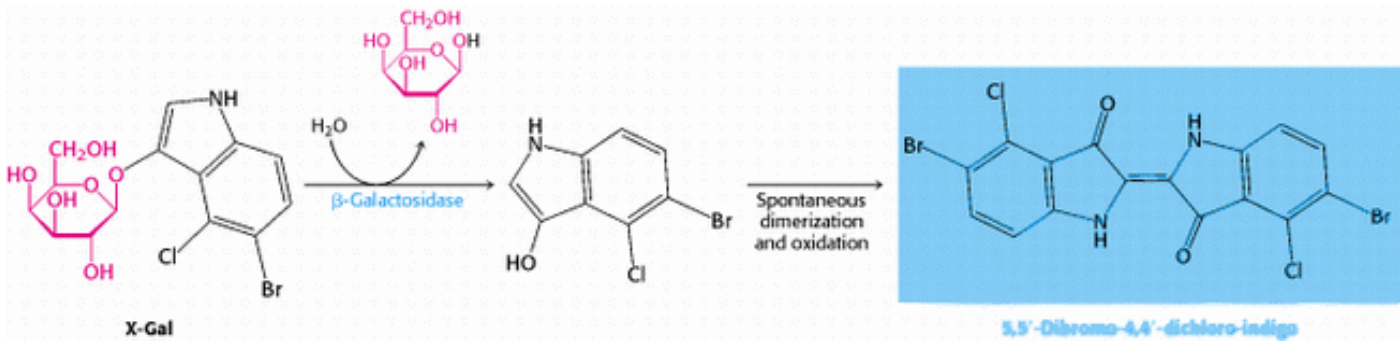


Figure 31.1. Following the β-Galactosidase Reaction. The galactoside substrate X-Gal produces a colored product on cleavage by β-galactosidase. The appearance of this colored product provides a convenient means for monitoring the amount of the enzyme both in vitro and in vivo.

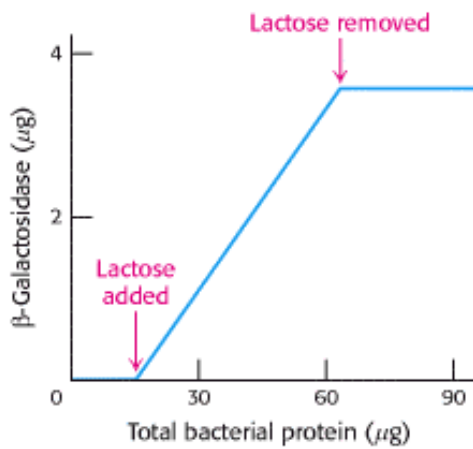


Figure 31.2. β-Galactosidase Induction. The addition of lactose to an *E. coli* culture causes the production of β-galactosidase to increase from very low amounts to much larger amounts. The increase in the amount of enzyme parallels the increase in the number of cells in the growing culture. β-Galactosidase constitutes 6.6% of the total protein synthesized in the presence of lactose.



Figure 31.3. Operons. (A) The general structure of an operon as conceived by Jacob and Monod. (B) The structure of the lactose operon. In addition to the promoter (p) in the operon, a second promoter is present in front of the regulator gene (i) to drive the synthesis of the regulator.

5'-...TGTGTGGAATTGTGAGCGGATAACAATTCACACA...3'
 3'-...ACACACCTTAACACTCGCCTAATGTTAAAGTGTGT...5'

Figure 31.4. The *LAC* Operator. The nucleotide sequence of the *lac* operator shows a nearly perfect inverted repeat, corresponding to twofold rotational symmetry in the DNA. Parts of the sequences that are related by this symmetry are shown in the same color.

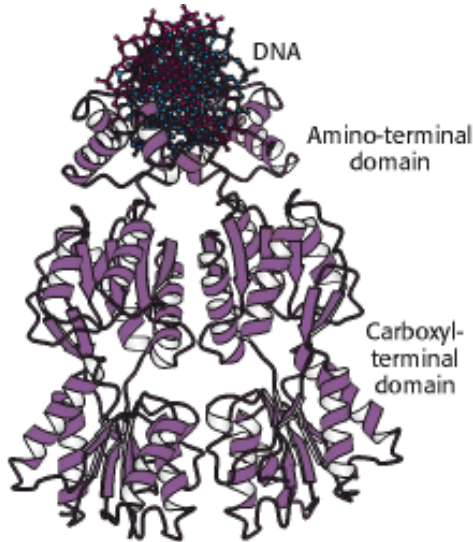


Figure 31.5. Structure of the *LAC* Repressor. A *lac* repressor dimer is shown bound to DNA. A part of the structure that mediates the formation of *lac* repressor tetramers is not shown.

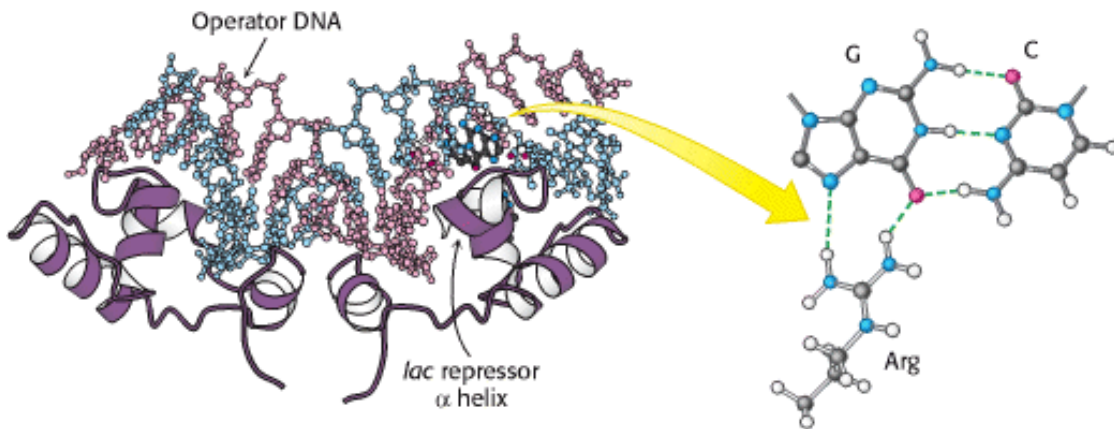


Figure 31.6. *LAC* Repressor-DNA Interactions. The *lac* repressor DNA-binding domain inserts an α helix into the major groove of operator DNA. A specific contact between an arginine residue of the repressor and a G-C base pair is shown at the right.

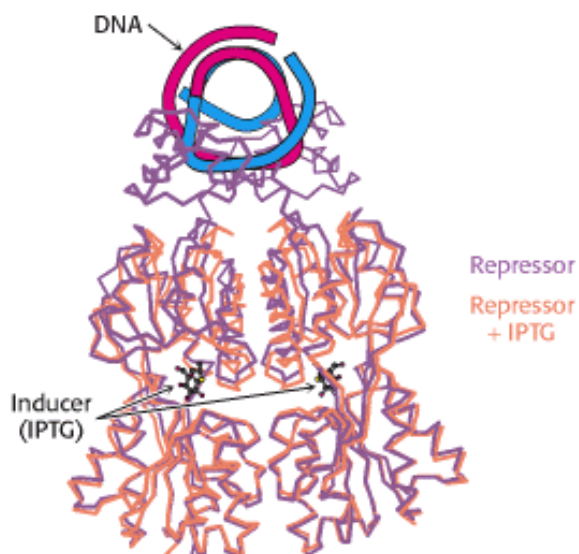


Figure 31.7. Effects of IPTG On LAC Repressor Structure. The structure of the *lac* repressor bound to the inducer isopropylthiogalactoside (IPTG), shown in orange, is superimposed on the structure of the *lac* repressor bound to DNA, shown in purple. The binding of IPTG induces structural changes that alter the relation between the two DNA-binding domains so that they cannot interact effectively with DNA. The DNA-binding domains of the *lac* repressor bound to IPTG are not shown, because these regions are not well ordered in the crystals studied.

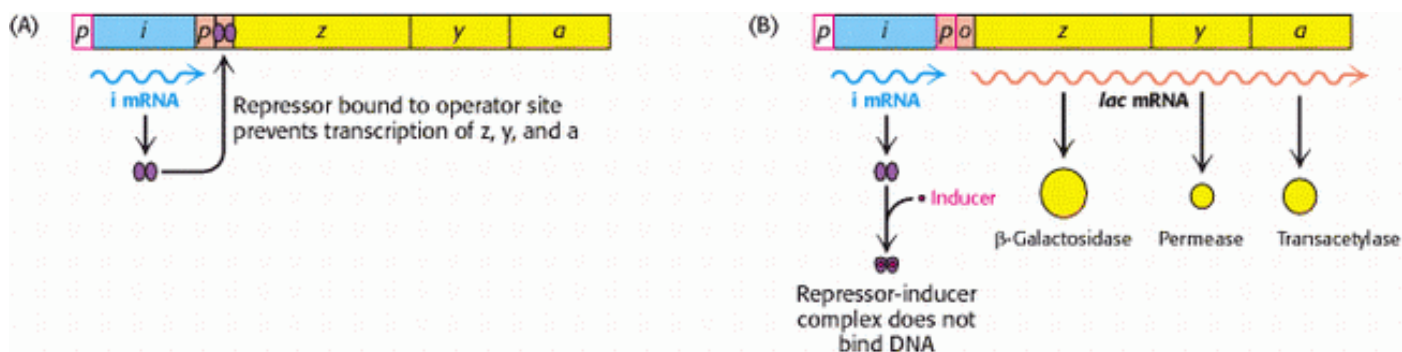


Figure 31.8. Induction of the LAC Operon. (A) In the absence of lactose, the *lac* repressor binds DNA and represses transcription from the *lac* operon. (B) Allolactose or another inducer binds to the *lac* repressor, leading to its dissociation from DNA and to the production of *lac* mRNA.

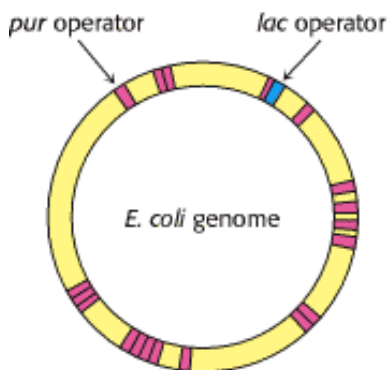


Figure 31.9. Binding-Site Distributions. The *E. coli* genome contains only a single region that closely matches the sequence of the *lac* operator (shown in blue). In contrast, 20 sites match the sequence of the *pur* operator (shown in red). Thus, the *pur* repressor regulates the expression of many more genes than does the *lac* repressor.

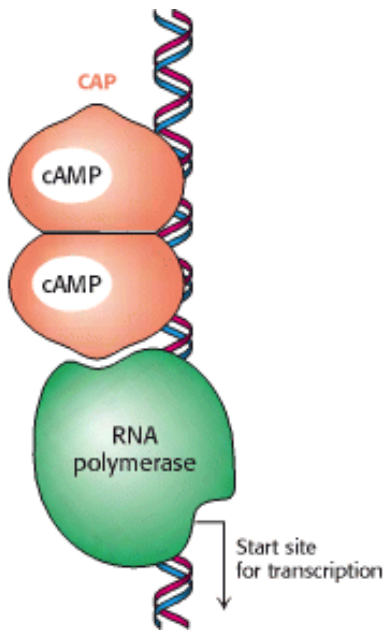


Figure 31.10. Binding Site for Catabolite Activator Protein (CAP). This protein binds as a dimer to an inverted repeat that is at the position -61 relative to the start site of transcription. The CAP binding site on DNA is adjacent to the position at which RNA polymerase binds.

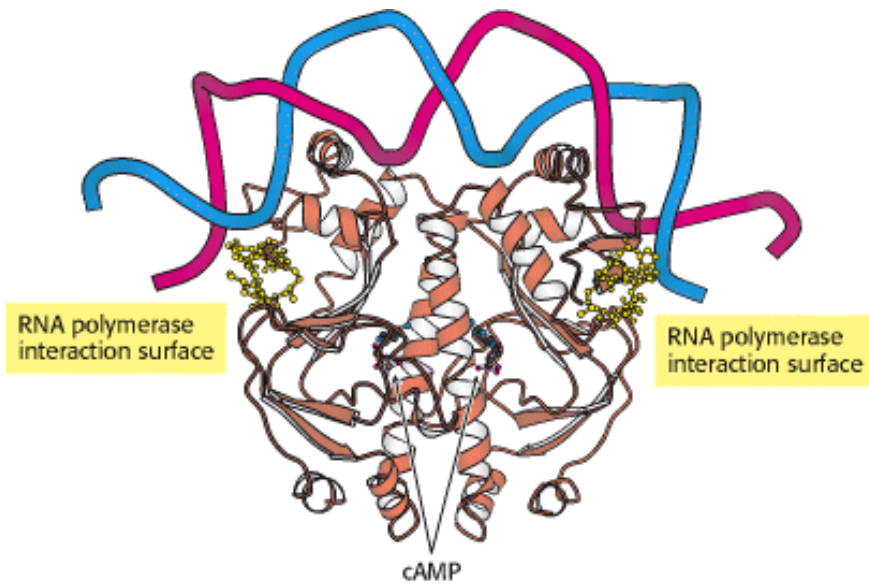


Figure 31.11. Structure of a Dimer of CAP Bound to DNA. The residues in each CAP monomer that have been implicated in direct interactions with RNA polymerase are shown in yellow.

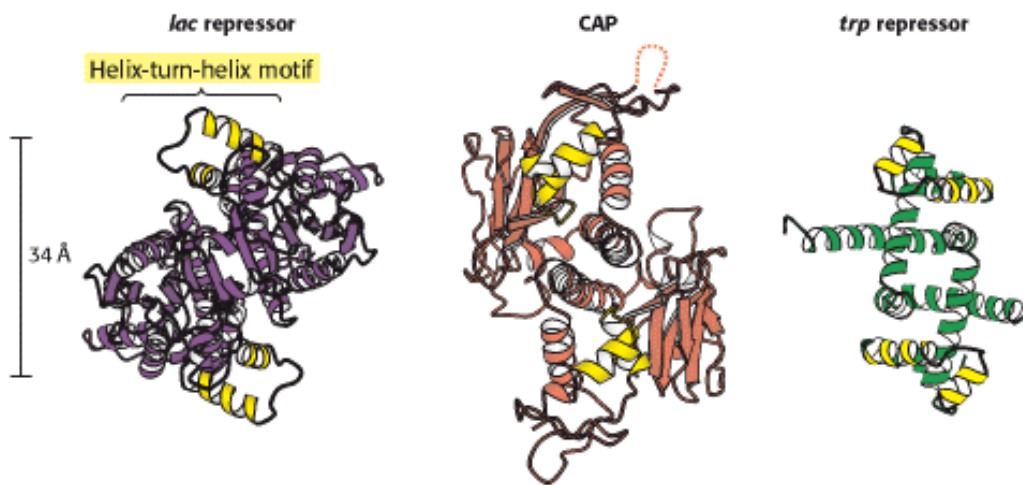


Figure 31.12. Helix-Turn-Helix Motif. These structures show three sequence-specific DNA-binding proteins that interact with DNA through a helix-turn-helix motif (highlighted in yellow). In each case, the helix-turn-helix units within a protein dimer are approximately 34 Å apart, corresponding to one full turn of DNA.

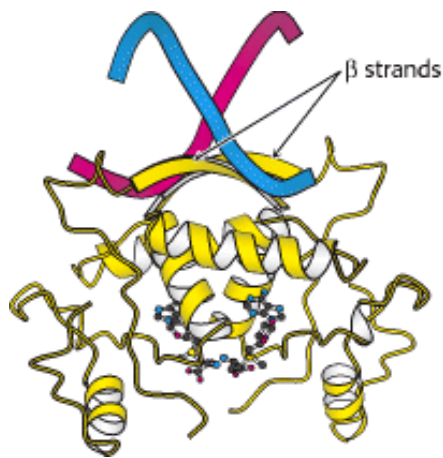


Figure 31.13. DNA Recognition Through β Strands. The structure of the methionine repressor bound to DNA reveals that residues in β strands, rather than α helices, participate in the crucial interactions between the protein and DNA.

31.2. The Greater Complexity of Eukaryotic Genomes Requires Elaborate Mechanisms for Gene Regulation

Gene regulation is significantly more complex in eukaryotes than in prokaryotes for a number of reasons. First, the genome being regulated is significantly larger. The *E. coli* genome consists of a single, circular chromosome containing 4.6 Mb. This genome encodes approximately 2000 proteins. In comparison, one of the simplest eukaryotes, *Saccharomyces cerevisiae* (baker's yeast), contains 16 chromosomes ranging in size from 0.2 to 2.2 Mb (Figure 31.14). The yeast genome totals 17 Mb and encodes approximately 6000 proteins. The genome within a human cell contains 23 pairs of chromosomes ranging in size from 50 to 250 Mb. Approximately 40,000 genes are present within the 3000 Mb of human DNA. It would be very difficult for a DNA-binding protein to recognize a unique site in this vast array of DNA sequences. Consequently, more-elaborate mechanisms are required to achieve specificity.

Megabase (Mb)

A length of DNA consisting of 10^6 base pairs (if double stranded) or 10^6 bases (if single stranded).

$$1 \text{ Mb} = 10^3 \text{ kb} = 10^6 \text{ bases}$$

Another source of complexity in eukaryotic gene regulation is the many different cell types present in most eukaryotes. Liver and pancreatic cells, for example, differ dramatically in the genes that are highly expressed (see [Table 31.1](#)). Moreover, eukaryotic genes are not generally organized into operons. Instead, genes that encode proteins for steps within a given pathway are often spread widely across the genome. Finally, transcription and translation are uncoupled in eukaryotes, eliminating some potential gene-regulatory mechanisms.

31.2.1. Nucleosomes Are Complexes of DNA and Histones

The DNA in eukaryotic chromosomes is not bare. Rather eukaryotic DNA is tightly bound to a group of small basic proteins called *histones*. In fact, histones constitute half the mass of a eukaryotic chromosome. The entire complex of a cell's DNA and associated protein is called *chromatin*. Five major histones are present in chromatin: four histones, called H2A, H2B, H3, and H4, associate with one another; the other histone is called H1. Histones have strikingly basic properties because a quarter of the residues in each histone is either arginine or lysine.

In 1974, Roger Kornberg proposed that *chromatin is made up of repeating units, each containing 200 bp of DNA and two copies each of H2A, H2B, H3, and H4*, called the *histone octamer*. These repeating units are known as *nucleosomes*. Strong support for this model comes from the results of a variety of experiments, including observations of appropriately prepared samples of chromatin viewed by electron microscopy ([Figure 31.15](#)). Chromatin viewed with the electron microscope has the appearance of beads on a string; each bead has a diameter of approximately 100 Å. Partial digestion of chromatin with DNase yields the isolated beads. These particles consist of fragments of DNA ≈ 200 bp in length bound to the eight histones. More extensive digestion yields a reduced DNA fragment of 145 bp bound to the histone octamer. This smaller complex of the histone octamer and the 145-bp DNA fragment is the *nucleosome core particle*. The DNA connecting core particles in undigested chromatin is called *linker DNA*. Histone H1 binds, in part, to the linker DNA.

31.2.2. Eukaryotic DNA Is Wrapped Around Histones to Form Nucleosomes

The overall structure of the nucleosome was revealed through electron microscopic and x-ray crystallographic studies pioneered by Aaron Klug and his colleagues. More recently, the three-dimensional structure of a reconstituted nucleosome core ([Figure 31.16](#)) was determined to relatively high resolution by x-ray diffraction methods. As was shown by Evangelos Moudrianakis, the four types of histone that make up the protein core are homologous and similar in structure ([Figure 31.17](#)). The eight histones in the core are arranged into a $(\text{H3})_2(\text{H4})_2$ tetramer and a pair of H2A-H2B dimers. The tetramer and dimers come together to form a left-handed superhelical ramp around which the DNA wraps. In addition, each histone has an amino-terminal tail that extends out from the core structure. These tails are flexible and contain a number of lysine and arginine residues. As we shall see, *covalent modifications of these tails play an essential role in modulating the affinity of the histones for DNA and other properties*.

The DNA forms a left-handed superhelix as it wraps around the outside of the histone octamer. The protein core forms contacts with the inner surface of the superhelix at many points, particularly along the phosphodiester backbone and the minor groove of the DNA. Nucleosomes will form on almost all DNA sites, although some sequences are preferred because the dinucleotide steps are properly spaced to favor bending around the histone core. Histone H1, which has a different structure from the other histones, seals off the nucleosome at the location at which the linker DNA enters and leaves the nucleosome. The amino acid sequences of histones, including their amino-terminal tails, are remarkably

conserved from yeast through human beings.

The winding of DNA around the nucleosome core contributes to DNA's packing by decreasing its linear extent. An extended 200-bp stretch of DNA would have a length of about 680 Å. Wrapping this DNA around the histone octamer reduces the length to approximately 100 Å along the long dimension of the nucleosome. Thus the DNA is compacted by a factor of seven. However, human chromosomes in metaphase, which are highly condensed, are compacted by a factor of 10^4 . Clearly, the nucleosome is just the first step in DNA compaction. What is the next step? The nucleosomes themselves are arranged in a helical array approximately 360 Å across, forming a series of stacked layers approximately 110 Å apart (Figure 31.18). The folding of these fibers of nucleosomes into loops further compacts DNA.

The writhing of DNA around the histone core in a left-handed helical manner also stores negative supercoils; if the DNA in a nucleosome is straightened out, it will be underwound (Section 23.3.2). This underwinding is exactly what is needed to separate the two DNA strands during replication and transcription (Sections 27.5 and 28.1.5).

31.2.3. The Control of Gene Expression Requires Chromatin Remodeling

Does chromatin structure play a role in the control of gene expression? Early observations suggested that it does indeed. The treatment of cell nuclei with the nonspecific DNA-cleaving enzyme DNase I revealed that regions adjacent to genes that are being actively transcribed are more sensitive to cleavage than are other sites in the genome, suggesting that the DNA in these regions is less compacted than it is elsewhere in the genome and more accessible to proteins. In addition, some sites, usually within 1 kb of the start site of an active gene, are exquisitely sensitive to DNase I and other nucleases. These *hypersensitive sites* correspond to regions that have few nucleosomes or have nucleosomes in an altered conformational state. *Hyper-sensitive sites are cell-type specific and developmentally regulated.* For example, globin genes in the precursors of erythroid cells from 20-hour-old chicken embryos are insensitive to DNase I. However, when hemoglobin synthesis begins at 35 hours, regions adjacent to these genes become highly susceptible to digestion. In tissues such as the brain that produce no hemoglobin, the globin genes remain resistant to DNase I throughout development and into adulthood. The results of these studies suggest that a prerequisite for gene expression is a relaxing of the chromatin structure.

Recent experiments even more clearly revealed the role of chromatin structure in regulating access to DNA binding sites. Genes required for galactose utilization in yeast are activated by a DNA-binding protein called GAL4, which recognizes DNA binding sites with two 5'-CGG-3' sequences separated by 11 base pairs (Figure 31.19). Approximately 4000 potential GAL4 binding sites of the form 5'-CGG(N)₁₁CCG-3' are present in the yeast genome, but only 10 of them regulate genes necessary for galactose metabolism. What fraction of the potential binding sites are actually bound by GAL4? This question is addressed through the use of a technique called *chromatin immunoprecipitation (ChIP)*. GAL4 is first cross-linked to the DNA to which it is bound in chromatin. The DNA is then fragmented into small pieces, and antibodies to GAL4 are used to isolate the chromatin fragments containing GAL4. The cross-linking is reversed, and the DNA is isolated and characterized. The results of these studies reveal that only approximately 10 of the 4000 potential GAL4 sites are occupied by GAL4 when the cells are growing on galactose; more than 99% of the sites appear to be blocked. Thus, whereas in prokaryotes all sites appear to be equally accessible, chromatin structure shields a large number of the potential binding sites in eukaryotic cells. GAL4 is thereby prevented from binding to sites that are unimportant in galactose metabolism.

These lines of evidence and others reveal that chromatin structure is altered in active genes compared with inactive ones. How is chromatin structure modified? As we shall see in Section 31.3.4, specific covalent modifications of histone proteins are crucial. In addition, the binding of specific proteins to DNA sequences called *enhancers* at specific sites in the genome plays a role.

31.2.4. Enhancers Can Stimulate Transcription by Perturbing Chromatin Structure

We can now understand the action of *enhancers*, already introduced in Section 28.2.6. Recall that these DNA sequences, although they have no promoter activity of their own, greatly increase the activities of many promoters in eukaryotes,

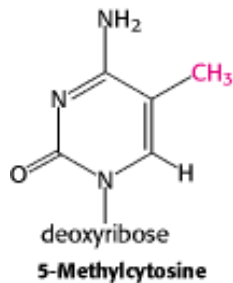
even when the enhancers are located at a distance of several thousand base pairs from the gene being expressed.


Enhancers function by serving as binding sites for specific regulatory proteins. (Figure 31.20). An enhancer is effective only in the specific cell types in which appropriate regulatory proteins are expressed. In many cases, these DNA-binding proteins influence transcription initiation by perturbing the local chromatin structure to expose a gene or its regulatory sites rather than by direct interactions with RNA polymerase. This mechanism accounts for the ability of enhancers to act at a distance.

The properties of enhancers are illustrated by studies of the enhancer controlling the muscle isoform of creatine kinase (Section 14.1.5). The results of mutagenesis and other studies revealed the presence of an enhancer located between 1350 and 1050 base pairs upstream of the start site of the gene for this enzyme. Experimentally inserting this enhancer near a gene not normally expressed in muscle cells is sufficient to cause the gene to be expressed at high levels in muscle cells, but not other cells (Figure 31.21).

31.2.5. The Modification of DNA Can Alter Patterns of Gene Expression

The modification of DNA provides another mechanism, in addition to packaging with histones, for inhibiting inappropriate gene expression in specific cell types. Approximately 70% of the 5'-CpG-3' sequences in mammalian genomes are methylated at the C-5 position of cytosine by specific methyltransferases. However, the distribution of these methylated cytosines varies, depending on the cell type. Consider, again, the globin genes. In cells that are actively expressing hemoglobin, the region from approximately 1 kb upstream of the start site of the β -globin gene to approximately 100 bp downstream of the start site contains fewer 5-methylcytosine residues than does the corresponding region in cells that do not express these genes. The relative absence of 5-methylcytosines near the start site is referred to as *hypomethylation*. The methyl group of 5-methylcytosine protrudes into the major groove where it could easily interfere with the binding of proteins that stimulate transcription.



 The distribution of CpG sequences in mammalian genomes is not uniform. The deamination of 5-methylcytosine produces thymine; so CpG sequences are subject to mutation to TpG. Many CpG sequences have been converted into TpG through this mechanism. However, sites near the 5' ends of genes have been maintained because of their role in gene expression. Thus, most genes are found in *CpG islands*, regions of the genome that contain approximately four times as many CpG sequences as does the remainder of the genome. Note that methylation is not a universal regulatory device, even in multicellular eukaryotes. For example, *Drosophila* DNA is not methylated at all.

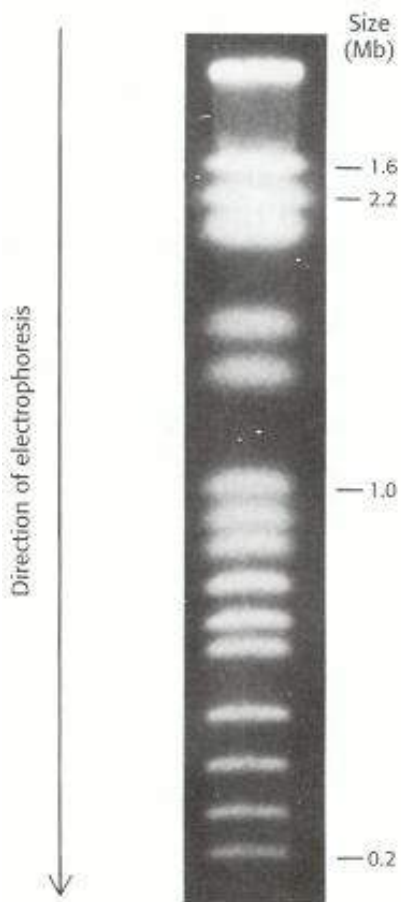


Figure 31.14. Yeast Chromosomes. Pulsed-field electrophoresis allows the separation of 16 yeast chromosomes. [From G. Chu, D. Wollrath, and R. W. Davis. *Science* 234(1986):1583.]

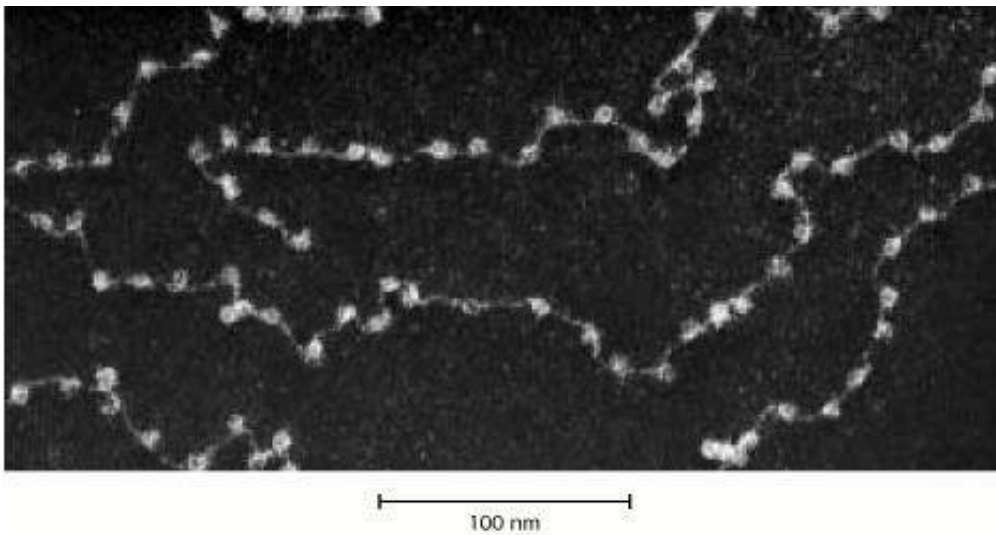


Figure 31.15. Chromatin Structure. An electron micrograph of chromatin showing its "beads on a string" character. [Courtesy of Dr. Ada Olins and Dr. Donald Olins.]

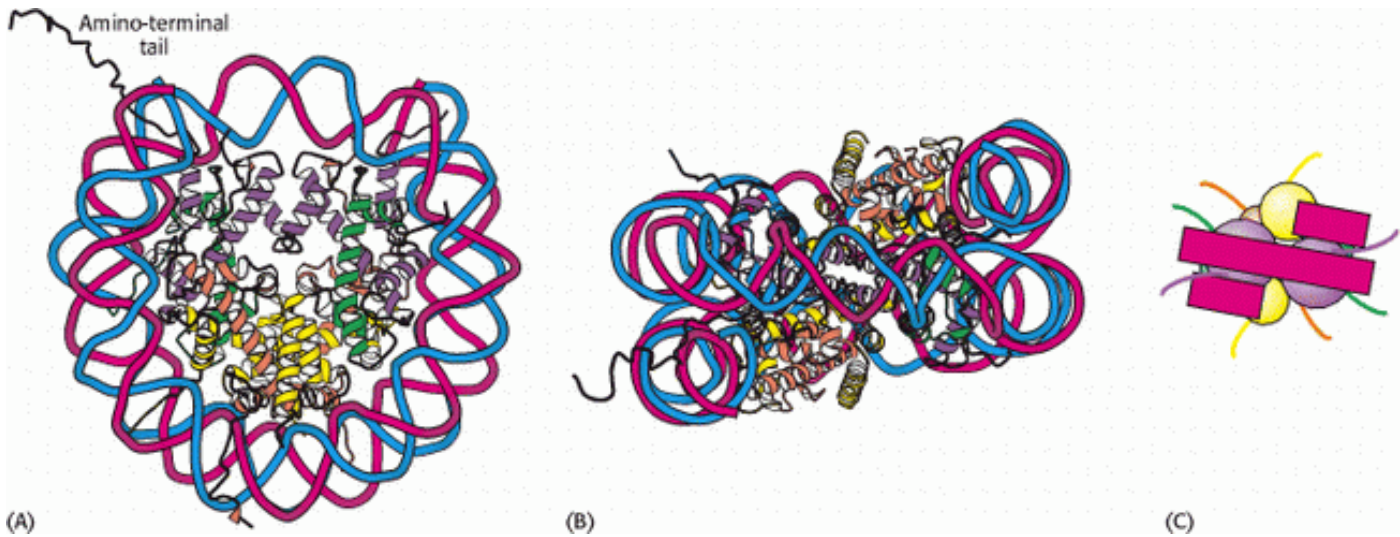


Figure 31.16. Nucleosome Core Particle. The structure consists of a core of eight histone proteins surrounded by DNA.

(A) A view showing the DNA wrapping around the histone core. (B) A view related to that in part A by a 90-degree rotation shows that the DNA forms a left-handed superhelix as it wraps around the core. (C) A schematic view.

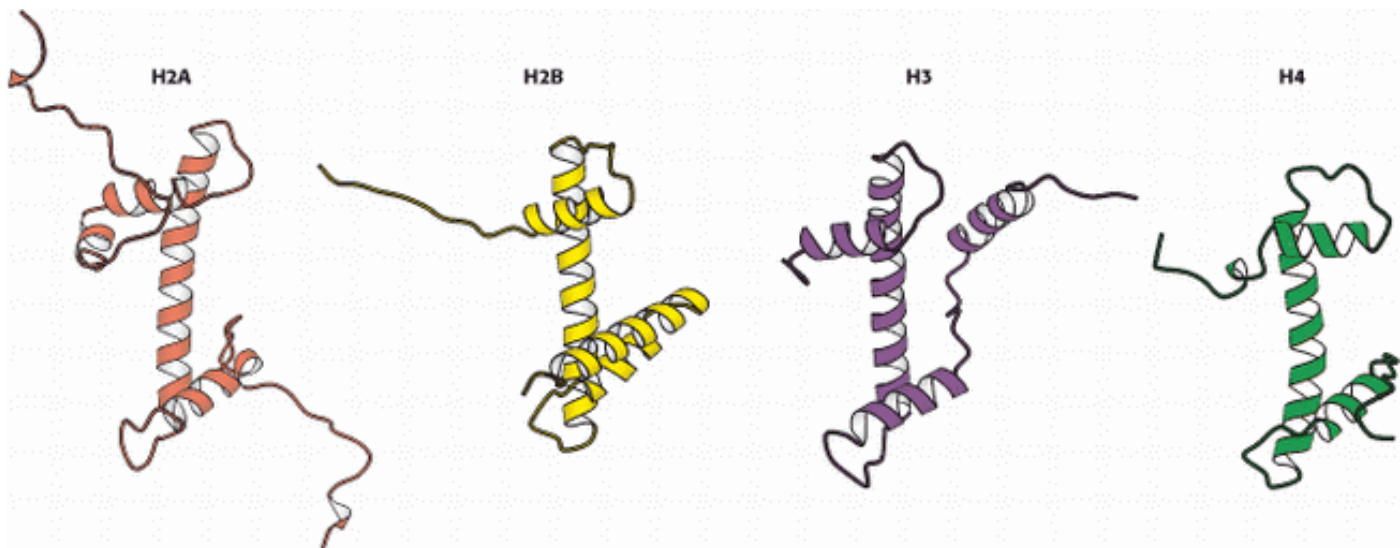


Figure 31.17. Homologous Histones. Histones H2A, H2B, H3, and H4 each adopt a similar three-dimensional structure

as a consequence of common ancestry. Some parts of the tails present at the termini of the proteins are not shown.

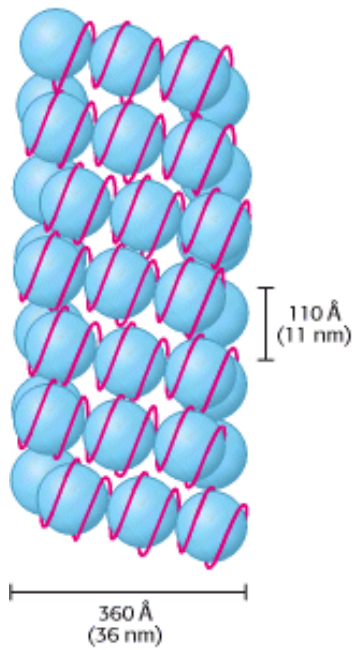


Figure 31.18. Higher-Order Chromatin Structure. A proposed model for chromatin arranged in a helical array consisting of six nucleosomes per turn of helix. The DNA double helix (shown in red) is wound around each histone octamer (shown in blue). [After J. T. Finch and A. Klug, *Proc. Natl. Acad. Sci. USA* 73(1976):1900.]

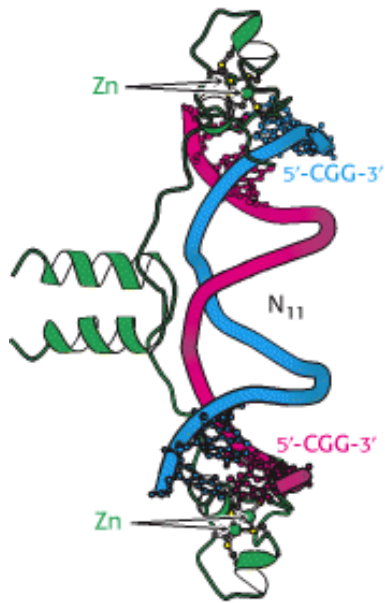


Figure 31.19. Gal4 Binding Sites. The yeast transcription factor GAL4 binds to DNA sequences of the form 5'-CGG(N)11CCG-3'. Two zinc-based domains are present in the DNA-binding region of this protein. These domains contact the 5'-CGG-3' sequences, leaving the center of the site uncontacted.

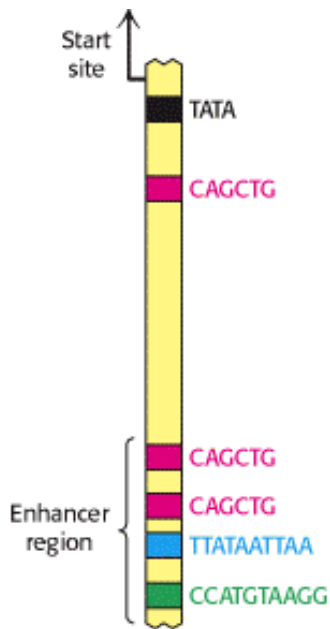


Figure 31.20. Enhancer Binding Sites. A schematic structure for the region 1 kb upstream of the start site for the muscle creatine kinase gene. One binding site of the form 5'-CAGCTG-3' is present near the TATA box. The enhancer region farther upstream contains two binding sites for the same protein and two additional binding sites for other proteins.

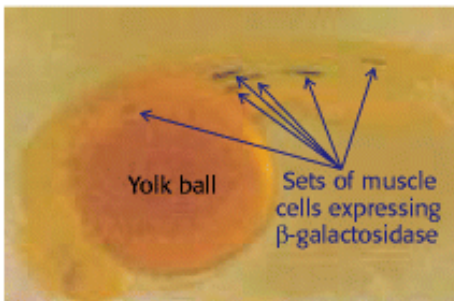


Figure 31.21. An Experimental Demonstration of Enhancer Function. A promoter for muscle creatine kinase artificially drives the transcription of β -galactosidase in a zebrafish embryo. Only specific sets of muscle cells produce β -galactosidase, as visualized by the formation of the blue product on treatment of the embryo with X-Gal. [From F. Müller, D. W. Williamson, J. Kobilák, L. Gauvry, G. Goldspink, L. Orbán, and N. MacLean. *Molecular Reproduction and Development* 47(1997): 404.]

31.3. Transcriptional Activation and Repression Are Mediated by Protein-Protein Interactions

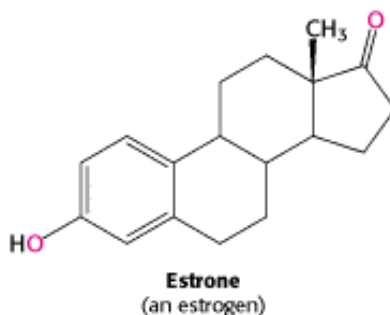
We have seen how interactions between DNA-binding proteins such as CAP and RNA polymerase can activate transcription in prokaryotic cells (Section 31.1.6). Such protein-protein interactions play a dominant role in eukaryotic gene regulation. In contrast with those of prokaryotic transcription, few eukaryotic transcription factors have any effect on transcription on their own. Instead, *each factor recruits other proteins to build up large complexes that interact with the transcriptional machinery to activate or repress transcription.*

A major advantage of this mode of regulation is that a given regulatory protein can have different effects, depending on what other proteins are present in the same cell. This phenomenon, called *combinatorial control*, is crucial to

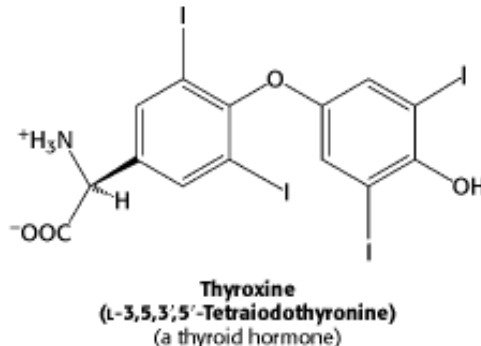
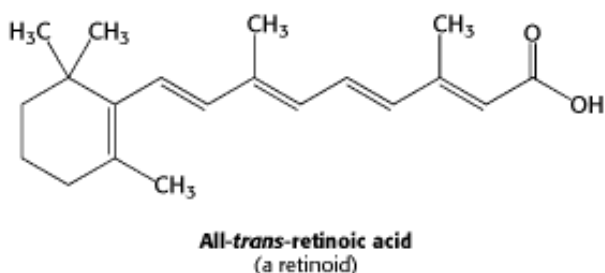
multicellular organisms that have many different cell types. Even in unicellular eukaryotes such as yeast, combinatorial control allows the generation of distinct cell types.

31.3.1. Steroids and Related Hydrophobic Molecules Pass Through Membranes and Bind to DNA-Binding Receptors

Just as prokaryotes can adjust their patterns of gene expression in response to chemicals in their environment, eukaryotes have many systems for responding to specific molecules with which they come in contact. We first consider a system that detects and responds to estrogens. Synthesized and released by the ovaries, *estrogens*, such as estrone, are cholesterol-derived, steroid hormones (Section 26.4). They are required for the development of female secondary sex characteristics and, along with progesterone, participate in the ovarian cycle.



Because they are hydrophobic molecules, estrogens easily diffuse across cell membranes. When inside a cell, estrogens bind to highly specific, soluble receptor proteins. Estrogen receptors are members of a large family of proteins that act as receptors for a wide range of hydrophobic molecules, including other steroid hormones, thyroid hormones, and retinoids.



These receptors all have a similar mode of action. On binding of the signal molecule (called, generically, a *ligand*), the ligand-receptor complex modifies the expression of specific genes by binding to control elements in the DNA. The human genome encodes approximately 50 members of this family, often referred to as *nuclear hormone receptors*. The genomes of other multicellular eukaryotes encode similar numbers of nuclear hormone receptors, although they are absent in yeast. A comparison of the amino acid sequences of members of this family reveals two highly conserved domains: a DNA-binding domain and a ligand-binding domain (Figure 31.22). The DNA-binding domain lies toward the center of the molecule and includes nine conserved cysteine residues. This domain provides these receptors with sequence-specific DNA activity. Eight of the cysteine residues are conserved because of their role in binding zinc ions: the first four cysteine residues bind one zinc ion, and the second four bind a second zinc ion (see Figure 31.22). The zinc ions stabilize the structure of this small domain; without the bound zinc ions, the domains unfold. Such domains are often referred to as *zinc finger domains*.

The structure of the zinc-binding region of a steroid receptor includes an α helix that begins at the end of the first zinc finger domain. This helix lies in the major groove in the specific DNA complexes formed by estrogen receptors and binds with specific DNA sequences, analogously to prokaryotic DNA-binding proteins. Estrogen receptors bind to

specific DNA sites (referred to as *estrogen response elements* or *EREs*) that contain the consensus sequence 5'-AGGTCANNNTGACCT-3'. As expected from the symmetry of this sequence, an estrogen receptor binds to such sites as a dimer.

The second highly conserved region of the nuclear receptor proteins lies near the carboxyl terminus and is the ligand-binding site. This domain folds into a structure that consists almost entirely of α helices, arranged in three layers. The ligand binds in a hydrophobic pocket that lies in the center of this array of helices (Figure 31.23). The ligand-binding domain also participates in receptor dimerization.


A comparison of the structures of the ligand-binding domains with and without bound ligand reveals that ligand binding leads to substantial structural rearrangement. In particular, the last α helix (referred to as helix 12), which has hydrophobic residues lining one face but extends out from the receptor in the ligand-free form, folds into a shallow groove on the side of the receptor on ligand binding (see Figure 31.23). How does ligand binding lead to changes in gene expression? The simplest model would have the binding of ligand alter the DNA-binding properties of the receptor, analogously to the *lac* repressor in prokaryotes. However, the results of experiments with purified nuclear hormone receptors revealed that ligand binding does *not* significantly alter DNA-binding affinity and specificity. Another mechanism must be operative.

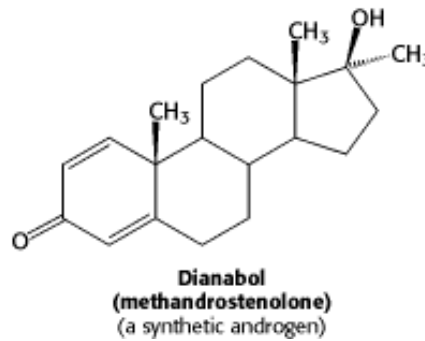
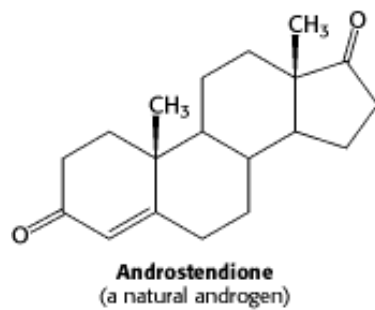
31.3.2. Nuclear Hormone Receptors Regulate Transcription by Recruiting Coactivators and Corepressors to the Transcription Complex

Because ligand binding does not alter the ability of nuclear hormone receptors to bind DNA, investigators sought to determine whether specific proteins might bind to the nuclear hormone receptors only in the presence of ligand. Such searches led to the identification of several related proteins called *coactivators*, such as SRC-1 (steroid receptor coactivator-1), GRIP-1 (glucocorticoid receptor interacting protein-1), and NcoA-1 (nuclear hormone receptor coactivator-1). These coactivators, referred to as the p160 family because of their size, have a common modular structure (Figure 31.24). Each coactivator protein contains three sequences of the form Leu-X-X-Leu-Leu within a central region of 200 amino acids. These sequences form short α helices that bind to a hydrophobic patch on the surface of the ligand-binding domains of a nuclear hormone receptor (Figure 31.25). The binding site for the coactivator is fully formed only when ligand is bound, inasmuch as it is adjacent to helix 12. It is likely that a coactivator molecule binds to the ligand-binding domains of a receptor dimer through two of its three Leu-X-X-Leu-Leu sequences. Thus, the binding of ligand to the receptor induces a conformational change that allows the recruitment of a coactivator (Figure 31.26).

Some members of the nuclear hormone receptor family, such as the receptors for thyroid hormone and retinoic acid, repress transcription in the absence of ligand. This repression also is mediated by the ligand-binding domain. In their unbound forms, the ligand-binding domains of these receptors bind to *corepressor proteins*. Members of this family of proteins include SMRT (Silencing mediator for retinoid and thyroid hormone receptors) and N-CoR (nuclear hormone receptor corepressor). Such a corepressor binds to a site in the ligand-binding domain that overlaps the coactivator binding site; ligand binding triggers the release of the corepressor and frees the ligand-binding domain for binding to a coactivator.

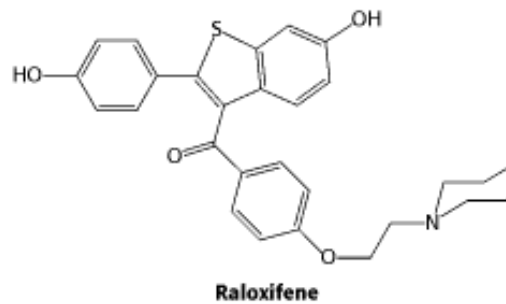
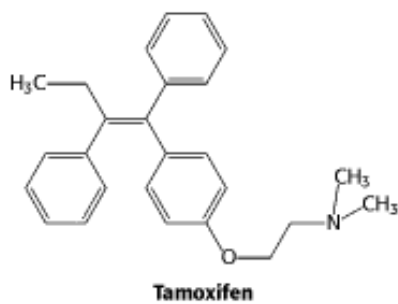
31.3.3. Steroid-Hormone Receptors Are Targets for Drugs

 Molecules such as estradiol that bind to a receptor and trigger signaling pathways are called *agonists*. Athletes sometimes take natural and synthetic agonists of the androgen receptor, a member of the nuclear hormone receptor family, because their binding to the androgen receptor stimulates the expression of genes that enhance the development of lean muscle mass.



Referred to as *anabolic steroids*, such compounds used in excess are not without side effects. In men, excessive use leads to a decrease in the secretion of testosterone, to testicular atrophy, and sometimes to breast enlargement (gynecomastia) if some of the excess androgen is converted into estrogen. In women, excess testosterone causes a decrease in ovulation and estrogen secretion; it also causes breast regression and growth of facial hair.

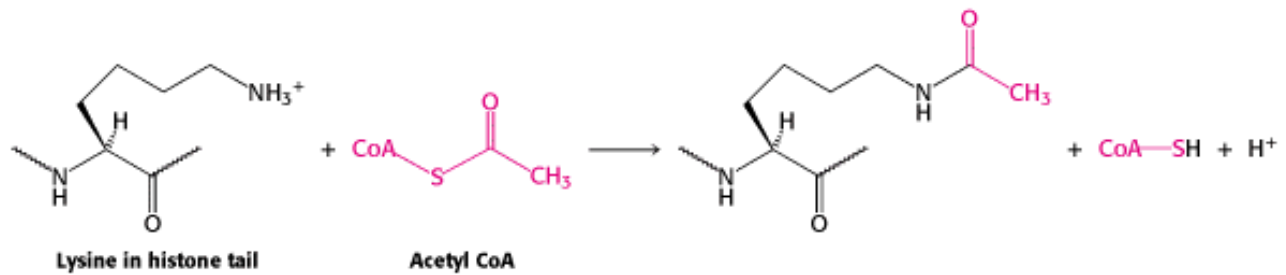
Other molecules bind to nuclear hormone receptors but do not effectively trigger signaling pathways. Such compounds are called *antagonists* and are, in many ways, like competitive inhibitors of enzymes. Some important drugs are antagonists that target the estrogen receptor. For example, *tamoxifen* and *raloxifene* are used in the treatment and prevention of breast cancer, because some breast tumors rely on estrogen-mediated pathways for growth. These compounds are sometimes called *selective estrogen receptor modulators (SERMs)*.



The determination of the structures of complexes between the estrogen receptor and these drugs revealed the basis for their antagonist effect (Figure 31.27). Tamoxifen binds to the same site as estradiol. However, tamoxifen (and other antagonists) have groups that extend out of the normal ligand-binding pocket. These groups prevent helix 12 from binding in its usual position; instead, this helix binds to the site normally occupied by the coactivator. Tamoxifen blocks the binding of coactivators and thus inhibits the activation of gene expression.

31.3.4. Chromatin Structure Is Modulated Through Covalent Modifications of Histone Tails

We have seen that nuclear receptors respond to signal molecules by recruiting coactivators and corepressors to the chromatin. Now we can ask, How do coactivators and corepressors modulate transcriptional activity? Much of their effectiveness appears to result from their ability to covalently modify the amino-terminal tails of histones and perhaps other proteins. Some of the p160 coactivators and, in addition, the proteins that they recruit catalyze the transfer of acetyl groups from acetyl CoA to specific lysine residues in the amino-terminal tails of histones.



Enzymes that catalyze such reactions are called *histone acetyltransferases (HATs)*. The histone tails are readily extended; so they can fit into the HAT active site and become acetylated ([Figure 31.28](#)).

What are the consequences of histone acetylation? Lysine bears a positively charged ammonium group at neutral pH. The addition of an acetyl group generates an uncharged amide group. This change dramatically reduces the affinity of the tail for DNA and modestly decreases the affinity of the entire histone complex for DNA. The loosening of the histone complex from the DNA exposes additional DNA regions to the transcription machinery. In addition, the acetylated lysine residues interact with a specific *acetyllysine-binding domain* that is present in many proteins that regulate eukaryotic transcription. This domain, termed a *bromodomain*, comprises approximately 110 amino acids that form a four-helix bundle containing a peptide-binding site at one end ([Figure 31.29](#)).

Bromodomain-containing proteins are components of two large complexes essential for transcription. One is a complex of more than 10 polypeptides that binds to the *TATA-box-binding protein*. Recall that the TATA-box-binding protein is an essential transcription factor for many genes ([Section 28.2.4](#)). Proteins that bind to the TATA-box-binding protein are called *TAFs* (for *TATA-box-binding protein associated factors*). In particular, TAFII250 (named for its participation in RNA polymerase II transcription and its apparent molecular weight of 250 kd) contains a pair of bromodomains near its carboxyl terminus. The two domains are oriented so that each can bind one of two acetyllysine residues at positions 5 and 12 in the histone H4 tail. Thus, *acetylation of the histone tails provides a mechanism for recruiting other components of the transcriptional machinery*.

Bromodomains are also present in some components of large complexes known as *chromatin-remodeling engines*. These complexes, which also contain domains homologous to those of helicases ([Section 27.2.5](#)), utilize the free energy of ATP hydrolysis to shift the positions of nucleosomes along the DNA and to induce other conformational changes in chromatin ([Figure 31.30](#)). Histone acetylation can lead to reorganization of the chromatin structure, potentially exposing binding sites for other factors. *Thus, histone acetylation can activate transcription through a combination of three mechanisms: by reducing the affinity of the histones for DNA, by recruiting other components of the transcriptional machinery, and by initiating the active remodeling of the chromatin structure.*

31.3.5. Histone Deacetylases Contribute to Transcriptional Repression

Just as in prokaryotes, some changes in a cell's environment lead to the repression of genes that had been active. The modification of histone tails again plays an important role. However, in repression, a key reaction appears to be the deacetylation of acetylated lysine, catalyzed by specific *histone deacetylase* enzymes.

In many ways, the acetylation and deacetylation of lysine residues in histone tails (and, likely, in other proteins) is analogous to the phosphorylation and dephosphorylation of serine, threonine, and tyrosine residues in other stages of signaling processes. Like the addition of phosphoryl groups, the addition of acetyl groups can induce conformational changes and generate novel binding sites. Without a means of removing these groups, however, these signaling switches will become stuck in one position and lose their effectiveness. Like phosphatases, deacetylases help reset the switches.

31.3.6. Ligand Binding to Membrane Receptors Can Regulate Transcription Through Phosphorylation Cascades

In [Chapter 15](#), we examined several signaling pathways that begin with the binding of molecules to receptors in the cell membrane. Some of these pathways lead to the regulation of gene expression. Let us review the pathway initiated by epinephrine. The binding of epinephrine to a 7TM receptor results in the activation of a G protein. The activated G protein, in turn, binds to and activates adenylate cyclase, increasing the intracellular concentration of cAMP. This cAMP binds to the regulatory subunit of protein kinase A (PKA), activating the enzyme. We have previously examined the role of phosphorylation by PKA of a variety of enzymes—for example, those controlling glycogen metabolism. PKA also phosphorylates the *cyclic AMP-response element binding protein (CREB)*, a transcription factor that binds specific DNA sequences as a dimer. Each monomer contributes a long α helix; together, the two helices grab the DNA in the manner of a pair of chopsticks ([Figure 31.31](#)).

How does the phosphorylation of CREB affect its ability to activate transcription? Phosphorylation does not appear to alter the DNA-binding properties of this protein. Instead, phosphorylated CREB binds a coactivator protein termed CBP, for *CREB-binding protein*. CBP possesses a highly revealing domain structure ([Figure 31.32](#)).

Its domains include a KIX domain (for *kinase-inducible interaction*) that binds to the phosphorylated region of CREB ([Figure 31.33](#)); a bromodomain that binds acetylated histone tails, and two TAZ domains, zinc-binding domains that facilitate the binding of CBP to a variety of proteins through a remarkable triangular structure (see [Figure 31.32](#)). Thus, the pathway initiated by epinephrine binding induces the phosphorylation of a transcription factor, the recruitment of a coactivator, and the assembly of complexes that participate in chromatin remodeling and transcription initiation.

31.3.7. Chromatin Structure Effectively Decreases the Size of the Genome

The transcriptional regulatory mechanisms utilized by prokaryotes and eukaryotes have some significant differences, many of which are related to the significant difference in genome sizes between these classes of organisms. However, much of the DNA in a eukaryotic cell is stably assembled into chromatin. The packaging of DNA with chromatin renders many potential binding sites for transcription factors inaccessible—in effect, reducing the size of the genome. Thus, rather than scanning through the entire genome, a *eukaryotic DNA-binding protein scans a set of accessible binding sites that is close in size to the genome of a prokaryote*. The cell type is determined by the genes that are accessible.

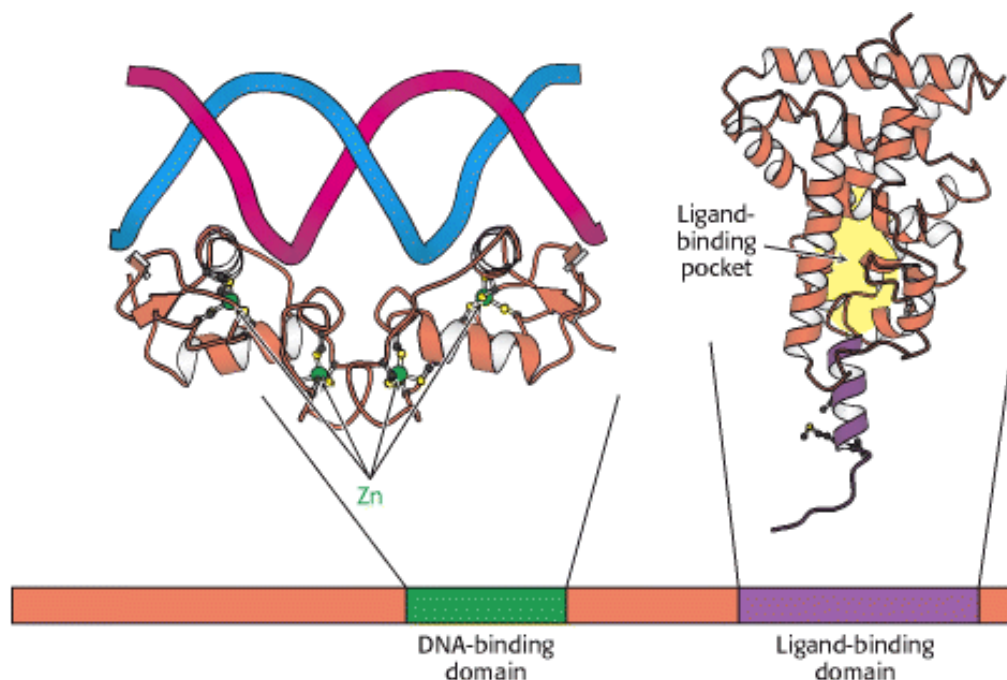


Figure 31.22. Structure of Two Nuclear Hormone Receptor Domains. Nuclear hormone receptors contain two crucial

conserved domains: (1) a DNA-binding domain toward the center of the sequence and (2) a ligand-binding domain toward the carboxyl terminus. The structure of a dimer of the DNA-binding domain bound to DNA is shown, as is one monomer of the normally dimeric ligand-binding domain.

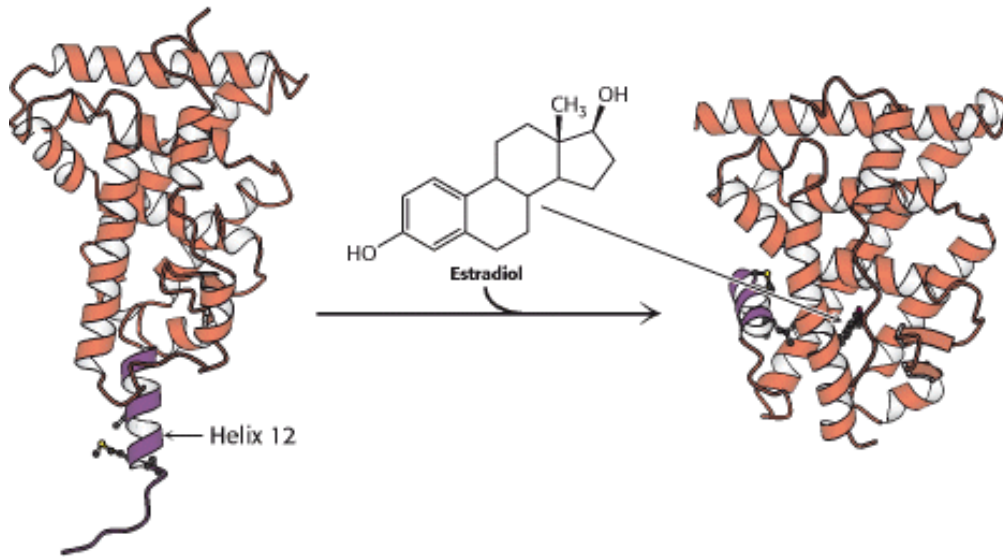


Figure 31.23. Ligand Binding to Nuclear Hormone Receptor. The ligand lies completely surrounded within a pocket in the ligand-binding domain. The last α helix, helix 12 (shown in purple), folds into a groove on the side of the structure on ligand binding.

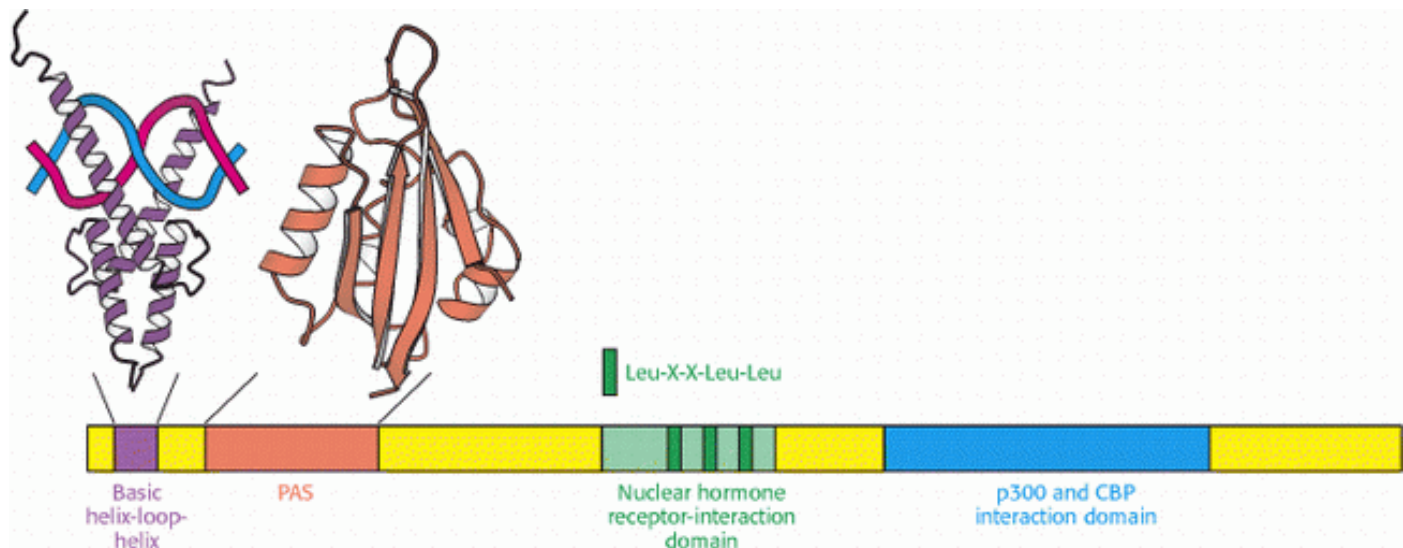


Figure 31.24. Coactivator Structure. The p160 family of coactivators includes a number of domains that can be recognized at the amino acid sequence level, including a basic helix-loop-helix domain that takes part in DNA binding, a PAS domain that participates in protein-protein interactions, a central domain that contacts the ligand-binding domain of the nuclear hormone receptors, and a domain that interacts with additional coactivators such as p300 and CREB-binding protein (CBP). (CREB stands for cyclic AMP-response element binding protein.) The nuclear hormone receptor interaction domain includes three Leu-X-X-Leu-Leu sequences.

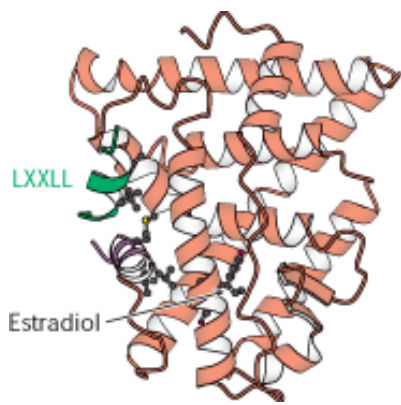


Figure 31.25. Coactivator-Nuclear Hormone Receptor Interactions. The structure of a complex between the ligand-binding domain of the estrogen receptor with estradiol bound and a peptide from a coactivator reveals that the Leu-X-X-Leu-Leu (LXXLL) sequence forms a helix that binds in a groove on the surface of the ligand-binding domain.

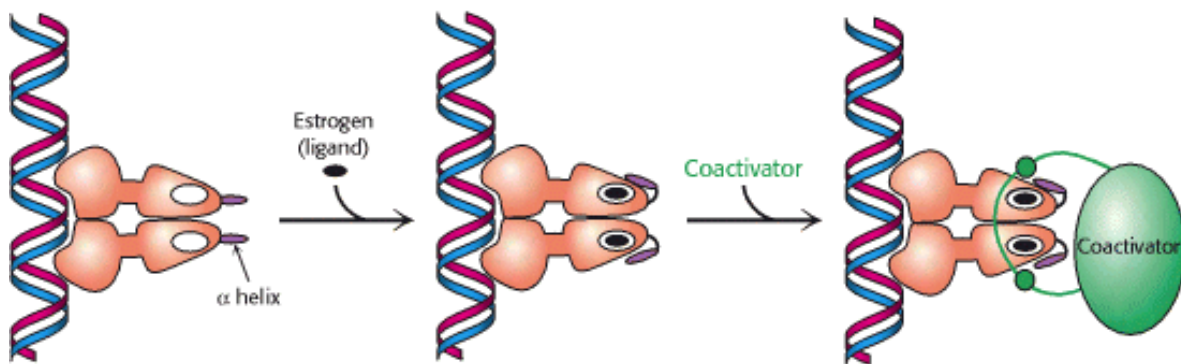


Figure 31.26. Coactivator Recruitment. The binding of ligand to a nuclear hormone receptor induces a conformational change in the ligand-binding domain. This change in conformation generates favorable sites for the binding of a coactivator.

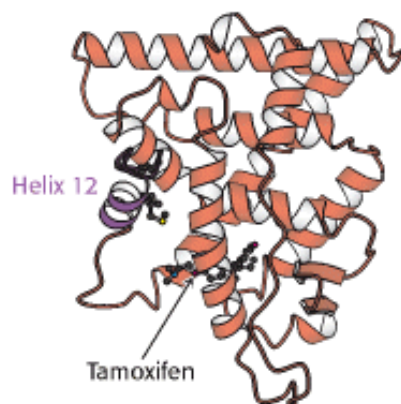


Figure 31.27. Estrogen Receptor-Tamoxifen Complex. Tamoxifen binds in the pocket normally occupied by estrogen. However, part of the tamoxifen structure extends from this pocket, and so helix 12 cannot pack in its usual position. Instead, this helix blocks the coactivator-binding site.

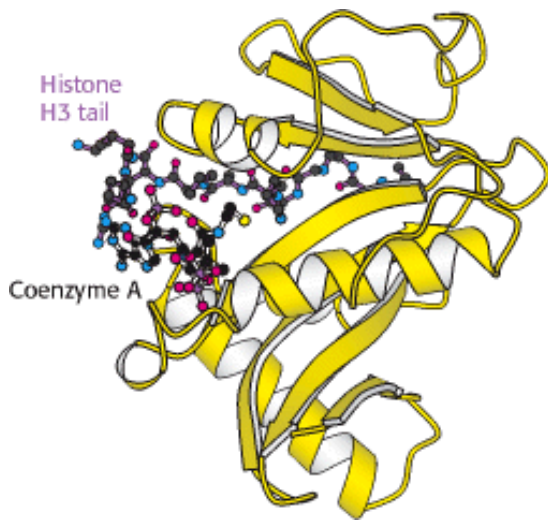




Figure 31.28. Structure of Histone Acetyltransferase. The amino-terminal tail of histone H3 extends into a pocket in  which a lysine side chain can accept an acetyl group from acetyl CoA bound in an adjacent site.



Figure 31.29. Structure of a Bromodomain. This four-helix-bundle domain binds peptides containing acetyllysine. An  acetylated peptide of histone H4 is bound in the structure shown.

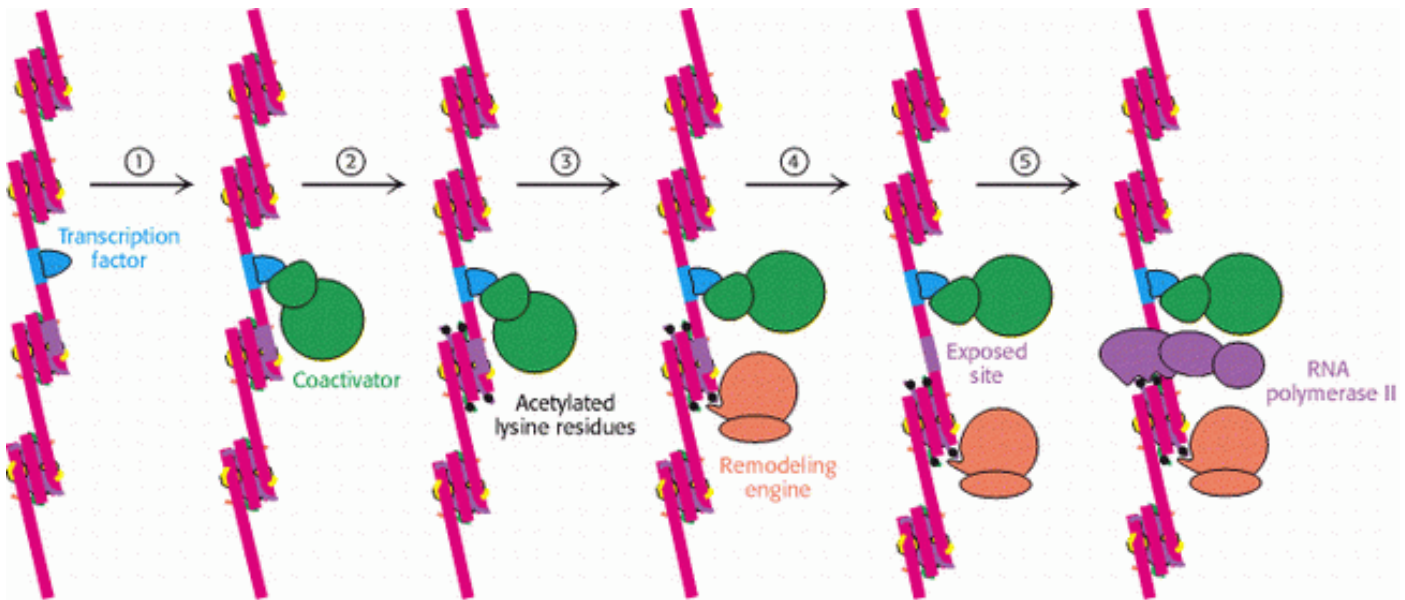


Figure 31.30. Chromatin Remodeling. Eukaryotic gene regulation begins with an activated transcription factor bound to a specific site on DNA. One scheme for the initiation of transcription by RNA polymerase II requires five steps: (1) recruitment of a coactivator, (2) acetylation of lysine residues in the histone tails, (3) binding of a remodeling engine complex to the acetylated lysine residues, (4) ATP-dependent remodeling of the chromatin structure to expose a binding site for RNA polymerase or for other factors, and (5) recruitment of RNA polymerase. Only two subunits are shown for each complex, although the actual complexes are much larger. Other schemes are possible.

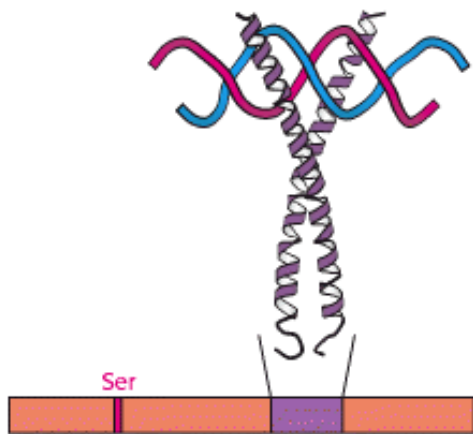


Figure 31.31. Cyclic AMP-Response Element Binding Protein (CREB). Each of two CREB subunits contributes a long α helix. The two helices coil around each other to form a dimeric DNA-binding unit. CREB is phosphorylated on a specific serine residue by protein kinase A.

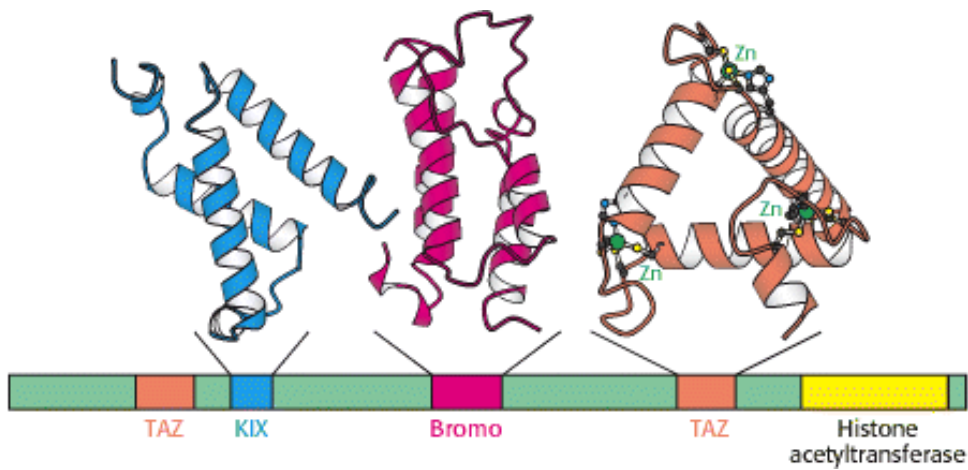


Figure 31.32. Domain Structure of CREB-Binding Protein (CBP). The CREB-binding protein includes at least three types of protein-protein interaction domains in addition to a histone acetyltransferase domain that lies near the carboxyl terminus. The kinase-inducible interaction (KIX) domain interacts specifically with a region of CREB in its phosphorylated form.

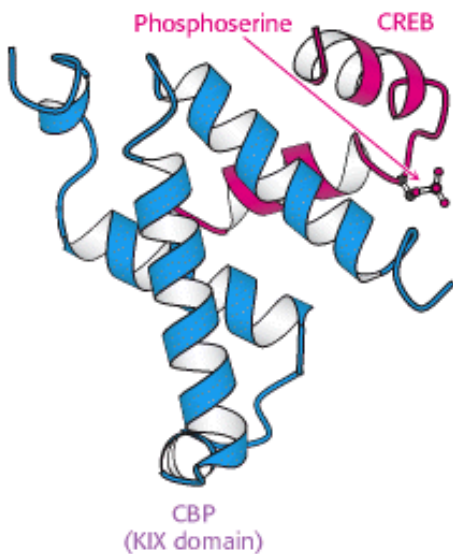


Figure 31.33. Interaction between CBP and CREB. The KIX domain of CBP binds a region of CREB in its phosphorylated form.

31.4. Gene Expression Can Be Controlled at Posttranscriptional Levels

Modulation of the rate of transcriptional initiation is the most common mechanism of gene regulation. However, other stages of transcription also are targets for regulation in some cases. In addition, the process of translation provides other points of intervention for regulating the level of a protein produced in a cell. These mechanisms are quite distinct in prokaryotic and eukaryotic cells because prokaryotes and eukaryotes differ greatly in how transcription and translation are coupled and in how translation is initiated. We shall consider two important examples of posttranscriptional regulation: one from prokaryotes and the other from eukaryotes. In both examples, regulation depends on the formation of distinct secondary structures in mRNA.

31.4.1. Attenuation Is a Prokaryotic Mechanism for Regulating Transcription Through Modulation of Nascent RNA Secondary Structure

A novel mechanism for regulating transcription in bacteria was discovered by Charles Yanofsky and his colleagues as a result of their studies of the tryptophan operon. The 7-kb mRNA transcript from this operon encodes five enzymes that convert chorismate into tryptophan (Section 24.2.10). The mode of regulation of this operon is called *attenuation*, and it depends on features at the 5' end of the mRNA product (Figure 31.34). Upstream of the coding regions for the enzymes responsible for tryptophan biosynthesis lies a short open reading frame encoding a 14-amino-acid leader peptide. Following this open reading frame is a region of RNA, called an *attenuator*, that is capable of forming several alternative structures. Recall that transcription and translation are tightly coupled in bacteria. Thus, the translation of the *trp* mRNA begins soon after the ribosome-binding site has been synthesized.

A ribosome is able to translate the leader region of the mRNA product only in the presence of adequate concentrations of tryptophan. When enough tryptophan is present, a stem-loop structure forms in the attenuator region, which leads to the release of RNA polymerase from the DNA (Figure 31.35). However, when tryptophan is scarce, transcription is terminated less frequently. How does the level of tryptophan alter transcription of the *trp* operon? An important clue was the finding that the 14-amino-acid leader peptide includes two adjacent tryptophan residues. When tryptophan is scarce, little tryptophanyl-tRNA is present. Thus, the ribosome stalls at the tandem UGG codons encoding tryptophan. This delay leaves the adjacent region of the mRNA exposed as transcription continues. An alternative RNA structure that does not function as a terminator is formed and transcription continues into and through the coding regions for the enzymes. Thus, attenuation provides an elegant means of sensing the supply of tryptophan required for protein synthesis.

Several other operons for the biosynthesis of amino acids in *E. coli* also are regulated by attenuator sites. The leader peptide of each contains an abundance of the amino acid residues of the type controlled by the operon (Figure 31.36). For example, the leader peptide for the phenylalanine operon includes 7 phenylalanine residues among 15 residues. The threonine operon encodes enzymes required for the synthesis of both threonine and isoleucine; the leader peptide contains 8 threonine and 4 isoleucine residues in a 16-residue sequence. The leader peptide for the histidine operon includes 7 histidine residues in a row. In each case, low levels of the corresponding charged tRNA causes the ribosome to stall, trapping the nascent mRNA in a state that can form a structure that allows RNA polymerase to read through the attenuator site.


31.4.2. Genes Associated with Iron Metabolism Are Translationally Regulated in Animals

RNA secondary structure plays a role in the regulation of iron metabolism in eukaryotes. Iron is an essential nutrient, required for the synthesis of hemoglobin, cytochromes, and many other proteins. However, excess iron can be quite harmful because, untamed by a suitable protein environment, iron can initiate a range of free-radical reactions that damage proteins, lipids, and nucleic acids. Animals have evolved sophisticated systems for the accumulation of iron in times of scarcity and for the safe storage of excess iron for later use. Key proteins include *transferrin*, a transport protein that carries iron in the serum, *transferrin receptor*, a membrane protein that binds iron-loaded transferrin and initiates its entry into cells, and *ferritin*, an impressively efficient iron-storage protein found primarily in the liver and kidneys. Twenty-four ferritin polypeptides form a nearly spherical shell that encloses as many as 2400 iron atoms, a ratio of one iron atom per amino acid (Figure 31.37).

Ferritin and transferrin receptor expression levels are reciprocally related in their responses to changes in iron levels. When iron is scarce, the amount of transferrin receptor increases and little or no new ferritin is synthesized. Interestingly, the extent of mRNA synthesis for these proteins does not change correspondingly. Instead, regulation takes place at the level of translation.

Consider ferritin first. Ferritin mRNA includes a stem-loop structure termed an *iron-response element (IRE)* in its 5' untranslated region (Figure 31.38). This stem-loop binds a 90-kd protein, called an *IRE-binding protein (IRE-BP)*, that blocks the initiation of translation. When the iron level increases, the IRE-BP binds iron as a 4Fe-4S cluster. The IRE-BP bound to iron cannot bind RNA, because the binding sites for iron and RNA substantially overlap. Thus, in the presence of iron, ferritin mRNA is released from the IRE-BP and translated to produce ferritin, which sequesters the excess iron.

An examination of the nucleotide sequence of transferrin-receptor mRNA reveals the presence of several IRE-like regions. However, these regions are located in the 3' untranslated region rather than in the 5' untranslated region (Figure 31.39). Under low-iron conditions, IRE-BP binds to these IREs. However, given the location of these binding sites, the transferrin-receptor mRNA can still be translated. What happens when the iron level increases and the IRE-BP no longer binds transferrin-receptor mRNA? Freed from the IRE-BP, transferrin-receptor mRNA is rapidly degraded. Thus, an increase in the cellular iron level leads to the destruction of transferrin-receptor mRNA and, hence, a reduction in the production of transferrin-receptor protein.

 The purification of the IRE-BP and the cloning of its cDNA were sources of truly remarkable insight into evolution. The IRE-BP was found to be approximately 30% identical in amino acid sequence with the citric acid cycle enzyme aconitase from mitochondria. Further analysis revealed that the IRE-BP is, in fact, an active aconitase enzyme; it is a cytosolic aconitase that had been known for a long time, but its function was not well understood (Figure 31.40). The iron-sulfur center at the active site of the IRE-BP is rather unstable, and loss of the iron triggers significant changes in protein conformation. Thus, this protein can serve as an iron-sensing factor.

Other mRNAs, including those taking part in heme synthesis, have been found to contain IREs. Thus, genes encoding proteins required for iron metabolism acquired sequences that, when transcribed, provided binding sites for the iron-sensing protein. An environmental signal—the concentration of iron—controls the translation of proteins required for the metabolism of this metal. The IREs have evolved appropriately in the untranslated regions of mRNAs to lead to beneficial regulation by iron levels.

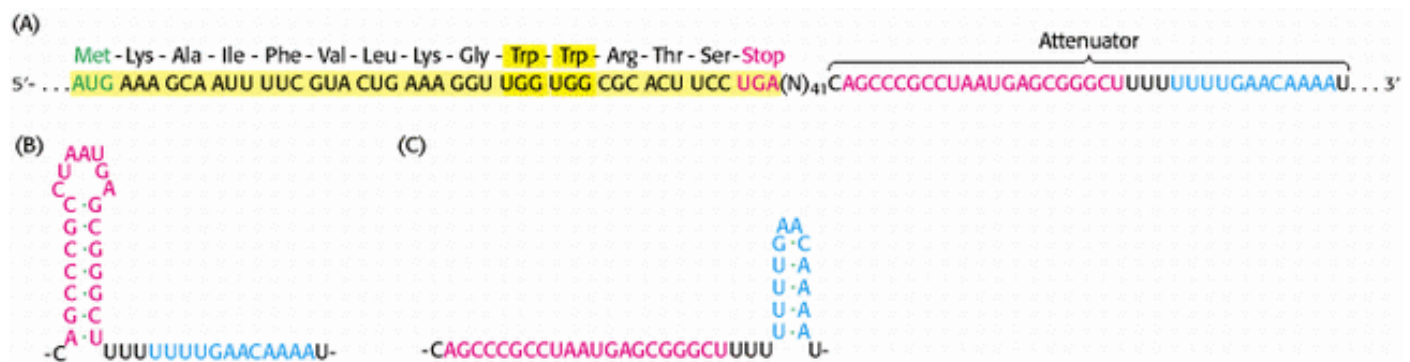


Figure 31.34. Leader Region of *TRP* mRNA. (A) The nucleotide sequence of the 5' end of *trp* mRNA includes a short open reading frame that encodes a peptide comprising 14 amino acids; the leader encodes two tryptophan residues and has an untranslated attenuator region (blue and red nucleotides). (B and C) The attenuator region can adopt two distinct stem-loop structures.

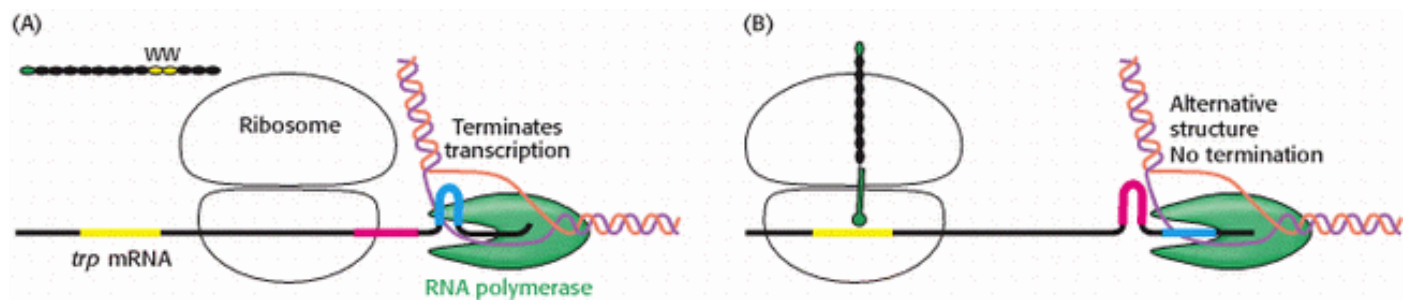


Figure 31.35. Attenuation. (A) In the presence of adequate concentrations of tryptophan (and, hence, Trp-tRNA), translation proceeds rapidly and an RNA structure forms that terminates transcription. (B) At low concentrations of tryptophan, translation stalls awaiting Trp-tRNA, giving time for an alternative RNA structure to form that does not terminate transcription efficiently.

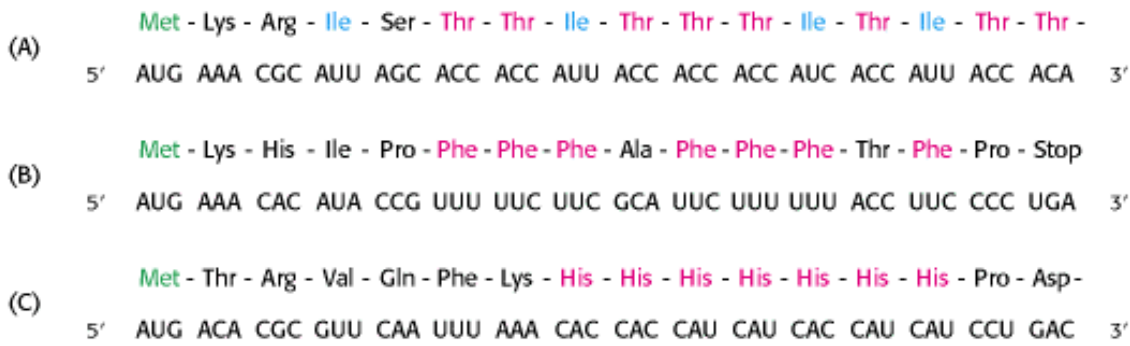


Figure 31.36. Leader Peptide Sequences. Amino acid sequences and the corresponding mRNA nucleotide sequences of the (A) threonine operon, (B) phenylalanine operon, and (C) histidine operon. In each case, an abundance of one amino acid in the leader peptide sequence leads to attenuation.

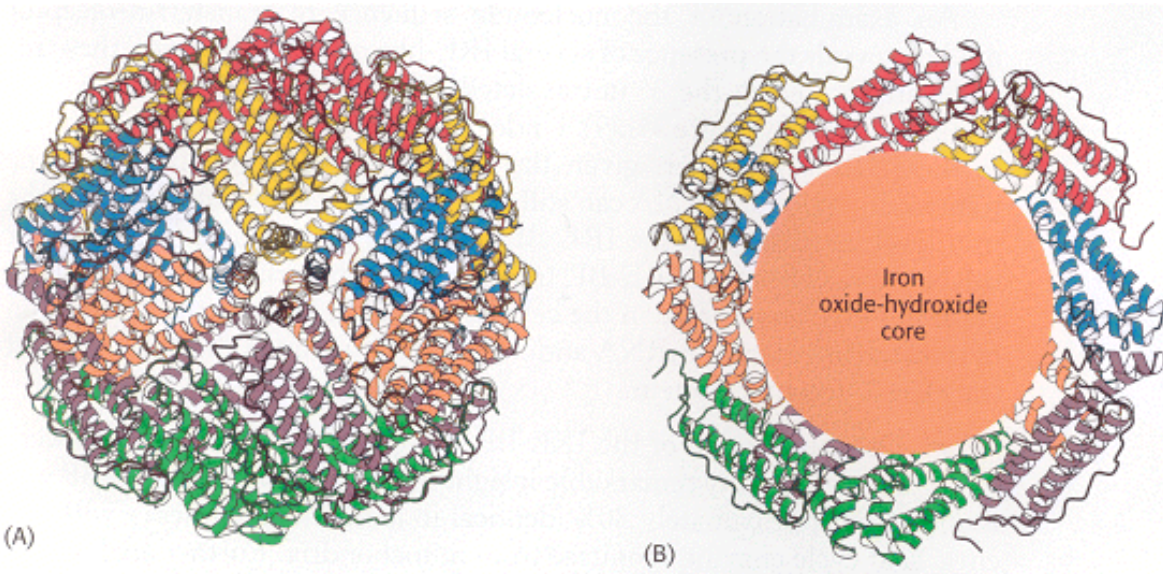


Figure 31.37. Structure of Ferritin. (A) Twenty-four ferritin polypeptides form a nearly spherical shell. (B) A cutaway view reveals the core that stores iron as an iron oxide-hydroxide complex.

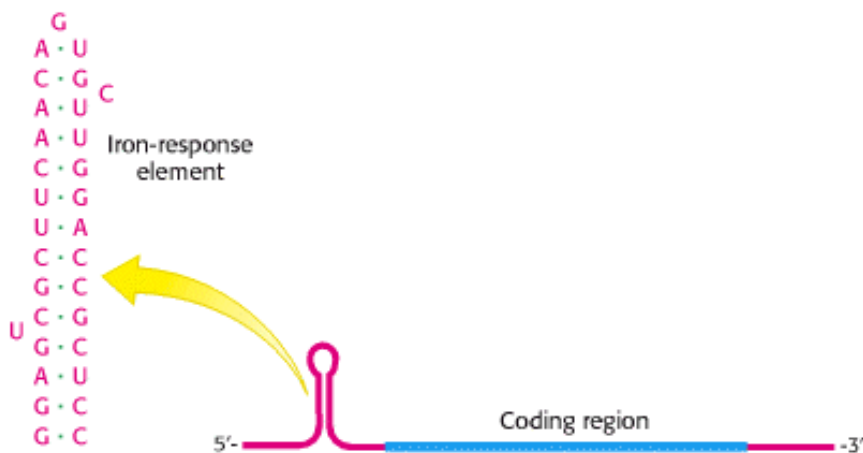


Figure 31.38. Iron-Response Element. Ferritin mRNA includes a stem-loop structure, termed an iron-response element (IRE), in its 5' untranslated region. The IRE binds a specific protein that blocks the translation of this mRNA under low

iron conditions.

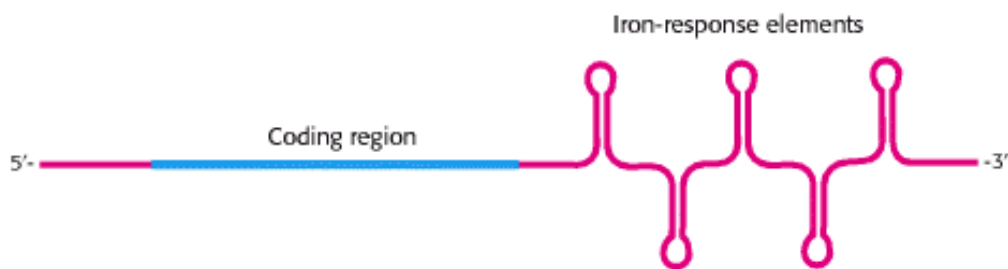


Figure 31.39. Transferrin-receptor mRNA. This mRNA has a set of iron response elements (IREs) in its 3' untranslated region. The binding of the IRE-binding protein to these elements stabilizes the mRNA but does not interfere with translation.

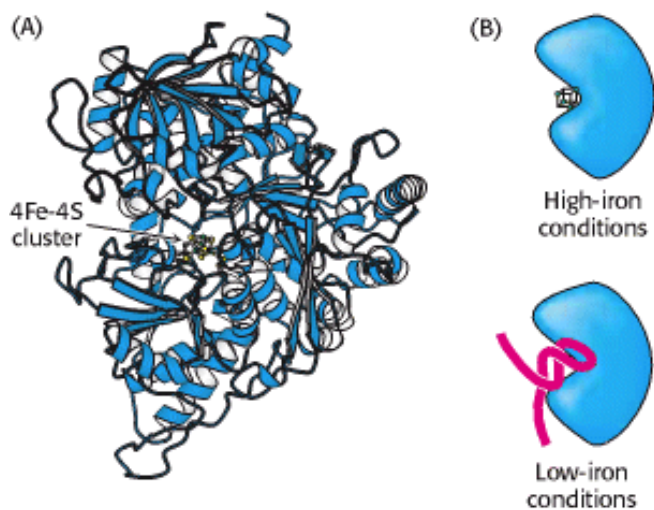



Figure 31.40. The IRE-BP Is an Aconitase. (A) Aconitase contains a relatively unstable 4Fe-4S cluster at its center.

 (B) Under conditions of low iron, the 4Fe-4S cluster dissociates and appropriate RNA molecules can bind in its place.

Summary

Prokaryotic DNA-Binding Proteins Bind Specifically to Regulatory Sites in Operons

In prokaryotes, many genes are clustered into operons, which are units of coordinated genetic expression. An operon consists of control sites (an operator and a promoter) and a set of structural genes. In addition, regulator genes encode proteins that interact with the operator and promoter sites to stimulate or inhibit transcription. The treatment of *E. coli* with lactose induces an increase in the production of β -galactosidase and two additional proteins that are encoded in the lactose operon. In the absence of lactose or a similar galactoside inducer, the *lac* repressor protein binds to an operator site on the DNA and blocks transcription. The binding of allolactose, a derivative of lactose, to the *lac* repressor induces a conformational change that leads to dissociation from DNA. RNA polymerase can then move through the operator to transcribe the *lac* operon.

Some proteins activate transcription by directly contacting RNA polymerase. For example, cyclic AMP, a hunger signal, stimulates the transcription of many catabolic operons by binding to the catabolite activator protein. The binding of the cAMP-CAP complex to a specific site in the promoter region of an inducible catabolic operon enhances the binding of RNA polymerase and the initiation of transcription. Many, but not all, prokaryotic DNA-binding proteins interact with DNA through helix-turn-helix motifs.

The Greater Complexity of Eukaryotic Genomes Requires Elaborate Mechanisms for Gene Regulation

Eukaryotic DNA is tightly bound to basic proteins called histones; the combination is called chromatin. DNA wraps around an octamer of core histones to form a nucleosome, blocking access to many potential DNA binding sites. Changes in chromatin structure play a major role in regulating gene expression. Enhancers can modulate gene expression from more than 1000 bp away from the start site of transcription by perturbing local chromatin structure. Enhancers are often specific for certain cell types, depending on which DNA-binding proteins are present.

Transcriptional Activation and Repression Are Mediated by Protein-Protein Interactions

Steroids such as estrogens bind to eukaryotic transcription factors called nuclear hormone receptors. These proteins are capable of binding DNA whether or not ligands are bound. The binding of ligands induces a conformational change that allows the recruitment of additional proteins called coactivators. Among the most important functions of coactivators is catalysis of the addition of acetyl groups to lysine residues in the tails of histone proteins. Histone acetylation decreases the affinity of the histones for DNA, making additional genes accessible for transcription. In addition, acetylated histones are targets for proteins containing specific binding units called bromodomains. Two classes of large complexes are eventually recruited: chromatin remodeling engines and RNA polymerase II and its associated factors. These complexes open up sites on chromatin and initiate transcription.

Gene Expression Can Be Controlled at Posttranscriptional Levels

Gene expression can also be regulated at the level of translation. In prokaryotes, many operons important in amino acid biosynthesis are regulated by attenuation, a process that depends on the formation of alternative structures in mRNA, one of which favors transcriptional termination. Attenuation is mediated by the translation of a leader region of mRNA. A ribosome stalled by the absence of an aminoacyl-tRNA needed to translate the leader mRNA alters the structure of mRNA so that RNA polymerase transcribes the operon beyond the attenuator site.

In eukaryotes, genes encoding proteins that transport and store iron are regulated at the translational level. Iron-response elements, structures that are present in certain mRNAs, are bound by an IRE-binding protein when this protein is not binding iron. Whether the expression of a gene is stimulated or inhibited in response to changes in the iron status of a cell depends on the location of the IRE within the mRNA.

Key Terms

cell type

β -galactosidase

operon model

repressor

lac operator

lac repressor

inducer

isopropylthiogalactoside (IPTG)

pur repressor

corepressor

catabolite activator protein (CAP)

catabolite repression

helix-turn-helix motif

histone

chromatin

nucleosome

nucleosome core particle

hypersensitive site

chromatin immunoprecipitation (ChIP)

enhancer

hypomethylation

CpG island

combinatorial control

nuclear hormone receptor

zinc-finger domain

estrogen response element (ERE)

coactivator

corepressor protein

agonist

anabolic steroid

antagonist

selective estrogen modulator (SERM)

histone acetyltransferase (HAT)

acetyllysine-binding domain

bromodomain

TATA-box-binding protein associated factor (TAF)

chromatin-remodeling engine

histone deacetylase

cyclic AMP-response element binding protein (CREB)

attenuation

transferrin

transferrin receptor

ferritin

iron-response element (IRE)

IRE-binding protein (IRE-BP)

Problems

1. *Missing genes.* Predict the effects of deleting the following regions of DNA:

(a) The gene encoding *lac* repressor

(b) The *lac* operator

(c) The gene encoding CAP

See answer

2. *Minimal concentration.* Calculate the concentration of *lac* repressor, assuming that one molecule is present per cell. Assume that each *E. coli* cell has a volume of 10^{-12} cm³. Would you expect the single molecule to be free or bound to DNA?

See answer

3. *Counting sites.* Calculate the expected number of times that a given 8-base-pair DNA site should be present in the *E. coli* genome. Assume that all four bases are equally probable. Repeat for a 10-base-pair site and a 12-base-pair site.

See answer

4. *Charge neutralization.* Given the histone amino acid sequences illustrated on the next page, estimate the charge of a histone octamer at pH 7. Assume that histidine residues are uncharged at this pH. How does this charge compare with the charge on 150 base pairs of DNA?

Histone H2A

MSGRGKQGKAKAKTRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAADVLEYLTAEILELAGNA
ARDNKKTRIIIPRHLQLAIRNDEELNKLGRVTIAQGGVLPNIQAVLLPKKTESHHKAKGK

Histone H2B

MPEPAKSAPAPKKGSKKAVTKAQKKGKRRKRSRKEYSVYVYKVLKQVHPDTGISSKAMGIMNSPVNDI
FERIAGEASRLAHYNNRSTITSREIQTAVRLLLPCELAKHAVSEGKAVTKYTSSK

Histone H3

MARTKQTARKSTGGKAPRKLATKAARKSAPSTGGVKKPHRYRPGTVLREIRRYQKSTELLIRKLPFQR
LVREIAQDFKTDLRFQSAIGALQEASEAYLVGLFEDTNLCAIHAKRVTIMPKDIQLARRIRGERA

Histone H4

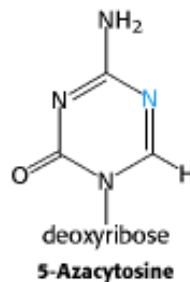
MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDA
VITYEHAKRKTVTAMDVVYALKRQGRITLYGFGG

See answer

5. *Chromatin immunoprecipitation.* You have used the technique of chromatin immunoprecipitation to isolate DNA fragments containing a DNA-binding protein of interest. Suppose that you wish to know whether a particular known DNA fragment is present in the isolated mixture. How might you detect its presence? How many different fragments would you expect if you used antibodies to the *lac* repressor to perform a chromatin immunoprecipitation experiment in *E. coli*? If you used antibodies to the *pur* repressor?

See answer

6. *Nitrogen substitution.* Growth of mammalian cells in the presence of 5-azacytidine results in the activation of some normally inactive genes. Propose an explanation.



See answer

7. *A new domain.* A protein domain has been characterized that recognizes 5-methylcytosine in the context of double-stranded DNA. What role might proteins containing such a domain play in regulating gene expression? Where on a double-stranded DNA molecule would you expect such a domain to bind?

See answer

8. *The same but not the same.* The *lac* repressor and the *pur* operator are homologous proteins with very similar threedimensional structures, yet they have different effects on gene expression. Describe two important ways in which the gene-regulatory properties of these proteins differ.

See answer

9. *The opposite direction.* Some compounds called antiinducers bind to repressors such as the *lac* repressor and inhibit the action of inducers—that is, transcription is repressed and higher concentrations of inducer are required to induce transcription. Propose a mechanism of action for anti-inducers.

See answer

10. *Inverted repeats.* Suppose that a nearly perfect inverted repeat is observed in a DNA sequence over 20 base pairs. Provide two possible explanations.

See answer

Mechanism Problem

11. *Acetyltransferases.* Propose a mechanism for the transfer of an acetyl group from acetyl CoA to the amino group of lysine.

See answer

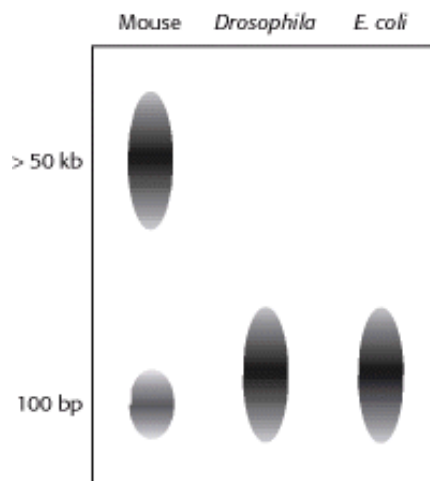
Chapter Integration Problem

12. *The biochemistry of memory.* Long-term memory requires the synthesis of new proteins. The neurotransmitter serotonin plays a key role in stimulating neurons that participate in the storage of long-term memory. Remarkably, microinjection of DNA fragments that contain binding sites for CREB, the cAMP response element binding protein, blocks long-term memory. Propose an explanation for this effect. On the basis of your answer, propose a signaling pathway initiated by the arrival of serotonin at the surface of a neuron cell membrane and ending with the synthesis of new proteins having roles in memory storage.

See answer

Data Interpretation Problem

13. *Limited restriction.* The restriction enzyme HpaII is a powerful tool for analyzing DNA methylation. This enzyme cleaves sites of the form 5'-CCGG-3' but will not cleave such sites if the DNA is methylated on any of the cytosine residues. Genomic DNA from different organisms is treated with HpaII and the results are analyzed by gel electrophoresis (see the adjoining patterns). Provide an explanation for the observed patterns.



See answer

Selected Readings

Where to start

C.O. Pabo and R.T. Sauer. 1984. Protein-DNA recognition *Annu. Rev. Biochem.* 53: 293-321. ([PubMed](#))

K. Struhl. 1989. Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins *Trends Biochem. Sci.* 14: 137-140. ([PubMed](#))

K. Struhl. 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes *Cell* 98: 1-4. ([PubMed](#))

E. Kozmus, J. Torchia, D.W. Rose, L. Xu, R. Kurokawa, E.M. McInerney, T.M. Mullen, C.K. Glass, and M.G. Rosenfeld. 1998. Transcription factor-specific requirements for coactivators and their acetyltransferase functions *Science*. 279: 703-707. ([PubMed](#))

J.D. Aalfs and R.E. Kingston. 2000. What does "chromatin remodeling" mean? *Trends Biochem. Sci.* 25: 548-555. ([PubMed](#))

Books

Ptashne, M., 1992. *A Genetic Switch: Phage λ and Higher Organisms* (2nd ed.). Cell Press and Blackwell Scientific.

McKnight, S. L., and Yamamoto, K. R. (Eds.), 1992. *Transcriptional Regulation* (vols. 1 and 2). Cold Spring Harbor Laboratory Press.

Latchman, D. S., 1999. *Eukaryotic Transcription Factors* (3rd ed.). Academic Press.

Wolffe, A., 1992. *Chromatin Structure and Function*. Academic Press.

Lodish, H., Baltimore, D., Berk, A., Zipursky, S. L., Matsudaira, P., and Darnell, J., 1999. *Molecular Cell Biology* (4th ed.). Scientific American Books.

Prokaryotic gene regulation

C.E. Bell and M. Lewis. 2001. The lac repressor: A second generation of structural and functional studies *Curr. Opin. Struct. Biol.* 11: 19-25. ([PubMed](#))

M. Lewis, G. Chang, N.C. Horton, M.A. Kercher, H.C. Pace, M.A. Schumacher, R.G. Brennan, and P. Lu. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer *Science* 271: 1247-1254. ([PubMed](#))

W. Niu, Y. Kim, G. Tau, T. Heyduk, and R.H. Ebright. 1996. Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase *Cell* 87: 1123-1134. ([PubMed](#))

S.C. Schultz, G.C. Shields, and T.A. Steitz. 1991. Crystal structure of a CAP-DNA complex: The DNA is bent by 90 degrees *Science* 253: 1001-1007. ([PubMed](#))

G. Parkinson, C. Wilson, A. Gunasekera, Y.W. Ebright, R.E. Ebright, and H.M. Berman. 1996. Structure of the CAP-DNA complex at 2.5 angstroms resolution: A complete picture of the protein-DNA interface *J. Mol. Biol.* 260: 395-408. ([PubMed](#))

S. Busby and R.H. Ebright. 1999. Transcription activation by catabolite activator protein (CAP) *J. Mol. Biol.* 293: 199-213. ([PubMed](#))

W.S. Somers and S.E. Phillips. 1992. Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals

DNA recognition by beta-strands *Nature* 359: 387-393. ([PubMed](#))

Nucleosomes and histones

K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, and T.J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution *Nature* 389: 251-260. ([PubMed](#))

G. Arents and E.N. Moudrianakis. 1995. The histone fold: A ubiquitous architectural motif utilized in DNA compaction and protein dimerization *Proc. Natl. Acad. Sci. USA* 92: 11170-11174. ([PubMed](#)) ([Full Text in PMC](#))

A.D. Baxevanis, G. Arents, E.N. Moudrianakis, and D. Landsman. 1995. A variety of DNA-binding and multimeric proteins contain the histone fold motif *Nucleic Acids Res.* 23: 2685-2691. ([PubMed](#))

Clements, A., Rojas, J. R., Trievel, R. C., Wang, L., Berger, S. L., and Marmorstein, R., 1999. Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A. *EMBO J.* 18:3521 – 3532.

J. Deckert and K. Struhl. 2001. Histone acetylation at promoters is differentially affected by specific activators and repressors *Mol. Cell. Biol.* 21: 2726-2735. ([PubMed](#)) ([Full Text in PMC](#))

R.N. Dutnall, S.T. Tafrov, R. Sternglanz, and V. Ramakrishnan. 1998. Structure of the histone acetyltransferase Hat1: A paradigm for the GCN5-related *N*-acetyltransferase superfamily *Cell* 94: 427-438. ([PubMed](#))

M.S. Finnin, J.R. Donigian, A. Cohen, V.M. Richon, R.A. Rifkind, P.A. Marks, R. Breslow, and N.P. Pavletich. 1999. Structures of a histone deacetylase homologue bound to the TSA and SAHA inhibitors *Nature* 401: 188-193. ([PubMed](#))

M.S. Finnin, J.R. Donigian, and N.P. Pavletich. 2001. Structure of the histone deacetylase SIR2 *Nat. Struct. Biol.* 8: 621-625. ([PubMed](#))

R.H. Jacobson, A.G. Ladurner, D.S. King, and R. Tjian. 2000. Structure and function of a human TAFII250 double bromodomain module *Science* 288: 1422-1425. ([PubMed](#))

J.R. Rojas, R.C. Trievel, J. Zhou, Y. Mo, X. Li, S.L. Berger, C.D. Allis, and R. Marmorstein. 1999. Structure of *Tetrahymena* GCN5 bound to coenzyme A and a histone H3 peptide *Nature* 401: 93-98. ([PubMed](#))

Nuclear hormone receptors

R.M. Evans. 1988. The steroid and thyroid hormone receptor superfamily *Science* 240: 889-895. ([PubMed](#))

K.R. Yamamoto. 1985. Steroid receptor regulated transcription of specific genes and gene networks *Annu. Rev. Genet.* 19: 209-252. ([PubMed](#))

D.M. Tanenbaum, Y. Wang, S.P. Williams, and P.B. Sigler. 1998. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains *Proc. Natl. Acad. Sci. USA* 95: 5998-6003. ([PubMed](#)) ([Full Text in PMC](#))

J.W. Schwabe, L. Chapman, J.T. Finch, and D. Rhodes. 1993. The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: How receptors discriminate between their response elements *Cell* 75: 567-578. ([PubMed](#))

A.K. Shiau, D. Barstad, P.M. Loria, L. Cheng, P.J. Kushner, D.A. Agard, and G.L. Greene. 1998. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen *Cell* 95: 927-937. ([PubMed](#))

T.N. Collingwood, F.D. Urnov, and A.P. Wolffe. 1999. Nuclear receptors: Coactivators, corepressors and chromatin remodeling in the control of transcription *J. Mol. Endocrinol.* 23: 255-275. ([PubMed](#))

Chromatin and chromatin remodeling

S.C. Elgin. 1981. DNAase I-hypersensitive sites of chromatin *Cell* 27: 413-415. ([PubMed](#))

H. Weintraub, A. Larsen, and M. Groudine. 1981. α -Globin-gene switching during the development of chicken embryos: Expression and chromosome structure *Cell* 24: 333-344. ([PubMed](#))

B. Ren, F. Robert, J.J. Wyrick, O. Aparicio, E.G. Jennings, I. Simon, J. Zeitlinger, J. Schreiber, N. Hannett, E. Kanin, T. L. Volkert, C.J. Wilson, S.P. Bell, and R.A. Young. 2000. Genome-wide location and function of DNA-binding proteins *Science* 290: 2306-2309. ([PubMed](#))

J.A. Goodrich and R. Tjian. 1994. TBP-TAF complexes: Selectivity factors for eukaryotic transcription *Curr. Opin. Cell. Biol.* 6: 403-409. ([PubMed](#))

A.P. Bird and A.P. Wolffe. 1999. Methylation-induced repression: Belts, braces, and chromatin *Cell* 99: 451-454. ([PubMed](#))

B.R. Cairns. 1998. Chromatin remodeling machines: Similar motors, ulterior motives *Trends Biochem. Sci.* 23: 20-25. ([PubMed](#))

S.R. Albright and R. Tjian. 2000. TAFs revisited: More data reveal new twists and confirm old ideas *Gene* 242: 1-13. ([PubMed](#))

F.D. Urnov and A.P. Wolffe. 2001. Chromatin remodeling and transcriptional activation: The cast (in order of appearance) *Oncogene* 20: 2991-3006. ([PubMed](#))

Posttranscriptional regulation

R. Kolter and C. Yanofsky. 1982. Attenuation in amino acid biosynthetic operons *Annu. Rev. Genet.* 16: 113-134. ([PubMed](#))

C. Yanofsky. 1981. Attenuation in the control of expression of bacterial operons *Nature* 289: 751-758. ([PubMed](#))

T.A. Rouault, C.D. Stout, S. Kaptain, J.B. Harford, and R.D. Klausner. 1991. Structural relationship between an iron-regulated RNA-binding protein (IRE-BP) and aconitase: Functional implications *Cell* 64: 881-883. ([PubMed](#))

R.D. Klausner, T.A. Rouault, and J.B. Harford. 1993. Regulating the fate of mRNA: The control of cellular iron metabolism *Cell* 72: 19-28. ([PubMed](#))

M.J. Gruer, P.J. Artymiuk, and J.R. Guest. 1997. The aconitase family: Three structural variations on a common theme *Trends Biochem. Sci.* 22: 3-6. ([PubMed](#))

E.C. Theil. 1994. Iron regulatory elements (IREs): A family of mRNA non-coding sequences *Biochem. J.* 304: 1-11. ([PubMed](#))

Historical aspects

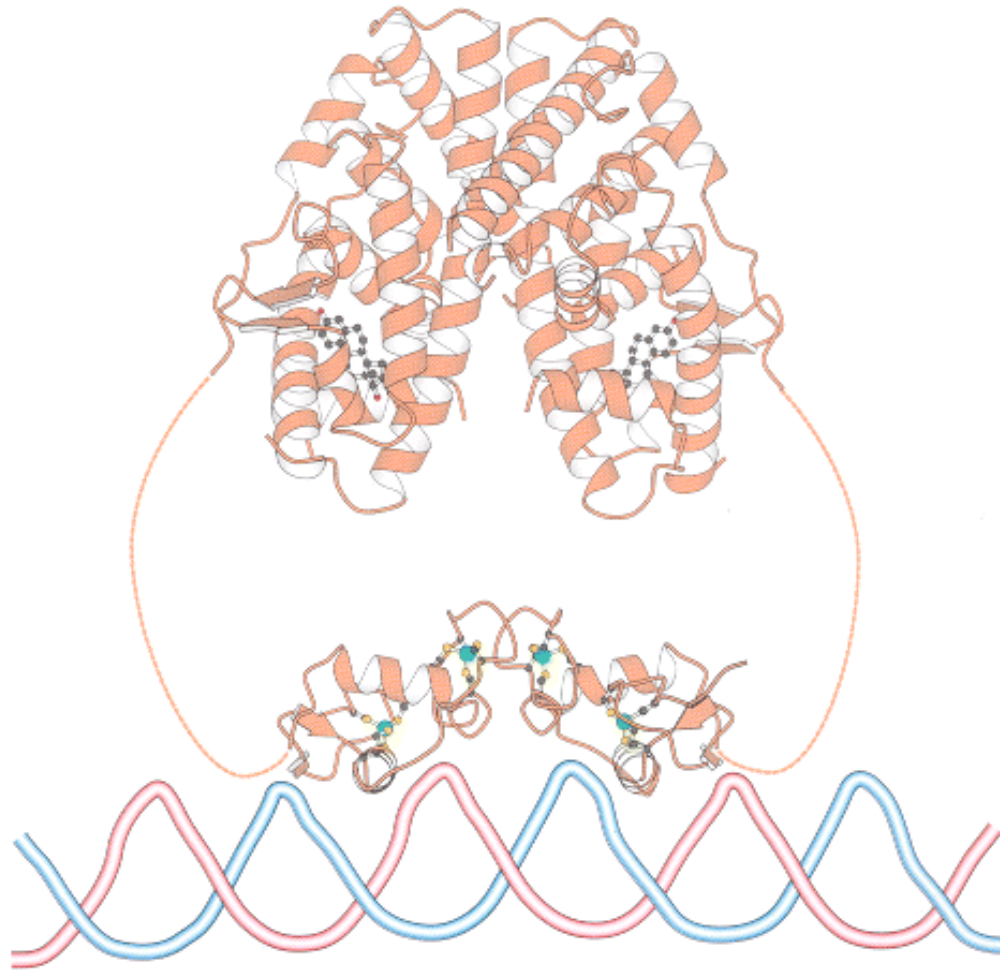
F. Jacob and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins *J. Mol. Biol.* 3: 318-356.

M. Ptashne and W. Gilbert. 1970. Genetic repressors *Sci. Am.* 222: (6) 36-44. ([PubMed](#))

Lwoff, A., and Ullmann, A. (Eds.), 1979. *Origins of Molecular Biology: A Tribute to Jacques Monod*. Academic Press.

Judson, H., 1996. *The Eighth Day of Creation: Makers of the Revolution in Biology*. Cold Spring Harbor Laboratory Press.

IV. Responding to Environmental Changes



Estrogen receptor bound to DNA. Chemical signals in the environment regulate many biochemical processes, including the selective expression of genes. The binding of hormones such as estradiol to the estrogen receptor (orange ribbon diagram) leads to gene expression. The hormone-receptor complex binds to specific DNA sites adjacent to the DNA sequence to be expressed.

32. Sensory Systems

Our senses provide us with means for detecting a diverse set of external signals, often with incredible sensitivity and specificity. For example, when fully adapted to a darkened room, our eyes allow us to sense very low levels of light, down to a *limit of less than ten photons*. With more light, we are able to distinguish millions of colors. Through our senses of smell and taste, we are able to detect thousands of chemicals in our environment and sort them into categories: pleasant or unpleasant? healthful or toxic? Finally, we can perceive mechanical stimuli in the air and around us through our senses of hearing and touch.

How do our sensory systems work? How are the initial stimuli detected? How are these initial biochemical events transformed into perceptions and experiences? We have previously encountered systems that sense and respond to chemical signals—namely, receptors that bind to growth factors and hormones. Our knowledge of these receptors and their associated signal-transduction pathways provides us with concepts and tools for unraveling some of the workings of sensory systems. For example, 7TM receptors (seven-transmembrane receptors, [Section 15.1](#)) play key roles in olfaction,

taste, and vision. Ion channels that are sensitive to mechanical stress are essential for hearing and touch.

In this chapter, we shall focus on the five major sensory systems found in human beings and other mammals: olfaction (the sense of smell; i.e., the detection of small molecules in the air), taste or gustation (the detection of selected organic compounds and ions by the tongue), vision (the detection of light), hearing (the detection of sound, or pressure waves in the air), and touch (the detection of changes in pressure, temperature, and other factors by the skin). Each of these primary sensory systems contains specialized sensory neurons that transmit nerve impulses to the central nervous system (Figure 32.1). In the central nervous system, these signals are processed and combined with other information to yield a perception that may trigger a change in behavior. By these means, our senses allow us to detect changes in our environments and to adjust our behavior appropriately.

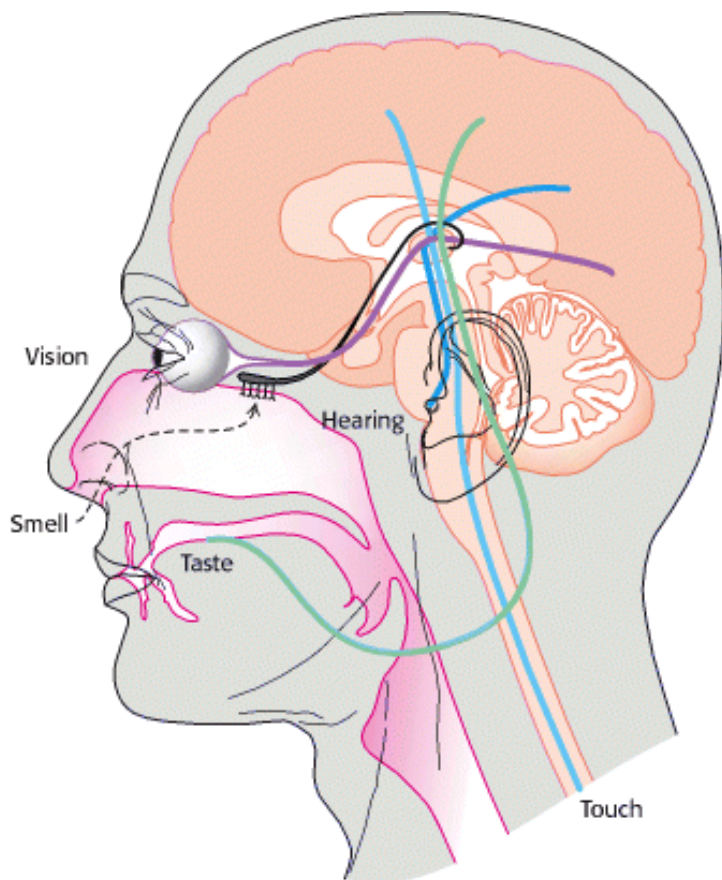
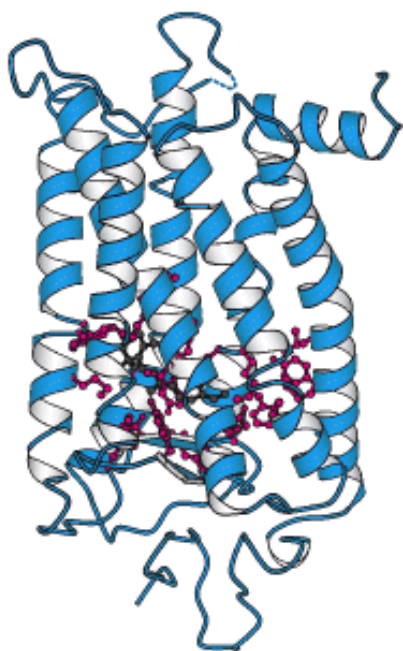


Figure 32.1. Sensory Connections to the Brain. Sensory nerves connect sensory organs to the brain and spinal cord.

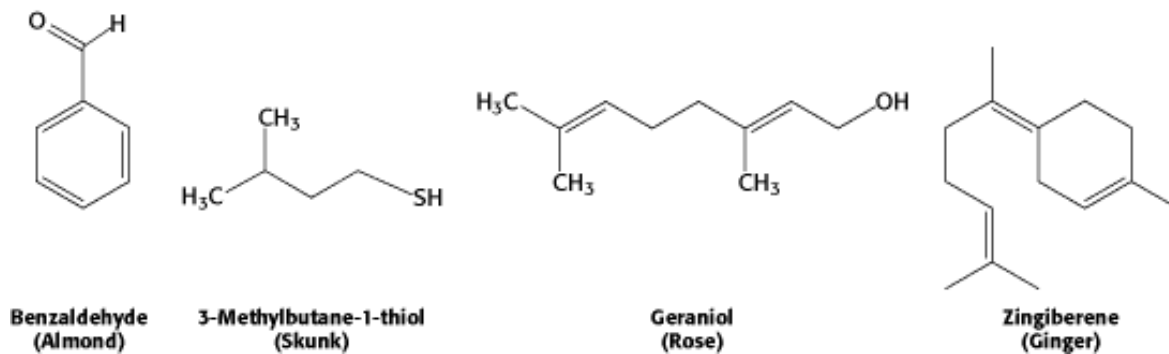




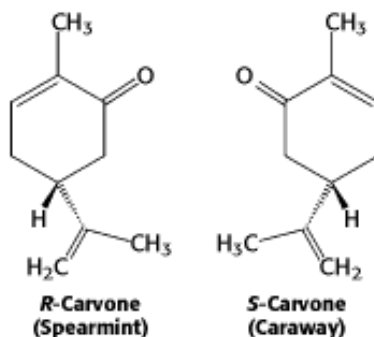
Color Perception. The photoreceptor rhodopsin (bottom), which absorbs light in the process of vision, consists of the protein opsin and a bound vitamin A derivative, retinal. The amino acids (shown in red) that surround the retinal determine the color of light that is most efficiently absorbed. Individuals lacking a light-absorbing photoreceptor for the color green will see a colorful fruit stand (top) as mostly yellows (middle). (Top, middle) from L. T. Sharpe, A. Stockman, H. Jagle, and J. Nathans. (1999) -Opsin genes, cone photopigments, color vision, and color blindness, in *Color Vision: from Genes to Perception*, pp. 3–51. K. Gegenfurtner, L. T. Sharpe, eds. Cambridge University Press.]

32.1. A Wide Variety of Organic Compounds Are Detected by Olfaction

Human beings can detect and distinguish thousands of different compounds by smell, often with considerable sensitivity and specificity. Most odorants are relatively small organic compounds with sufficient volatility that they can be carried as vapors into the nose. For example, a major component responsible for the smell of almonds is the simple aromatic compound benzaldehyde, whereas the sulfhydryl compound 3-methylbutane-1-thiol is a major component of the smell of skunks.




What properties of these molecules are responsible for their smells? First, *the shape of the molecule rather than its other physical properties is crucial*. We can most clearly see the importance of shape by comparing molecules such as those responsible for the smells of spearmint and caraway. These compounds are identical in essentially all physical properties such as hydrophobicity because they are exact mirror images of one another. Thus, the smell produced by an odorant depends not on a physical property but on the compound's interaction with a specific binding surface, most likely a protein receptor. Second, some human beings (and other animals) suffer from *specific anosmias*; that is, they are incapable of smelling specific compounds even though their olfactory systems are otherwise normal. Such anosmias are often inherited. These observations suggest that mutations in individual receptor genes lead to the loss of the ability to detect a small subset of compounds.



32.1.1. Olfaction Is Mediated by an Enormous Family of Seven-Transmembrane-Helix Receptors

Odorants are detected in a specific region of the nose, called the *main olfactory epithelium*, that lies at the top of the nasal cavity (Figure 32.2). Approximately 1 million sensory neurons line the surface of this region. Cilia containing the odorant-binding protein receptors project from these neurons into the mucous lining of the nasal cavity.

Biochemical studies in the late 1980s examined isolated cilia from rat olfactory epithelium that had been treated with odorants. Exposure to the odorants increased the cellular level of cAMP, and this increase was observed only in the presence of GTP. On the basis of what was known about signal-transduction systems, *the participation of cAMP and GTP strongly suggested the involvement of a G protein and, hence, 7TM receptors*. Indeed, Randall Reed purified and cloned a G protein α subunit, termed $G_{(olf)}$, which is uniquely expressed in olfactory cilia. The involvement of 7TM receptors suggested a strategy for identifying the olfactory receptors themselves. cDNAs were sought that (1) were expressed primarily in the sensory neurons lining the nasal epithelium, (2) encoded members of the 7TM receptor family, and (3) were present as a large and diverse family to account for the range of odorants. Through the use of these criteria, cDNAs for odorant receptors from rats were identified in 1991 by Richard Axel and Linda Buck.

 The odorant receptor (hereafter, OR) family is even larger than expected: *more than 1000 OR genes are present in the mouse and the rat, whereas the human genome encodes between an estimated 500 and 750 ORs*. The OR family is thus one of the largest gene families in human beings. However, more than half the human odorant receptor genes appear to be pseudogenes that is, they contain mutations that prevent the generation of a full-length, proper odorant receptor. In contrast, essentially all rodent OR genes are fully functional. Further analysis of primate OR genes

reveals that the fraction of pseudogenes is greater in species more closely related to human beings (Figure 32.3). Thus, we may have a glimpse at the evolutionary loss of acuity in the sense of smell as higher mammals presumably became less dependent on this sense for survival.

The OR proteins are typically 20% identical in sequence to the β -adrenergic receptor (Section 15.1) and from 30 to 60% identical with each other. Several specific sequence features are present in most or all OR family members (Figure 32.4). The central region, particularly transmembrane helices 4 and 5, is highly variable, suggesting that this region is the site of odorant binding. That site must be different in odorant receptors that bind distinct odorant molecules.

What is the relation between OR gene expression and the individual neuron? Interestingly, *each olfactory neuron expresses only a single OR gene*, among hundreds available. Apparently, the precise OR gene expressed is determined largely at random. The mechanism by which all other OR genes are excluded from expression remains to be elucidated. The binding of an odorant to an OR on the neuronal surface initiates a signal-transduction cascade that results in an action potential (Figure 32.5). The ligand-bound OR activates $G_{(olf)}$, the specific G protein mentioned earlier. $G_{(olf)}$ is initially in its GDP-bound form. When activated, it releases GDP, binds GTP, and releases its associated $\beta \gamma$ subunits. The α subunit then activates a specific adenylate cyclase, increasing the intracellular concentration of cAMP. The rise in the intracellular concentration of cAMP activates a nonspecific cation channel that allows calcium and other cations into the cell. The flow of cations through the channel depolarizes the neuronal membrane and initiates an action potential. This action potential, combined with those from other olfactory neurons, leads to the perception of a specific odor.



Conceptual Insights, Signaling Pathways: Response and Recovery presents an animated version of Figure 32.5 and a comparison to visual signal transduction (Figure 32.5).

32.1.2. Odorants Are Decoded by a Combinatorial Mechanism

An obvious challenge presented to the investigator by the large size of the OR family is to match up each OR with the one or more odorant molecules to which it binds. Exciting progress has been made in this regard. Initially, an OR was matched with odorants by overexpressing a single, specific OR gene in rats. This OR responded to straight-chain aldehydes, most favorably to *n*-octanal and less strongly to *n*-heptanal and *n*-hexanal. More dramatic progress was made by taking advantage of our knowledge of the OR signal-transduction pathway and the power of PCR (Section 6.1.5). A section of nasal epithelium from a mouse was loaded with the calcium-sensitive dye Fura-2 (Section 15.3.1). The tissue was then treated with different odorants, one at a time, at a specific concentration. If the odorant bound to and activated an OR, that neuron could be detected under a microscope by the change in fluorescence caused by the influx of calcium that occurs as part of the signal-transduction process. To determine which OR was responsible for the response, cDNA was generated from mRNA that had been isolated from single identified neurons. The cDNA was then subjected to PCR with the use of primers that are effective in amplifying most or all OR genes. The sequence of the PCR product from each neuron was then determined and analyzed.

Using this approach, investigators analyzed the responses of neurons to a series of compounds having varying chain lengths and terminal functional groups (Figure 32.6). The results of these experiments appear surprising at first glance (Figure 32.7). Importantly, there is not a simple 1:1 correspondence between odorants and receptors. *Almost every odorant activates a number of receptors* (usually to different extents) and *almost every receptor is activated by more than one odorant*. Note, however, that each odorant activates a unique combination of receptors. In principle, this combinatorial mechanism allows even a relatively small array of receptors to distinguish a vast number of odorants.

How is the information about which receptors have been activated transmitted to the brain? Recall that each neuron expresses only one OR and that the pattern of expression appears to be largely random. A substantial clue to the connections between receptors and the brain has been provided by the creation of mice that express a gene for an easily detectable colored marker in conjunction with a specific OR gene. Olfactory neurons that express the OR-marker protein combination were traced to their destination in the brain, a structure called the olfactory bulb (Figure 32.8). The processes from neurons that express the same OR gene were found to connect to the same location in the olfactory bulb.

Moreover, this pattern of neuronal connection was found to be identical in all mice examined. Thus, *neurons that express specific ORs are linked with specific sites in the brain*. This property creates a spatial map of odorant-responsive neuronal activity within the olfactory bulb.

Can such a combinatorial mechanism truly distinguish many different odorants? An "electronic nose" that functions by the same principles provides compelling evidence that it can (Figure 32.9). The receptors for the electronic nose are polymers that bind a range of small molecules. Each polymer binds every odorant, but to varying degrees. Importantly, the electrical properties of these polymers change on odorant binding. A set of 32 of these polymer sensors, wired together so that the pattern of responses can be evaluated, is capable of distinguishing individual compounds such as *n*-pentane and *n*-hexane as well as complex mixtures such as the odors of fresh and spoiled fruit.

32.1.3. Functional Magnetic Resonance Imaging Reveals Regions of the Brain Processing Sensory Information

Can we extend our understanding of how odorants are perceived to events in the brain? Biochemistry has provided the basis for powerful methods for examining responses within the brain. One method, called *functional magnetic resonance imaging (fMRI)*, takes advantage of two key observations. The first is that, when a specific part of the brain is active, blood vessels relax to allow more blood flow to the active region. Thus, a more active region of the brain will be richer in oxyhemoglobin. The second observation is that the iron center in hemoglobin undergoes substantial structural changes on binding oxygen (Section 10.4.1). These changes are associated with a rearrangement of electrons such that the iron in deoxyhemoglobin acts as a strong magnet, whereas the iron in oxyhemoglobin does not. The difference between the magnetic properties of these two forms of hemoglobin can be used to image brain activity.

Nuclear magnetic resonance techniques (Section 4.5.1) detect signals that originate primarily from the protons in water molecules but are altered by the magnetic properties of hemoglobin. With the use of appropriate techniques, images can be generated that reveal differences in the relative amounts of deoxy- and oxyhemoglobin and thus the relative activity of various parts of the brain.

These noninvasive methods reveal areas of the brain that process sensory information. For example, subjects have been imaged while breathing air that either does or does not contain odorants. When odorants are present, the fMRI technique detects an increase in the level of hemoglobin oxygenation (and, hence, brain activity) in several regions of the brain (Figure 32.10). Such regions include those in the primary olfactory cortex as well as other regions in which secondary processing of olfactory signals presumably takes place. Further analysis reveals the time course of activation of particular regions and other features. Functional MRI shows tremendous potential for mapping regions and pathways engaged in processing sensory information obtained from all the senses. Thus, *a seemingly incidental aspect of the biochemistry of hemoglobin has yielded the basis for observing the brain in action*.

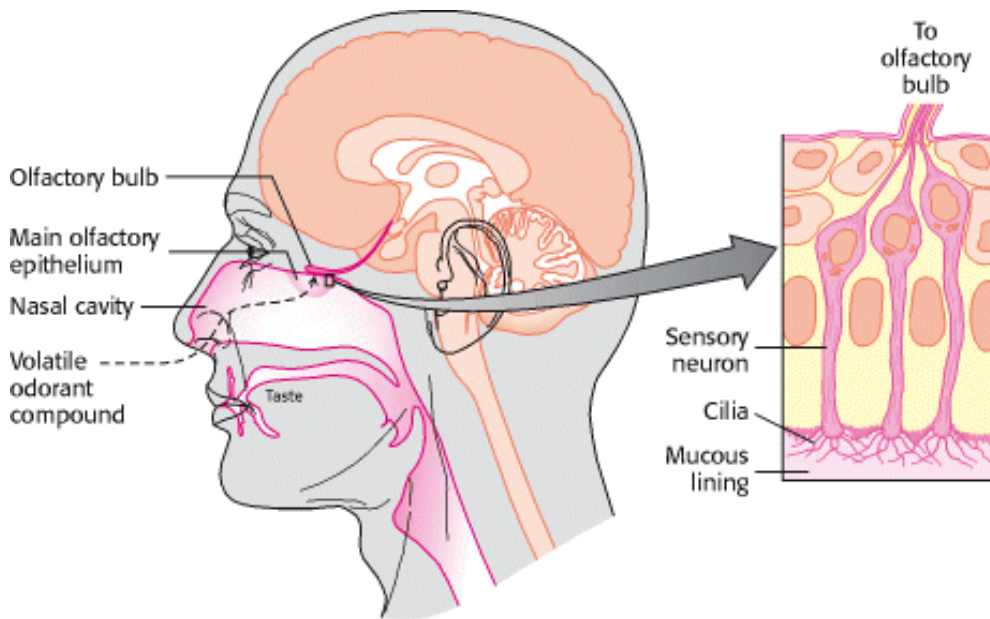


Figure 32.2. The Main Nasal Epithelium. This region of the nose, which lies at the top of the nasal cavity, contains approximately 1 million sensory neurons. Nerve impulses generated by odorant molecules binding to receptors on the cilia travel from the sensory neurons to the olfactory bulb.



Figure 32.3. Evolution of Odorant Receptors. Odorant receptors appear to have lost function through conversion into pseudogenes in the course of primate evolution. The percentage of OR genes that appear to be functional for each species is shown in parentheses.

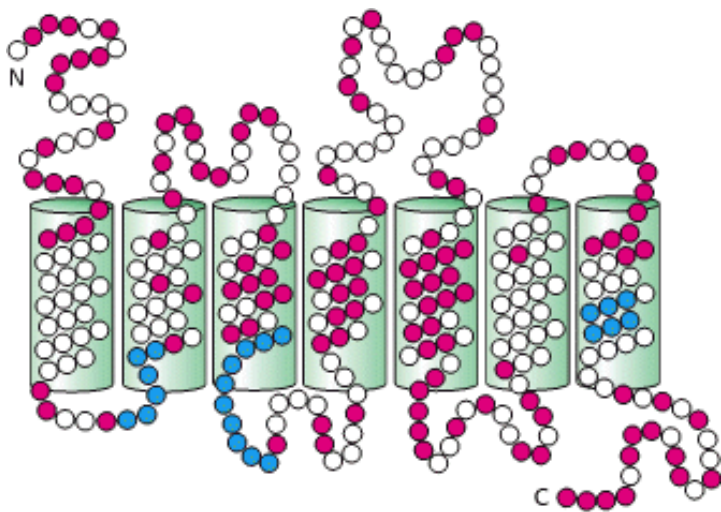


Figure 32.4. Conserved and Variant Regions in Odorant Receptors. Odorant receptors are members of the 7TM receptor family. The green cylinders represent the seven presumed transmembrane helices. Strongly conserved residues characteristic of this protein family are shown in blue, whereas highly variable residues are shown in red.

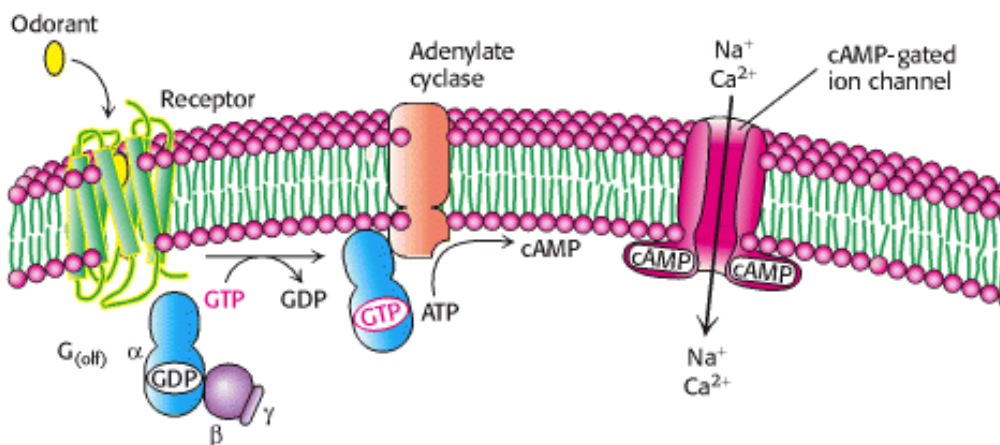


Figure 32.5. The Olfactory Signal-Transduction Cascade. The binding of odorant to the olfactory receptor activates a signaling pathway similar to those initiated in response to the binding of some hormones to their receptors (see [Section 15.1](#)). The final result is the opening of cAMP-gated ion channels and the initiation of an action potential.

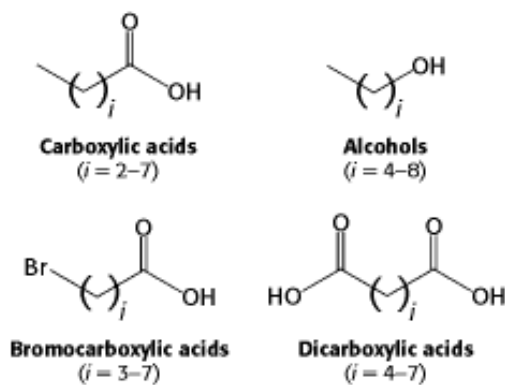


Figure 32.6. Four Series of Odorants Tested for Olfactory Receptor Activation.

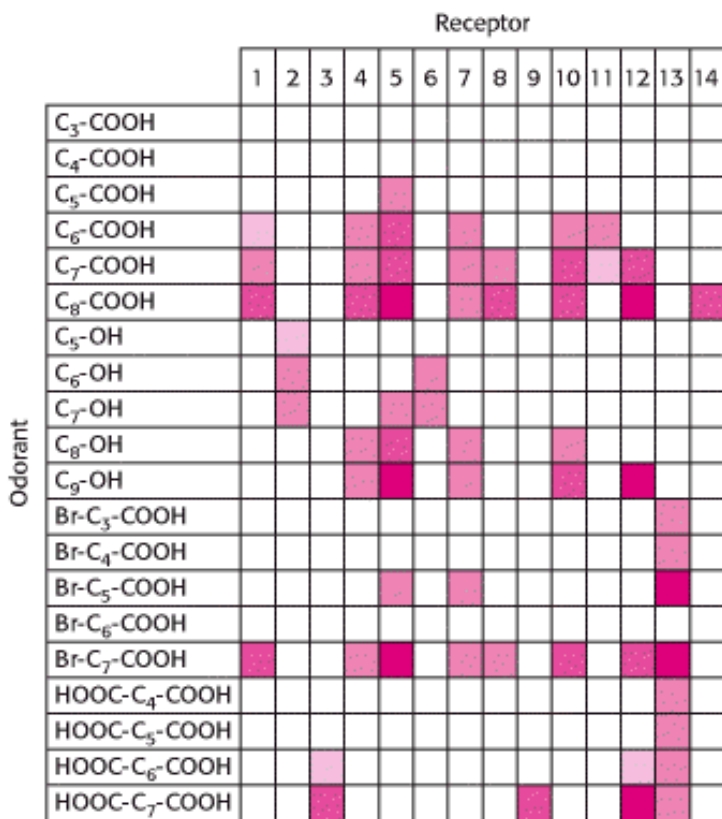


Figure 32.7. Patterns of Olfactory Receptor Activation. Fourteen different receptors were tested for responsiveness to the compounds shown in Figure 32.6. A colored box indicates that the receptor at the top responded to the compound at the left. Darker colors indicate that the receptor was activated at a lower concentration of odorant.



Figure 32.8. Converging Olfactory Neurons. This section of the nasal cavity is stained to reveal processes from sensory neurons expressing the same olfactory receptor. The processes converge to a single location in the olfactory bulb. [From P. Mombaerts, F. Wang, C. Dulac, S. K. Chao, A. Nemes, M. Mendelsohn, J. Edmondson, and R. Axel. *Cell* 87(1996):675–689.]



Figure 32.9. The Cyrano 320. The electronic nose may find uses in the food industry, animal husbandry, law enforcement, and medicine. [Courtesy of Cyrano Sciences.]

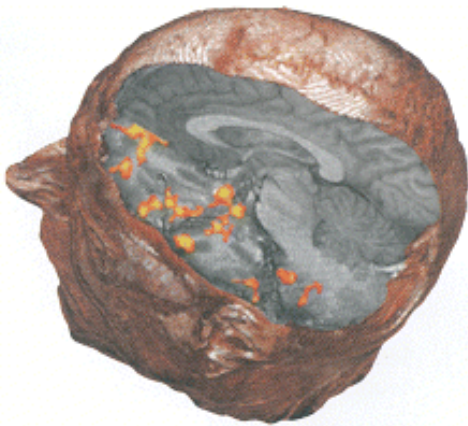


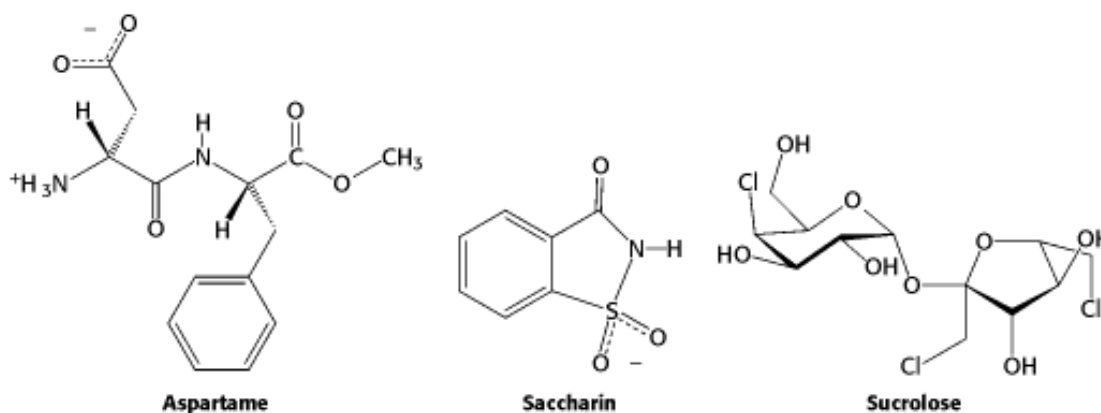
Figure 32.10. Brain Response to Odorants. A functional magnetic resonance image reveals brain response to odorants. The light spots indicate regions of the brain activated by odorants. [N. Sobel et al., *J. Neurophysiol.* 83:537–551 2000 537; courtesy of Nathan Sobel.]

32.2. Taste Is a Combination of Senses that Function by Different Mechanisms

The inability to taste food is a common complaint when nasal congestion reduces the sense of smell. Thus, smell greatly augments our sense of taste (also known as *gustation*), and taste is, in many ways, the sister sense to olfaction. Nevertheless, the two senses differ from each other in several important ways. First, we are able to sense several classes of compounds by taste that we are unable to detect by smell; salt and sugar have very little odor, yet they are primary stimuli of the gustatory system. Second, whereas we are able to discriminate thousands of odorants, discrimination by taste is much more modest. Five primary tastes are perceived: *bitter*, *sweet*, *sour*, *salty*, and *umami* (the taste of glutamate from the Japanese word for "deliciousness"). These five tastes serve to classify compounds into potentially

nutritive and beneficial (sweet, salty, umami) or potentially harmful or toxic (bitter, sour). Tastants (the molecules sensed by taste) are quite distinct for the different groups ([Figure 32.11](#)).

The simplest tastant, the hydrogen ion, is perceived as sour. Other simple ions, particularly sodium ion, are perceived as salty. The taste called umami is evoked by the amino acid glutamate, often encountered as the flavor enhancer monosodium glutamate (MSG). In contrast, *tastants perceived as bitter or sweet are extremely diverse*. Many bitter compounds are alkaloids or other plant products of which many are toxic. However, they do not have any common structural elements or other common properties. Carbohydrates such as glucose and sucrose are perceived as sweet, as are other compounds including some simple peptide derivatives, such as aspartame, and even some proteins.

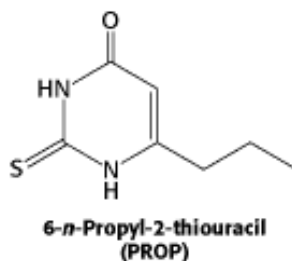


These differences in specificity among the five tastes are due to differences in their underlying biochemical mechanisms. The sense of taste is, in fact, a number of independent senses all utilizing the same organ, the tongue, for their expression.

Tastants are detected by specialized structures called *taste buds*, which contain approximately 150 cells, including sensory neurons ([Figure 32.12](#)). Fingerlike projections called *microvilli*, which are rich in taste receptors, project from one end of each sensory neuron to the surface of the tongue. Nerve fibers at the opposite end of each neuron carry electrical impulses to the brain in response to stimulation by tastants. Structures called *taste papillae* contain numerous taste buds.

32.2.1. Sequencing the Human Genome Led to the Discovery of a Large Family of 7TM Bitter Receptors

Just as in olfaction, a number of clues pointed to the involvement of G proteins and, hence, 7TM receptors in the detection of bitter and sweet tastes. The evidence included the isolation of a specific G protein α subunit termed *gustducin*, which is expressed primarily in taste buds ([Figure 32.13](#)). How could the 7TM receptors be identified? The ability to detect some compounds depends on specific genetic loci in both human beings and mice. For instance, the ability to taste the bitter compound 6-*n*-propyl-2-thiouracil (PROP) was mapped to a region on human chromosome 5 by comparing DNA markers of persons who vary in sensitivity to this compound.



This observation suggested that this region might encode a 7TM receptor that responded to PROP. Approximately 450 kilobases in this region had been sequenced early in the human genome project. This sequence was searched by

computer for potential 7TM receptor genes, and, indeed, one was detected and named *T2R-1*. Additional database searches for sequences similar to this one detected 12 genes encoding full-length receptors as well as 7 pseudogenes within the sequence of the human genome known at the time. The encoded proteins were between 30 and 70% identical with T2R-1 (Figure 32.14). *Further analysis suggests that there are from 50 to 100 members of this family of 7TM receptors in the entire human genome.* Similar sequences have been detected in the mouse and rat genomes.

Are these proteins, in fact, bitter receptors? Several lines of evidence suggest that they are. First, their genes are expressed in taste-sensitive cells—in fact, in many of the same cells that express gustducin. Second, cells that express individual members of this family respond to specific bitter compounds. For example, cells that express a specific mouse receptor (mT2R-5) responded when exposed specifically to cycloheximide. Third, mice that had been found unresponsive to cycloheximide were found to have point mutations in the gene encoding mT2R-5. Finally, cycloheximide specifically stimulates the binding of GTP analogs to gustducin in the presence of the mT2R-5 protein (Figure 32.15).

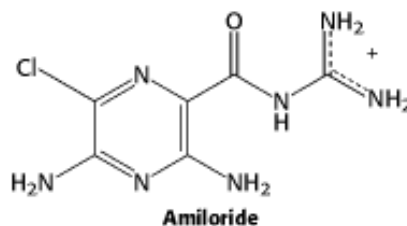
Importantly, each taste receptor cell expresses many different members of the T2R family. This pattern of expression stands in sharp contrast to the pattern of one receptor type per cell that characterizes the olfactory system (Figure 32.16). The difference in expression patterns accounts for the much greater specificity of our perceptions of smells compared with tastes. *We are able to distinguish among subtly different odors because each odorant stimulates a unique pattern of neurons. In contrast, many tastants stimulate the same neurons.* Thus, we perceive only "bitter" without the ability to discriminate cycloheximide from quinine.

32.2.2. A Family of 7TM Receptors Almost Certainly Respond to Sweet Compounds

Most sweet compounds are carbohydrates, energy rich and easily digestible. Some noncarbohydrate compounds such as saccharin and aspartame also taste sweet. The structural diversity among sweet-tasting compounds, though less than that among bitter compounds, strongly suggested that a family of receptors detects these compounds. The observation that mice in which the gene for gustducin was disrupted lost much of their ability to sense sweet, as well as bitter, compounds strongly suggested that the sweet receptors would belong to the 7TM receptor superfamily. Recently, a small group of 7TM receptors that respond to sweet compounds has been identified. Interestingly, simultaneous expression of two members of the family in the same cell is required for the cells to respond to sweet compounds. The biochemical explanation for this observation remains to be elucidated.

32.2.3. Salty Tastes Are Detected Primarily by the Passage of Sodium Ions Through Channels

Salty tastants are not detected by 7TM receptors. Rather, they are detected directly by their passage through ion channels expressed on the surface of cells in the tongue. Evidence for the role of these ion channels comes from examining known properties of sodium channels characterized in other biological contexts. One class of channels, characterized first for their role in salt reabsorption, are thought to be important in salt taste detection because they are sensitive to the compound *amiloride*, which mutes the taste of salt and significantly lowers sensory neuron activation in response to sodium.



An *amiloride-sensitive sodium channel* comprises four subunits that may be either identical or distinct but in any case are homologous. An individual subunit ranges in length from 500 to 1000 amino acids and includes two presumed membrane-spanning helices as well as a large extracellular domain in between them (Figure 32.17). The extracellular

region includes two (or, sometimes, three) distinct regions rich in cysteine residues (and, presumably, disulfide bonds). A region just ahead of the second membrane-spanning helix appears to form part of the pore in a manner analogous to the structurally characterized potassium channel ([Section 13.5.6](#)). The members of the amiloride-sensitive sodium-channel family are numerous and diverse in their biological roles. We shall encounter them again in the context of the sense of touch.


Sodium ions passing through these channels produce a significant transmembrane current. Amiloride blocks this current, accounting for its effect on taste. However, about 20% of the response to sodium remains even in the presence of amiloride, suggesting that other ion channels also contribute to salt detection.

32.2.4. Sour Tastes Arise from the Effects of Hydrogen Ions (Acids) on Channels

Like salty tastes, *sour tastes are also detected by direct interactions with ion channels*, but the incoming ions are hydrogen ions (in high concentrations) rather than sodium ions. For example, in the absence of high concentrations of sodium, hydrogen ion flow can induce substantial transmembrane currents through amiloride-sensitive sodium channels. However, hydrogen ions are also sensed by mechanisms other than their direct passage through membranes. Binding by hydrogen ions blocks some potassium channels and activates other types of channels. Together, these mechanisms lead to changes in membrane polarization in sensory neurons that produce the sensation of sour taste.

32.2.5. Umami, the Taste of Glutamate, Is Detected by a Specialized Form of Glutamate Receptor

Glutamate is an abundant amino acid that is present in protein-rich foods as well as in the widely used flavor enhancer monosodium glutamate. This amino acid has a taste, termed *umami*, that is distinct from the other four basic tastes. Adults can detect glutamate at a concentration of approximately 1 mM. Glutamate is also a widely used neurotransmitter, and thus, not surprisingly, several classes of receptors for glutamate have been identified in the nervous system. One class, called *metabotropic glutamate receptors*, are 7TM receptors with large amino-terminal domains of approximately 600 amino acids. Sequence analysis reveals that the first half of the aminoterminal region is most likely a ligand-binding domain, because it is homologous to such domains found in the Lac repressor ([Section 31.1.3](#)) and other bacterial ligand-binding proteins.

 One glutamate receptor gene, encoding a protein called the metabotropic glutamate receptor 4 (mGluR4), has been found to be expressed in taste buds. Further analysis of the mRNA that is expressed in taste buds reveals that this mRNA lacks the region encoding the first 309 amino acids in brain mGluR4, which includes most of the high-affinity glutamate-binding domain ([Figure 32.18](#)). The glutamate receptor found in taste buds shows a lowered affinity for glutamate that is appropriate to glutamate levels in the diet. Thus, *the receptor responsible for the perception of glutamate taste appears to have evolved simply by changes in the expression of an existing glutamate-receptor gene*. We shall consider an additional receptor related to taste, that responsible for the "hot" taste of spicy food, when we deal with mechanisms of touch perception.

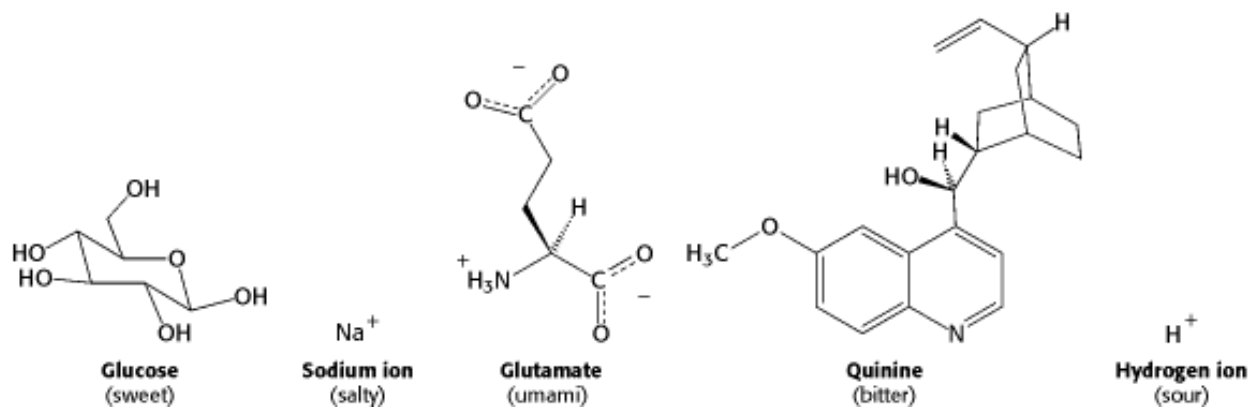


Figure 32.11. Examples of Tastant Molecules. Tastants fall into five groups: sweet, salty, umami, bitter, and sour.

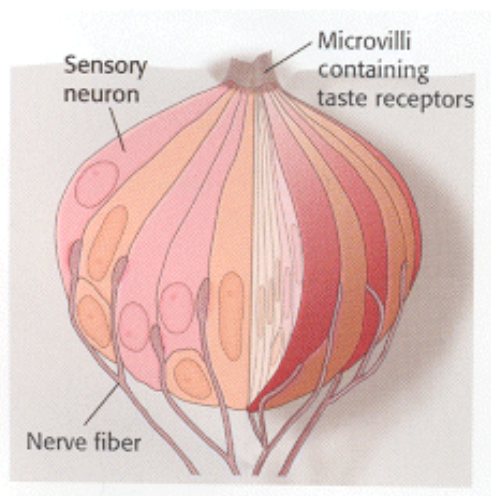
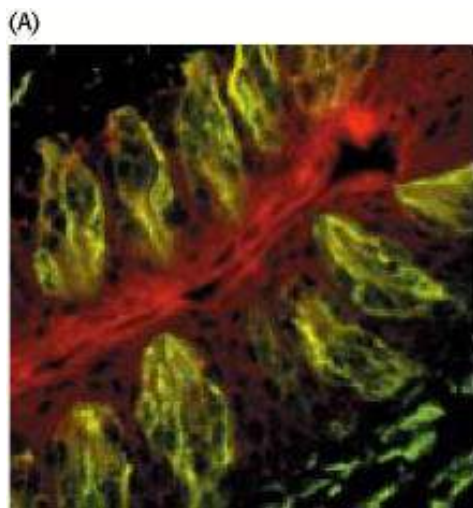


Figure 32.12. A Taste Bud. Each taste bud contains sensory neurons that extend microvilli to the surface of the tongue, where they interact with tastants.



(B)

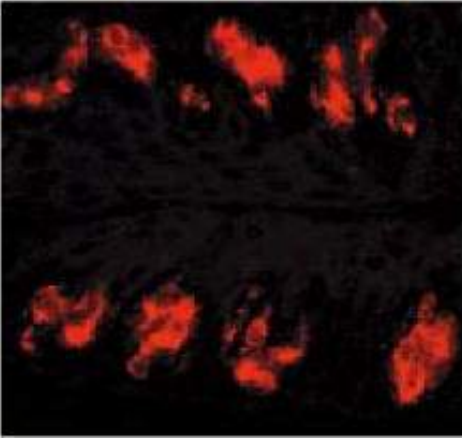


Figure 32.13. Expression of Gustducin in the Tongue. (A) A section of tongue stained with a fluorescent antibody reveals the position of the taste buds. (B) The same region stained with a antibody directed against gustducin reveals that this G protein is expressed in taste buds. [Courtesy of Charles S. Zuker.]

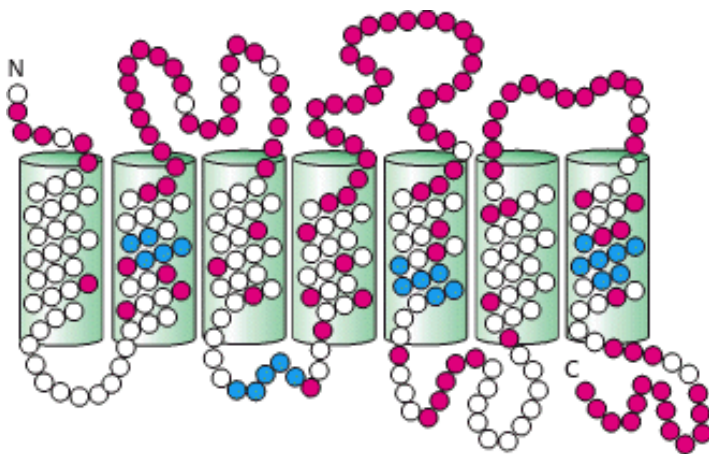


Figure 32.14. Conserved and Variant Regions in Bitter Receptors. The bitter receptors are members of the 7TM receptor family. Strongly conserved residues characteristic of this protein family are shown in blue, and highly variable residues are shown in red.

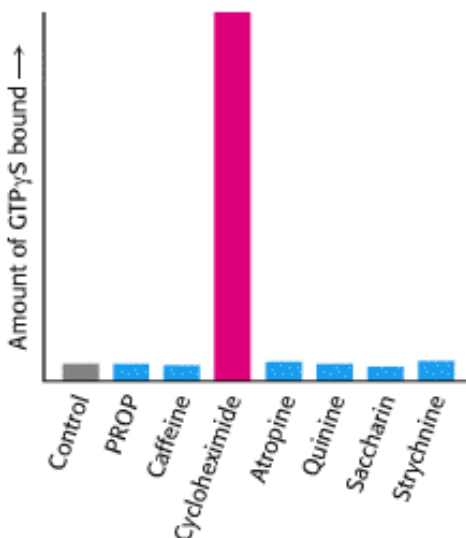


Figure 32.15. Evidence that T2R Proteins Are Bitter Taste Receptors. Cycloheximide uniquely stimulates the

binding of the GTP analog GTP γ S to gustducin in the presence of the mT2R protein. [Adapted from J. Chandrashekar, K. L. Mueller, M. A. Hoon, E. Adler, L. Feng, W. Guo, C. S. Zuker, and N. J. Ryba. *Cell* 100(2000):703.]

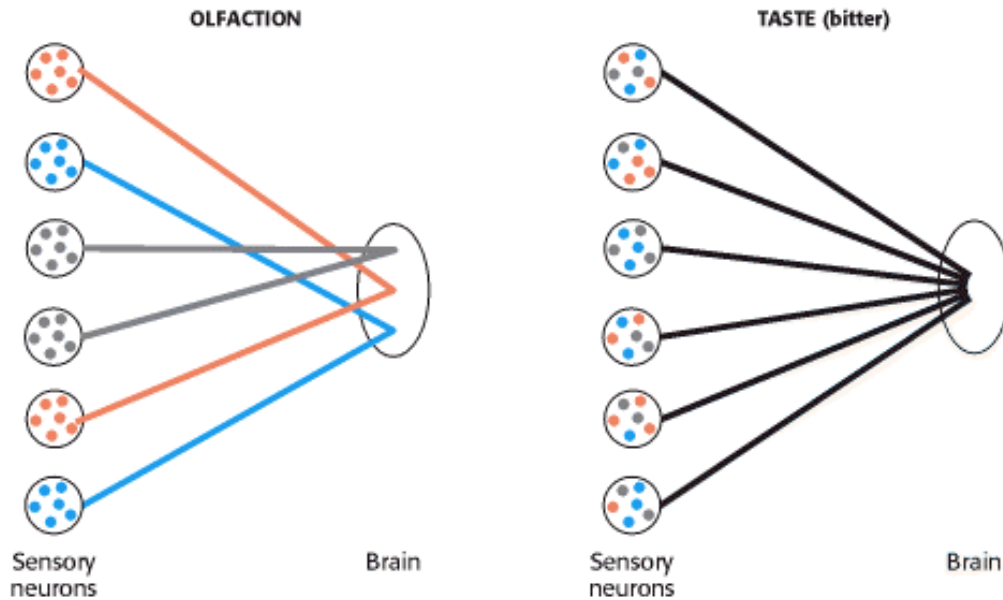


Figure 32.16. Differing Gene Expression and Connection Patterns in Olfactory and Bitter Taste Receptors. In olfaction, each neuron expresses a single OR gene, and the neurons expressing the same OR converge to specific sites in the brain, enabling specific perception of different odorants. In gustation, each neuron expresses many bitter receptor genes, so the identity of the tastant is lost in transmission.

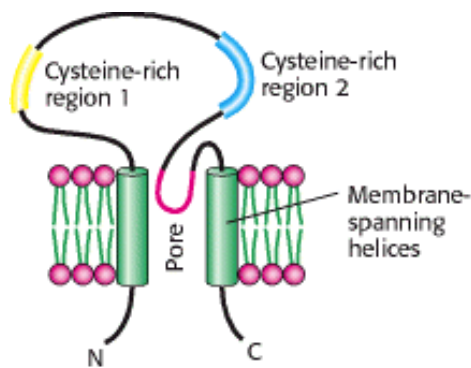


Figure 32.17. Schematic Structure of the Amiloride-Sensitive Sodium Channel. Only one of the four subunits that constitute the functional channel is illustrated. The amiloride-sensitive sodium channel belongs to a superfamily having common structural features, including two hydrophobic membrane-spanning regions, intracellular amino and carboxyl termini; and a large, extracellular region with conserved cysteine-rich domains.

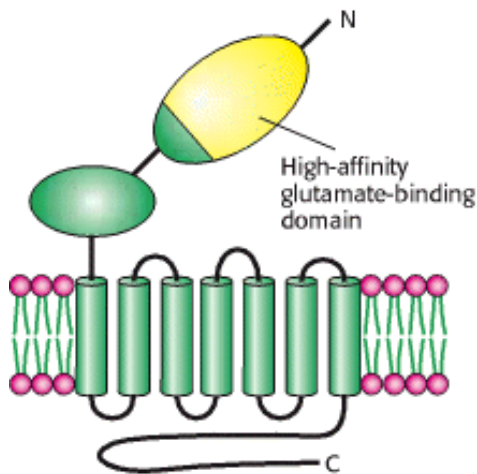


Figure 32.18. Schematic Structure of a Metabotropic Glutamate Receptor. The umami receptor is a variant of a brain glutamate receptor. A substantial part of the high-affinity glutamate-binding domain (shown in yellow) is missing in the form expressed in the tongue.

32.3. Photoreceptor Molecules in the Eye Detect Visible Light

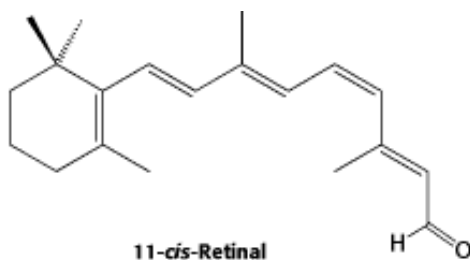
Vision is based on the absorption of light by photoreceptor cells in the eye. These cells are sensitive to light in a relatively narrow region of the electromagnetic spectrum, the region with wavelengths between 300 and 850 nm (Figure 32.19). Vertebrates have two kinds of photoreceptor cells, called *rods* and *cones* because of their distinctive shapes. Cones function in bright light and are responsible for color vision, whereas rods function in dim light but do not perceive color. A human retina contains about 3 million cones and 100 million rods. Remarkably, a rod cell can respond to a single photon, and the brain requires fewer than 10 such responses to register the sensation of a flash of light.

32.3.1. Rhodopsin, a Specialized 7TM Receptor, Absorbs Visible Light



Structural Insights, Rhodopsin: A G Protein Coupled 7TM Receptor offers a more detailed look at rhodopsin structure and function (Figure 32.5).

Rods are slender elongated structures; the outer segment is specialized for photoreception (Figure 32.20). It contains a stack of about 1000 discs, which are membrane-enclosed sacs densely packed with photoreceptor molecules. The photosensitive molecule is often called a *visual pigment* because it is highly colored owing to its ability to absorb light. The photoreceptor molecule in rods is *rhodopsin* (Section 15.1), which consists of the protein *opsin* linked to *11-cis-retinal*, a prosthetic group.



Rhodopsin absorbs light very efficiently in the middle of the visible spectrum, its absorption being centered on 500 nm, which nicely matches the solar output (Figure 32.21). A rhodopsin molecule will absorb a high percentage of the photons of the correct wavelength that strike it, as indicated by the extinction coefficient of $40,000 \text{ M}^{-1}\text{cm}^{-1}$ at 500 nm. The extinction coefficient for rhodopsin is more than an order of magnitude greater than that for tryptophan, the most efficient absorber in proteins that lack prosthetic groups.

Opsin, the protein component of rhodopsin, is a member of the 7TM receptor family. Indeed, rhodopsin was the first member of this family to be purified, its gene was the first to be cloned and sequenced, and its three-dimensional structure was the first to be determined. The color of rhodopsin and its responsiveness to light depend on the presence of the light-absorbing group (*chromophore*) 11-*cis*-retinal. This compound is a powerful absorber of light because it is a polyene; its six alternating single and double bonds constitute a long, unsaturated electron network. Recall that alternating single and double bonds account for the chromophoric properties of chlorophyll (Section 19.2). The aldehyde group of 11-*cis*-retinal forms a Schiff base (Figure 32.22) with the ϵ -amino group of lysine residue 296, which lies in the center of the seventh transmembrane helix. Free retinal absorbs maximally at 370 nm, and its unprotonated Schiff-base adduct absorbs at 380 nm, whereas the protonated Schiff base absorbs at 440 nm or longer wavelengths. Thus, *the 500-nm absorption maximum for rhodopsin strongly suggests that the Schiff base is protonated*; additional interactions with opsin shift the absorption maximum farther toward the red. The positive charge of the protonated Schiff base is compensated by the negative charge of glutamate 113 located in helix 2; the glutamate residue closely approaches the lysine-retinal linkage in the three-dimensional structure of rhodopsin.

32.3.2. Light Absorption Induces a Specific Isomerization of Bound 11-*cis*-Retinal

How does the absorption of light by the retinal Schiff base generate a signal? George Wald and his coworkers discovered that *light absorption results in the isomerization of the 11-*cis*-retinal group of rhodopsin to its all-*trans* form* (Figure 32.23). This isomerization causes the Schiff-base nitrogen atom to move approximately 5 Å, assuming that the cyclohexane ring of the retinal group remains fixed. In essence, *the light energy of a photon is converted into atomic motion*. The change in atomic positions, like the binding of a ligand to other 7TM receptors, sets in train a series of events that lead to the closing of ion channels and the generation of a nerve impulse.

The isomerization of the retinal Schiff base takes place within a few picoseconds of a photon being absorbed. The initial product, termed *bathorhodopsin*, contains a strained all-*trans*-retinal group. Within approximately 1 millisecond, this intermediate is converted through several additional intermediates into *metarhodopsin II*. In *metarhodopsin II*, the Schiff base is deprotonated and the opsin protein has undergone significant reorganization.

Metarhodopsin II (also referred to as R*) is analogous to the ligand-bound state of 7TM receptors such as the β_2 -adrenergic receptor (Section 15.1) and the odorant and tastant receptors heretofore discussed (Figure 32.24). Like these receptors, this form of rhodopsin activates a heterotrimeric G protein that propagates the signal. The G protein associated with rhodopsin is called *transducin*. Metarhodopsin II triggers the exchange of GDP for GTP by the α subunit of transducin (Figure 32.25). On the binding of GTP, the β γ subunits of transducin are released and the α subunit switches on a *cGMP phosphodiesterase* by binding to and removing an inhibitory subunit. The activated phosphodiesterase is a potent enzyme that rapidly hydrolyzes cGMP to GMP. The reduction in cGMP concentration causes cGMP-gated ion channels to close, leading to hyperpolarization of the membrane and neuronal signaling. *At each step in this process, the initial signal—the absorption of a single photon—is amplified so that it leads to sufficient membrane hyperpolarization to result in signaling.*



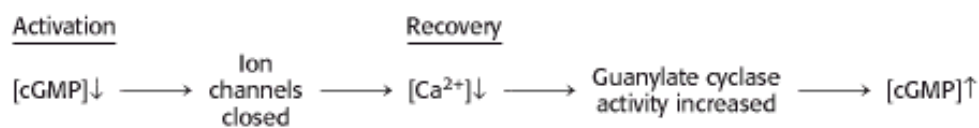
Conceptual Insights, Signaling Pathways: Response and Recovery presents an animated version of Figure 32.25 and a comparison to olfactory signal transduction (Figure 32.5).

32.3.3. Light-Induced Lowering of the Calcium Level Coordinates Recovery

As we have seen, the visual system responds to changes in light and color within a few milliseconds, quickly enough that we are able to perceive continuous motion at nearly 1000 frames per second. To achieve a rapid response, the signal must also be terminated rapidly and the system must be returned to its initial state. First, activated rhodopsin must be blocked from continuing to activate transducin. *Rhodopsin kinase* catalyzes the phosphorylation of the carboxyl terminus of R* at multiple serine and threonine residues. *Arrestin*, an inhibitory protein (Section 15.1.4), then binds phosphorylated R* and prevents additional interaction with transducin.

Second, the α subunit of transducin must be returned to its inactive state to prevent further signaling. Like other G proteins, the α subunit possesses built-in GTPase activity that hydrolyzes bound GTP to GDP. Hydrolysis takes place in less than a second when transducin is bound to the phosphodiesterase. The GDP form of transducin then leaves the phosphodiesterase and reassociates with the $\beta \gamma$ subunits, and the phosphodiesterase returns to its inactive state. Third, the level of cGMP must be raised to reopen the cGMP-gated ion channels. *The action of guanylate cyclase accomplishes this third step by synthesizing cGMP from GTP.*

Calcium ion plays an essential role in controlling guanylate cyclase because it markedly inhibits the activity of the enzyme. In the dark, Ca^{2+} as well as Na^+ enter the rod outer segment through the cGMP-gated channels. Calcium ion influx is balanced by its efflux through an exchanger, a transport system that uses the thermodynamically favorable flow of four Na^+ ions into the cell and one K^+ ion out of the cell to extrude one Ca^{2+} ion. After illumination, the entry of Ca^{2+} through the cGMP-gated channels stops, but its export through the exchanger continues. Thus, the cytosolic Ca^{2+} level drops from 500 nM to 50 nM after illumination. This drop markedly stimulates guanylate cyclase, rapidly restoring the concentration of cGMP to reopen the cGMP-gated channels.



By controlling the rate of cGMP synthesis, Ca^{2+} levels govern the speed with which the system is restored to its initial state.

32.3.4. Color Vision Is Mediated by Three Cone Receptors That Are Homologs of Rhodopsin




Structural Insights, Rhodopsin: A G Protein Coupled 7TM Receptor

explores the structural basis of color vision and night blindness in more detail.

Cone cells, like rod cells, contain visual pigments. Like rhodopsin, these photoreceptor proteins are members of the 7TM receptor family and utilize 11-*cis*-retinal as their chromophore. In human cone cells, there are three distinct photoreceptor proteins with absorption maxima at 426, 530, and ~ 560 nm (Figure 32.26). *These absorbances correspond to (in fact, define) the blue, green, and red regions of the spectrum.* Recall that the absorption maximum for rhodopsin is 500 nm.


The amino acid sequences of the cone photoreceptors have been compared with each other and with rhodopsin. The result is striking. Each of the cone photoreceptors is approximately 40% identical in sequence with rhodopsin. Similarly, the blue photoreceptor is 40% identical with each of the green and red photoreceptors. The green and red photoreceptors, however, are > 95% identical with each other, differing in only 15 of 364 positions (Figure 32.27).

 These observations are sources of insight into photoreceptor evolution. First, the green and red photoreceptors are clearly products of a recent evolutionary event (Figure 32.28). The green and red pigments appear to have diverged in the primate lineage approximately 35 million years ago. Mammals, such as dogs and mice, that diverged from primates earlier have only two cone photoreceptors, blue and green. They are not sensitive to light as far toward the infrared region as we are, and they do not discriminate colors as well. In contrast, birds such as chickens have a total of six pigments: rhodopsin, four cone pigments, and a pineal visual pigment called *pinopsin*. Birds have highly acute color perception.

Second, the high level of similarity between the green and red pigments has made it possible to identify the specific

amino acid residues that are responsible for spectral tuning. Three residues (at positions 180, 277, and 285) are responsible for most of the difference between the green and red pigments. In the green pigment, these residues are alanine, phenylalanine, and alanine, respectively; in the red pigment, they are serine, tyrosine, and threonine. A hydroxyl group has been added to each amino acid in the red pigment. The hydroxyl groups can interact with the photoexcited state of retinal and lower its energy, leading to a shift toward the lower-energy (red) region of the spectrum.

32.3.5. Rearrangements in the Genes for the Green and Red Pigments Lead to "Color Blindness"

 The genes for the green and red pigments lie adjacent to each other on the human X chromosome. These genes are more than 98% identical in nucleotide sequence, including introns and untranslated regions as well as the protein-coding region. Regions with such high similarity are very susceptible to unequal homologous recombination.

Homologous recombination

The exchange of DNA segments at equivalent positions between chromosomes with substantial sequence similarity.

Recombination can take place either between or within transcribed regions of the gene (Figure 32.29). If recombination takes place between transcribed regions, the product chromosomes will differ in the number of pigment genes that they carry. One chromosome will lose a gene and thus may lack the gene for, say, the green pigment; the other chromosome will gain a gene. Consistent with this scenario, approximately 2% of human X chromosomes carry only a single color pigment gene, approximately 20% carry two, 50% carry three, 20% carry four, and 5% carry five or more. A person lacking the gene for the green pigment will have trouble distinguishing red and green color, characteristic of the most common form of color blindness. Approximately 5% of males have this form of color blindness. Recombination can also take place within the transcription units, resulting in genes that encode hybrids of the green and red photoreceptors. The absorption maximum of such a hybrid lies between that of the red and green pigments. A person with such hybrid genes who also lacks either a functional red or a functional green pigment gene does not discriminate color well.

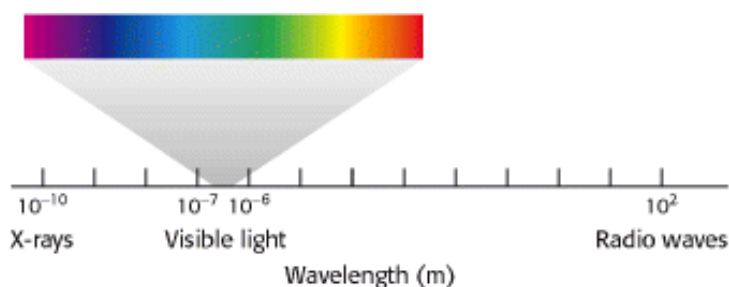


Figure 32.19. The Electromagnetic Spectrum. Visible light has wavelengths between 300 and 850 nanometers.

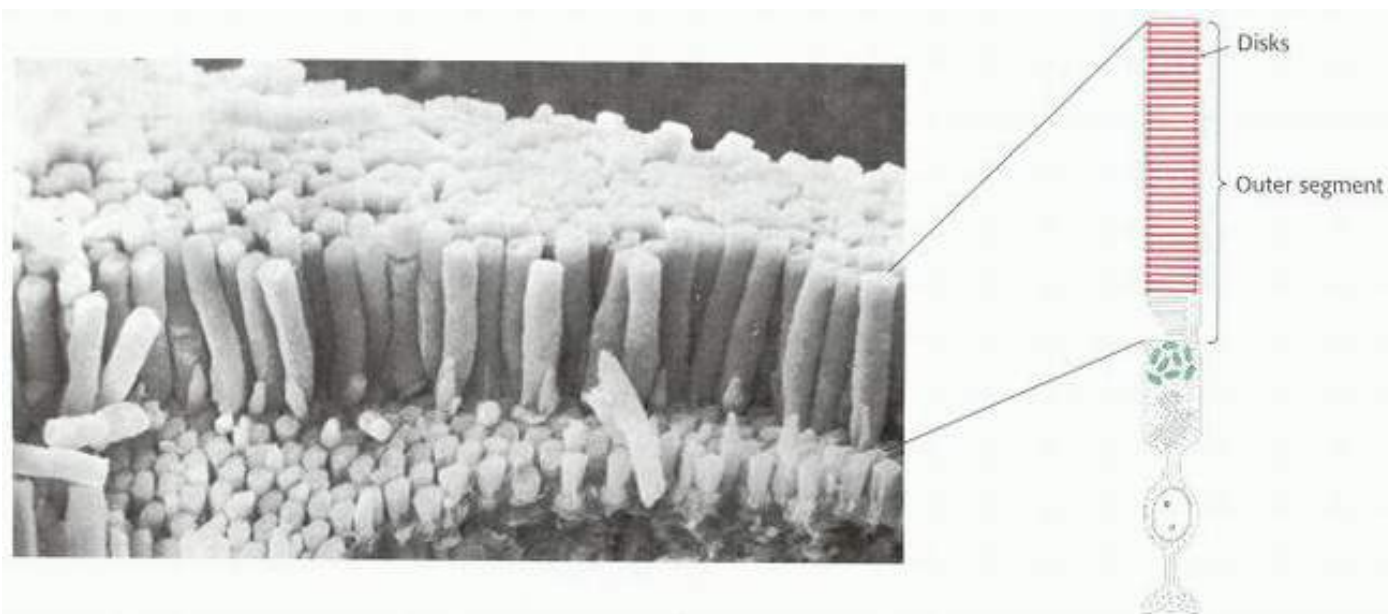


Figure 32.20. The Rod Cell. (Left) Scanning electron micrograph of retinal rod cells. (Right) Schematic representation of a rod cell. [Photograph courtesy of Dr. Deric Bownds.]

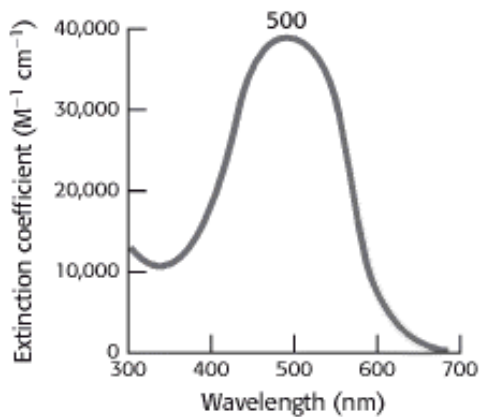


Figure 32.21. Rhodopsin Absorption Spectrum.

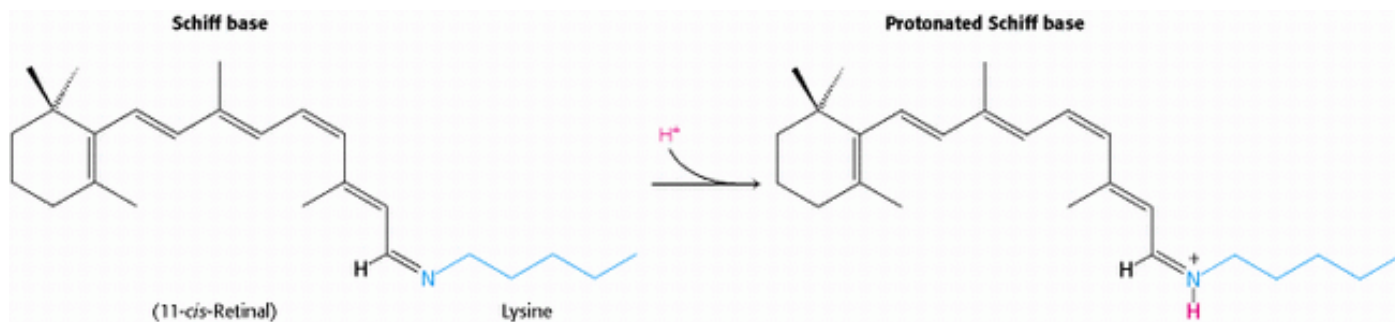


Figure 32.22. Retinal-Lysine Linkage. Retinal is linked to lysine 296 in opsin by a Schiff-base linkage. In the resting state of rhodopsin, this Schiff base is protonated.

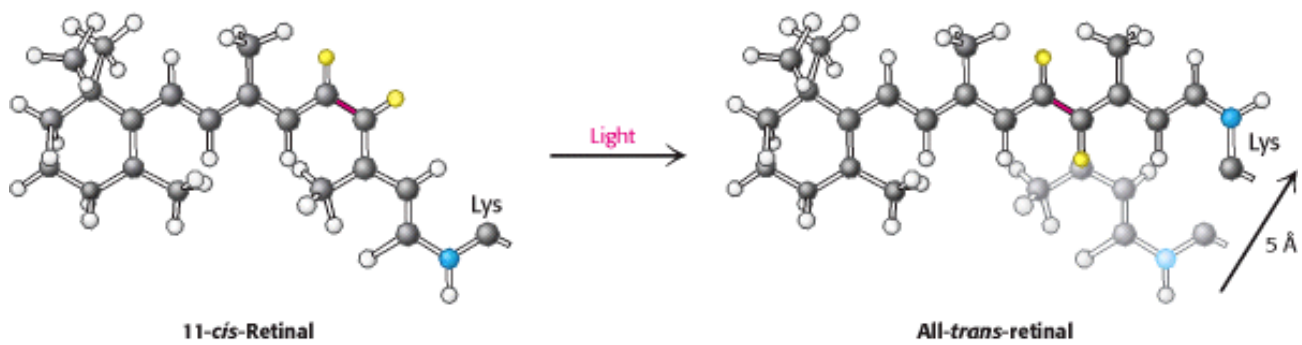


Figure 32.23. Atomic Motion in Retinal. The Schiff-base nitrogen atom moves 5 Å as a consequence of the light-induced isomerization of 11-*cis*-retinal to all-*trans*-retinal by rotation about the bond shown in red.

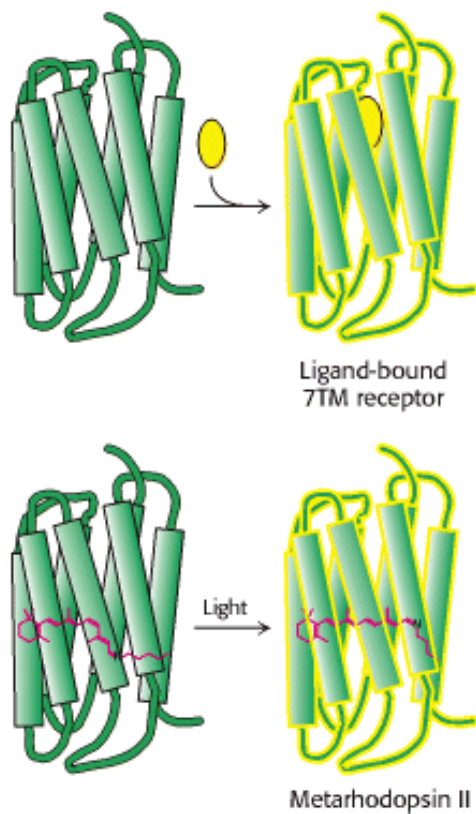


Figure 32.24. Analogous 7TM Receptors. The conversion of rhodopsin into metarhodopsin II activates a signal-transduction pathway analogously to the activation induced by the binding of other 7TM receptors to appropriate ligands.

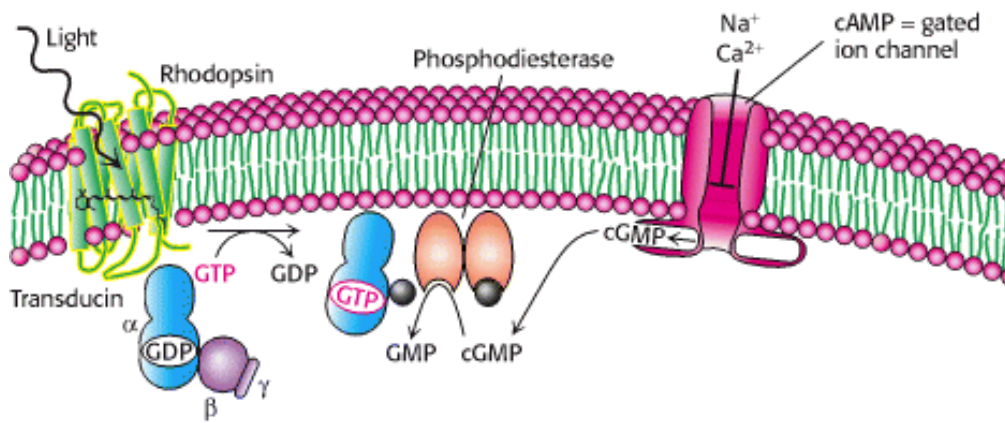


Figure 32.25. Visual Signal Transduction. The light-induced activation of rhodopsin leads to the hydrolysis of cGMP, which in turn leads to ion channel closing and the initiation of an action potential.

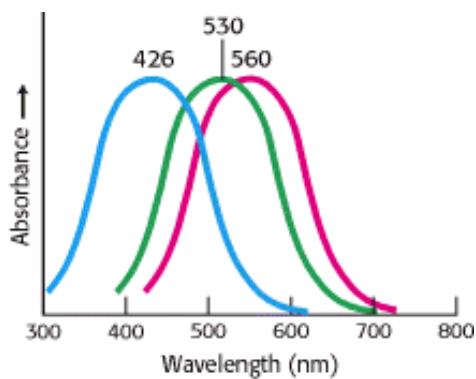


Figure 32.26. Cone-Pigment Absorption Spectra. The absorption spectra of the cone visual pigment responsible for color vision.

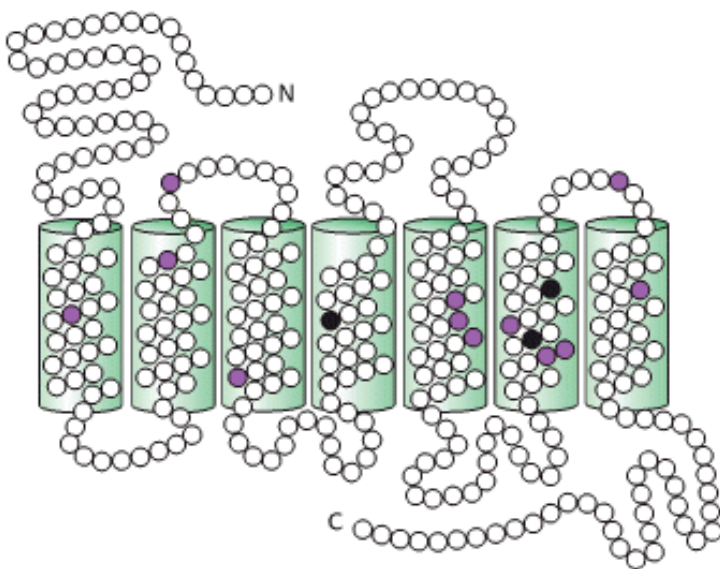


Figure 32.27. Comparison of the Amino Acid Sequences of the Green and Red Photoreceptors. Open circles correspond to identical residues, whereas colored circles mark residues that are different. The differences in the three black positions are responsible for most of the difference in their absorption spectra.

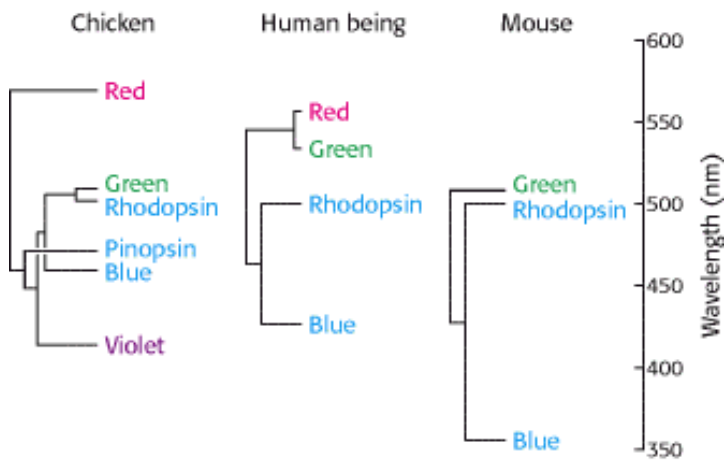


Figure 32.28. Evolutionary Relationships among Visual Pigments. Visual pigments have evolved by gene duplication along different branches of the animal evolutionary tree. The branch lengths of the "trees" correspond to the percentage of amino acid divergence. [Adapted from Nathans, J. *Neuron* 24(1999):299; by permission of Cell Press.]

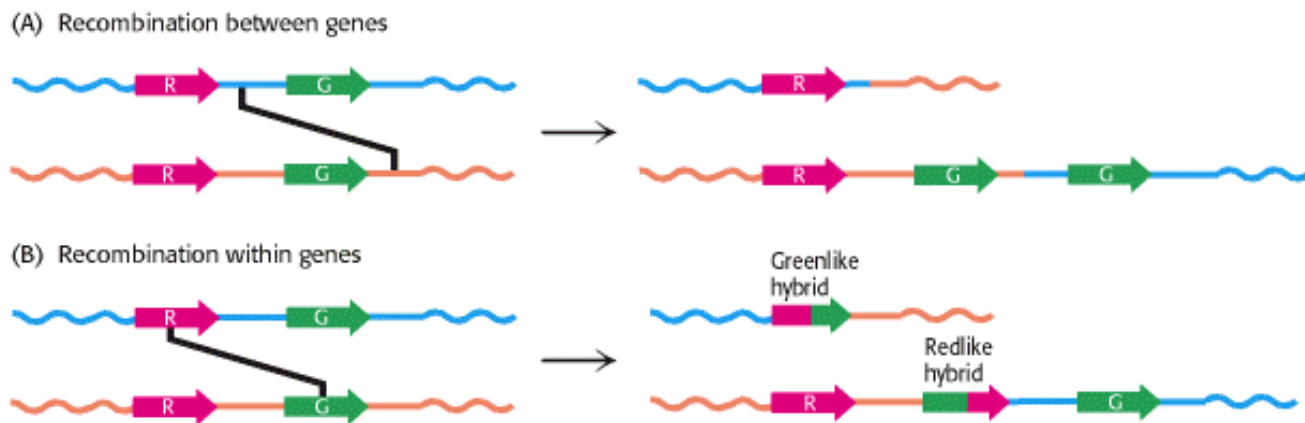


Figure 32.29. Recombination Pathways Leading to Color Blindness. Rearrangements in the course of DNA replication may lead to (A) the loss of visual pigment genes or (B) the formation of hybrid pigment genes that encode photoreceptors with anomalous absorption spectra. Because the amino acids most important for determining absorption spectra are in the carboxyl-terminal half of each photoreceptor protein, the part of the gene that encodes this region most strongly affects the absorption characteristics of hybrid receptors. [Adapted from J. Nathans. *Neuron* 24(1999):299–312; by permission of Cell Press.]

32.4. Hearing Depends on the Speedy Detection of Mechanical Stimuli

Hearing and touch are based on the detection of mechanical stimuli. Although the proteins of these senses have not been as well characterized as those of the senses already discussed, anatomical, physiological, and biophysical studies have elucidated the fundamental processes. *A major clue to the mechanism of hearing is its speed.* We hear frequencies ranging from 200 to 20,000 Hz (cycles per second), corresponding to times of 5 to 0.05 ms. Furthermore, our ability to locate sound sources, one of the most important functions of hearing, depends on the ability to detect the time delay between the arrival of a sound at one ear and its arrival at the other. Given the separation of our ears and the speed of sound, we must be able to accurately sense time differences of 0.7 ms. In fact, human beings can locate sound sources associated with temporal delays as short as 0.02 ms. This high time resolution implies that hearing must employ direct transduction mechanisms that do not depend on second messengers. Recall that, in vision, for which speed also is important, the signal-transduction processes take place in milliseconds.

32.4.1. Hair Cells Use a Connected Bundle of Stereocilia to Detect Tiny Motions

Sound waves are detected inside the cochlea of the inner ear. The *cochlea* is a fluid-filled, membranous sac that is coiled like a snail shell. The primary detection is accomplished by specialized neurons inside the cochlea called *hair cells* (Figure 32.30). Each cochlea contains approximately 16,000 hair cells, and each hair cell contains a hexagonally shaped bundle of 20 to 300 hairlike projections called *stereocilia* (Figure 32.31). These stereocilia are graded in length across the bundle. Mechanical deflection of the hair bundle, as occurs when a sound wave arrives at the ear, creates a change in the membrane potential of the hair cell.

Micromanipulation experiments have directly probed the connection between mechanical stimulation and membrane potential. Displacement toward the direction of the tallest part of the hair bundle results in depolarization of the hair cell, whereas displacement in the opposite direction results in hyperpolarization (Figure 32.32). Motion perpendicular to the hair-length gradient does not produce any change in resting potential. Remarkably, *displacement of the hair bundle by as little as 3 Å (0.3 nm) results in a measurable (and functionally important) change in membrane potential*. This motion of 0.003 degree corresponds to a 1-inch movement of the top of the Empire State Building.

How does the motion of the hair bundle create a change in membrane potential? The rapid response, within microseconds, suggests that the movement of the hair bundle acts on ion channels directly. An important observation is that adjacent stereocilia are linked by individual filaments called *tip links* (Figure 32.33).

The presence of these tip links suggests a simple mechanical model for transduction by hair cells (Figure 32.34). The tip links are coupled to ion channels in the membranes of the stereocilia that are gated by mechanical stress. In the absence of a stimulus, approximately 15% of these channels are open. When the hair bundle is displaced toward its tallest part, the stereocilia slide across one another and the tension on the tip links increases, causing additional channels to open. The flow of ions through the newly opened channels depolarizes the membrane. Conversely, if the displacement is in the opposite direction, the tension on the tip links decreases, the open channels close, and the membrane hyperpolarizes. *Thus, the mechanical motion of the hair bundle is directly converted into current flow across the hair-cell membrane.*

32.4.2. Mechanosensory Channels Have Been Identified in *Drosophila* and Bacteria

Although the ion channel that functions in human hearing has not been identified, other mechanosensory channels in other organisms have been. *Drosophila* have sensory bristles used for detecting small air currents. These bristles respond to mechanical displacement in ways similar to those of hair cells; displacement of a bristle in one direction leads to substantial transmembrane current. Strains of mutant fruit flies that show uncoordinated motion and clumsiness have been examined for their electrophysiological responses to displacement of the sensory bristles. In one set of strains, transmembrane currents were dramatically reduced. The mutated gene in these strains was found to encode a protein of 1619 amino acids, called NompC for *no mechanoreceptor potential*.

The carboxyl-terminal 469 amino acids of NompC resemble a class of ion channel proteins called TRP (*transient receptor potential*) channels. This region includes six putative transmembrane helices with a porelike region between the fifth and sixth helices. The amino-terminal 1150 amino acids consist almost exclusively of 29 *ankyrin repeats* (Figure 32.35). Ankyrin repeats are structural motifs formed by 33 amino acids folded into a hairpin loop followed by a helix-turn-helix. Importantly, in other proteins, regions with tandem arrays of these motifs mediate protein-protein interactions, suggesting that these arrays couple the motions of other proteins to the activity of the NompC channel.

Prokaryotes such as *E. coli* have ion channels in their membranes that open in response to mechanical changes. These channels play a role in regulating the osmotic pressure within the bacteria. The three-dimensional structure of one such channel, that from *Mycobacterium tuberculosis*, has been determined. The channel is constructed of five identical subunits arranged such that an alpha helix from each subunit lines the inner surface of the pore. Further studies should reveal whether the transduction channel in hearing is homologous to either of these classes of mechanosensory channels or represents a novel class.

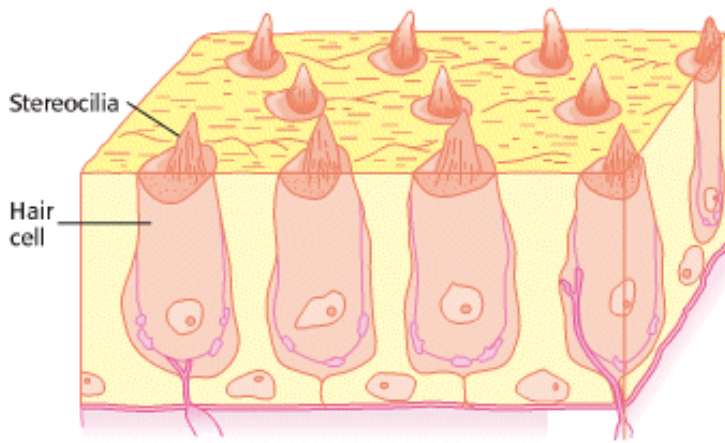


Figure 32.30. Hair Cells, the Sensory Neurons Crucial for Hearing. [Adapted from Hudspeth, A. J. *Nature* 341 (1989):397.]

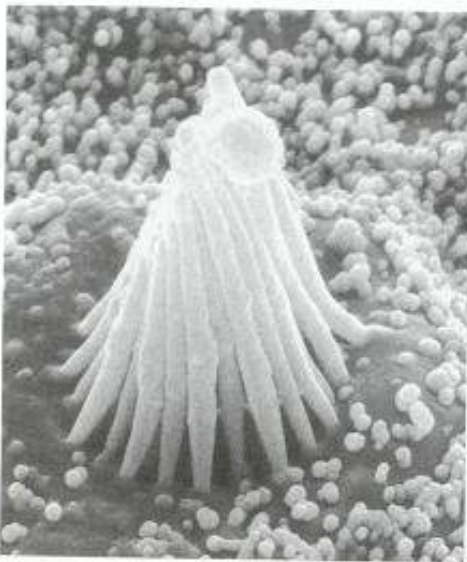


Figure 32.31. An Electron Micrograph of a Hair Bundle. [Courtesy of A. Jacobs and A. J. Hudspeth.]

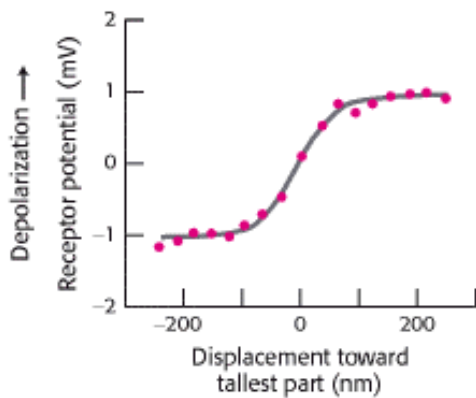


Figure 32.32. Micromanipulation of a Hair Cell. Movement toward the tallest part of the bundle depolarizes the cell as measured by the microelectrode. Movement toward the shortest part hyperpolarizes the cell. Lateral movement has no effect. [Adapted from Hudspeth, A. J. *Nature* 341(1989):397.]



Figure 32.33. Electron Micrograph of Tip Links. The tip link between two hair fibers is marked by an arrow. [Courtesy of A. Jacobs and A. J. Hudspeth.]

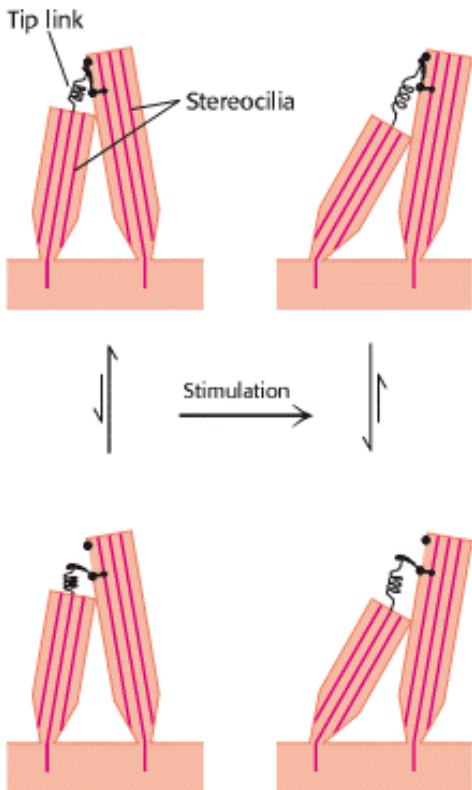


Figure 32.34. Model for Hair-Cell Transduction. When the hair bundle is tipped toward the tallest part, the tip link pulls on and opens an ion channel. Movement in the opposite direction relaxes the tension in the tip link, increasing the probability that any open channels will close. [Adapted from A. J. Hudspeth. *Nature* 341 (1989):397.]



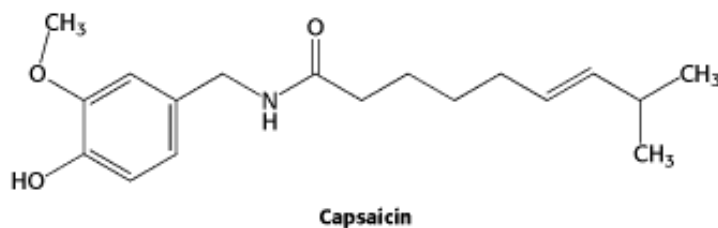
Figure 32.35. Ankyrin Repeat Structure. Four ankyrin repeats are shown with one shown in red. These domains interact with other proteins, primarily through their loops.

32.5. Touch Includes the Sensing of Pressure, Temperature, and Other Factors

Like taste, touch is a combination of sensory systems that are expressed in a common organ—in this case, the skin. The detection of pressure and the detection of temperature are two key components. Amiloride-sensitive sodium channels, homologous to those of taste, appear to play a role. Other systems are responsible for detecting painful stimuli such as high temperature, acid, or certain specific chemicals. Although our understanding of this sensory system is not as advanced as that of the other sensory systems, recent work has revealed a fascinating relation between pain and taste sensation, a relation well known to anyone who has eaten "spicy" food.

32.5.1. Studies of Capsaicin, the Active Ingredient in "Hot" Peppers, Reveal a Receptor for Sensing High Temperatures and Other Painful Stimuli


Our sense of touch is intimately connected with the sensation of pain. Specialized neurons, termed *nociceptors*, transmit signals to pain-processing centers in the spinal cord and brain in response to the onset of tissue damage. What is the molecular basis for the sensation of pain? An intriguing clue came from the realization that *capsaicin*, the chemical responsible for the "hot" taste of spicy food, activates nociceptors.



Early research suggested that capsaicin would act by opening ion channels that are expressed in nociceptors. Thus, a cell that expresses the *capsaicin receptor* should take up calcium on treatment with the molecule. This insight led to the isolation of the capsaicin receptor with the use of cDNA from cells expressing this receptor. Such cells had been detected by their fluorescence when loaded with the calcium-sensitive compound Fura-2 and then treated with capsaicin or related molecules. Cells expressing the capsaicin receptor, which is called VR1 (for vanilloid receptor 1), respond to capsaicin below a concentration of 1 μM . The deduced 838-residue sequence of VR1 revealed it to be a member of the TRP channel family (Figure 32.36). The amino-terminal region of VR1 includes three ankyrin repeats.

Currents through VR1 are also induced by temperatures above 40°C and by exposure to dilute acid, with a midpoint for activation at pH 5.4 (Figure 32.37). Temperatures and acidity in these ranges are associated with infection and cell

injury. The responses to capsaicin, temperature, and acidity are not independent. The response to heat is greater at lower pH, for example. Thus, *VR1 acts to integrate several noxious stimuli*. We feel these responses as pain and act to avoid the potentially destructive conditions that caused the unpleasant sensation. Mice that do not express VR1 suggest that this is the case; such mice do not mind food containing high concentrations of capsaicin and are, indeed, less responsive than control mice to normally noxious heat. Plants such as chili peppers presumably gained the ability to synthesize capsaicin and other "hot" compounds to protect themselves from being consumed by mammals. Birds, which play the beneficial role of spreading pepper seeds into new territory, do not appear to respond to capsaicin.

 Because of its ability to simulate VR1, capsaicin is used in pain management for arthritis, neuralgia, and other neuropathies. How can a compound that induces pain assist in its alleviation? Chronic exposure to capsaicin overstimulates pain-transmitting neurons, leading to their desensitization.

32.5.2. Subtle Sensory Systems Detect Other Environmental Factors Such as Earth's Magnetic Field

In addition to the five primary senses, human beings may have counterparts to less-familiar sensory systems characterized in other organisms. These sensory systems respond to environmental factors other than light, molecular shape, or air motion. For example, some species of bacteria are magnetotactic; that is, they move in directions dictated by Earth's magnetic field (Figure 32.38). In the Northern Hemisphere, Earth's magnetic field points northward but also has a component directed downward, toward Earth's center. Magnetotactic bacteria not only swim northward but also swim downward, away from the surface and the presence of high levels of oxygen, toxic to these bacteria. Remarkably, these bacteria synthesize intracellular chains of small particles containing a magnetic ore called magnetite (Fe_3O_4) that run through the center of each bacterium. Such chains are called *magnetosomes*. The magnetic force exerted by these particles is sufficiently strong in relation to the size of the bacterium that it causes the bacterium to become passively aligned with Earth's magnetic field. Intriguingly, similar magnetite particles have been detected in the brains of birds, fish, and even human beings, although their role in sensing magnetic fields has not yet been established.

There may exist other subtle senses that are able to detect environmental signals that then influence our behavior. The biochemical basis of these senses is now under investigation. One such sense is our ability to respond, often without our awareness, to chemical signals called pheromones, released by other persons. Another is our sense of time, manifested in our daily (circadian) rhythms of activity and restfulness. Daily changes in light exposure strongly influence these rhythms. The foundations for these senses have been uncovered in other organisms; future studies should reveal to what extent these mechanisms apply to human beings as well.

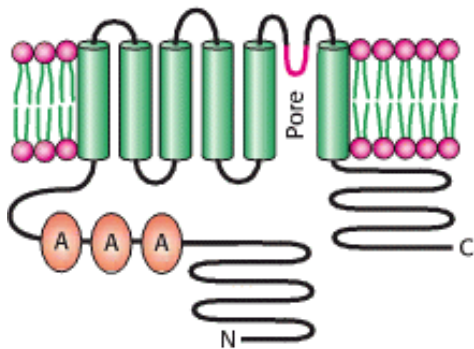


Figure 32.36. The Membrane Topology Deduced for VR1, the Capsaicin Receptor. The proposed site of the membrane pore is indicated in red, and the three ankyrin (A) repeats are shown in orange. The active receptor comprises four of these subunits. [Adapted from Caterina, M. J., Schumacher, M. A., Tominaga, M., Rosen, T. A., Levine, J. D., and Julius, D. *Nature* 389 (1997):816.]

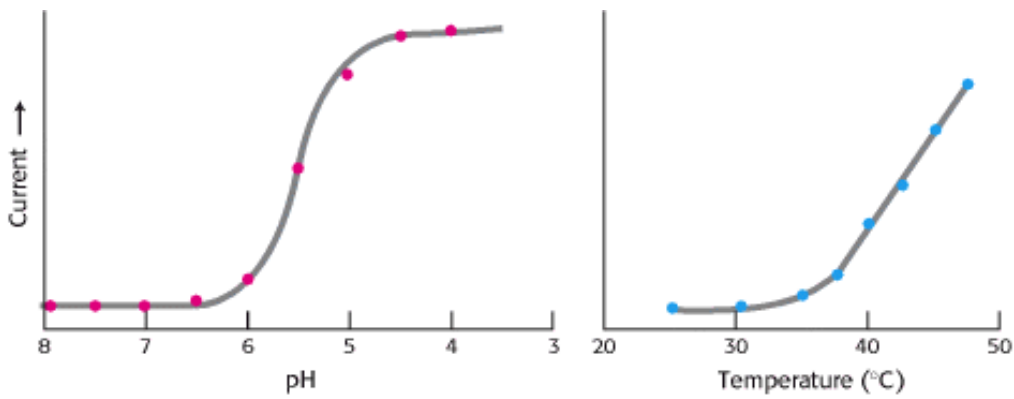


Figure 32.37. Response of the Capsaicin Receptor to pH and Temperature. [Adapted from Tominaga, M., Caterina, M. J., Malmberg, A. B., Rosen, T. A., Gilbert, H., Skinner, K., Raumann, B. E., Basbaum, A. I., and Julius, D. *Neuron* 21 (1998):531.]

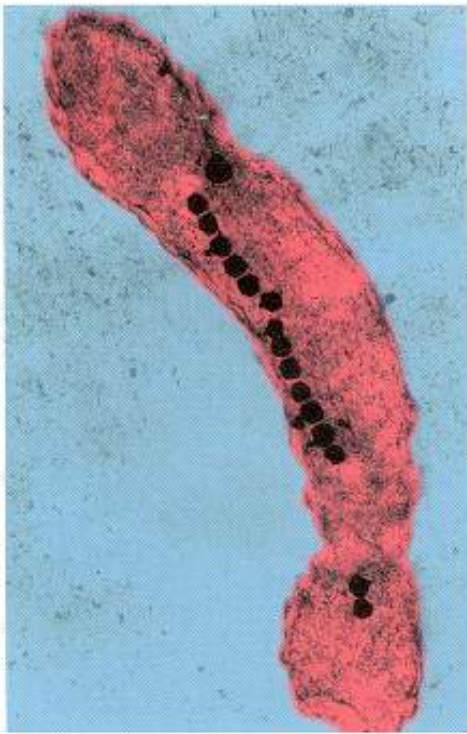


Figure 32.38. Magnetotactic Bacterium. The magnetosome, visible as a chain of opaque membrane-bound magnetite crystals, acts as a compass to orient the bacteria with the earth's magnetic field. The bacterium is artificially colored. [Courtesy of Richard B. Frankel, California Polytechnic State University, San Luis Obispo, California.]

Summary

Smell, Taste, Vision, Hearing, and Touch Are Based on Signal-Transduction Pathways Activated by Signals from the Environment.

These sensory systems function similarly to the signal-transduction pathways for many hormones. These intercellular signaling pathways appear to have been appropriated and modified to process environmental information.

A Wide Variety of Organic Compounds Are Detected by Olfaction

The sense of smell, or olfaction, is remarkable in its specificity—it can, for example, discern stereoisomers of small organic compounds as distinct aromas. The 7TM receptors that detect these odorants operate in conjunction with $G_{(olf)}$, a G protein that activates a cAMP cascade resulting in the opening of an ion channel and the generation of a nerve impulse. An outstanding feature of the olfactory system is its ability to detect a vast array of odorants. Each olfactory neuron expresses only one type of receptor and connects to a particular region of the olfactory bulb. Odors are decoded by a combinatorial mechanism—each odorant activates a number of receptors, each to a different extent, and most receptors are activated by more than one odorant.

Taste Is a Combination of Senses That Function by Different Mechanisms

We can detect only five tastes: bitter, sweet, salt, sour, and umami. The transduction pathways that detect taste are, however, diverse. Bitter and sweet tastants are experienced through 7TM receptors acting through a special G protein called gustducin. Salty and sour tastants act directly through membrane channels. Salt tastants are detected by passage through sodium channels, whereas sour taste results from the effects of hydrogen ions on a number of types of channels. The end point is the same in all cases—membrane polarization that results in the transmission of a nerve impulse. Umami, the taste of glutamate, is detected by a receptor that is a modified form of a brain receptor that responds to glutamate as a neurotransmitter rather than as a tastant.

Photoreceptor Molecules in the Eye Detect Visible Light

Vision is perhaps the best understood of the senses. Two classes of photoreceptor cells exist: cones, which respond to bright lights and colors, and rods, which respond only to dim light. The photoreceptor in rods is rhodopsin, a 7TM receptor that is a complex of the protein opsin and the chromophore 11-*cis*-retinal. Absorption of light by 11-*cis*-retinal changes its structure into that of all-*trans*-retinal, setting in motion a signal-transduction pathway that leads to the breakdown of cGMP, to membrane hyperpolarization, and to a subsequent nerve impulse. Color vision is mediated by three distinct 7TM photoreceptors that employ 11-*cis*-retinal as a chromophore and absorb light in the blue, green, and red parts of the spectrum.

Hearing Depends on the Speedy Detection of Mechanical Stimuli

The immediate receptors for hearing are found in the hair cells of the cochleae, which contain bundles of stereocilia. When the stereocilia move in response to sound waves, cation channels will open or close, depending on the direction of movement. The mechanical motion of the cilia is converted into current flow and then into a nerve impulse.

Touch Includes the Sensing of Pressure, Temperature, and Other Factors

Touch, detected by the skin, senses pressure, temperature, and pain. Specialized nerve cells called nociceptors transmit signals that are interpreted in the brain as pain. A receptor responsible for the perception of pain has been isolated on the

basis of its ability to bind capsaicin, the molecule responsible for the hot taste of spicy food. The capsaicin receptor, also called VR1, functions as a cation channel that initiates a nerve impulse.

Key Terms

main olfactory epithelium

$G_{(olf)}$

functional magnetic resonance imaging (fMRI)

gustducin

amiloride-sensitive sodium channel

metabotropic glutamate receptor

rod

cone

rhodopsin

opsin

retinal

chromophore

transducin

cGMP phosphodiesterase

cGMP-gated calcium channel

rhodopsin kinase

arrestin

guanylate cyclase

hair cell

stereocilium

tip link

nociceptor

capsaicin receptor

Problems

1. *Of mice and rats.* As noted in [Section 32.1.2](#), one of the first odorant receptors to be matched with its ligand was a rat receptor that responded best to *n*-octanal. The sequence of the corresponding mouse receptor differed from the rat receptor at 15 positions. Surprisingly, the mouse receptor was found to respond best to *n*-heptanal rather than *n*-octanal. The substitution of isoleucine at position 206 in the mouse for valine at this position in the rat receptor was found to be important in determining the specificity for *n*-heptanal. Propose an explanation.

[See answer](#)

2. *Olfaction in worms.* Unlike the olfactory neurons in the mammalian systems discussed herein, olfactory neurons in the nematode *C. elegans* express multiple olfactory receptors. In particular, one neuron (called AWA) expresses receptors for compounds to which the nematode is attracted, whereas a different neuron (called AWB) expresses receptors for compounds that the nematode avoids. Suppose that a transgenic nematode is generated such that one of the receptors for an attractant is expressed in AWB rather than AWA. What behavior would you expect in the presence of the corresponding attractant?

[See answer](#)

3. *Odorant matching.* A mixture of two of the compounds illustrated in [Figure 32.6](#) is applied to a section of olfactory epithelium. Only receptors 3, 5, 9, 12, and 13 are activated, according to [Figure 32.7](#). Identify the likely compounds in the mixture.

[See answer](#)

4. *Timing.* Compare the aspects of taste (bitter, sweet, salty, sour) in regard to their potential for rapid time resolution.

[See answer](#)

5. *Two ears.* Our ability to determine the direction from which a sound is coming is partly based on the difference in time at which our two ears detect the sound. Given the speed of sound (350 meter/second) and the separation between our ears (0.15 meter), what difference is expected in the times at which a sound arrives at our two ears? How does this difference compare with the time resolution of the human hearing system? Would a sensory system that utilized 7TM receptors and G proteins be capable of adequate time resolution?

[See answer](#)

6. *Constitutive mutants.* What effect within the olfactory system would you expect for a mutant in which adenylyl cyclase is always fully active? What effect within the visual system would you expect for a mutant in which guanylyl cyclase is always fully active?

[See answer](#)

7. *Bottle choice.* A widely used method for quantitatively monitoring rodent behavior with regard to taste is the bottle-choice assay. An animal is placed in a cage with two water bottles, one of which contains a potential tastant. After a fixed period of time (24–48 hours), the amount of water remaining in each bottle is measured. Suppose that much less water remains in the bottle with the tastant after 48 hours. Do you suspect the tastant to be sweet or bitter?

[See answer](#)

8. *It's better to be bitter.* Some nontoxic plants taste very bitter to us. Suggest one or more explanations.

See answer

9. *Unexpected consequences.* Sildenafil (Viagra) is a drug widely used to treat male impotence. Sildenafil exerts its effect by inhibiting a cGMP phosphodiesterase isozyme (PDE5) that is especially prevalent in smooth muscle. Interestingly, certain airlines restrict pilots from flying for 24 hours after using sildenafil. Suggest a reason for this restriction.

See answer

Chapter Integration Problem

10. *Energy and information.* The transmission of sensory information requires the input of free energy. For each sensory system (olfaction, gustation, vision, hearing, and touch), identify mechanisms for the input of free energy that allow the transmission of sensory information.


See answer

Mechanism Problem

11. *Schiff-base formation.* Propose a mechanism for the reaction between opsin and 11-*cis*-retinal.

See answer

Media Problems

12.  *Homologous proteins, analogous binding?* Odorants bind to 7TM receptors, but where they bind (and whether all bind in the same way) is unclear. Odorants might, for example, bind on the extracellular surface, or, like retinal, they might bind in the interior of the transmembrane region. Problem 1 of this chapter presents evidence for the direct involvement of residue 206 in odorant binding in receptors from mouse and rat. While these receptors' structures are not known in detail, their sequences are similar enough to rhodopsin's that the rhodopsin structure can be used to infer the approximate location of residue 206. To see the likely location, look in the **Structural Insights** module on rhodopsin. Where do you think the mouse and rat receptors bind their odorants?
13. *Deodorant?* A cAMP phosphodiesterase has been discovered that is found predominantly in olfactory sensory neurons (Yan et al., 1995, Proc. Natl. Acad. Sci. 10:9677). The enzyme is activated by Ca^{2+} . What do you think this enzyme does, and why do you think it is regulated by calcium? (Hint: Study the response and recovery animations in the **Conceptual Insights** module on signaling pathways.)

Selected Readings

Where to start

R. Axel. 1995. The molecular logic of smell *Sci. Am.* 273: (4) 154-159. ([PubMed](#))

C. Dulac. 2000. The physiology of taste, vintage 2000 *Cell* 100: 607-610. ([PubMed](#))

L. Stryer. 1996. Vision: From photon to perception *Proc. Natl. Acad. Sci. U. S. A.* 93: 557-559. ([PubMed](#)) ([Full Text in PMC](#))

A.J. Hudspeth. 1989. How the ear's works work *Nature* 341: 397-404. ([PubMed](#))

Olfaction

L. Buck and R. Axel. 1991. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition *Cell* 65: 175-187. ([PubMed](#))

B. Malnic, J. Hirono, T. Sato, and L.B. Buck. 1999. Combinatorial receptor codes for odors *Cell* 96: 713-723. ([PubMed](#))

P. Mombaerts, F. Wang, C. Dulac, S.K. Chao, A. Nemes, M. Mendel-sohn, J. Edmondson, and R. Axel. 1996. Visualizing an olfactory sensory map *Cell* 87: 675-686. ([PubMed](#))

P. Mombaerts. 1999. Molecular biology of odorant receptors in vertebrates *Annu. Rev. Neurosci.* 22: 487-509. ([PubMed](#))

L. Belluscio, G.H. Gold, A. Nemes, and R. Axel. 1998. Mice deficient in G(olf) are anosmic *Neuron* 20: 69-81. ([PubMed](#))

L.B. Vosshall, A.M. Wong, and R. Axel. 2000. An olfactory sensory map in the fly brain *Cell* 102: 147-159. ([PubMed](#))

Taste

M.S. Herness and T.A. Gilbertson. 1999. Cellular mechanisms of taste transduction *Annu. Rev. Physiol.* 61: 873-900. ([PubMed](#))

E. Adler, M.A. Hoon, K.L. Mueller, J. Chandrashekar, N.J. Ryba, and C.S. Zuker. 2000. A novel family of mammalian taste receptors *Cell* 100: 693-702. ([PubMed](#))

J. Chandrashekar, K.L. Mueller, M.A. Hoon, E. Adler, L. Feng, W. Guo, C.S. Zuker, and N.J. Ryba. 2000. T2Rs function as bitter taste receptors *Cell* 100: 703-711. ([PubMed](#))

I. Mano and M. Driscoll. 1999. DEG/ENaC channels: A touchy superfamily that watches its salt *Bioessays* 21: 568-578. ([PubMed](#))

D.J. Benos and B.A. Stanton. 1999. Functional domains within the degenerin/epithelial sodium channel (Deg/ENaC) superfamily of ion channels *J. Physiol. (Lond.)* 520: (part 3) 631-644. ([PubMed](#))

S.K. McLaughlin, P.J. McKinnon, and R.F. Margolskee. 1992. Gustducin is a taste-cell-specific G protein closely related to the transducins *Nature* 357: 563-569. ([PubMed](#))

N. Chaudhari, A.M. Landin, and S.D. Roper. 2000. A metabotropic glutamate receptor variant functions as a taste receptor *Nat. Neurosci.* 3: 113-119. ([PubMed](#))

Vision

L. Stryer. 1988. Molecular basis of visual excitation *Cold Spring Harbor Symp. Quant. Biol.* 53: 283-294. ([PubMed](#))

G. Wald. 1968. The molecular basis of visual excitation *Nature* 219: 800-807. ([PubMed](#))

J.B. Ames, A.M. Dizhoor, M. Ikura, K. Palczewski, and L. Stryer. 1999. Three-dimensional structure of guanylyl cyclase activating protein-2, a calcium-sensitive modulator of photoreceptor guanylyl cyclases *J. Biol. Chem.* 274: 19329-19337. ([PubMed](#))

J. Nathans. 1994. In the eye of the beholder: Visual pigments and inherited variation in human vision *Cell* 78: 357-360. ([PubMed](#))

J. Nathans. 1999. The evolution and physiology of human color vision: Insights from molecular genetic studies of visual pigments *Neuron* 24: 299-312. ([PubMed](#))

K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. LeTrong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, and M. Miyano. 2000. Crystal structure of rhodopsin: A G protein-coupled receptor *Science* 289: 739-745. ([PubMed](#))

Hearing

A.J. Hudspeth. 1997. How hearing happens *Neuron* 19: 947-950. ([PubMed](#))

J.O. Pickles and D.P. Corey. 1992. Mechanoelectrical transduction by hair cells *Trends Neurosci.* 15: 254-259. ([PubMed](#))

R.G. Walker, A.T. Willingham, and C.S. Zuker. 2000. A *Drosophila* mechanosensory transduction channel *Science* 287: 2229-2234. ([PubMed](#))

Touch and pain reception

A. Franco-Obregon and D.E. Clapham. 1998. Touch channels sense blood pressure *Neuron* 21: 1224-1226. ([PubMed](#))

M.J. Caterina, M.A. Schumacher, M. Tominaga, T.A. Rosen, J.D. Levine, and D. Julius. 1997. The capsaicin receptor: A heat-activated ion channel in the pain pathway *Nature* 389: 816-824. ([PubMed](#))

M. Tominaga, M.J. Caterina, A.B. Malmberg, T.A. Rosen, H. Gilbert, K. Skinner, B.E. Raumann, A.I. Basbaum, and D. Julius. 1998. The cloned capsaicin receptor integrates multiple pain-producing stimuli *Neuron* 21: 531-543. ([PubMed](#))

M.J. Caterina and D. Julius. 1999. Sense and specificity: A molecular identity for nociceptors *Curr. Opin. Neurobiol.* 9: 525-530. ([PubMed](#))

Other sensory systems

R.B. Frankel. 1984. Magnetic guidance of organisms *Annu. Rev. Biophys. Bioeng.* 13: 85-103. ([PubMed](#))

J.L. Kirschvink, A. Kobayashi-Kirschvink, and B.J. Woodford. 1992. Magnetite biomineralization in the human brain *Proc. Natl. Acad. Sci. U. S. A.* 89: 7683-7687. ([PubMed](#)) ([Full Text in PMC](#))

C. Dulac and R. Axel. 1995. A novel family of genes encoding putative pheromone receptors in mammals *Cell* 83: 195-206. ([PubMed](#))

33. The Immune System

Dedicated to the memory of Don Wiley, a pioneer in unraveling the structural basis of immune-system function

We are constantly exposed to an incredible diversity of bacteria, viruses, and parasites, many of which would flourish in our cells or extracellular fluids were it not for our immune system. Remarkably, we are often even able to defend ourselves against organisms that we have never before encountered. How does the immune system protect us? The key is our ability to produce more than 10^8 distinct *antibodies* and more than 10^{12} *T-cell receptors*, each of which presents a different surface for specifically binding a molecule from a foreign organism and initiating the destruction of the invader.

The presence of this remarkable repertoire of defensive molecules poses a challenge. What prevents the immune system from attacking cells that express molecules normally present in our bodies; that is, *how does the immune system distinguish between nonself and self?* We shall examine these questions, focusing first on the structures of the proteins participating in the molecular recognition processes and then on the mechanisms for selecting cells that express molecules useful for protecting us from a specific pathogen. Emphasis will be on the modular construction of the proteins of the immune system — identifying structural motifs and considering how spectacular diversity can arise from modular construction.

33.0.1. The Immune System Adapts, Using the Principles of Evolution

The immune system comprises two parallel but interrelated systems. In the *humoral immune response*, soluble proteins called *antibodies (immunoglobulins)* function as recognition elements that bind to foreign molecules and serve as markers signaling foreign invasion (Figure 33.1). Antibodies are secreted by *plasma cells*, which are derived from *B lymphocytes (B cells)*. A foreign macromolecule that binds selectively to an antibody is called an *antigen*. In a physiological context, if the binding of the foreign molecule stimulates an immune response, that molecule is called an *immunogen*. The specific affinity of an antibody is not for the entire macromolecular antigen but for a particular site on the antigen called the *epitope* or *antigenic determinant*.

In the *cellular immune response*, cells called *cytotoxic T lymphocytes* (also commonly called *killer T cells*) kill cells that display foreign motifs on their surfaces. Another class of T cells called *helper T lymphocytes* contributes to both the humoral and the cellular immune responses by stimulating the differentiation and proliferation of appropriate B cells and cytotoxic T cells. The cellular immune response is mediated by specific receptors that are expressed on the surfaces of the T cells.

The remarkable ability of the immune system to adapt to an essentially limitless set of potential pathogens requires a powerful system for transforming the immune cells and molecules present in our systems in response to the presence of pathogens. *This adaptive system operates through the principles of evolution, including reproduction with variation followed by selection of the most well suited members of a population.*

If the human genome contains, by the latest estimates, only 40,000 genes, how can the immune system generate more than 10^8 different antibody proteins and 10^{12} T-cell receptors? The answer is found in a novel mechanism for generating a highly diverse set of genes from a limited set of genetic building blocks. Linking different sets of DNA regions in a combinatorial manner produces many distinct protein-encoding genes that are not present in the genome. A rigorous selection process then leaves for proliferation only cells that synthesize proteins determined to be useful in the immune response. The subsequent reproduction of these cells without additional recombination serves to enrich the cell population with members expressing a particular protein species.

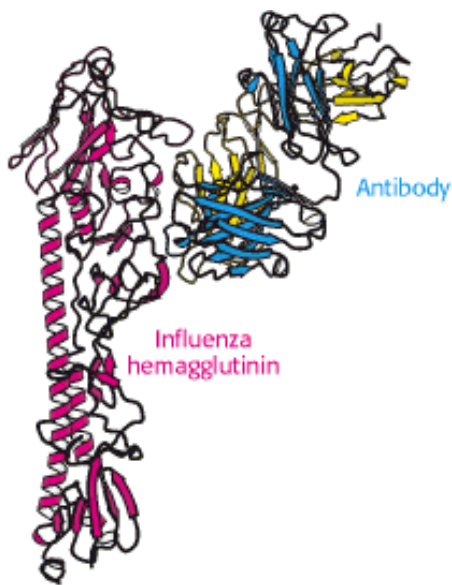
Critical to the development of the immune response is the selection process, which determines which cells will reproduce. The process comprises several stages. In the early stages of the development of an immune response, cells expressing molecules that bind tightly to self-molecules are destroyed or silenced, whereas cells expressing molecules

that do not bind strongly to self-molecules and that have the potential for binding strongly to foreign molecules are preserved. The appearance of an immunogenic invader at a later time will stimulate cells expressing immunoglobulins or T-cell receptors that bind specifically to elements of that pathogen to reproduce — in evolutionary terms, such cells are selected for. Thus, the immune response is based on the selection of cells expressing molecules that are specifically effective against a particular invader; the response evolves from a population with wide-ranging specificities to a more-focused collection of cells and molecules that are well suited to defend the host when confronted with that particular challenge.



Figure 33.1. Immunoglobulin Production. An electron micrograph of a plasma cell shows the highly developed rough endoplasmic reticulum necessary for antibody secretion. [Courtesy of Lynne Mercer.]





Just as Medieval defenders used their weapons and the castle walls to defend their city, the immune system constantly battles against foreign invaders such as viruses, bacteria, and parasites to defend the organism.

Antibody molecules provide a key element in the immune system's defensive arsenal. For example, specific antibodies can bind to molecules on the surfaces of viruses and prevent the viruses from infecting cells. Above, an antibody binds to one subunit on hemagglutinin from the surface of influenza virus. [(Left) The Granger Collection.]

33.1. Antibodies Possess Distinct Antigen-Binding and Effector Units

Antibodies are central molecular players in the immune response, and we examine them first. A fruitful approach in studying proteins as large as antibodies is to split the protein into fragments that retain activity. In 1959, Rodney Porter showed that *immunoglobulin G (IgG)*, the major antibody in serum, can be cleaved into three 50-kd fragments by the limited proteolytic action of papain. Two of these fragments bind antigen. They are called F_{ab} (F stands for fragment, *ab* for antigen binding). The other fragment, called F_c because it crystallizes readily, does not bind antigen, but it has other important biological activities, including the mediation of responses termed *effector functions*. These functions include the initiation of the *complement cascade*, a process that leads to the lysis of target cells. Although such effector functions are crucial to the functioning of the immune system, they will not be discussed further here.

How do these fragments relate to the three-dimensional structure of whole IgG molecules? Immunoglobulin G consists of two kinds of polypeptide chains, a 25-kd *light (L) chain* and a 50-kd *heavy (H) chain* (Figure 33.2). The subunit composition is L_2H_2 . Each L chain is linked to an H chain by a disulfide bond, and the H chains are linked to each other by at least one disulfide bond. Examination of the amino acid sequences and three-dimensional structures of IgG molecules reveals that each L chain comprises two homologous domains, termed *immunoglobulin domains*, to be described in detail in Section 33.2. Each H chain has four immunoglobulin domains. Overall, the molecule adopts a conformation that resembles the letter Y, in which the stem, corresponding to the F_c fragment obtained by cleavage with papain, consists of the two carboxyl-terminal immunoglobulin domains of each H chain and in which the two arms of the Y, corresponding to the two F_{ab} fragments, are formed by the two amino-terminal domains of each H chain and the two amino-terminal domains of each L chain. The linkers between the stem and the two arms consist of relatively extended polypeptide regions within the H chains and are quite flexible.

Papain cleaves the H chains on the carboxyl-terminal side of the disulfide bond that links each L and H chain (Figure 33.3). Thus, each F_{ab} consists of an entire L chain and the amino-terminal half of an H chain, whereas F_c consists of the carboxyl-terminal halves of both H chains. Each F_{ab} contains a single antigen-binding site. Because an intact IgG molecule contains two F_{ab} components and therefore has two binding sites, it can cross-link multiple antigens (Figure 33.4). Furthermore, the F_c and the two F_{ab} units of the intact IgG are joined by flexible polypeptide regions that allow

facile variation in the angle between the F_{ab} units through a wide range (Figure 33.5). This kind of mobility, called *segmental flexibility*, can enhance the formation of an antibody-antigen complex by enabling both combining sites on an antibody to bind an antigen that possesses multiple binding sites, such as a viral coat composed of repeating identical monomers or a bacterial cell surface. The combining sites at the tips of the F_{ab} units simply move to match the distance between specific determinants on the antigen.

Immunoglobulin G is the antibody present in highest concentration in the serum, but other classes of immunoglobulin also are present (Table 33.1). Each class includes an L chain (either κ or λ) and a distinct H chain (Figure 33.6). The heavy chains in IgG are called γ chains, whereas those in immunoglobulins A, M, D, and E are called α , μ , δ , and ϵ , respectively. *Immunoglobulin M (IgM)* is the first class of antibody to appear in the serum after exposure to an antigen. The presence of 10 combining sites enables IgM to bind especially tightly to antigens containing multiple identical epitopes. The strength of an interaction comprising multiple independent binding interactions between partners is termed *avidity* rather than *affinity*, which denotes the binding strength of a single combining site. The presence of 10 combining sites in IgM compared with 2 sites in IgG enables IgM to bind many multivalent antigens that would slip away from IgG.

Immunoglobulin A (IgA) is the major class of antibody in external secretions, such as saliva, tears, bronchial mucus, and intestinal mucus. Thus, IgA serves as a first line of defense against bacterial and viral antigens. The role of *immunoglobulin D (IgD)* is not yet known. *Immunoglobulin E (IgE)* is important in conferring protection against parasites, but IgE also causes allergic reactions. IgE-antigen complexes form cross-links with receptors on the surfaces of mast cells to trigger a cascade that leads to the release of granules containing pharmacologically active molecules. Histamine, one of the agents released, induces smooth muscle contraction and stimulates the secretion of mucus.

A comparison of the amino acid sequences of different IgG antibodies from human beings or mice shows that the carboxyl-terminal half of the L chains and the carboxyl-terminal three-quarters of the H chains are very similar in all of the antibodies. Importantly, the amino-terminal domain of each chain is more variable, including three stretches of approximately 7 to 12 amino acids within each chain that are hypervariable, as shown for the H chain in Figure 33.7. The amino-terminal immunoglobulin domain of each chain is thus referred to as the *variable region*, whereas the remaining immunoglobulin domains are much more similar in all antibodies and are referred to as *constant regions* (Figure 33.8).

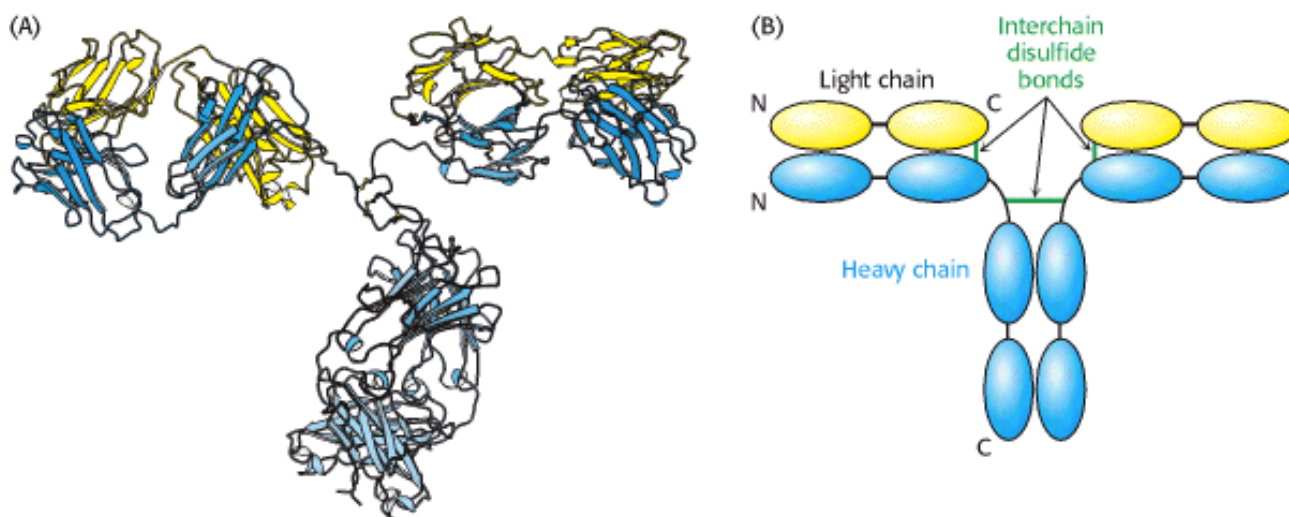


Figure 33.2. Immunoglobulin G Structure. (A) The three-dimensional structure of an IgG molecule showing the light chains in yellow and the heavy chains in blue. (B) A schematic view of an IgG molecule indicating the positions of the interchain disulfide bonds. N, amino terminus; C, carboxyl terminus.

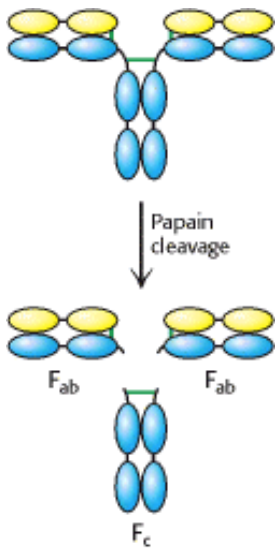


Figure 33.3. Immunoglobulin G Cleavage. Treatment of intact IgG molecules with the protease papain results in the formation of three large fragments: two F_{ab} fragments that retain antigen-binding capability and one F_c fragment that does not.

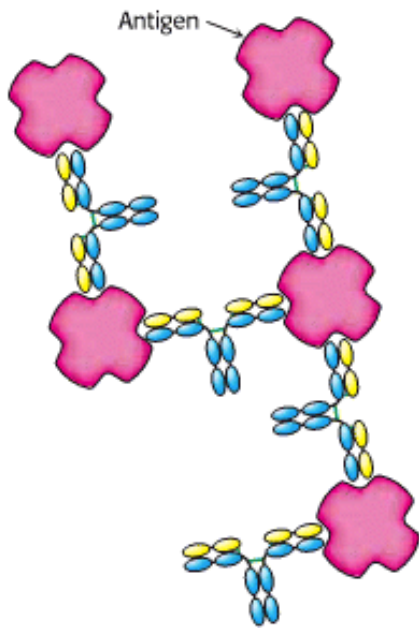


Figure 33.4. Antigen Cross-Linking. Because IgG molecules include two antigen-binding sites, antibodies can cross-link multivalent antigens such as viral surfaces.

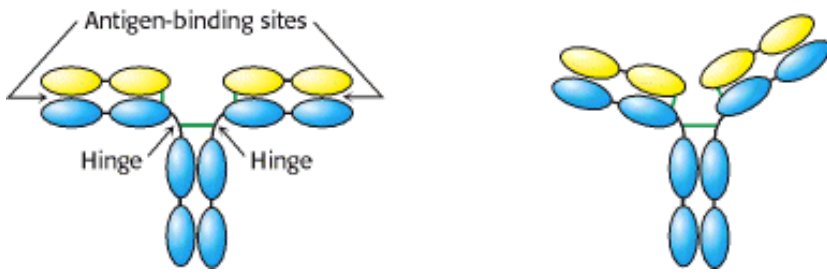


Figure 33.5. Segmental Flexibility. The linkages between the F_{ab} and the F_c regions of an IgG molecule are flexible, allowing the two antigen-binding sites to adopt a range of orientations with respect to one another. This flexibility allows effective interactions with a multivalent antigen without requiring that the epitopes on the target be a precise distance apart.

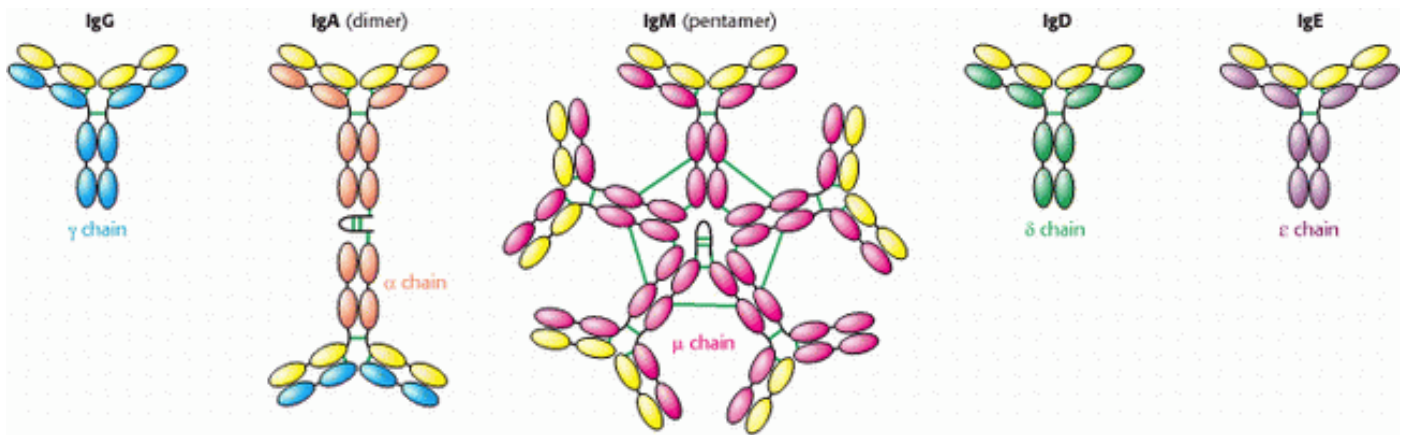


Figure 33.6. Classes of Immuno-Globulin. Each of five classes of immuno-globulin has the same light chain (shown in yellow) combined with a different heavy chain (γ , α , μ , δ , or ϵ). Disulfide bonds are indicated by green lines. The IgA dimer and the IgM pentamer have a small polypeptide chain in addition to the light and heavy chains.

Table 33.1. Properties of immunoglobulin classes

Class	Serum concentration (mg/ml)	Mass (kd)	Sedimentation coefficient(s)	Light chains	Heavy chains	Chain structure
IgG	12	150	7	κ or λ	γ	$\kappa_2 \gamma_2$ or $\lambda_2 \gamma_2$
IgA	3	180 – 500	7, 10, 13	κ or λ	α	$(\kappa_2 \alpha_2)_n$ or $(\lambda_2 \alpha_2)_n$
IgM	1	950	18 – 20	κ or λ	μ	$(\kappa_2 \mu_2)_5$ or $(\lambda_2 \mu_2)_5$
IgD	0.1	175	7	κ or λ	δ	$\kappa_2 \delta_2$ or $\lambda_2 \delta_2$
IgE	0.001	200	8	κ or λ	ϵ	$\kappa_2 \epsilon_2$ or $\lambda_2 \epsilon_2$

Note: $n = 1, 2,$ or 3 . IgM and oligomers of IgA also contain J chains that connect immunoglobulin molecules. IgA in secretions has an additional component.

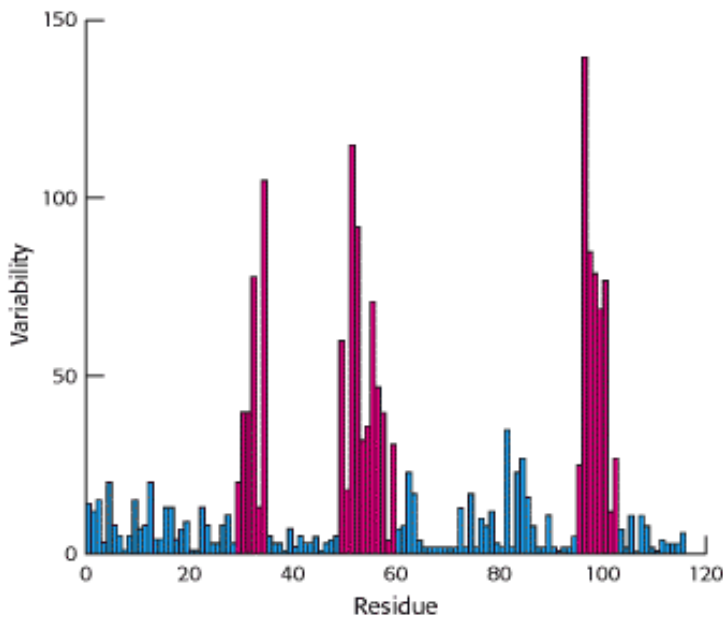


Figure 33.7. Immunoglobulin Sequence Diversity. A plot of sequence variability as a function of position along the sequence of the amino-terminal immunoglobulin domain of the H chain of human IgG molecules. Three regions (in red) show remarkably high levels of variability. These hypervariable regions correspond to three loops in the immunoglobulin domain structure. [After R. A. Goldsby, T. J. Kindt, and B. A. Osborne, *Kuby Immunology*, 4th ed. (W. H. Freeman and Company, 2000), p. 91.]

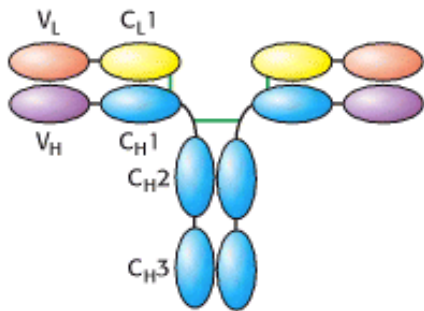



Figure 33.8. Variable and Constant Regions. Each L and H chain includes one immunoglobulin domain at its amino terminus that is quite variable from one antibody to another. These domains are referred to as V_L and V_H . The remaining domains are more constant from one antibody to another and are referred to as constant domains (C_L1 , C_H1 , C_H2 , and C_H3).

33.2. The Immunoglobulin Fold Consists of a Beta-Sandwich Framework with Hypervariable Loops

An IgG molecule consists of a total of 12 immunoglobulin domains. These domains have many sequence features in common and adopt a common structure, the *immunoglobulin fold* (Figure 33.9). Remarkably, this same structural domain is found in many other proteins that play key roles in the immune system.

The immunoglobulin fold consists of a pair of β sheets, each built of antiparallel β strands, that surround a central hydrophobic core. A single disulfide bond bridges the two sheets. Two aspects of this structure are particularly important for its function. First, three loops present at one end of the structure form a potential binding surface. These loops contain the hypervariable sequences present in antibodies and in T-cell receptors (see Sections 33.3 and 33.5.2). Variation of the amino acid sequences of these loops provides the major mechanism for the generation of the vastly diverse set of antibodies and T-cell receptors expressed by the immune system. These loops are referred to as *hypervariable loops* or *complementaritydetermining regions (CDRs)*. Second, the amino terminus and the carboxyl terminus are at opposite ends of the structure, which allows structural domains to be strung together to form chains, as in the L and H chains of antibodies. Such chains are present in several other key molecules in the immune system.

 The immunoglobulin fold is one of the most prevalent domains encoded by the human genome — more than 750 genes encode proteins with at least one immunoglobulin fold recognizable at the level of amino acid sequence. Such domains are also common in other multicellular animals such as flies and nematodes. However, from inspection of amino acid sequence alone, immunoglobulin-fold domains do not appear to be present in yeast or plants. However, structurally similar domains are present in these organisms, including the key photosynthetic electron-transport protein plastocyanin in plants (Section 19.3.2). Thus, the immunoglobulin-fold family appears to have expanded greatly along evolutionary branches leading to animals — particularly, vertebrates.

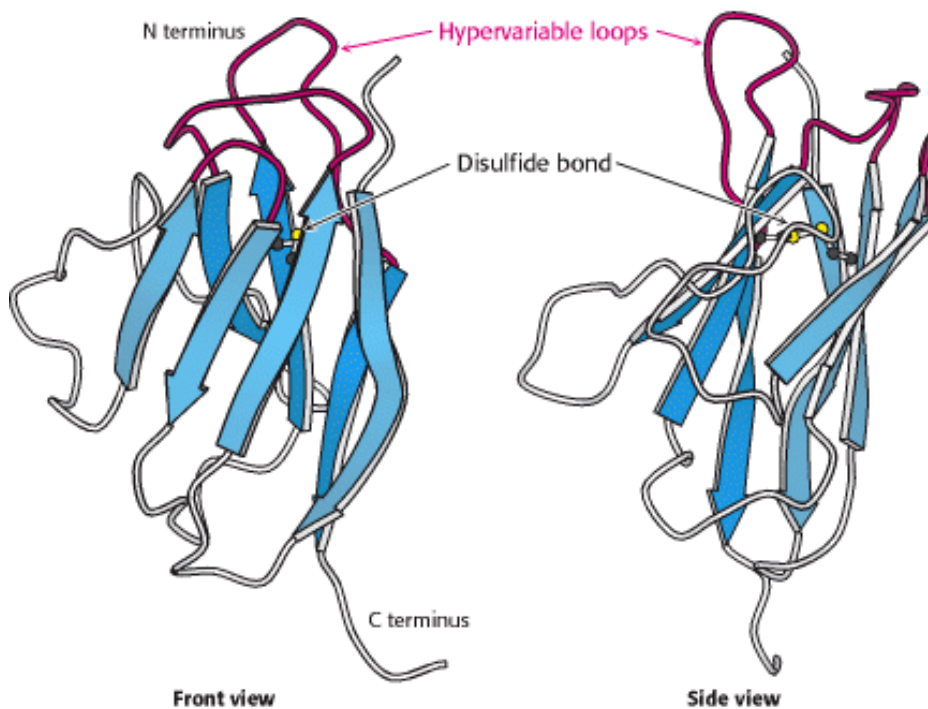


Figure 33.9. Immunoglobulin Fold. An immunoglobulin domain consists of a pair of β -sheets linked by a disulfide bond and hydrophobic interactions. Three hypervariable loops lie at one end of the structure.



33.3. Antibodies Bind Specific Molecules Through Their Hypervariable Loops

For each class of antibody, the amino-terminal immunoglobulin domains of the L and H chains (the variable domains, designated V_L and V_H) come together at the ends of the arms extending from the structure. The positions of the complementarity-determining regions are striking. These hypervariable sequences, present in three loops of each domain, come together so that all six loops form a single surface at the end of each arm (Figure 33.10). Because virtually any V_L can pair with any V_H , *a very large number of different binding sites can be constructed by their combinatorial association.*

33.3.1. X-Ray Analyses Have Revealed How Antibodies Bind Antigens

The results of x-ray crystallographic studies of many large and small antigens bound to F_{ab} molecules have been sources of much insight into the structural basis of antibody specificity. The binding of antigens to antibodies is governed by the same principles that govern the binding of substrates to enzymes. The apposition of complementary shapes results in numerous contacts between amino acids at the binding surfaces of both molecules. Numerous hydrogen bonds, electrostatic interactions, and van der Waals interactions, reinforced by hydrophobic interactions, combine to give specific and strong binding.

A few aspects of antibody binding merit specific attention, inasmuch as they relate directly to the structure of immunoglobulins. The binding site on the antibody has been found to incorporate some or all of the CDRs in the variable domains of the antibody. Small molecules (e.g., octapeptides) are likely to make contact with fewer CDRs, with perhaps 15 residues of the antibody participating in the binding interaction. Macromolecules often make more extensive contact, interacting with all six CDRs and 20 or more residues of the antibody. Small molecules often bind in a cleft of the antigen-binding region. Macromolecules such as globular proteins tend to interact across larger, fairly flat apposed surfaces bearing complementary protrusions and depressions.

A well-studied case of small-molecule binding is seen in an example of phosphorylcholine bound to F_{ab} . Crystallographic analysis revealed phosphorylcholine bound to a cavity lined by residues from five CDRs — two from the L chain and three from the H chain (Figure 33.11). The positively charged trimethylammonium group of phosphorylcholine is buried inside the wedge-shaped cavity, where it interacts electrostatically with two negatively charged glutamate residues. The negatively charged phosphate group of phosphorylcholine binds to the positively charged guanidinium group of an arginine residue at the mouth of the crevice and to a nearby lysine residue. The phosphate group is also hydrogen bonded to the hydroxyl group of a tyrosine residue and to the guanidinium group of the arginine side chain. Numerous van der Waals interactions, such as those made by a tryptophan side chain, also stabilize this complex.

The binding of phosphorylcholine does not significantly change the structure of the antibody, yet induced fit plays a role in the formation of many antibody-antigen complexes. A malleable binding site can accommodate many more kinds of ligands than can a rigid one. Thus, induced fit increases the repertoire of antibody specificities.

33.3.2. Large Antigens Bind Antibodies with Numerous Interactions

How do large antigens interact with antibodies? A large collection of antibodies raised against hen egg-white lysozyme has been structurally characterized in great detail (Figure 33.12). Each different antibody binds to a distinct surface of lysozyme. Let us examine the interactions present in one of these complexes in detail. This antibody binds two polypeptide segments that are widely separated in the primary structure, residues 18 to 27 and 116 to 129 (Figure 33.13).

All six CDRs of the antibody make contact with this epitope. The region of contact is quite extensive (about $30 \times 20 \text{ \AA}$). The apposed surfaces are rather flat. The only exception is the side chain of glutamine 121 of lysozyme, which

penetrates deeply into the antibody binding site, where it forms a hydrogen bond with a main-chain carbonyl oxygen atom and is surrounded by three aromatic side chains. The formation of 12 hydrogen bonds and numerous van der Waals interactions contributes to the high affinity ($K_d = 20 \text{ nM}$) of this antibody-antigen interaction. Examination of the F_{ab} molecule without bound protein reveals that the structures of the V_L and V_H domains change little on binding, although they slide 1 \AA apart to allow more intimate contact with lysozyme.

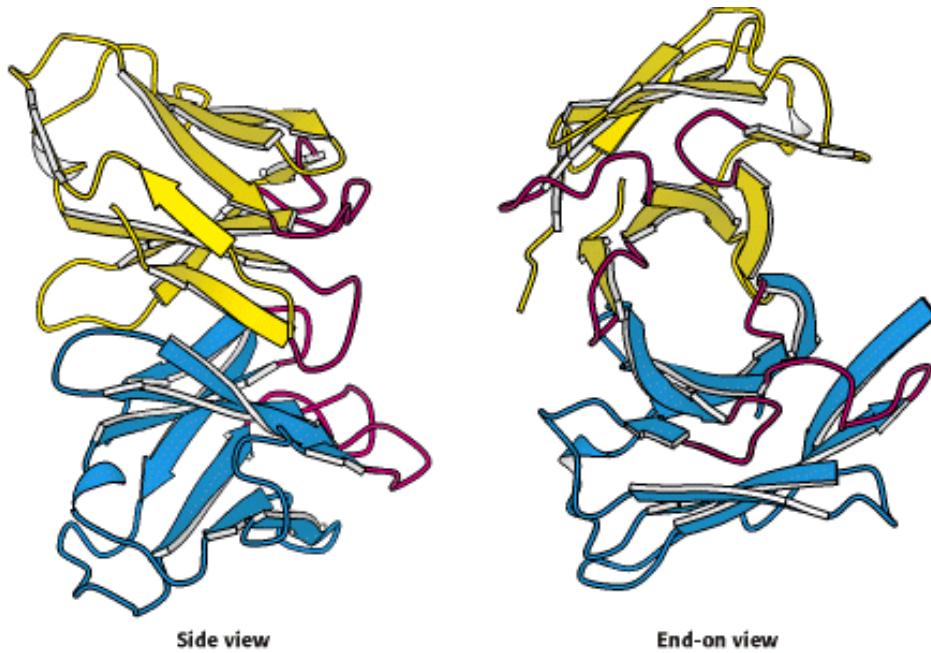


Figure 33.10. Variable Domains. Two views of the variable domains of the L chain (yellow) and the H chain (blue); the complementarity-determining regions (CDRs) are shown in red. The six CDRs come together to form a binding surface. The specificity of the surface is determined by the sequences and structures of the CDRs.

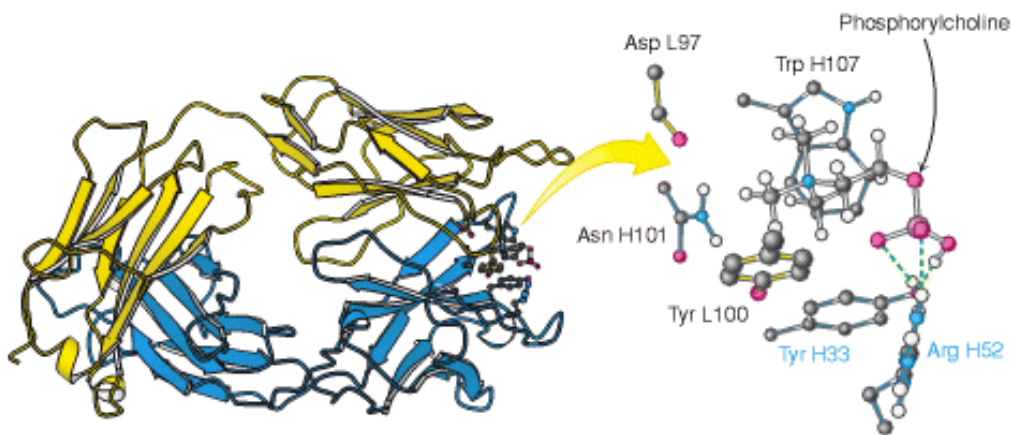


Figure 33.11. Binding of a Small Antigen. The structure of a complex between an F_{ab} fragment of an antibody and its target — in this case, phosphorylcholine. Residues from the antibody interact with phosphorylcholine through hydrogen bonding and electrostatic and van der Waals interactions.

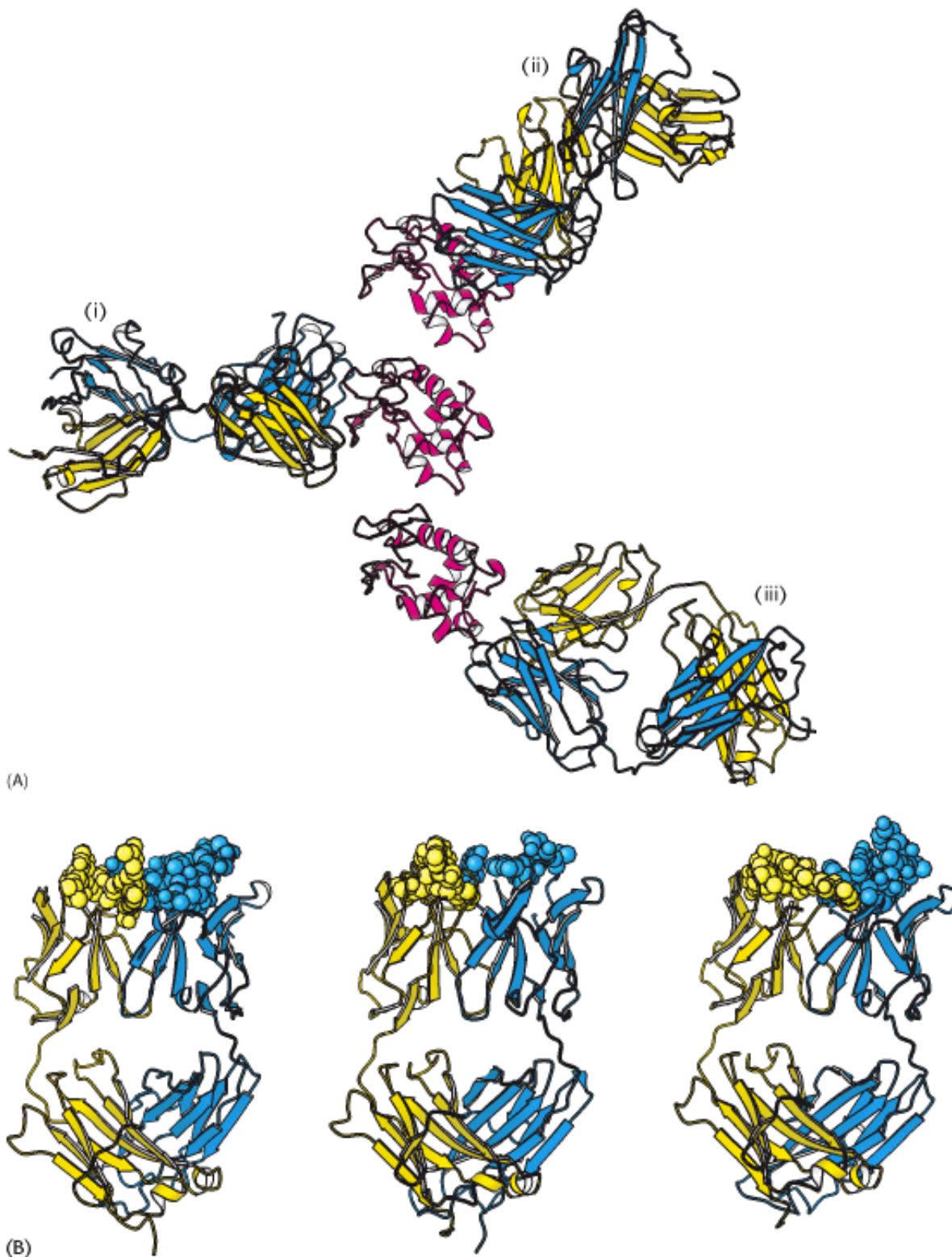


Figure 33.12. Antibodies Against Lysozyme. (A) The structures of three complexes (i, ii, iii) between F_{ab} fragments (blue and yellow) and hen egg-white lysozyme (red) shown with lysozyme in the same orientation in each case. The three antibodies recognize completely different epitopes on the lysozyme molecule. (B) The F_{ab} fragments from part A with points of contact highlighted as space-filling models, revealing the different shapes of the antigen-binding sites.

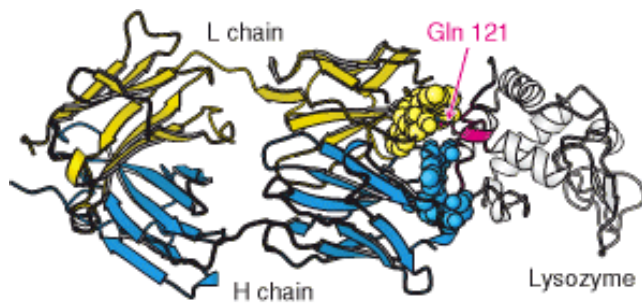


Figure 33.13. Antibody - Protein Interactions. The structure of a complex between an F_{ab} fragment and lysozyme reveals that the binding surfaces are complementary in shape over a large area. A single residue of lysozyme, glutamine 121, penetrates more deeply into the antibody combining site.

33.4. Diversity Is Generated by Gene Rearrangements

A mammal such as a mouse or a human being can synthesize large amounts of specific antibody against virtually any foreign determinant within a matter of days of being exposed to it. We have seen that antibody specificity is determined by the amino acid sequences of the variable regions of both light and heavy chains, which brings us to the key question: How are different variable-region sequences generated?

The discovery of distinct variable and constant regions in the L and H chains raised the possibility that the genes that encode immunoglobulins have an unusual architecture that facilitates the generation of a diverse set of polypeptide products. In 1965, William Dreyer and Claude Bennett proposed that multiple *V* (*variable*) genes are separate from a single *C* (*constant*) gene in embryonic (germ-line) DNA. According to their model, one of these *V* genes becomes joined to the *C* gene in the course of differentiation of the antibody-producing cell. A critical test of this novel hypothesis had to await the isolation of pure immunoglobulin mRNA and the development of techniques for analyzing mammalian genomes. Twenty years later, Susumu Tonegawa found that *V* and *C* genes are indeed far apart in embryonic DNA but are closely associated in the DNA of antibody-producing cells. Thus, immunoglobulin genes are rearranged in the differentiation of lymphocytes.

33.4.1. J (Joining) Genes and D (Diversity) Genes Increase Antibody Diversity

Sequencing studies carried out by Susumu Tonegawa, Philip Leder, and Leroy Hood revealed that *V* genes in embryonic cells do not encode the entire variable region of L and H chains. Consider, for example, the region that encodes the κ light-chain family. A tandem array of 40 segments, each of which encodes approximately the first 97 residues of the variable domain of the L chain, is present on human chromosome 2 (Figure 33.14).

However, the variable region of the L chain extends to residue 110. Where is the DNA that encodes the last 13 residues of the *V* region? For L chains in undifferentiated cells, this stretch of DNA is located in an unexpected place: near the *C* gene. It is called the *J* gene because it joins the *V* and *C* genes in a differentiated cell. In fact, a tandem array of five *J* genes is located near the *C* gene in embryonic cells. In the differentiation of an antibody-producing cell, a *V* gene becomes spliced to a *J* gene to form a complete gene for the variable region (Figure 33.15). RNA splicing generates an mRNA molecule for the complete L chain by linking the coding regions for the rearranged *VJ* unit with that for the *C* unit (Figure 33.16).

J genes are important contributors to antibody diversity because they encode part of the last hypervariable segment (CDR3). In forming a continuous variable-region gene, any of the 40 *V* genes can become linked to any of five *J* genes. Thus, somatic recombination of these gene segments amplifies the diversity already present in the germ line. The linkage between *V* and *J* is not precisely controlled. Recombination between these genes can take place at one of several bases

near the codon for residue 95, generating additional diversity. A similar array of V and J genes encoding the λ light chain is present on human chromosome 22. This region includes 30 V_{λ} gene segments and four J_{λ} segments. In addition, this region includes four distinct C genes, in contrast with the single C gene in the κ locus.

In human beings, the genes encoding the heavy chain are present on chromosome 14. Remarkably, the variable domain of heavy chains is assembled from *three* rather than two segments. In addition to V_H genes that encode residues 1 to 94 and J_H segments that encode residues 98 to 113, this chromosomal region includes a distinct set of segments that encode residues 95 to 97 ([Figure 33.17](#)). These gene segments are called D for *diversity*. Some 27 D segments lie between 51 V_H and 6 J_H segments. The recombination process first joins a D segment to a J_H segment; a V_H segment is then joined to DJ_H . A greater variety of antigen-binding patches and clefts can be formed by the H chain than by the L chain because the H chain is encoded by three rather than two gene segments. Moreover, CDR3 of the H chain is diversified by the action of terminal deoxyribonucleotidyl transferase, a special DNA polymerase that requires no template. This enzyme inserts extra nucleotides between V_H and D. The *V(D)J recombination* of both the L and the H chains is executed by specific enzymes present in immune cells. These proteins, called *RAG-1* and *RAG-2*, recognize specific DNA sequences called *recombination signal sequences (RSSs)* adjacent to the V, D, and J segments and facilitate the cleavage and religation of the DNA segments.

33.4.2. More Than 10^8 Antibodies Can Be Formed by Combinatorial Association and Somatic Mutation

Let us recapitulate the sources of antibody diversity. The germ line contains a rather large repertoire of variable-region genes. For κ light chains, there are about 40 V-segment genes and five J-segment genes. Hence, a total of $40 \times 5 = 200$ kinds of complete V_{κ} genes can be formed by the combinations of V and J. A similar analysis suggests that at least 120 different λ light chains can be generated. A larger number of heavy-chain genes can be formed because of the role of the D segments. For 51 V, 27 D, and 6 J gene segments, the number of complete V_H genes that can be formed is 8262. The association of 320 kinds of L chains with 8262 kinds of H chains would yield 2.6×10^6 different antibodies. Variability in the exact points of segment joining and other mechanisms increases this value by at least two orders of magnitude.

Even more diversity is introduced into antibody chains by *somatic mutation* — that is, the introduction of mutations into the recombined genes. In fact, a 1000-fold increase in binding affinity is seen in the course of a typical humoral immune response, arising from somatic mutation, a process called *affinity maturation*. The generation of an expanded repertoire leads to the selection of antibodies that more precisely fit the antigen. Thus, nature draws on each of three sources of diversity — a germ-line repertoire, somatic recombination, and somatic mutation — to form the rich variety of antibodies that protect an organism from foreign incursions.


33.4.3. The Oligomerization of Antibodies Expressed on the Surface of Immature B Cells Triggers Antibody Secretion

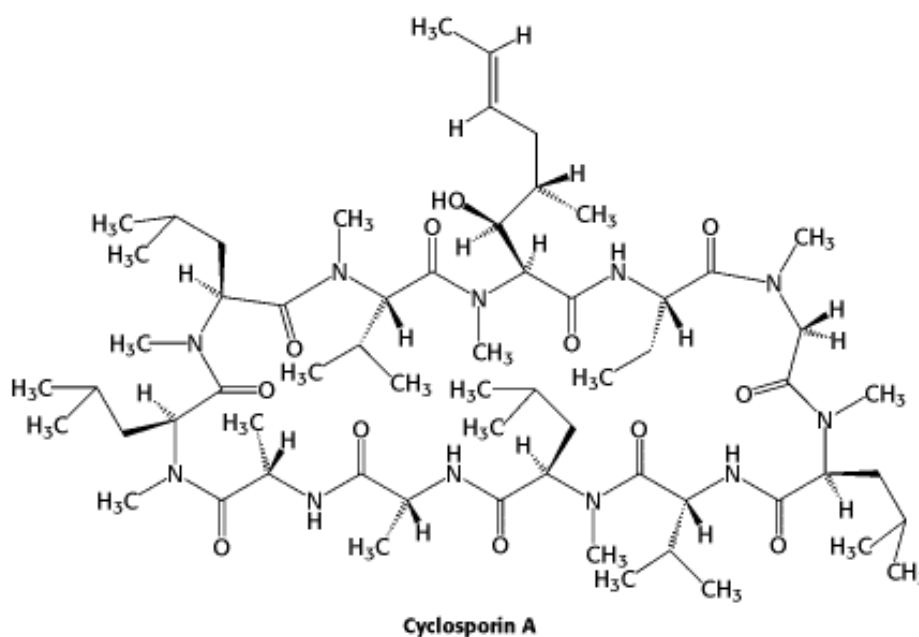
The processes heretofore described generate a highly diverse set of antibody molecules — a key first step in the generation of an immune response. The next stage is the selection of a particular set of antibodies directed against a specific invader. How does this selection occur? Each immature B cell, produced in the bone marrow, expresses a monomeric form of IgM attached to its surface ([Figure 33.18](#)). Each cell expresses approximately 10^5 IgM molecules, but *all of these molecules are identical in amino acid sequence and, hence, in antigen-binding specificity*. Thus, the selection of a particular immature B cell for growth will lead to the amplification of an antibody with a unique specificity. The selection process begins with the binding of an antigen to the membrane-bound antibody.

Associated with each membrane-linked IgM molecule are two molecules of a heterodimeric membrane protein called $Ig\text{-}\alpha$ - $Ig\text{-}\beta$ (see [Figure 33.18](#)). Examination of the amino acid sequences of $Ig\text{-}\alpha$ and $Ig\text{-}\beta$ is highly instructive. The amino terminus of each protein lies outside the cell and corresponds to a single immunoglobulin, and the carboxyl terminus,

which lies inside the cell, includes a sequence of 18 amino acids called an *immunoreceptor tyrosine-based activation motif (ITAM)* (see [Figure 33.18](#)). As its name suggests, each ITAM includes key tyrosine residues, which are subject to phosphorylation by particular protein kinases present in immune-system cells.

A fundamental observation with regard to the mechanism by which the binding of antigen to membrane-bound antibody triggers the subsequent steps of the immune response is that *oligomerization or clustering of the antibody molecules is required* ([Figure 33.19](#)). The requirement for oligomerization is reminiscent of the dimerization of receptors triggered by growth hormone and epidermal growth factor encountered in [Sections 15.4](#) and [15.4.1](#); indeed, the associated signaling mechanisms appear to be quite similar. The oligomerization of the membrane-bound antibodies results in the phosphorylation of the tyrosine residues within the ITAMs by protein tyrosine kinases including Lyn, a homolog of Src ([Section 15.5](#)). The phosphorylated ITAMs serve as docking sites for a protein kinase termed spleen tyrosine kinase (Syk), which has two SH2 domains that interact with the pair of phosphorylated tyrosines in each ITAM. Syk, when activated by phosphorylation, proceeds to phosphorylate other signal-transduction proteins including an inhibitory subunit of a transcription factor called NF- κ B and an isoform of phospholipase C. The signaling processes continue downstream to activate gene expression, leading to the stimulation of cell growth and initiating further B-cell differentiation.

 Drugs that modulate the immune system have served as sources, of insight into immune-system signaling pathways. For example, *cyclosporin*, a powerful suppressor of the immune system, acts by blocking a phosphatase called *calcineurin*, which normally activates a transcription factor called NF-AT by dephosphorylating it.



The potent immune suppression that results reveals how crucial the activity of this transcription factor is to the development of an immune response. Without drugs such as cyclosporin, organ transplantation would be extremely difficult because transplanted tissue expresses a wide range of foreign antigens, which causes the immune system to reject the new tissue.

The role of oligomerization in the B-cell signaling pathway is illuminated when we consider the nature of many antigens presented by pathogens. The surfaces of many viruses, bacteria, and parasites are characterized by arrays of identical membrane proteins or membrane-linked carbohydrates. Thus, most pathogens present multiple binding surfaces that will naturally cause membrane-associated antibodies to oligomerize as they bind adjacent epitopes. In addition, the mechanism accounts for the observation that most small molecules do not induce an immune response; however, coupling multiple copies of the small molecule to a large oligomeric protein such as keyhole limpet hemocyanin (KLH), which has a molecular mass of close to 1 million daltons or more, promotes antibody oligomerization and, hence, the production of antibodies against the small-molecule epitope. The large protein is called the *carrier* of the attached chemical group, which is called a *haptenic determinant*. The small foreign molecule by itself is called a *hapten*.

Antibodies elicited by attached haptens will bind unattached haptens as well.

33.4.4. Different Classes of Antibodies Are Formed by the Hopping of V_H Genes

The development of an effective antibody-based immune response depends on the secretion into the blood of antibodies that have appropriate effector functions. At the beginning of this response, an alternative mRNA splicing pathway is activated so that the production of membrane-linked IgM is supplanted by the synthesis of secreted IgM. As noted in Section 33.1, secreted IgM is pentameric and has a relatively high avidity for multivalent antigens. Later, the antibody-producing cell makes either IgG, IgA, IgD, or IgE of the same specificity as the initially secreted IgM. In this switch, the light chain is unchanged, as is the variable region of the heavy chain. Only the constant region of the heavy chain changes. This step in the differentiation of an antibody-producing cell is called *class switching* (Figure 33.20). In undifferentiated cells, the genes for the constant region of each class of heavy chain, called C_μ , C_δ , C_γ , C_ϵ , and C_α , are next to each other. There are eight in all, including four genes for the constant regions of γ chains. A complete gene for the heavy chains of IgM antibody is formed by the translocation of a V_H gene segment to a DJ_H gene segment.

How are other heavy chains formed? Class switching is mediated by a gene-rearrangement process that moves a VDJ gene from a site near one C gene to a site near another C gene. Importantly, *the antigen-binding specificity is conserved in class switching because the entire V_HDJ_H gene is translocated in an intact form*. For example, the antigen-combining specificity of IgA produced by a particular cell is the same as that of IgM synthesized at an earlier stage of its development. The biological significance of C_H switching is that a whole recognition domain (the variable domain) is shifted from the early constant region (C_μ) to one of several other constant regions that mediate different effector functions.



Figure 33.14. The κ Light-Chain Locus. This part of human chromosome 2 includes an array of 40 segments that encode the variable (V) region (approximately residues 1 – 97) of the light chain, an array of 5 segments that encode the joining (J) region (residues 98 – 110), and a single region that encodes the constant (C) region.

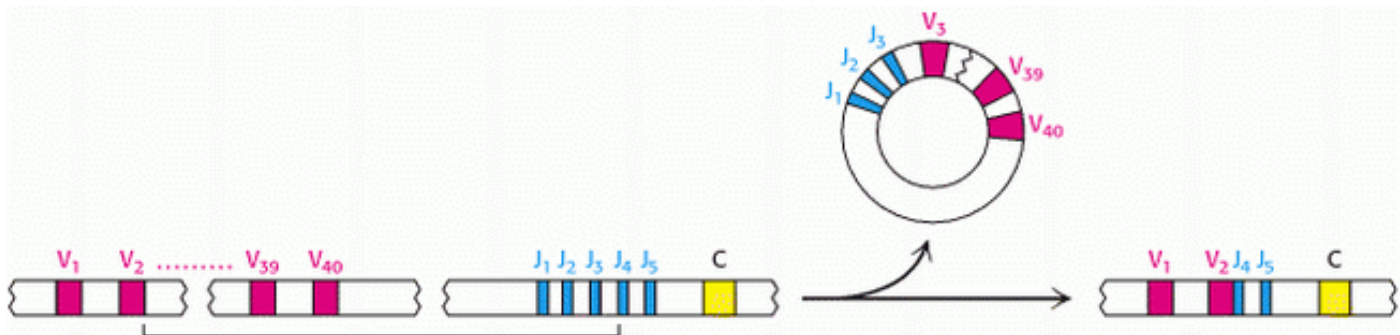


Figure 33.15. VJ Recombination. A single V gene (in this case, V_2) is linked to a J gene (here, J_4) to form an intact VJ region. The intervening DNA is released in a circular form. Because the V and J regions are selected at random and the joint between them is not always in exactly the same place, many VJ combinations can be generated by this process.

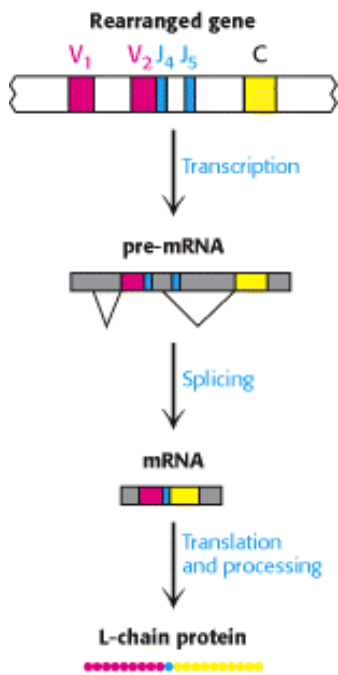


Figure 33.16. Light-Chain Expression. The light-chain protein is expressed by transcription of the rearranged gene to produce a pre-RNA molecule with the VJ and C regions separated. RNA splicing removes the intervening sequences to produce an mRNA molecule with the VJ and C regions linked. Translation of the mRNA and processing of the initial protein product produces the light chain.

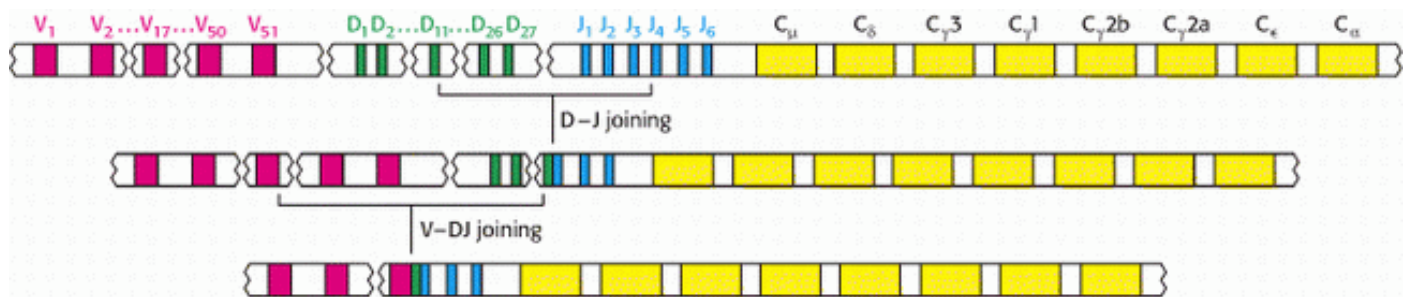


Figure 33.17. V(D)J Recombination. The heavy-chain locus includes an array of 51 V segments, 27 D segments, and 6 J segments. Gene rearrangement begins with D-J joining, followed by further rearrangement to link the V segment to the DJ segment.

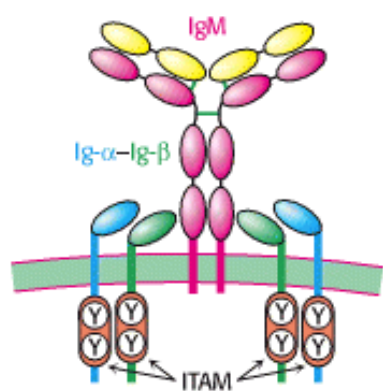


Figure 33.18. B-Cell Receptor. This complex consists of a membrane-bound IgM molecule noncovalently bound to two

Ig- α -Ig- β heterodimers. The intracellular domains of each of the Ig- α and Ig- β chains include an immunoreceptor tyrosine-based activation motif (ITAM).

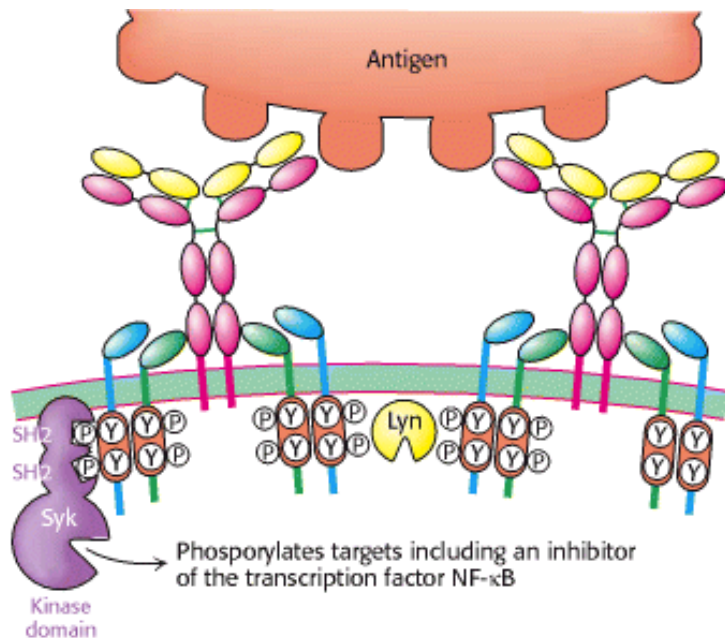


Figure 33.19. B-Cell Activation. The binding of multivalent antigen such as bacterial or viral surfaces links membrane-bound IgM molecules. This oligomerization triggers the phosphorylation of tyrosine residues in the ITAM sequences by protein tyrosine kinases such as Lyn. After phosphorylation, the ITAMs serve as docking sites for Syk, a protein kinase that phosphorylates a number of targets, including transcription factors.

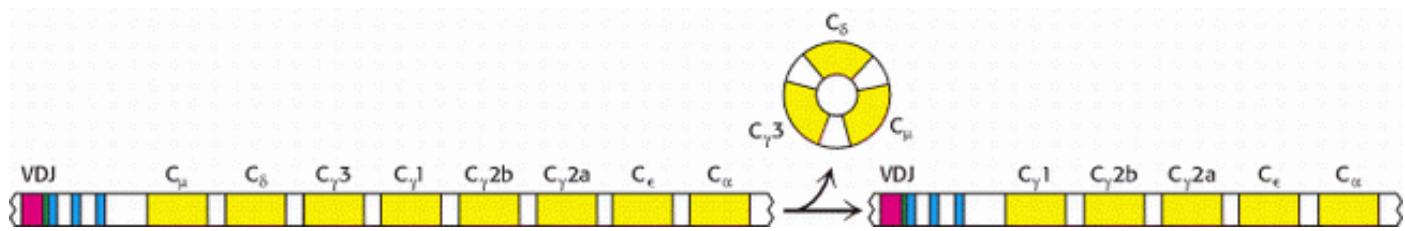


Figure 33.20. Class Switching. Further rearrangement of the heavy-chain locus results in the generation of genes for anti-body classes other than IgM. In the case shown, rearrangement places the VDJ region next to the C γ 1 region, resulting in the production of IgG1. Note that no further rearrangement of the VDJ region takes place, so the specificity of the anti-body is not affected.

33.5. Major-Histocompatibility-Complex Proteins Present Peptide Antigens on Cell Surfaces for Recognition by T-Cell Receptors

Soluble antibodies are highly effective against extracellular pathogens, but they confer little protection against microorganisms that are predominantly intracellular, such as viruses and mycobacteria (which cause tuberculosis and leprosy). These pathogens are shielded from antibodies by the host-cell membrane (Figure 33.21). A different and more subtle strategy, *cell-mediated immunity*, evolved to cope with intracellular pathogens. *T cells* continually scan the surfaces of all cells and kill those that exhibit foreign markings. The task is not simple; intracellular microorganisms are not so obliging as to intentionally leave telltale traces on the surface of their host. Quite the contrary, successful pathogens are masters of the art of camouflage. Vertebrates have devised an ingenious mechanism — cut and display — to reveal the presence of stealthy intruders. Nearly all vertebrate cells exhibit on their surfaces a sample of peptides derived from the digestion of proteins in their cytosol. These peptides are displayed by integral membrane proteins that

are encoded by the *major histocompatibility complex (MHC)*. Specifically, peptides derived from cytosolic proteins are bound to *class I MHC proteins*.

How are these peptides generated and delivered to the plasma membrane? The process starts in the cytosol with the degradation of proteins, self proteins as well as those of pathogens ([Figure 33.22](#)). Digestion is carried out by proteasomes ([Section 23.2.2](#)). The resulting peptide fragments are transported from the cytosol into the lumen of the endoplasmic reticulum by an ATP-driven pump. In the ER, peptides combine with nascent class I MHC proteins; these complexes are then targeted to the plasma membrane.

MHC proteins embedded in the plasma membrane tenaciously grip their bound peptides so that they can be touched and scrutinized by T-cell receptors on the surface of a killer cell. Foreign peptides bound to class I MHC proteins signal that a cell is infected and mark it for destruction by cytotoxic T cells. An assembly consisting of the foreign peptide-MHC complex, the T-cell receptor, and numerous accessory proteins triggers a cascade that induces apoptosis in the infected cell. Strictly speaking, infected cells are not killed but, instead, are triggered to commit suicide to aid the organism.

33.5.1. Peptides Presented by MHC Proteins Occupy a Deep Groove Flanked by Alpha Helices

The three-dimensional structure of a large fragment of a human MHC class I protein, *human leukocyte antigen A2 (HLA-A2)*, was solved in 1987 by Don Wiley and Pamela Bjorkman. Class I MHC proteins consist of a 44-kd α chain noncovalently bound to a 12-kd polypeptide called β_2 -*microglobulin*. The α chain has three extracellular domains (α_1 , α_2 , and α_3), a transmembrane segment, and a tail that extends into the cytosol ([Figure 33.23](#)). Cleavage by papain of the HLA α chain several residues before the transmembrane segment yielded a soluble heterodimeric fragment. The β_2 -microglobulin subunit and the α_3 domains have immunoglobulin folds, although the pairing of the two domains differs from that in antibodies. The α_1 and α_2 domains exhibit a novel and remarkable architecture. They associate intimately to form a deep groove that serves as the peptide-binding site ([Figure 33.24](#)). The floor of the groove, which is about 25 Å long and 10 Å wide, is formed by eight β strands, four from each domain. A long helix contributed by the α_1 domain forms one side, and a helix contributed by the α_2 domain forms the other side. *This groove is the binding site for the presentation of peptides.*

The groove can be filled by a peptide from 8 to 10 residues long in an extended conformation. As we shall see ([Section 33.5.6](#)), MHC proteins are remarkably diverse in the human population; each person expresses as many as six distinct class I MHC proteins and many different forms are present in different people. The first structure determined, HLA-A2, binds peptides that almost always have leucine in the second position and valine in the last position ([Figure 33.25](#)). Side chains from the MHC molecule interact with the amino and carboxyl termini and with the side chains in these two key positions. These residues are often referred to as the *anchor residues*. The other residues are highly variable. Thus, many millions of different peptides can be presented by this particular class I MHC protein; the identities of only two of the nine residues are crucial for binding. Each class of MHC molecules requires a unique set of anchor residues. Thus, a tremendous range of peptides can be presented by these molecules. Note that *one face of the bound peptide is exposed to solution where it can be examined by other molecules, particularly T-cell receptors*. An additional remarkable feature of MHC-peptide complexes is their kinetic stability; once bound, a peptide is not released, even over a period of days.

33.5.2. T-Cell Receptors Are Antibody-like Proteins Containing Variable and Constant Regions

We are now ready to consider the receptor that recognizes peptides displayed by MHC proteins on target cells. The *T-cell receptor* consists of a 43-kd α chain (T_α) joined by a disulfide bond to a 43-kd β chain (T_β ; [Figure 33.26](#)). Each chain spans the plasma membrane and has a short carboxyl-terminal region on the cytosolic side. A small proportion of T cells express a receptor consisting of γ and δ chains in place of α and β . T_α and T_β , like immunoglobulin L and H

chains, consist of *variable* and *constant* regions. Indeed, *these domains of the T-cell receptor are homologous to the V and C domains of immunoglobulins*. Furthermore, hypervariable sequences present in the V regions of T_α and T_β form the binding site for the epitope.

The genetic architecture of these proteins is similar to that of immunoglobulins. The variable region of T_α is encoded by about 50 V-segment genes and 70 J-segment genes. T_β is encoded by two D-segment genes in addition to 57 V and 13 J-segment genes. Again, the diversity of component genes and the use of slightly imprecise modes of joining them increase the number of distinct proteins formed. *At least 10¹² different specificities could arise from combinations of this repertoire of genes*. Thus, T-cell receptors, like immunoglobulins, can recognize a very large number of different epitopes. All the receptors on a particular T cell have the same specificity.

How do T cells recognize their targets? The variable regions of the α and β chains of the T-cell receptor form a binding site that recognizes a combined epitope-foreign peptide bound to an MHC protein (Figure 33.27). Neither the foreign peptide alone nor the MHC protein alone forms a complex with the T-cell receptor. Thus, fragments of an intracellular pathogen are presented in a context that allows them to be detected, leading to the initiation of an appropriate response.

33.5.3. CD8 on Cytotoxic T Cells Acts in Concert with T-Cell Receptors

The T-cell receptor does not act alone in recognizing and mediating the fate of target cells. Cytotoxic T cells also express a protein termed *CD8* on their surfaces that is crucial for the recognition of the class I MHC-peptide complex. The abbreviation CD stands for *cluster of differentiation*, referring to a cell-surface marker that is used to identify a lineage or stage of differentiation. Antibodies specific for particular CD proteins have been invaluable in following the development of leukocytes and in discovering new interactions between specific cell types.

Each chain in the CD8 dimer contains a domain that resembles an immunoglobulin variable domain (Figure 33.28). CD8 interacts primarily with the relatively constant α₃ domain of class I MHC proteins. This interaction further stabilizes the interactions between the T cell and its target. The cytosolic tail of CD8 contains a docking site for Lck, a cytosolic tyrosine kinase akin to Src. The T-cell receptor itself is associated with six polypeptides that form the CD3 complex (Figure 33.29). The γ, δ, and ε chains of CD3 are homologous to Ig-α and Ig-β associated with the B-cell receptor (Section 33.4.3); each chain consists of an extracellular immunoglobulin domain and an intracellular ITAM region. These chains associate into CD3 γε and CD3 δε heterodimers. An additional component, the CD3 ζ chain, has only a small extracellular domain and a larger intracellular domain containing three ITAM sequences.

On the basis of these components, a model for T-cell activation can be envisaged that is closely parallel to the pathway for B-cell activation (Section 33.3; Figure 33.30). The binding of the T-cell receptor with the class I MHC-peptide complex and the concomitant binding of CD8 from the T-cell with the MHC molecule results in the association of the kinase Lck with the ITAM substrates of the components of the CD3 complex. Phosphorylation of the tyrosine residues in the ITAM sequences generates docking sites for a protein kinase called ZAP-70 (for 70-kd zeta-associated protein) that is homologous to Syk in B cells. Docked by its two SH2 domains, ZAP-70 phosphorylates downstream targets in the signaling cascade. Additional molecules, including a membrane-bound protein phosphatase called CD45 and a cell-surface protein called CD28, play ancillary roles in this process.

T-cell activation has two important consequences. First, the activation of cytotoxic T cells results in the secretion of *perforin*. This 70-kd protein makes the cell membrane of the target cell permeable by polymerizing to form transmembrane pores 10 nm wide (Figure 33.31). The cytotoxic T cell then secretes proteases called *granzymes* into the target cell. These enzymes initiate the pathway of apoptosis (Section 18.6.6), leading to the death of the target cell and the fragmentation of its DNA, including any viral DNA that may be present. Second, after it has stimulated its target cell to commit suicide, the activated T cell disengages and is stimulated to reproduce. Thus, additional T cells that express the same T-cell receptor are generated to continue the battle against the invader after these T cells have been identified as a suitable weapon.

33.5.4. Helper T Cells Stimulate Cells That Display Foreign Peptides Bound to Class II MHC Proteins

Not all T cells are cytotoxic. *Helper T cells, a different class, stimulate the pro-liferation of specific B lymphocytes and cytotoxic T cells and thereby serve as partners in determining the immune responses that are produced.* The importance of helper T cells is graphically revealed by the devastation wrought by AIDS, a condition that destroys these cells. Helper T cells, like cytotoxic T cells, detect foreign peptides that are presented on cell surfaces by MHC proteins. However, the source of the peptides, the MHC proteins that bind them, and the transport pathway are different.

Helper T cells recognize peptides bound to MHC molecules referred to as class II. Their helping action is focused on B cells, macrophages, and dendritic cells. *Class II MHC proteins* are expressed only by these *antigen-presenting cells*, unlike class I MHC proteins, which are expressed on nearly all cells. The peptides presented by class II MHC proteins do not come from the cytosol. Rather, *they arise from the degradation of proteins that have been internalized by endocytosis.* Consider, for example, a virus particle that is captured by membrane-bound immunoglobulins on the surface of a B cell (Figure 33.32). This complex is delivered to an endosome, a membrane-enclosed acidic compartment, where it is digested. The resulting peptides become associated with class II MHC proteins, which move to the cell surface. Peptides from the cytosol cannot reach class II proteins, whereas peptides from endosomal compartments cannot reach class I proteins. This segregation of displayed peptides is biologically critical. The association of a foreign peptide with a class II MHC protein signals that a cell has *encountered* a pathogen and serves as a call for *help*. In contrast, association with a class I MHC protein signals that a cell has *succumbed* to a pathogen and is a call for *destruction*.

33.5.5. Helper T Cells Rely on the T-Cell Receptor and CD4 to Recognize Foreign Peptides on Antigen-Presenting Cells

The overall structure of a class II MHC molecule is remarkably similar to that of a class I molecule. Class II molecules consist of a 33-kd α chain and a noncovalently bound 30-kd β chain (Figure 33.33).


Each contains two extracellular domains, a transmembrane segment, and a short cytosolic tail. The peptide-binding site is formed by the α_1 and β_1 domains, each of which contributes a long helix and part of a β sheet. Thus, the same structural elements are present in class I and class II MHC molecules, but they are combined into polypeptide chains in different ways. Class II MHC molecules appear to form stable dimers, unlike class I molecules, which are monomeric. The peptide-binding site of a class II molecule is open at both ends, and so this groove can accommodate longer peptides than can be bound by class I molecules; typically, peptides between 13 and 18 residues long are bound. The peptide-binding specificity of each class II molecule depends on binding pockets that recognize particular amino acids in specific positions along the sequence.

Helper T cells express T-cell receptors that are produced from the same genes as those on cytotoxic T cells. These T-cell receptors interact with class II MHC molecules in a manner that is analogous to T-cell-receptor interaction with class I MHC molecules. Nonetheless, helper T cells and cytotoxic T cells are distinguished by other proteins that they express on their surfaces. In particular, helper T cells express a protein called CD4 instead of expressing CD8. *CD4* consists of four immunoglobulin domains that extend from the T-cell surface, as well as a small cytoplasmic region (Figure 33.34). The amino-terminal immunoglobulin domains of CD4 interact with the base of the class II MHC molecule. Thus, helper T cells bind cells expressing class II MHC specifically because of the interactions with CD4 (Figure 33.35).

When a helper T cell binds to an antigen-presenting cell expressing an appropriate class II MHC-peptide complex, signaling pathways analogous to those in cytotoxic T cells are initiated by the action of the kinase Lck on ITAMs in the CD3 molecules associated with the T-cell receptor. However, rather than triggering events leading to the death of the attached cell, *these signaling pathways result in the secretion of cytokines from the helper cell.* Cytokines are a family of molecules that include, among others, interleukin-2 and interferon- γ . Cytokines bind to specific receptors on the antigen-presenting cell and stimulate growth, differentiation, and in regard to plasma cells, which are derived from B cells, antibody secretion (Figure 33.36). Thus, the internalization and presentation of parts of a foreign pathogen help to generate a local environment in which cells taking part in the defense against this pathogen can flourish through the

action of helper T cells.

33.5.6. MHC Proteins Are Highly Diverse

 MHC class I and II proteins, the presenters of peptides to T cells, were discovered because of their role in *transplantation rejection*. A tissue transplanted from one person to another or from one mouse to another is usually rejected by the immune system. In contrast, tissues transplanted from one identical twin to another or between mice of an inbred strain are accepted. Genetic analyses revealed that rejection occurs when tissues are transplanted between individuals having different genes in the major histocompatibility complex, a cluster of more than 75 genes playing key roles in immunity. The 3500-kb span of the MHC is nearly the length of the entire *E. coli* chromosome. The MHC encodes class I proteins (presenters to cytotoxic T cells) and class II proteins (presenters to helper T cells), as well as class III proteins (components of the complement cascade) and many other proteins that play key roles in immunity.


Human beings express six different class I genes (three from each parent) and six different class II genes. The three loci for class I genes are called HLA-A, -B, and -C; those for class II genes are called HLA-DP, -DQ, and -DR. These loci are *highly polymorphic*: many alleles of each are present in the population. For example, more than 50 each of HLA-A, -B, and -C alleles are known; the numbers discovered increase each year. Hence, the likelihood that two unrelated persons have identical class I and II proteins is very small ($<10^{-4}$), accounting for transplantation rejection unless the genotypes of donor and acceptor are closely matched in advance.

Differences between class I proteins are located mainly in the α_1 and α_2 domains, which form the peptide-binding site (Figure 33.37). The α_3 domain, which interacts with a constant β_2 -microglobulin is largely conserved. Similarly, the differences between class II proteins cluster near the peptide-binding groove. Why are MHC proteins so highly variable? *Their diversity makes possible the presentation of a very wide range of peptides to T cells. A particular class I or class II molecule may not be able to bind any of the peptide fragments of a viral protein. The likelihood of a fit is markedly increased by having several kinds (usually six) of each class of presenters in each individual. If all members of a species had identical class I or class II molecules, the population would be much more vulnerable to devastation by a pathogen that had evolved to evade presentation. The evolution of the diverse human MHC repertoire has been driven by the selection for individual members of the species who resist infections to which other members of the population may be susceptible.*

33.5.7. Human Immunodeficiency Viruses Subvert the Immune System by Destroying Helper T Cells

In 1981, the first cases of a new disease now called *acquired immune deficiency syndrome (AIDS)* were recognized. The victims died of rare infections because their immune systems were crippled. The cause was identified two years later by Luc Montagnier and coworkers. AIDS is produced by *human immunodeficiency virus (HIV)*, of which two major classes are known: HIV-1 and the much less common HIV-2. Like other *retroviruses*, HIV contains a single-stranded RNA genome that is replicated through a double-stranded DNA intermediate. This viral DNA becomes integrated into the genome of the host cell. In fact, viral genes are transcribed only when they are integrated into the host DNA.

The HIV virion is enveloped by a lipid bilayer membrane containing two glycoproteins: gp41 spans the membrane and is associated with gp120, which is located on the external face (Figure 33.38). The core of the virus contains two copies of the RNA genome and associated transfer RNAs, and several molecules of reverse transcriptase. They are surrounded by many copies of two proteins called p18 and p24. *The host cell for HIV is the helper T cell.* The gp120 molecules on the membrane of HIV bind to CD4 molecules on the surface of the helper T cell (Figure 33.39). This interaction allows the associated viral gp41 to insert its amino-terminal head into the host-cell membrane. The viral membrane and the helper-cell membrane fuse, and the viral core is released directly into the cytosol. Infection by HIV leads to the destruction of helper T cells because the permeability of the host plasma membrane is markedly increased by the insertion of viral glycoproteins and the budding of virus particles. The influx of ions and water disrupts the ionic balance, causing osmotic lysis.

 The development of an effective AIDS vaccine is difficult owing to the antigenic diversity of HIV strains. Because its mechanism for replication is quite error prone, a population of HIV presents an everchanging array of coat proteins. Indeed, the mutation rate of HIV is more than 65 times as high as that of influenza virus. A major aim now is to define relatively conserved sequences in these HIV proteins and use them as immunogens.

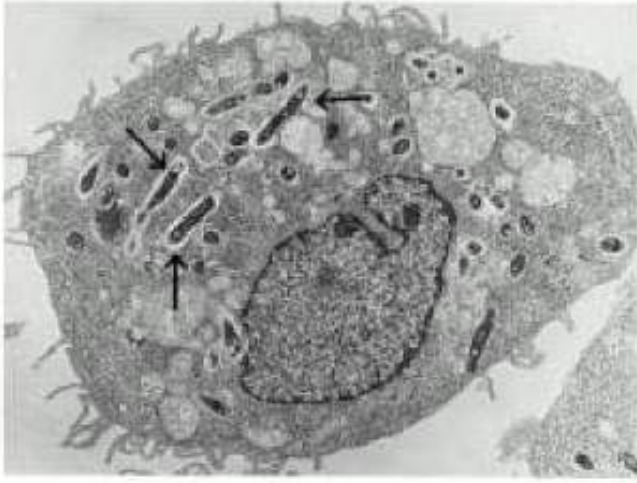


Figure 33.21. Intracellular Pathogen. An electron micrograph showing mycobacteria (arrows) inside an infected macrophage. [Courtesy of Dr. Stanley Falkow.]

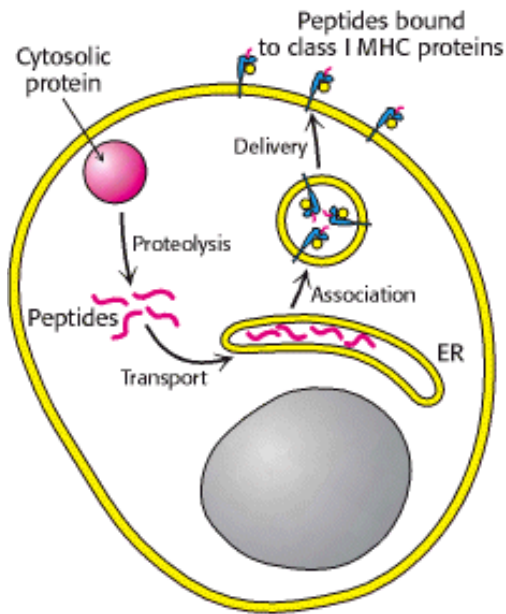


Figure 33.22. Presentation of Peptides from Cytosolic Proteins. Class I MHC proteins on the surfaces of most cells display peptides that are derived from cytosolic proteins by proteolysis.

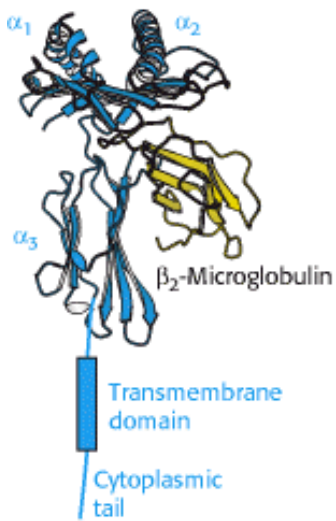


Figure 33.23. Class I MHC Protein. A protein of this class consists of two chains. The α chain begins with two domains that include α helices (α_1 , α_2), an immunoglobulin domain (α_3), a transmembrane domain, and a cytoplasmic tail. The second chain, β_2 -microglobulin, adopts an immunoglobulin fold.

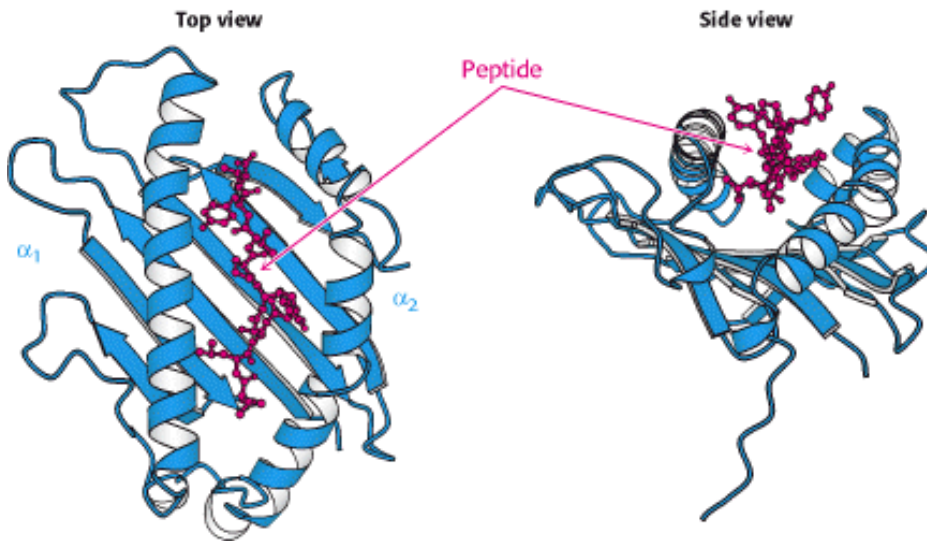


Figure 33.24. Class I MHC Peptide-Binding Site. The α_1 and α_2 domains come together to form a groove in which peptides are displayed. The two views shown reveal that the peptide is surrounded on three sides by a β sheet and two α helices, but it is accessible from the top of the structure.

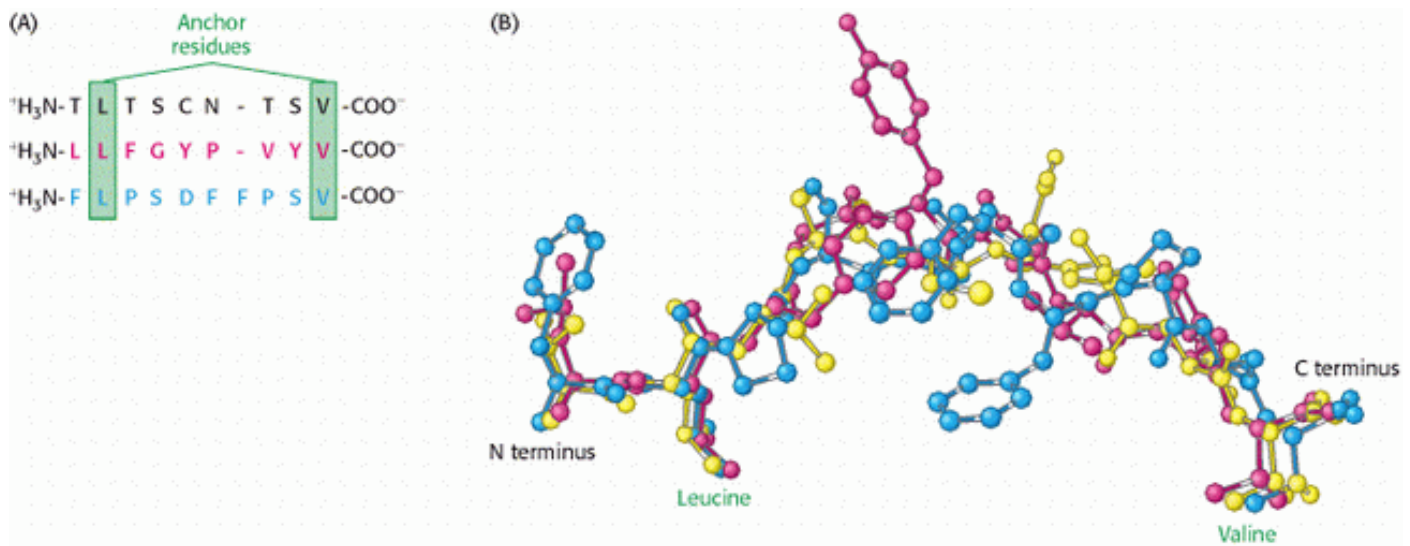


Figure 33.25. Anchor Residues. (A) The amino acid sequences of three peptides that bind to the class I MHC protein HLA-A2 are shown. Each of these peptides has leucine in the second position and valine in the carboxyl-terminal position. (B) Comparison of the structures of these peptides reveals that the amino and carboxyl termini as well as the side chains of the leucine and valine residues are in essentially the same position in each peptide, whereas the remainder of the structures are quite different.

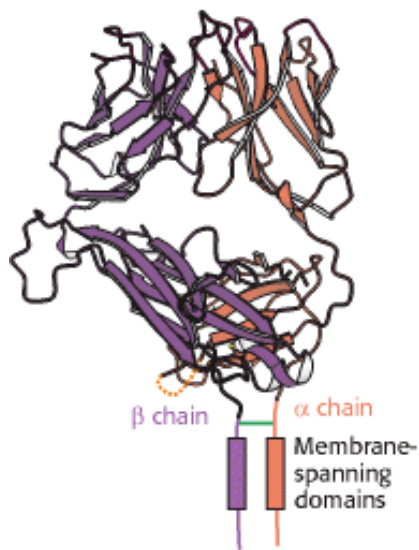


Figure 33.26. T-Cell Receptor. This protein consists of an α chain and a β chain, each of which consists of two immunoglobulin domains and a membrane-spanning domain. The two chains are linked by a disulfide bond.

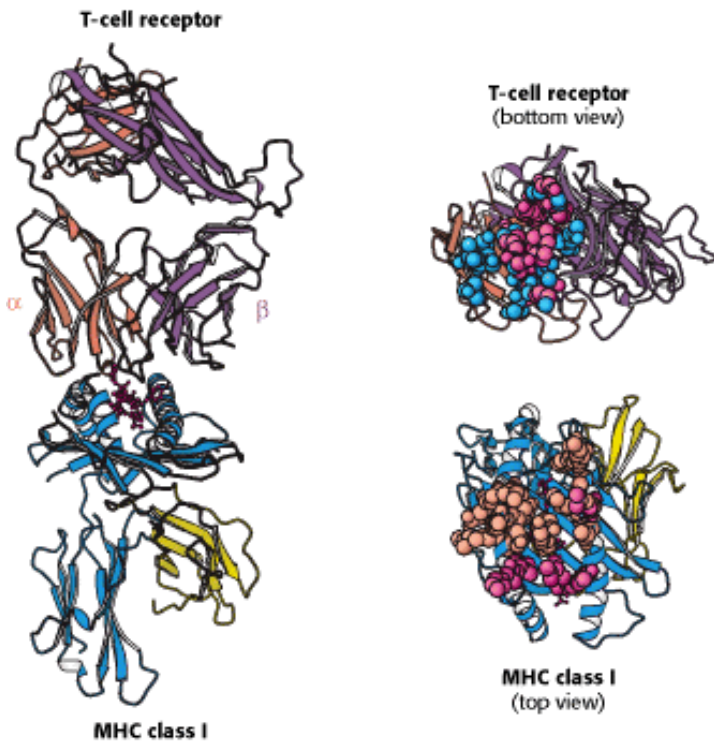


Figure 33.27. T-Cell Receptor Class I MHC Complex. The T-cell receptor binds to a class I MHC protein containing a bound peptide. The T-cell receptor contacts both the MHC protein and the peptide as shown by surfaces exposed when the complex is separated (right). These surfaces are colored according to the chain that they contact.

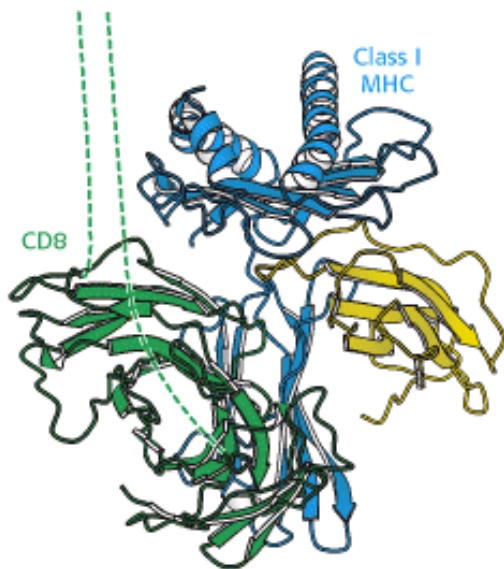


Figure 33.28. The Coreceptor CD8. This dimeric protein extends from the surface of a cytotoxic T cell and binds to class I MHC molecules that are expressed on the surface of the cell that is bound to the T cell. The dashed lines represent extended polypeptide chains that link the immunoglobulin domain of CD8 to the membrane.

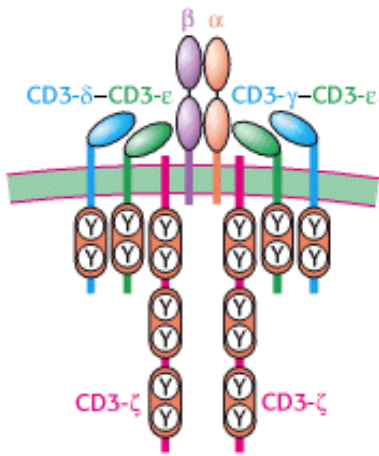


Figure 33.29. T-Cell Receptor Complex. The T-cell receptor is associated with six CD3 molecules: a CD3- γ - CD3- ϵ heterodimer, a CD3- δ - CD3- ϵ heterodimer, and two chains of CD3- ζ . Single ITAM sequences are present in the cytoplasmic domains of CD3- γ , CD3- δ , and CD3- ϵ whereas three such sequences are found in each CD3- ζ chain.

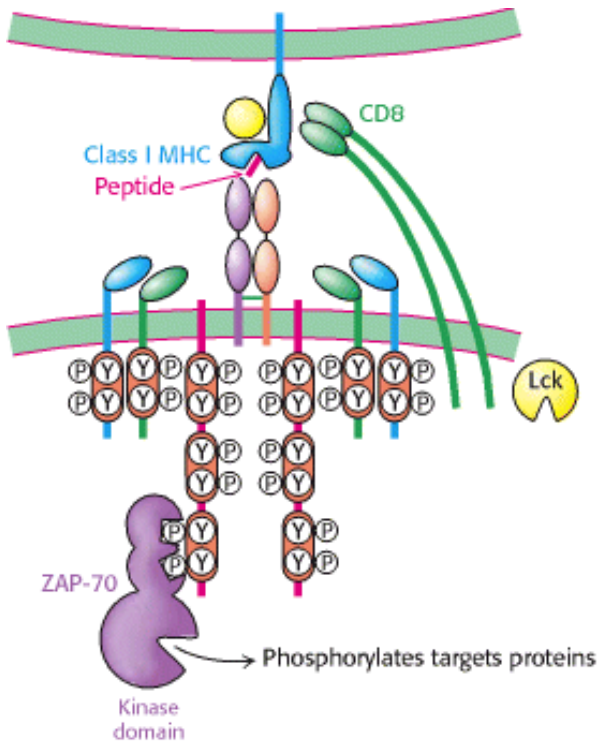


Figure 33.30. T-Cell Activation. The interaction between the T-cell receptor and a class I MHC-peptide complex results in the binding of CD8 to the MHC protein, the recruitment of the protein tyrosine kinase Lck, and the phosphorylation of tyrosine residues in the ITAM sequences of the CD3 chains. After phosphorylation, the ITAM regions serve as docking sites for the protein kinase ZAP-70, which phosphorylates protein targets to transmit the signal.

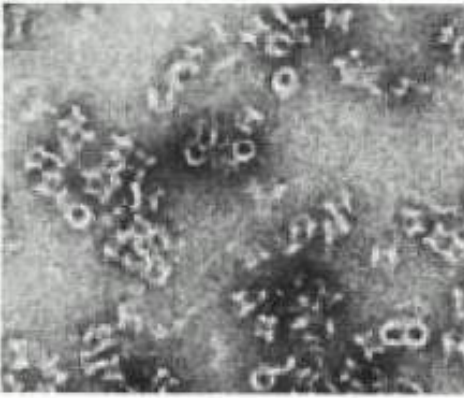


Figure 33.31. Consequences of Cytotoxic-T-Cell Action. An electron micrograph showing pores in the membrane of a cell that has been attacked by a cytotoxic T cell. The pores are formed by the polymerization of perforin, a protein secreted by the cytotoxic T cell. [Courtesy of Dr. Eckhard Podock.]

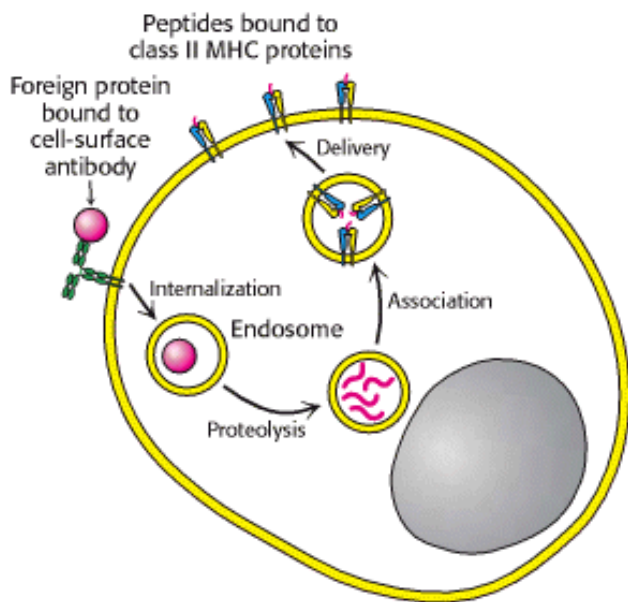


Figure 33.32. Presentation of Peptides from Internalized Proteins. Antigen-presenting cells bind and internalize foreign proteins and display peptides that are formed from the digestion of these proteins in Class II MHC proteins.

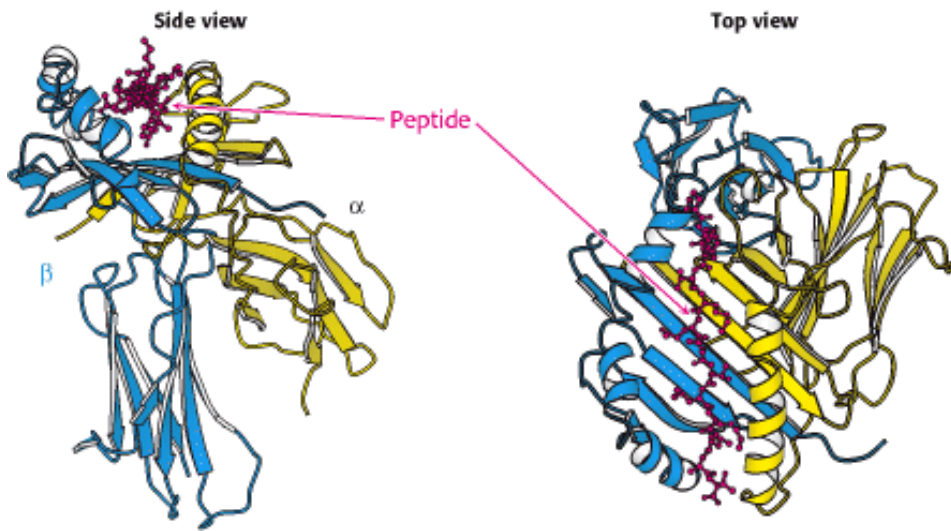


Figure 33.33. Class II MHC Protein. A class II MHC protein consists of homologous α and β chains, each of which has an amino-terminal domain that constitutes half of the peptide-binding structure, as well as a carboxyl-terminal immunoglobulin domain. The peptide-binding site is similar to that in class I MHC proteins except that it is open at both ends, allowing class II MHC proteins to bind longer peptides than those bound by class I.

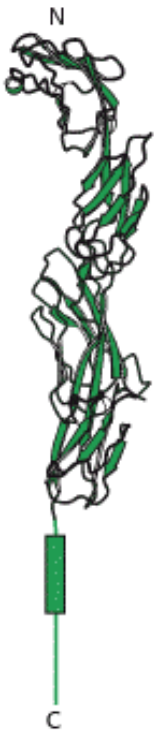


Figure 33.34. Coreceptor CD4. This protein comprises four tandem immunoglobulin domains that extend from the surface of helper T cells.

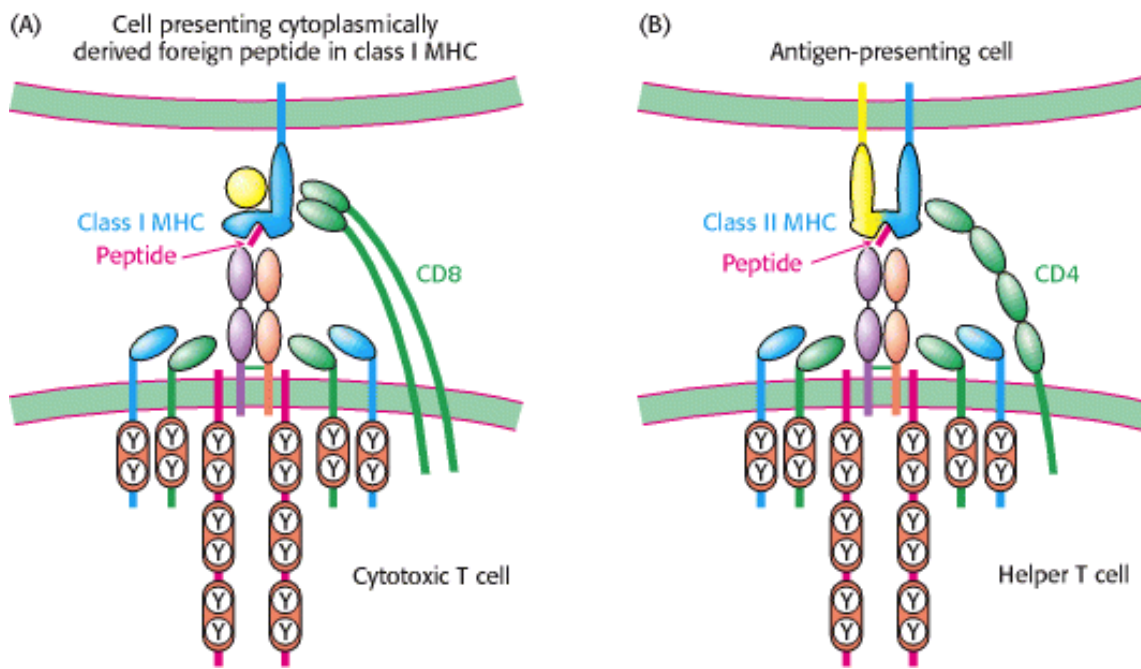


Figure 33.35. Variations on a Theme. (A) Cytotoxic T cells recognize foreign peptides presented in class I MHC proteins with the aid of the coreceptor CD8. (B) Helper T cells recognize peptides presented in class II MHC proteins by specialized antigen-presenting cells with the aid of the coreceptor CD4.

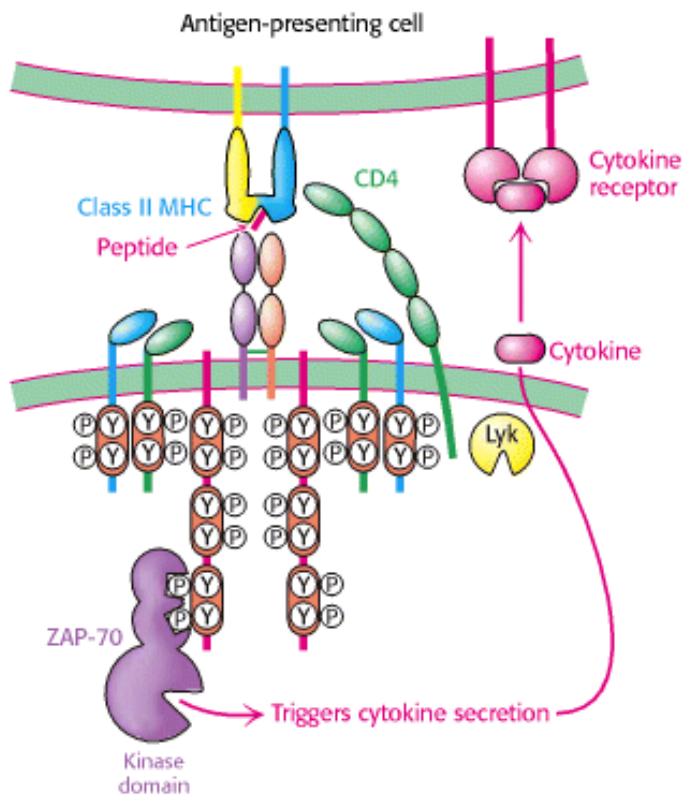


Figure 33.36. Helper T Cell Action. The engagement of the T-cell receptor in helper T cells results in the secretion of cytokines. These cytokines bind to cytokine receptors expressed on the surface of the antigen-presenting cell, stimulating cell growth, differentiation, and, in regard to a B cell, antibody secretion.

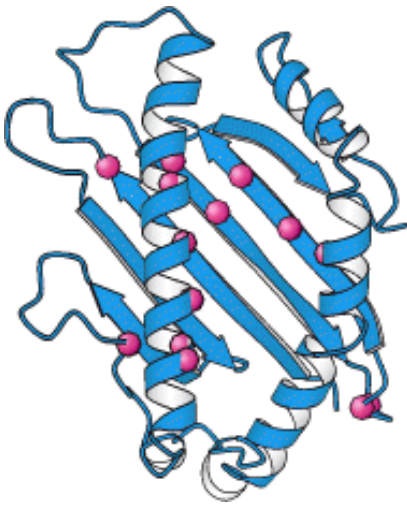


Figure 33.37. Polymorphism in Class I MHC Protein. The positions of sites with a high degree of polymorphism in the human population are displayed as red spheres on the structure of the amino-terminal part of a class I MHC protein.

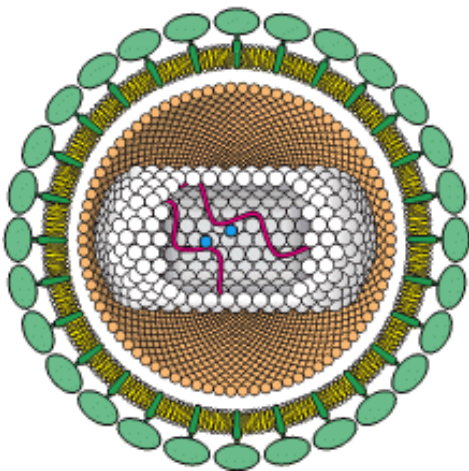


Figure 33.38. Human Immunodeficiency Virus. A schematic diagram of HIV reveals its proteins and nucleic acid components. The membrane-envelope glycoproteins gp41 and gp120 are shown in dark and light green. The viral RNA is shown in red, and molecules of reverse transcriptase are shown in blue. [After R. C. Gallo. The AIDS virus. Copyright © 1987 by Scientific American, Inc. All rights reserved.]

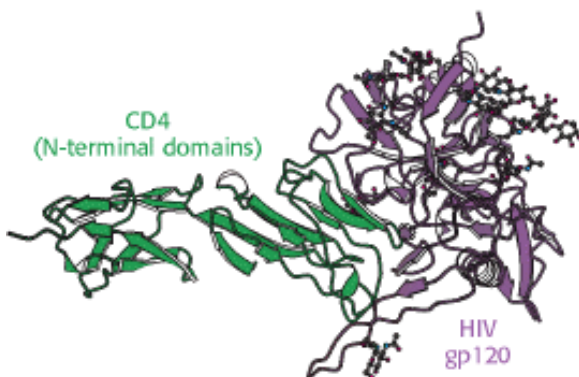



Figure 33.39. HIV Receptor. A complex between a modified form of the envelope glycoprotein gp120 from HIV and a

 peptide corresponding to the two amino-terminal domains from the helper T-cell protein CD4 reveals how viral infection of helper T cells is initiated.

33.6. Immune Responses Against Self-Antigens Are Suppressed

The primary function of the immune system is to protect the host from invasion by foreign organisms. But how does the immune system avoid mounting attacks against the host organism? In other words, how does the immune system distinguish between self and nonself? Clearly, proteins from the organism itself do not bear some special tag identifying them. Instead, selection processes early in the developmental pathways for immune cells kill or suppress those immune cells that react strongly with self-antigens. The evolutionary paradigm still applies; immune cells that recognize self-antigens are generated, but selective mechanisms eliminate such cells in the course of development.


33.6.1. T Cells Are Subject to Positive and Negative Selection in the Thymus

T cells derive their name from the location of their production — the thymus, a small organ situated just above the heart. Examination of the developmental pathways leading to the production of mature cytotoxic and helper T cells reveals the selection mechanisms that are crucial for distinguishing self from nonself. These selection criteria are quite stringent; approximately 98% of the thymocytes, the precursors of T cells, die before the completion of the maturation process.

Thymocytes produced in the bone marrow do not express the T-cell receptor complex, CD4, or CD8. On relocation to the thymus and rearrangement of the T-cell-receptor genes, the immature thymocyte expresses all of these molecules. These cells are first subjected to *positive selection* (Figure 33.40). Cells for which the T-cell receptor can bind with reasonable affinity to either class I or class II MHC molecules survive this selection; those for which the T-cell receptor does not participate in such an interaction undergo apoptosis and die. The affinities of interaction required to pass this selection are relatively modest, and so contacts between the T-cell receptor and the MHC molecules themselves are sufficient without any significant contribution from the bound peptides (which will be derived from proteins in the thymus). *The role of the positive selection step is to prevent the production of T cells that will not bind to any MHC complex present, regardless of the peptide bound.*


The cell population that survives positive selection is subjected to a second step, *negative selection*. Here, T cells that bind with high affinity to MHC complexes bound to self-peptides expressed on the surfaces of antigen-presenting cells in the thymus undergo apoptosis or are otherwise suppressed. Those that do not bind too avidly to any such MHC complex complete development and become mature cytotoxic T cells (which express only CD8) or helper T cells (which express only CD4). The negative selection step leads to *self tolerance*; cells that bind an MHC-self-peptide complex are removed from the T-cell population. Similar mechanisms apply to developing B cells, suppressing B cells that express antibodies that interact strongly with self-antigens.

33.6.2. Autoimmune Diseases Result from the Generation of Immune Responses Against Self-Antigens

 Although thymic selection is remarkably efficient in suppressing the immune response to self-antigens, failures do occur. Such failures result in *autoimmune diseases*. These diseases include relatively common illnesses such as insulin-dependent diabetes mellitus, multiple sclerosis, and rheumatoid arthritis. In these illnesses, immune responses against self-antigens result in damage to selective tissues that express the antigen (Figure 33.41).

In many cases, the cause of the generation of self-reactive antibodies or T cells is unclear. However, in other cases, infectious organisms such as bacteria or viruses may play a role. Infection leads to the generation of antibodies and T cells that react with many different epitopes from the infectious organism. If one of these antigens closely resembles a self-antigen, an autoimmune response can result. For example, *Streptococcus* infections sometimes lead to rheumatic fever owing to the production of antibodies to streptococcal antigens that cross-react with exposed molecules in heart muscle.

33.6.3. The Immune System Plays a Role in Cancer Prevention

 The development of immune responses against proteins encoded by our own genomes can be beneficial under some circumstances. Cancer cells have undergone significant changes that often result in the expression of proteins that are not normally expressed. For example, the mutation of genes can generate proteins that do not correspond in amino acid sequence to any normal protein. Such proteins may be recognized as foreign, and an immune response will be generated specifically against the cancer cell. Alternatively, cancer cells often produce proteins that are expressed during embryonic development but are not expressed or are expressed at very low levels after birth. For example, a membrane glycoprotein protein called *carcinoembryonic antigen (CEA)* appears in the gastrointestinal cells of developing fetuses but is not normally expressed at significant levels after birth. More than 50% of patients with colorectal cancer have elevated serum levels of CEA. Immune cells recognizing epitopes from such proteins will not be subject to negative selection and, hence, will be present in the adult immune repertoire. These cells may play a cancer surveillance role, killing cells that overexpress antigens such as CEA and preventing genetically damaged cells from developing into tumors.

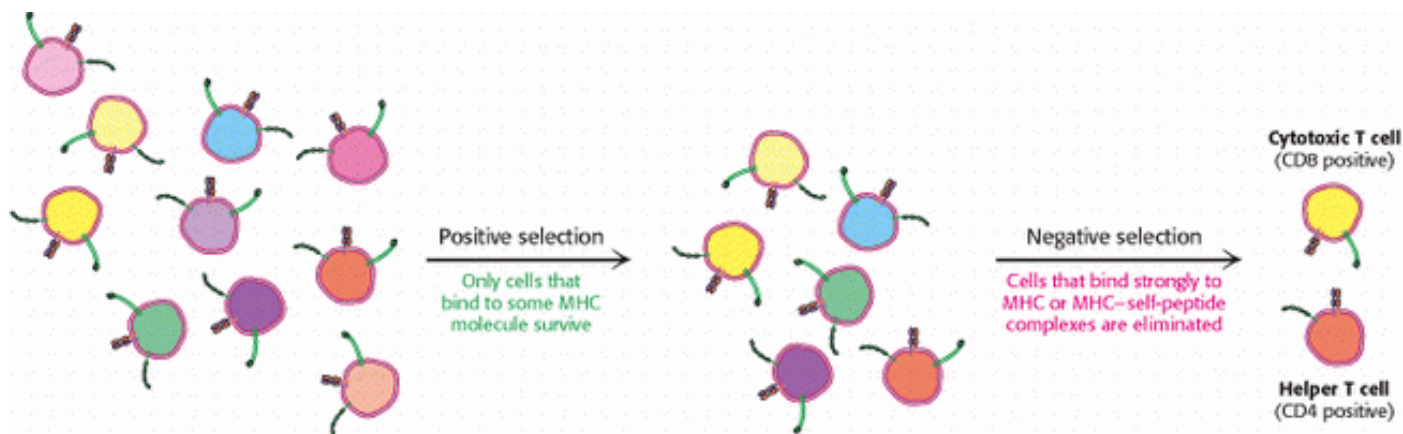
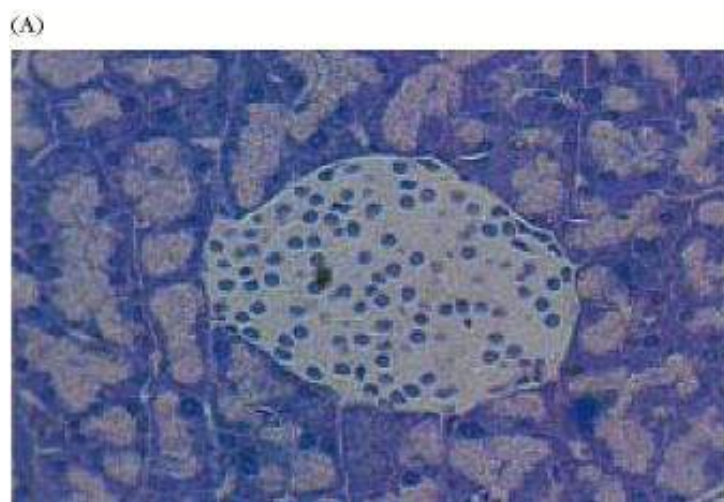


Figure 33.40. T-Cell Selection. A population of thymocytes is subjected first to positive selection to remove cells that express T-cell receptors that will not bind to MHC proteins expressed by the individual organism. The surviving cells are then subjected to negative selection to remove cells that bind strongly to MHC complexes bound to self-peptides.



(B)

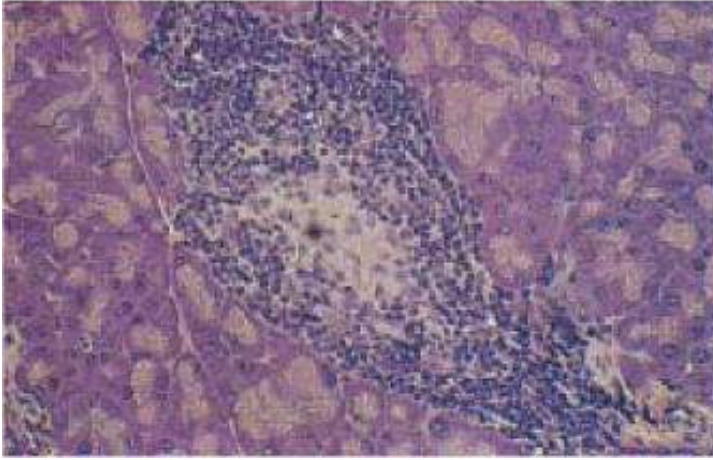


Figure 33.41. Consequences of Autoimmunity. Photo-micrographs of an islet of Langerhans (A) in the pancreas of a normal mouse and (B) in the pancreas of a mouse with an immune response against pancreatic β cells, which results in a disease resembling insulin-dependent diabetes mellitus in human beings. [From M. A. Atkinson and N. K. Maclaren. What causes diabetes? Copyright © 1990 by Scientific American, Inc. All rights reserved.]

Summary

To respond effectively to a vast array of pathogens, the immune system must be tremendously adaptable. Adaptation by the immune system follows the principles of evolution: an enormously diverse set of potentially useful proteins is generated; these proteins are then subjected to intense selection so that only cells that express useful proteins flourish and continue development, until an effective immune response to a specific invader is generated.

Antibodies Possess Distinct Antigen-Binding and Effector Units

The major immunoglobulin in the serum is immunoglobulin G. An IgG protein is a heterotetramer with two heavy chains and two light chains. Treatment of IgG molecules with proteases such as papain produces three fragments: two F_{ab} fragments that retain antigen-binding activity and an F_c fragment that retains the ability to activate effector functions such as the initiation of the complement cascade. The F_{ab} fragments include the L chain and the amino-terminal half of the H chain; the F_c domain is a dimer consisting of the carboxyl-terminal halves of two H chains. Five different classes of antibody — IgG, IgM, IgA, IgD, and IgE — differ in their heavy chains and, hence, in their effector functions.

The Immunoglobulin Fold Consists of a Beta-Sandwich Framework with Hypervariable Loops

One particular protein fold is found in many of the key proteins of the immune system. The immunoglobulin fold consists of a pair of β sheets that pack against one another, linked by a single disulfide bond. Loops projecting from one end of the structure form a binding surface that can be varied by changing the amino acid sequences within the loops. Domains with immunoglobulin folds are linked to form antibodies and other classes of proteins in the immune system including T-cell receptors.

Antibodies Bind Specific Molecules Through Their Hypervariable Loops

Two chains come together to form the binding surface of an antibody. Three loops from each domain, the complementarity-determining regions, form an essentially continuous surface that can vary tremendously in shape, charge, and other characteristics to allow particular antibodies to bind to molecules ranging from small molecules to

large protein surfaces.

Diversity Is Generated by Gene Rearrangements

The tremendous diversity of the amino acid sequences of antibodies is generated by segmental rearrangements of genes. For antibody κ light chains, one of 40 variable regions is linked to one of five joining regions. The combined VJ unit is then linked to the constant region. Thousands of different genes can be generated in this manner. Similar arrays are rearranged to form the genes for the heavy chains, but an additional region called the diversity region lies between the V and the J regions. The combination L and H chains, each obtained through such rearranged genes, can produce more than 10^8 distinct antibodies. Different classes of antibodies are also generated by gene rearrangements that lead to class switching. Oligomerization of membrane-bound antibody molecules initiates a signal-transduction cascade inside B cells. Key steps in this signaling process include the phosphorylation of specific tyrosine residues in sequences termed immunoreceptor tyrosine-based activation motifs (ITAMs), present in proteins that associate with the membrane-bound antibodies.

Major-Histocompatibility-Complex Proteins Present Peptide Antigens on Cell Surfaces for Recognition by T-Cell Receptors

Intracellular pathogens such as viruses and mycobacteria cannot be easily detected. Intracellular proteins are constantly being cut into small peptides by proteasomes and displayed in class I major-histocompatibility-complex proteins on cell surfaces. Such peptides lie in a groove defined by two helices in the class I MHC proteins. The combination of MHC protein and peptide can be bound by an appropriate T-cell receptor. T-cell receptors resemble the antigen-binding domains of antibodies in structure, and diversity in T-cell-receptor sequence is generated by V(D)J gene rearrangements. The T-cell receptor recognizes features of both the peptide and the MHC molecule that presents it. Cytotoxic T cells initiate apoptosis in cells to which they bind through T-cell receptor-class I MHC-peptide interactions aided by interactions with the coreceptor molecule CD8. Helper T cells recognize peptides presented in class II MHC proteins, a distinct type of MHC protein expressed only on antigen-presenting cells such as B cells and macrophages. Helper T cells express the coreceptor CD4, rather than CD8. CD4 interacts with class II MHC proteins present on antigen-presenting cells. Signaling pathways, analogous to those in B cells, are initiated by interactions between MHC-peptide complexes and T-cell receptors and the CD8 and CD4 coreceptors. Human immunodeficiency virus damages the immune system by infecting cells that express CD4, such as helper T cells.

Immune Responses Against Self-Antigens Are Suppressed

In principle, the immune system is capable of generating antibodies and T-cell receptors that bind to self-molecules; that is, molecules that are normally present in a healthy and uninfected individual organism. Selection mechanisms prevent such self-directed molecules from being expressed at high levels. The selection process includes both positive selection, to enrich the population of cells that express molecules that have the potential to bind foreign antigens in an appropriate context, and negative selection, which eliminates cells that express molecules with too high an affinity for self-antigens. Autoimmune diseases such as insulin-dependent diabetes mellitus can result from amplification of a response against a self-antigen.

Key Terms

humoral immune response

B lymphocyte (B cell)

antigen

antigenic determinant (epitope)

cellular immune response

cytotoxic T lymphocyte (killer T cell)

helper T lymphocyte

immunoglobulin G

F_{ab}

F_c

light chain

heavy chain

segmental flexibility

immunoglobulin M

immunoglobulin A

immunoglobulin D

immunoglobulin E

variable region

constant region

immunoglobulin fold

hypervariable loop

complementarity-determining region (CDR)

V(D)J recombination

immunoreceptor tyrosine-based activation motif (ITAM)

cyclosporin

hapten

class switching

T cell

major histocompatibility complex (MHC)

class I MHC protein

human leukocyte antigen (HLA)

β_2 -microglobulin

T-cell receptor

CD8

perforin

granzymes

helper T cell

class II MHC protein

CD4

human immunodeficiency virus (HIV)

positive selection

negative selection

autoimmune disease

carcinoembryonic antigen (CEA)

Problems

1. *Energetics and kinetics.* Suppose that the dissociation constant of an F_{ab} - hapten complex is 3×10^{-7} M at 25°C .

See answer

(a) What is the standard free energy of binding?

(b) Immunologists often speak of affinity (K_a), the reciprocal of the dissociation constant, in comparing antibodies. What is the affinity of this F_{ab} ?

(c) The rate constant of release of hapten from the complex is 120 s^{-1} . What is the rate constant for association? What does the magnitude of this value imply about the extent of structural change in the antibody on binding hapten?

2. *Sugar niche.* An antibody specific for dextran, a polysaccharide of glucose residues, was tested for its binding of glucose oligomers. Maximal binding affinity was obtained when the oligomer contained six glucose residues. How does the size of this site compare with that expected for the binding site on the surface of an antibody?

See answer

3. *A brilliant emitter.* Certain naphthalene derivatives exhibit a weak yellow fluorescence when they are in a highly polar environment (such as water) and an intense blue fluorescence when they are in a markedly nonpolar environment (such as hexane). The binding of ϵ -dansyl-lysine to specific antibody is accompanied by a marked increase in its fluorescence intensity and a shift in color from yellow to blue. What does this finding reveal about the hapten-antibody complex?

See answer

4. *Avidity versus affinity.* The standard free energy of binding of F_{ab} derived from an antiviral IgG is -7 kcal mol^{-1} (-29 kJ mol^{-1}) at 25°C .

(a) Calculate the dissociation constant of this interaction.

(b) Predict the dissociation constant of the intact IgG, assuming that both combining sites of the antibody can interact with viral epitopes and that the free-energy cost of assuming a favorable hinge angle is $+3 \text{ kcal mol}^{-1}$ (12.6 kJ mol^{-1}).

See answer

5. *Miniantibody.* The F_{ab} fragment of an antibody molecule has essentially the same affinity for a monovalent hapten as does intact IgG.

(a) What is the smallest unit of an antibody that can retain the specificity and binding affinity of the whole protein?

(b) Design a compact single-chain protein that is likely to specifically bind antigen with high affinity.

See answer

6. *Turning on B cells.* B lymphocytes, the precursors of plasma cells, are triggered to proliferate by the binding of multivalent antigens to receptors on their surfaces. The cell-surface receptors are transmembrane immunoglobulins. Univalent antigens, in contrast, do not activate B cells.

(a) What do these findings reveal about the mechanism of B-cell activation?

(b) How might antibodies be used to activate B cells?

See answer

7. *An ingenious cloning strategy.* In the cloning of the gene for the α chain of the T-cell receptor, T-cell cDNAs were hybridized with B-cell mRNAs. What was the purpose of this hybridization step? Can the principle be applied generally?

See answer

8. *Instruction.* Before the mechanism for generating antibody diversity had been established, a mechanism based on protein folding around an antigen was proposed, primarily by Linus Pauling. In this model, antibodies that had different specificities had the same amino acid sequence but were folded in different ways. Propose a test of this model.

See answer

9. *Dealing with nonsense.* Cells, including immune cells, degrade mRNA molecules in which no long open reading frame is present. The process is called nonsense-mediated RNA decay. Suggest a role for this process in immune cells.

See answer

10. *Crystallization.* The proteolytic digestion of a population of IgG molecules isolated from human serum results in the generation of F_{ab} and F_c fragments. Why do F_c fragments crystallize more easily than F_{ab} fragments generated from such a population?

See answer

11. *Presentation.* The amino acid sequence of a small protein is:

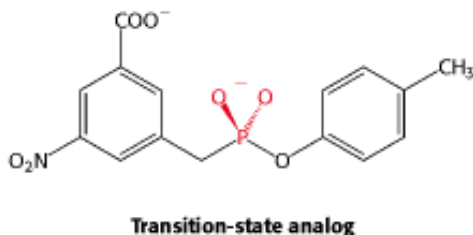
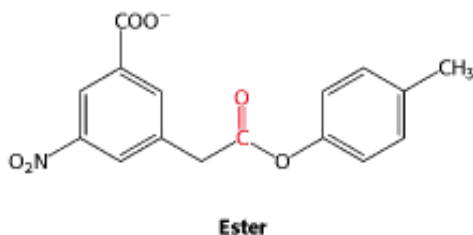
MSRLASKNLIRSDHAGLLQATYSAVSS-
IKNTMSFGAWSNAALNDSRDA

Predict the most likely peptide to be presented by the class I MHC molecule HLA-A2.

See answer

Mechanism Problem

12. *Catalytic antibody.* Antibody is generated against a transition state for the hydrolysis of the following ester.



Some of these antibodies catalyze the hydrolysis of the ester. What amino acid residue might you expect to find in the binding site on the antibody?

See answer

Chapter Integration Problem

13. *Signaling.* Protein tyrosine phosphatases, such as the molecule CD45 expressed in both B cells and T cells, play important roles in activating such protein tyrosine kinases as Fyn and Lck, which are quite similar to Src. Suggest a mechanism for the activation of such protein kinases by the removal of a phosphate from a phosphotyrosine residue.

See answer

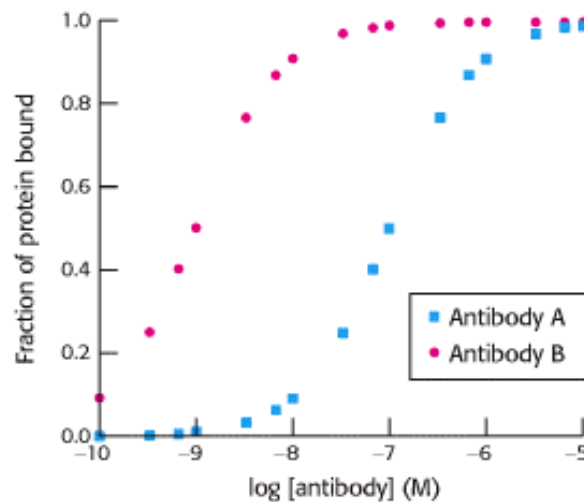
Data Interpretation Problem

14. *Affinity maturation.* A mouse is immunized with an oligomeric human protein. Shortly after immunization, a cell line

that expresses a single type of antibody molecule (antibody A) is derived. The ability of antibody A to bind the human protein is assayed with the results shown in the adjoining graph. After repeated immunizations with the same protein, another cell line is derived that expresses a different antibody (antibody B). The results of analyzing the binding of antibody B to the protein also are shown. From these data, estimate

(a) the dissociation constant (K_d) for the complex between the protein and antibody A.

(b) the dissociation constant for the complex between the protein and antibody B.



Comparison of the amino acid sequences of antibody A and antibody B reveals them to be identical except for a single amino acid. What does this finding suggest about the mechanism by which the gene encoding antibody B was generated?

See answer

Selected Readings

Where to start

G.J.V. Nossal. 1993. Life, death, and the immune system *Sci. Am.* 269: (3) 53-62.

S. Tonegawa. 1985. The molecules of the immune system *Sci. Am.* 253: (4) 122-131. ([PubMed](#))

P. Leder. 1982. The genetics of antibody diversity *Sci. Am.* 246: (5) 102-115. ([PubMed](#))

S.K. Bromley, W.R. Burack, K.G. Johnson, K. Somersalo, T.N. Sims, C. Sumen, M.M. Davis, A.S. Shaw, P.M. Allen, and M.L. Dustin. 2001. The immunological synapse *Annu. Rev. Immunol.* 19:: 375-396. ([PubMed](#))

Books

Goldsby R. A. Kindt T. J. Osborne B. A. 2000. *Kuby Immunology* (4th ed.). W. H. Freeman and Company.

Abbas, A. K., Lichtman, A. H., and Pober, J. S., 1992. *Cellular and Molecular Immunology* (2d ed). Saunders.

Cold Spring Harbor Symposia on Quantitative Biology, 1989. Volume 54. Immunological Recognition.

Nisino, A., 1985. *Introduction to Molecular Immunology* (2d ed.). Sinauer.

Weir, D. M. (Ed.), 1986. *Handbook of Experimental Immunology*. Oxford University Press.

Structure of antibodies and antibody-antigen complexes

D.R. Davies, E.A. Padlan, and S. Sheriff. 1990. Antibody-antigen complexes *Annu. Rev. Biochem.* 59:: 439-473. ([PubMed](#))

R.J. Poljak. 1991. Structure of antibodies and their complexes with antigens *Mol. Immunol.* 28:: 1341-1345. ([PubMed](#))

D.R. Davies and G.H. Cohen. 1996. Interactions of protein antigens with antibodies *Proc. Natl. Acad. Sci. USA* 93:: 7-12. ([PubMed](#)) ([Full Text in PMC](#))

M. Marquart, J. Deisenhofer, R. Huber, and W. Palm. 1980. Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution *J. Mol. Biol.* 141:: 369-391. ([PubMed](#))

E.W. Silverton, M.A. Navia, and D.R. Davies. 1977. Threedimensional structure of an intact human immunoglobulin *Proc. Natl. Acad. Sci. USA* 74:: 5140-5144. ([PubMed](#))

E.A. Padlan, E.W. Silverton, S. Sheriff, G.H. Cohen, G.S. Smith, and D.R. Davies. 1989. Structure of an antibody-antigen complex: Crystal structure of the HyHEL-10 F_{ab} lysozyme complex *Proc. Natl. Acad. Sci. USA* 86:: 5938-5942. ([PubMed](#))

J. Rini, U. Schultze-Gahmen, and I.A. Wilson. 1992. Structural evidence for induced fit as a mechanism for antibody-antigen recognition *Nature* 255:: 959-965.

T.O. Fischmann, G.A. Bentley, T.N. Bhat, G. Boulot, R.A. Mariuzza, S.E. Phillips, D. Tello, and R.J. Poljak. 1991. Crystallographic refinement of the three-dimensional structure of the FabD1.3-lysozyme complex at 2.5-Å resolution *J. Biol. Chem.* 266:: 12915-12920. ([PubMed](#))

D.R. Burton. 1990. Antibody: The flexible adaptor molecule *Trends Biochem. Sci.* 15:: 64-69. ([PubMed](#))

Generation of diversity

- S. Tonegawa. 1988. Somatic generation of immune diversity *Biosci. Rep.* 8:: 3-26. ([PubMed](#))
- T. Honjo and S. Habu. 1985. Origin of immune diversity: Genetic variation and selection *Annu. Rev. Biochem.* 54:: 803-830. ([PubMed](#))
- M. Gellert and J.F. McBlane. 1995. Steps along the pathway of VDJ recombination *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 347:: 43-47. ([PubMed](#))
- R.S. Harris, Q. Kong, and N. Maizels. 1999. Somatic hypermutation and the three R's: Repair, replication and recombination *Mutat. Res.* 436:: 157-178. ([PubMed](#))
- S.M. Lewis and G.E. Wu. 1997. The origins of V(D)J recombination *Cell* 88:: 159-162. ([PubMed](#))
- D.A. Ramsden, D.C. van Gent, and M. Gellert. 1997. Specificity in V(D)J recombination: New lessons from biochemistry and genetics *Curr. Opin. Immunol.* 9:: 114-120. ([PubMed](#))
- D.B. Roth and N.L. Craig. 1998. VDJ recombination: A transposase goes to work *Cell* 94:: 411-414. ([PubMed](#))
- M.J. Sadofsky. 2001. The RAG proteins in V(D)J recombination: More than just a nuclease *Nucleic Acids Res.* 29:: 1399-1409. ([PubMed](#)) ([Full Text in PMC](#))

MHC proteins and antigen processing

- P.J. Bjorkman and P. Parham. 1990. Structure, function, and diversity of class I major histocompatibility complex molecules *Annu. Rev. Biochem.* 59:: 253-288. ([PubMed](#))
- A.L. Goldberg and K.L. Rock. 1992. Proteolysis, proteasomes, and antigen presentation *Nature* 357:: 375-379. ([PubMed](#))
- D.R. Madden, J.C. Gorga, J.L. Strominger, and D.C. Wiley. 1992. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight binding to MHC *Cell* 70:: 1035-1048. ([PubMed](#))
- J.H. Brown, T.S. Jardetzky, J.C. Gorga, L.J. Stern, R.G. Urban, J.L. Strominger, and D.C. Wiley. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1 *Nature* 364:: 33-39. ([PubMed](#))
- M.A. Saper, P.J. Bjorkman, and D.C. Wiley. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution *J. Mol. Biol.* 219:: 277-319. ([PubMed](#))
- D.R. Madden, J.C. Gorga, J.L. Strominger, and D.C. Wiley. 1991. The structure of HLA-B27 reveals nonamer self-peptides bound in an extended conformation *Nature.* 353:: 321-325. ([PubMed](#))
- P. Cresswell, N. Bangia, T. Dick, and G. Diedrich. 1999. The nature of the MHC class I peptide loading complex *Immunol. Rev.* 172:: 21-28. ([PubMed](#))
- D.R. Madden, D.N. Garboczi, and D.C. Wiley. 1993. The antigenic identity of peptide-MHC complexes: A comparison of the conformations of five viral peptides presented by HLA-A2 *Cell* 75:: 693-708. ([PubMed](#))

T-cell receptors and signaling complexes

- J. Hennecke and D.C. Wiley. 2001. T cell receptor-MHC interactions up close *Cell* 104:: 1-4. ([PubMed](#))
- Y.H. Ding, K.J. Smith, D.N. Garboczi, U. Utz, W.E. Biddison, and D.C. Wiley. 1998. Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids *Immunity* 8:: 403-411. ([PubMed](#))

E.L. Reinherz, K. Tan, L. Tang, P. Kern, J. Liu, Y. Xiong, R.E. Hussey, A. Smolyar, B. Hare, R. Zhang, A. Joachimiak, H.C. Chang, G. Wagner, and J. Wang. 1999. The crystal structure of a T cell receptor in complex with peptide and MHC class II *Science* 286:: 1913-1921. ([PubMed](#))

J.R. Cochran, T.O. Cameron, and L.J. Stern. 2000. The relationship of MHC-peptide binding and T cell activation probed using chemically defined MHC class II oligomers *Immunity* 12:: 241-250. ([PubMed](#))

J.R. Cochran, T.O. Cameron, J.D. Stone, J.B. Lubetsky, and L.J. Stern. 2001. Receptor proximity, not intermolecular orientation, is critical for triggering T-cell activation *J. Biol. Chem.* 276:: 28068-28074. ([PubMed](#))

K.C. Garcia, L. Teyton, and I.A. Wilson. 1999. Structural basis of T cell recognition *Annu. Rev. Immunol.* 17:: 369-397. ([PubMed](#))

B.S. Gaul, M.L. Harrison, R.L. Geahlen, R.A. Burton, and C.B. Post. 2000. Substrate recognition by the Lyn protein-tyrosine kinase: NMR structure of the immunoreceptor tyrosine-based activation motif signaling region of the B cell antigen receptor *J. Biol. Chem.* 275:: 16174-16182. ([PubMed](#))

P.S. Kern, M.K. Teng, A. Smolyar, J.H. Liu, J. Liu, R.E. Hussey, R. Spoerl, H.C. Chang, E.L. Reinherz, and J.H. Wang. 1998. Structural basis of CD8 coreceptor function revealed by crystallographic analysis of a murine CD8 α domain fragment in complex with H-2Kb *Immunity* 9:: 519-530. ([PubMed](#))

R. Konig, S. Fleury, and R.N. Germain. 1996. The structural basis of CD4-MHC class II interactions: Coreceptor contributions to T cell receptor antigen recognition and oligomerization-dependent signal transduction *Curr. Top. Microbiol. Immunol.* 205:: 19-46. ([PubMed](#))

M. Krummel, C. Wulfig, C. Sumen, and M.M. Davis. 2000. Thirty-six views of T-cell recognition *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355:: 1071-1076. ([PubMed](#))

C.J. Janeway. 1992. The T cell receptor as a multicomponent signalling machine: CD4/CD8 coreceptors and CD45 in T cell activation *Annu. Rev. Immunol.* 10:: 645-674. ([PubMed](#))

E.R. Podack and A. Kupfer. 1991. T-cell effector functions: Mechanisms for delivery of cytotoxicity and help *Annu. Rev. Cell Biol.* 7:: 479-504. ([PubMed](#))

M.M. Davis. 1990. T cell receptor gene diversity and selection *Annu. Rev. Biochem.* 59:: 475-496. ([PubMed](#))

D.J. Leahy, R. Axel, and W.A. Hendrickson. 1992. Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution *Cell* 68:: 1145-1162. ([PubMed](#))

B. Lowin, M. Hahne, C. Mattmann, and J. Tschopp. 1994. Cytolytic T-cell cytotoxicity is mediated through perforin and Fas lytic pathways *Nature* 370:: 650-652. ([PubMed](#))

HIV and AIDS

A.S. Fauci. 1988. The human immunodeficiency virus: Infectivity and mechanisms of pathogenesis *Science* 239:: 617-622. ([PubMed](#))

R.C. Gallo and L. Montagnier. 1988. AIDS in 1988 *Sci. Am.* 259: (4) 41-48. ([PubMed](#))

P.D. Kwong, R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, and W.A. Hendrickson. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody *Nature* 393:: 648-659. ([PubMed](#))

Discovery of major concepts

G.L. Ada and G. Nossal. 1987. The clonal selection theory *Sci. Am.* 257: (2) 62-69. ([PubMed](#))

R.R. Porter. 1973. Structural studies of immunoglobulins *Science* 180:: 713-716. ([PubMed](#))

G.M. Edelman. 1973. Antibody structure and molecular immunology *Science* 180:: 830-840. ([PubMed](#))

G. Kohler. 1986. Derivation and diversification of monoclonal antibodies *Science* 233:: 1281-1286. ([PubMed](#))

C. Milstein. 1986. From antibody structure to immunological diversification of immune response *Science* 231:: 1261-1268. ([PubMed](#))

C.A. Janeway Jr. 1989. Approaching the asymptote? Evolution and revolution in immunology *Cold Spring Harbor Symp. Quant. Biol.* 54:: 1-13. ([PubMed](#))

34. Molecular Motors

Organisms, from human beings to bacteria, move to adapt to changes in their environments, navigating toward food and away from danger. Cells, themselves, are not static but are bustling assemblies of moving proteins, nucleic acids, and organelles ([Figure 34.1](#)). Remarkably, the fundamental biochemical mechanisms that produce contractions in our muscles are the same as those that propel organelles along defined paths inside cells. In fact, many of the proteins that play key roles in converting chemical energy in the form of ATP into kinetic energy, the energy of motion, are members of the same protein family, the P-loop NTPases. These molecular motors are homologous to proteins that we have encountered in other contexts, including the G proteins in protein synthesis, signaling, and other processes. Once again we see the economy of evolution in adapting an existing protein to perform new functions.

Molecular motors operate by small increments, converting changes in protein conformation into directed motion. Orderly motion across distances requires a track that steers the motion of the motor assembly. Indeed, we have previously encountered a class of molecular motors that utilize mechanisms that we will examine here — namely, the helicases that move along DNA and RNA tracks ([Section 28.1.7](#)). The proteins on which we will focus in this chapter move along actin and microtubules — protein filaments composed of repeating identical subunits. The motor proteins cycle between forms having high or low affinity for the filament tracks in response to ATP binding and hydrolysis, enabling a bind, pull, and release mechanism that generates motion.

We will also consider a completely different strategy for generating motion, one used by bacteria such as *E. coli*. A set of flagella act as propellers, rotated by a motor in the bacterial cell membrane. This rotary motor is driven by a proton gradient across the membrane, rather than by ATP hydrolysis. The mechanism for coupling the proton gradient to rotatory motion is analogous to that used by the F_0 subunit of ATP synthase ([Section 18.4.2](#)). Thus, both of the major modes for storing biochemical energy — namely, ATP and ion gradients — have been harnessed by evolution to drive organized molecular motion.

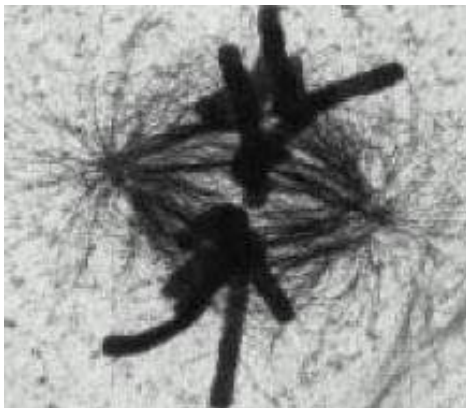
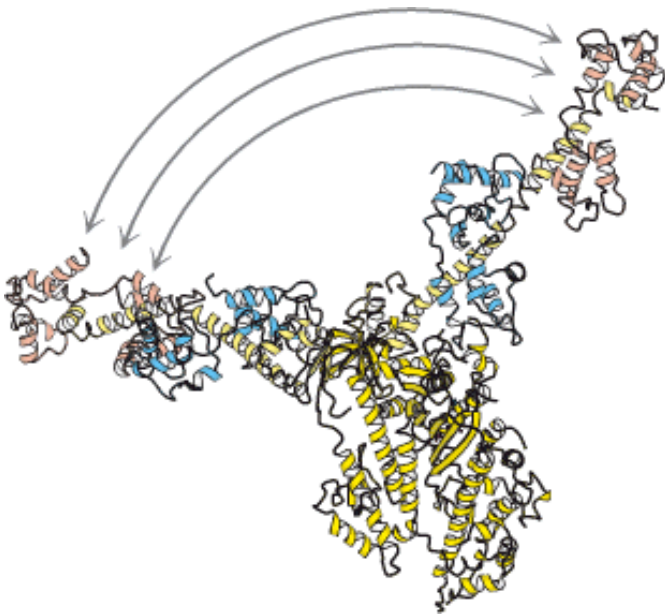


Figure 34.1. Motion Within Cells. This high-voltage electron micrograph shows the mitotic apparatus in a metaphase mammalian cell. The large cylindrical objects are chromosomes, and the threadlike structures stretched across the center are microtubules — tracks for the molecular motors that move chromosomes. Many processes, including chromosome segregation in mitosis, depend on the action of molecular-motor proteins. [Courtesy of J. R. McIntosh.]



The horse, like all animals, is powered by the molecular motor protein, myosin. A portion of myosin moves dramatically (as shown above) in response to ATP binding, hydrolysis, and release, propelling myosin along an actin filament. This molecular movement is translated into movement of the entire animal, excitingly depicted in da Vinci's rearing horse. [(Left) Leonardo da Vinci's "Study of a rearing horse" for the battle of Anghiari (c. 1504) from The Royal Collection © Her Royal Majesty Queen Elizabeth II.]

34.1. Most Molecular-Motor Proteins Are Members of the P-Loop NTPase Superfamily

Eukaryotic cells contain three major families of motor proteins: myosins, kinesins, and dyneins. At first glance, these protein families appear to be quite different from one another. *Myosin*, first characterized on the basis of its role in muscle (Section 34.2.1), moves along filaments of the protein actin. Muscle myosin consists of two copies each of a *heavy chain* with a molecular mass of 87 kd, an *essential light chain*, and a *regulatory light chain*. The human genome appears to encode more than 40 distinct myosins; some function in muscle contraction and others participate in a variety of other processes. *Kinesins*, which have roles in protein, vesicle, and organelle transport along microtubules, including chromosome segregation, often consist of two copies each of a heavy chain and a light chain. The heavy chain is approximately one-half the size of that for myosin. The human genome encodes more than 40 kinesins. *Dynein* powers the motion of cilia and flagella in some eukaryotic cells, among other roles. Dyneins are enormous, with heavy chains of molecular mass greater than 500 kd. The human genome appears to encode approximately 10 dyneins.

Comparison of the amino acid sequences of myosins, kinesins, and dyneins did not reveal significant relationships between these protein families but, after their three-dimensional structures were determined, members of the myosin and kinesin families were found to have remarkable similarities. In particular, both myosin and kinesin contain P-loop NTPase cores homologous to those found in G proteins. Sequence analysis of the dynein heavy chain reveals it to be a member of the AAA subfamily of P-loop NTPases that we encountered previously in the context of the 19S proteasome (Section 23.2.2). Dynein has six sequences encoding such P-loop NTPase domains arrayed along its length. Thus, we can draw on our knowledge of G proteins and other P-loop NTPases as we analyze the mechanisms of action of these motor proteins.

34.1.1. A Motor Protein Consists of an ATPase Core and an Extended Structure

Let us first consider the structure of myosin. The results of electron microscopic studies of skeletal muscle myosin show it to be a two-headed structure linked to a long stalk (Figure 34.2). As we saw in Chapter 33, limited proteolysis can be a powerful tool in probing the activity of large proteins. Treatment of myosin with trypsin and papain results in the formation of four fragments: two S1 fragments, an S2 fragment, also called heavy meromyosin (HMM), and a fragment called light meromyosin (LMM; Figure 34.3). Each *S1 fragment* corresponds to one of the heads from the intact structure and includes 850 amino-terminal amino acids from one of the two heavy chains as well as one copy of each of the light chains. Examination of the structure of an S1 fragment at high resolution reveals the presence of a P-loop NTPase-domain core that is the site of ATP binding and hydrolysis (Figure 34.4).

Extending away from this structure is a long α helix from the heavy chain. This helix is the binding site for the two light chains. The light chains are members of the EF-hand family, similar to calmodulin, although most of the EF hands in light chains do not bind metal ions (Figure 34.5). Like calmodulin, these proteins wrap around an α helix, serving to thicken and stiffen it. The remaining fragments of myosin — S2 and light meromyosin — are largely α helical, forming two-stranded coiled coils created by the remaining lengths of the two heavy chains wrapping around each other (Figure 34.6). These structures, together extending approximately 1700 Å, link the myosin heads to other structures. In muscle myosin, several LMM domains come together to form higher-order bundles.

Conventional kinesin, the first kinesin discovered, has a structure having several features in common with myosin (Figure 34.7). The dimeric protein has two heads, linked by an extended structure. The size of the head domain is approximately one-third of that of myosin. Determination of the three-dimensional structure of a kinesin fragment revealed that this motor protein also is built around a P-loop NTPase core (Figure 34.8). The myosin domain is so much larger than that of kinesin because of two large insertions in the myosin domain. For conventional kinesin, a region of approximately 500 amino acids follows the head domain. Like the corresponding region in myosin, the extended part of kinesin forms an α -helical coiled coil. Unlike myosin, the α -helical region directly adjacent to the head domain is not the binding site for kinesin light chains. Instead, kinesin light chains, if present, bind near the carboxyl terminus.

Dynein has a rather different structure. As noted earlier, the dynein heavy chain includes six regions that are homologous to the AAA subfamily of ATPase domains. Although no crystallographic data are yet available, the results of electron microscopic studies and comparison with known structures of other AAA ATPases have formed the basis for the construction of a model of the dynein head structure (Figure 34.9). The head domain is appended to a region of approximately 1300 amino acids that forms an extended structure that links dynein units together to form oligomers and interacts with other proteins.

34.1.2. ATP Binding and Hydrolysis Induce Changes in the Conformation and Binding Affinity of Motor Proteins

A key feature of P-loop NTPases such as G proteins is that they undergo structural changes induced by NTP binding and hydrolysis. Moreover, these structural changes alter their affinities for binding partners. Thus, it is not surprising that the NTPase domains of motor proteins display analogous responses to nucleotide binding. The S1 fragment of myosin from scallop muscle provides a striking example of the changes observed (Figure 34.10). The structure of the S1 fragment has been determined for S1 bound to a complex formed of ADP and vanadate (VO_4^{3-}), which is an analog of ATP, or, more precisely, the ATP-hydrolysis transition state. In the presence of the ADP - VO_4^{3-} complex, the long helix that binds the light chains (hereafter referred to as the *lever arm*) protrudes outward from the head domain. In the presence of ADP without VO_4^{3-} , the lever arm has rotated by nearly 90 degrees relative to its position in the ADP - VO_4^{3-} complex. How does the identity of the species in the nucleotide-binding site cause this dramatic transition? Two regions around the nucleotide-binding site, analogous to the switch regions of G proteins (Section 15.1.2), conform closely to the group in the position of the γ -phosphate group of ATP and adopt a looser conformation when such a group is absent (Figure 34.11). This conformational change allows a long α helix (termed the *relay helix*) to adjust its position. The carboxylterminal end of the relay helix interacts with structures at the base of the lever arm, and so a change in the position of the relay helix leads to a reorientation of the lever arm.

The binding of ATP significantly decreases the affinity of the myosin head for actin filaments. No structures of myosin - actin complexes have yet been determined at high resolution, so the mechanistic basis for this change remains to be elucidated. However, the amino-terminal end of the relay helix interacts with the domains of myosin that bind to actin, suggesting a clear pathway for the coupling of nucleotide binding to changes in actin affinity. The importance of the changes in actin-binding affinity will be clear later when we examine the role of myosin in generating directed motion (Section 34.2.4).

Analogous conformational changes take place in kinesin. The kinesins also have a relay helix that can adopt different configurations when kinesin binds different nucleotides. Kinesin lacks an α -helical lever arm, however. Instead, a relatively short segment termed the *neck linker* changes conformation in response to nucleotide binding (Figure 34.12). The neck linker binds to the head domain of kinesin when ATP is bound but is released when the nucleotide-binding site is vacant or occupied by ADP. Kinesin differs from myosin in that the binding of ATP to kinesin *increases* the affinity between kinesin and its binding partner, microtubules. The properties of myosin, kinesin, and a heterotrimeric G protein are compared in Table 34.1. Before turning to a discussion of how these properties are used to convert chemical energy into motion, we must consider the properties of the tracks along which these motors move.

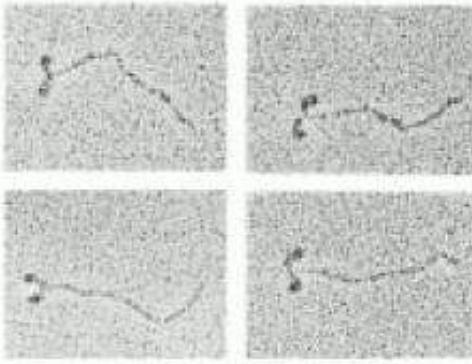


Figure 34.2. Myosin Structure at Low Resolution. Electron micrographs of myosin molecules reveal a two-headed structure with a long, thin tail. [Courtesy of Dr. Paula Flicker, Dr. Theo Walliman, and Dr. Peter Vibert.]

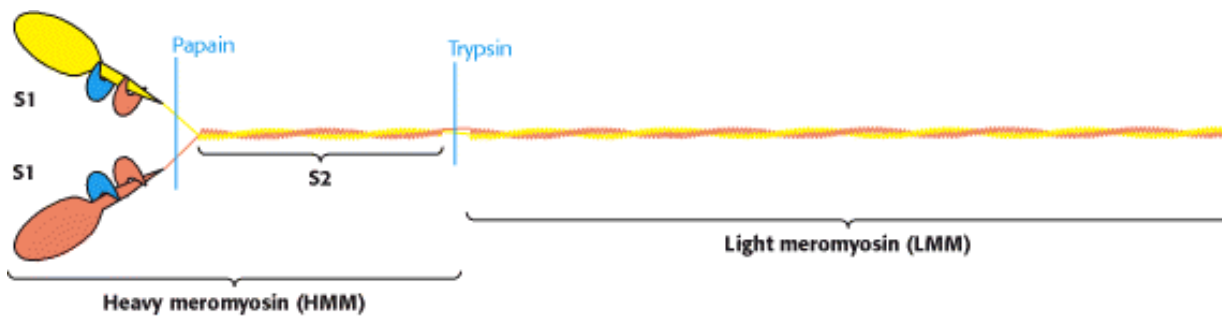


Figure 34.3. Myosin Dissection. Treatment of muscle myosin with proteases forms stable fragments, including subfragments S1 and S2 and light meromyosin. Each S1 fragment includes the head (shown in yellow and pink) from the heavy chain and one copy of each light chain (shown in blue and orange).

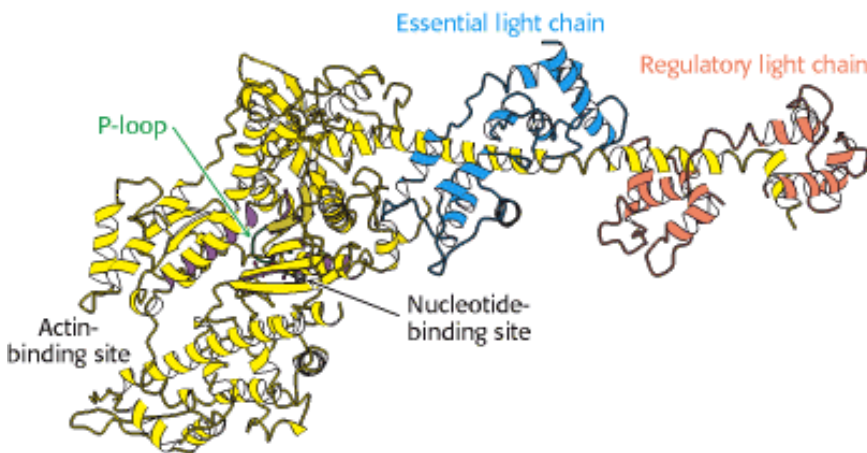


Figure 34.4. Myosin Structure at High Resolution. The structure of the S1 fragment from muscle myosin reveals the presence of a P-loop NTPase domain (shaded in purple). An α helix that extends from this domain is the binding site for the two light chains.

Essential light chain



Regulatory light chain



Calmodulin



Figure 34.5. Myosin Light Chains. The structures of the essential and regulatory light chains from muscle myosin are compared with the structure of calmodulin. Each of these homologous proteins binds an α helix (not shown) by wrapping around it.



Figure 34.6. Myosin Two-Stranded Coiled Coil. The two α helices form left-handed supercoiled structures that spiral around each other. Such structures are stabilized by hydrophobic residues at the contact points between the two helices.

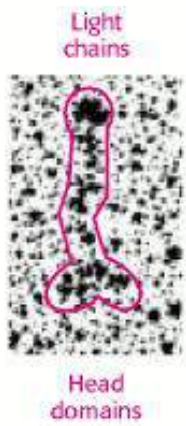


Figure 34.7. Kinesin at Low Resolution. An electron micrograph of conventional kinesin reveals an elongated structure with two heads at one end. The position of the light chains was confirmed through the use of antibody labels. [After N. Hirokawa, K. K. Pfister, H. Yorifuji, M. C. Wagner, S. T. Brady, and G. S. Broom. *Cell* 56 (1989):867.]



Figure 34.8. Structure of Head Domain of Kinesin at High Resolution. The head domain of kinesin has the structure of a P-loop NTPase core (indicated by purple shading).

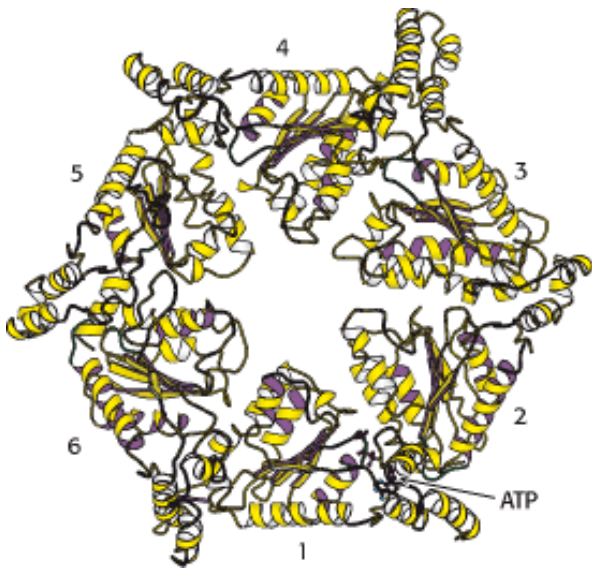


Figure 34.9. Dynein Head-Domain Model. ATP is bound in the first of six P-loop NTPase domains (numbered) in this model for the head domain of dynein. The model is based on electron micrographs and the structures of other members of the AAA ATPase family.

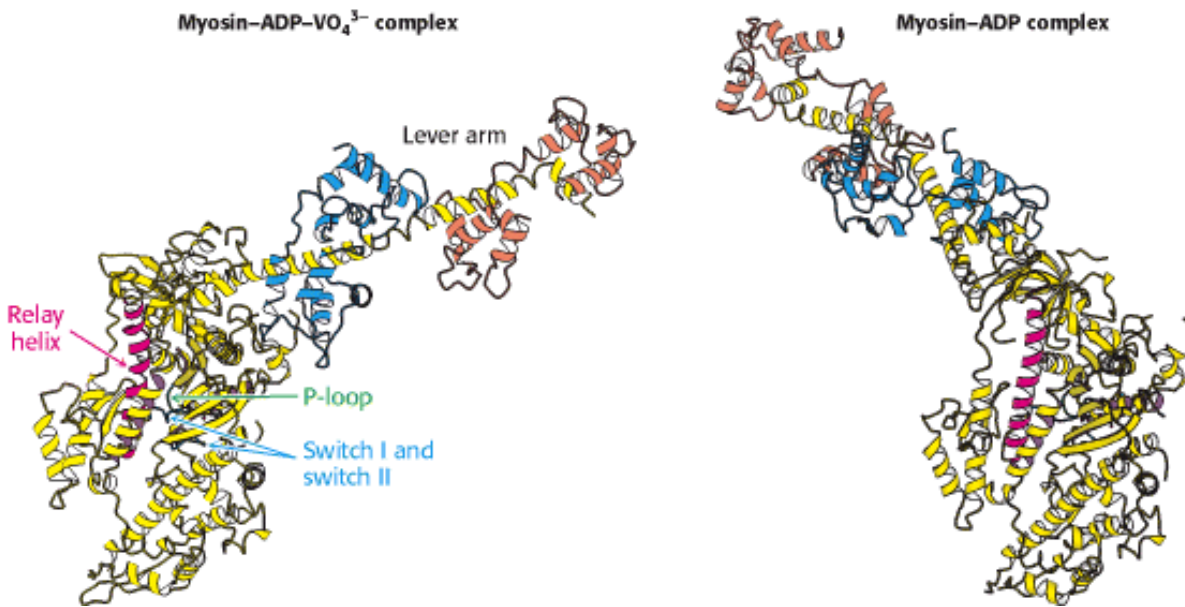


Figure 34.10. Lever-Arm Motion. Two forms of the S1 fragment of scallop muscle myosin. Dramatic conformational changes are observed when the identity of the bound nucleotide changes from ADP-VO₄³⁻ to ADP or vice versa, including a nearly 90-degree reorientation of the lever arm.

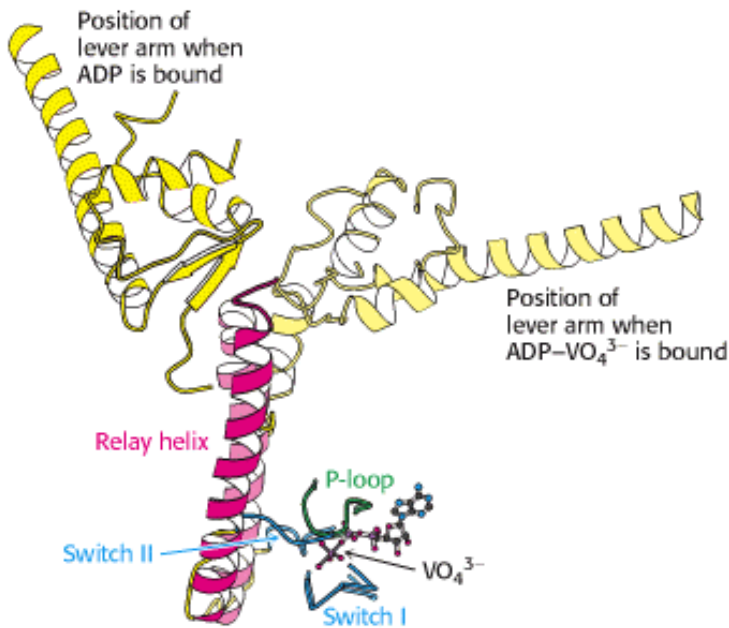


Figure 34.11. Relay Helix. A superposition of key elements in two forms of scallop myosin reveals the structural changes that are transmitted by the relay helix from the switch I and switch II loops to the base of the lever arm. The switch I and switch II loops interact with VO_4^{3-} in the position that would be occupied by the γ -phosphate group of ATP. The structure of the ADP - VO_4^{3-} myosin complex is shown in lighter colors.

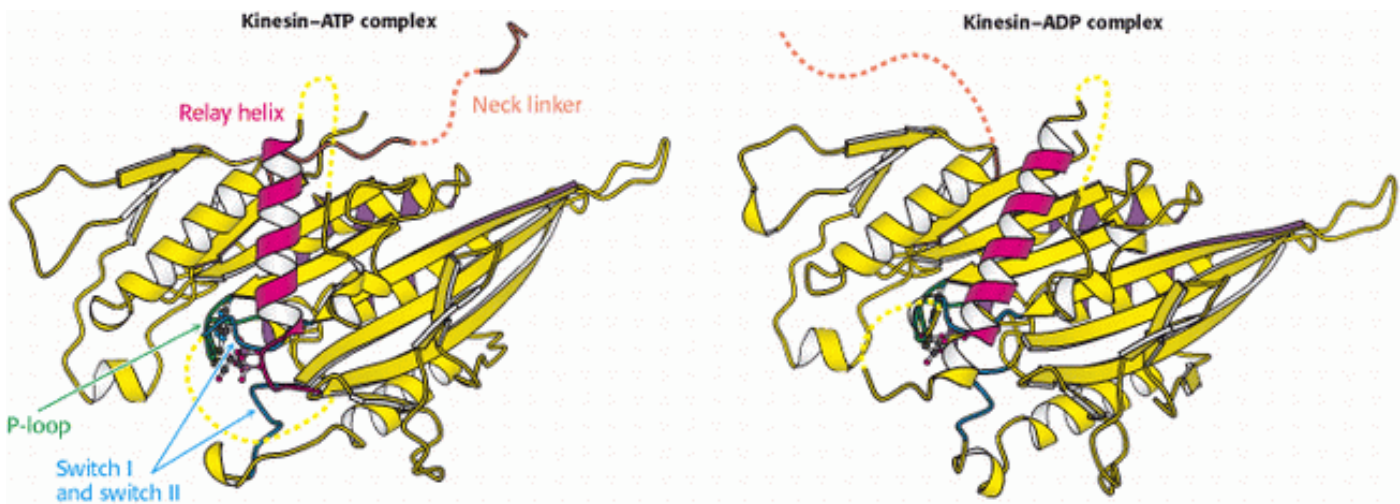


Figure 34.12. Neck Linker. A comparison of the structures of a kinesin bound to ADP and bound to an ATP analog. The neck linker (orange), which connects the head domain to the remainder of the kinesin molecule, is bound to the head domain in the presence of the ATP analog but is free in the presence of ADP only.

Table 34.1. Effect of nucleotide binding on protein affinity

	Bound to
Protein	NTP NDP
Myosin (ATP or ADP)	

Affinity for actin	Low High
Kinesin (ATP or ADP)	
Affinity for microtubules	High Low
Heterotrimeric G protein (α subunit) (GTP or GDP)	
Affinity for $\beta \gamma$ dimer	Low High
Affinity for effectors	High Low


34.2. Myosins Move Along Actin Filaments

Myosins, kinesins, and dyneins move by cycling between states with different affinities for the long, polymeric macromolecules that serve as their tracks. For myosin, the molecular track is a polymeric form of *actin*, a 42-kd protein that is one of the most abundant proteins in eukaryotic cells, typically accounting for as much as 10% of the total protein. Actin polymers are continually being assembled and disassembled in cells in a highly dynamic manner, accompanied by the hydrolysis of ATP. On the microscopic scale, actin filaments participate in the dynamic reshaping of the cytoskeleton and the cell itself and in other motility mechanisms that do not include myosin. In muscle, myosin and actin together are the key components responsible for muscle contraction.

34.2.1. Muscle Is a Complex of Myosin and Actin

Vertebrate muscle that is under voluntary control has a banded (striated) appearance when examined under a light microscope. It consists of multinucleated cells that are bounded by an electrically excitable plasma membrane. A muscle cell contains many parallel *myofibrils*, each about 1 μm in diameter. The functional unit, called a *sarcomere*, typically repeats every 2.3 μm (23,000 \AA) along the fibril axis in relaxed muscle (Figure 34.13). A dark *A band* and a light *I band* alternate regularly. The central region of the A band, termed the *H zone*, is less dense than the rest of the band. The I band is bisected by a very dense, narrow *Z line*.

The underlying molecular plan of a sarcomere is revealed by cross sections of a myofibril. These cross sections show the presence of two kinds of interacting protein filaments. The *thick filaments* have diameters of about 15 nm (150 \AA) and consist primarily of myosin. The *thin filaments* have diameters of approximately 9 nm (90 \AA) and consist of actin as well as *tropomyosin* and the *troponin complex*. Muscle contraction is achieved through the sliding of the thin filaments along the length of the thick filaments, driven by the hydrolysis of ATP (Figure 34.14). Tropomyosin and the troponin complex regulate this sliding in response to nerve impulses. Under resting conditions, tropomyosin blocks the intimate interaction between myosin and actin. A nerve impulse leads to an increase in calcium ion concentration within the muscle cell. A component of the troponin complex senses the increase in calcium and, in response, relieves the inhibition of myosin - actin interactions by tropomyosin.

 Although myosin was discovered through its role in muscle, other types of myosin play crucial roles in a number of biological contexts. Some defects in hearing in both mice and human beings have been linked to mutations in particular myosin homologs that are present in cells of the ear. For example, Usher syndrome in human beings and the shaker mutation in mice have been linked to myosin VIIa, expressed in hair cells (Section 32.4.1). The mutation of this myosin results in the formation of splayed stereocilia that do not function well. Myosin VIIa differs from muscle myosin in that its tail region possesses a number of amino acid sequences that correspond to domains known to mediate specific protein - protein interactions.

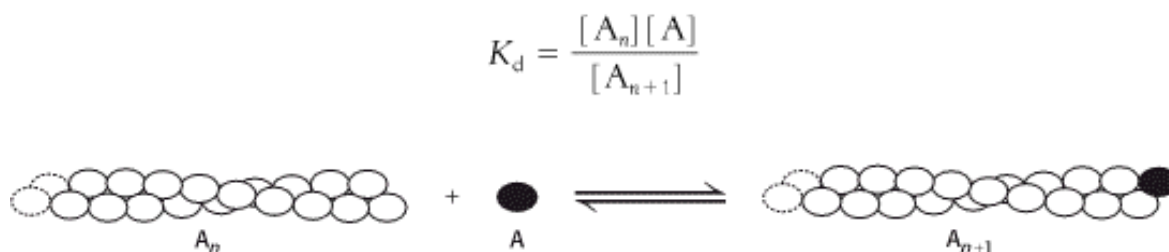
34.2.2. Actin Is a Polar, Self-Assembling, Dynamic Polymer

The structure of the actin monomer was determined to atomic resolution by x-ray crystallography and has been used to interpret the structure of actin filaments, already somewhat understood through electron microscopy studies at lower

resolution. Each actin monomer comprises four domains (Figure 34.15). These domains come together to surround a bound nucleotide, either ATP or ADP. The ATP form can be converted into the ADP form by hydrolysis.

Actin monomers (often called *G-actin* for globular) come together to form actin filaments (often called *F-actin*; see Figure 34.15). F-actin has a helical structure; each monomer is related to the preceding one by a translation of 27.5 Å and a rotation of 166 degrees around the helical axis. Because the rotation is nearly 180 degrees, F-actin resembles a two-stranded cable. Note that each actin monomer is oriented in the same direction along the F-actin filament, and so the structure is polar, with discernibly different ends. One end is called the barbed (plus) end, and the other is called the pointed (minus) end. The names "barbed" and "pointed" refer to the appearance of an actin filament when myosin S1 fragments are bound to it.

How are actin filaments formed? Like many biological structures, actin filaments self-assemble; that is, under appropriate conditions, actin monomers will come together to form well-structured, polar filaments. The aggregation of the first two or three monomers to form a filament is somewhat unfavorable. Once such a filament nucleus exists, the addition of subunits is more favorable. Let us consider the polymerization reaction in more detail. We designate an actin filament with n subunits A_n . This filament can bind an additional actin monomer, A , to form A_{n+1} .




The dissociation constant for this reaction, K_d , defines the monomer concentrations at which the polymerization reaction will take place, because the concentration of polymers of length $n + 1$ will be essentially equal to that for polymers of length n . Thus,

$$[A_n] \approx [A_{n+1}] \text{ and so } K_d = \frac{[A_n][A]}{[A_{n+1}]} \approx [A]$$

In other words, the polymerization reaction will proceed until the monomer concentration is reduced to the value of K_d . If the monomer concentration is below the value of K_d , the polymerization reaction will not proceed at all; indeed, existing filaments will depolymerize until the monomer concentration reaches the value of K_d . Because of these phenomena, K_d is referred to as the *critical concentration* for the polymer. Recall that actin contains a nucleotide-binding site that can contain either ATP or ADP. The critical concentration for the actin - ATP complex is approximately 20-fold lower than that for the actin - ADP complex; actin - ATP polymerizes more readily than does actin - ADP.

Actin filaments inside cells are highly dynamic structures that are continually gaining and losing monomers. The concentration of free actin monomers is controlled by several mechanisms. For example, actin-sequestering proteins such as β -thymosin bind to actin monomers and inhibit polymerization. Furthermore, the concentration and properties of actin filaments are closely regulated by proteins that sever an actin filament into two or that cap one of the ends of a filament. Regulated actin polymerization is central to the changes in cell shape associated with cell motility in amoebas as well as in human cells such as macrophages.

 A well-defined actin cytoskeleton is unique to eukaryotes; prokaryotes lack such structures. How did filamentous actin evolve? Comparison of the three-dimensional structure of G-actin with other proteins revealed remarkable similarity to several other proteins, including sugar kinases such as hexokinase (Figure 34.16; see also Section 16.1.1).

Notably, the nucleotide-binding site in actin corresponds to the ATP-binding site in hexokinase. Thus, actin evolved from an enzyme that utilized ATP as a substrate.

More recently, a closer prokaryotic homolog of actin was characterized. This protein, called MreB, plays an important role in determining cell shape in rod-shaped, filamentous, and helical bacteria. The internal structures formed by MreB are suggestive of the actin cytoskeleton of eukaryotic cells, although they are far less extensive. Even though this protein is only approximately 15% identical in sequence with actin, MreB folds into a very similar three-dimensional structure. It also polymerizes into structures that are similar to F-actin in a number of ways, including the alignment of the component monomers.

34.2.3. Motions of Single Motor Proteins Can Be Directly Observed

Muscle contraction is complex, requiring the action of many different myosin molecules. Studies of *single myosin molecules* moving relative to actin filaments have been sources of deep insight into the mechanisms underlying muscle contraction and other complex processes.

A powerful tool for these studies, called an *optical trap*, relies on highly focused laser beams ([Figure 34.17](#)). Small beads can be caught in these traps and held in place in solution.

The position of the beads can be monitored with nanometer precision. James Spudich and coworkers designed an experimental arrangement consisting of an actin filament that had a bead attached to each end. Each bead could be caught in an optical trap (one at each end of the filament) and the actin filament pulled taut over a microscope slide containing other beads that had been coated with fragments of myosin such as the heavy meromyosin fragment (see [Figure 34.17](#)). On the addition of ATP, transient displacements of the actin filament were observed along its long axis. The size of the displacement steps was fairly uniform with an average size of 11 nm.

The results of these studies, performed in the presence of varying concentrations of ATP, are interpreted as showing that individual myosin heads bind the actin filament and undergo a conformational change (the *power stroke*) that pulls the actin filament, leading to the displacement of the beads. After a period of time, the myosin head releases the actin, which then snaps back into place.

34.2.4. Phosphate Release Triggers the Myosin Power Stroke

How does ATP hydrolysis drive the power stroke? A key observation is that the addition of ATP to a complex of myosin and actin results in the dissociation of the complex. Thus, ATP binding and hydrolysis cannot be directly responsible for the power stroke. We can combine this fact with the structural observations described earlier to construct a mechanism for the motion of myosin along actin ([Figure 34.18](#)). Let us begin with myosin-ADP bound to actin. The release of ADP and the binding of ATP to actin result in the dissociation of myosin from actin. As we saw earlier, the binding of ATP with its γ -phosphate group to the myosin head leads to a significant conformational change, amplified by the lever arm. This conformational change moves the myosin head along the actin filament by approximately 110 Å. The ATP in the myosin is then hydrolyzed to ADP and P_i , which remain bound to myosin. The myosin head can then bind to the surface of actin, resulting in the dissociation of P_i from the myosin. Phosphate release, in turn, leads to a conformational change that increases the affinity of the myosin head for actin and allows the lever arm to move back to its initial position. *The conformational change associated with phosphate release corresponds to the power stroke.* After the release of P_i , the myosin remains tightly bound to the actin and the cycle can begin again.

How does this cycle apply to muscle contraction? Myosin molecules self-assemble into thick bipolar structures with the myosin heads protruding at both ends of a bare region in the center ([Figure 34.19](#)). Approximately 500 head domains line the surface of each thick filament. These domains are paired in myosin dimers, but the two heads within each dimer act independently. Actin filaments associate with each head-rich region, with the barbed ends of actin toward the Z-line. In the presence of normal levels of ATP, most of the myosin heads are detached from actin. Each head can independently hydrolyze ATP, bind to actin, release P_i , and undergo its power stroke. Because few other heads are

attached, the actin filament is relatively free to slide. Each head cycles approximately five times per second with a movement of 110 Å per cycle. However, because hundreds of heads are interacting with the same actin filament, the overall rate of movement of myosin relative to the actin filament may reach 80,000 Å per second, allowing a sarcomere to contract from its fully relaxed to its fully contracted form rapidly. Having many myosin heads briefly and independently attaching and moving an actin filament allows for much greater speed than could be achieved by a single motor protein.

34.2.5. The Length of the Lever Arm Determines Motor Velocity

A key feature of myosin motors is the role of the lever arm as an amplifier. The lever arm amplifies small structural changes at the nucleotide-binding site to achieve the 110-Å movement along the actin filament that takes place in each ATP hydrolysis cycle. A strong prediction of the mechanism proposed for the movement of myosin along actin is that the length traveled per cycle should depend on the length of this lever arm. Thus, the length of the lever arm should influence the overall rate at which actin moves relative to a collection of myosin heads.

This prediction was tested with the use of mutated forms of myosin with lever arms of different lengths. The lever arm in muscle myosin includes binding sites for two light chains (Section 34.1.2). Thus investigators shortened the lever arm by deleting the sequences that correspond to one or both of these binding sites. They then examined the rates at which actin filaments were transported along collections of these mutated myosins (Figure 34.20). As predicted, the rate decreased as the lever arm was shortened. A mutated form of myosin with an unusually long lever arm was generated by inserting 23 amino acids corresponding to the binding site for an additional regulatory light chain. Remarkably, this form was found to support actin movement that was *faster than the wild-type protein*. These results strongly support the proposed role of the lever arm in contributing to myosin motor activity.

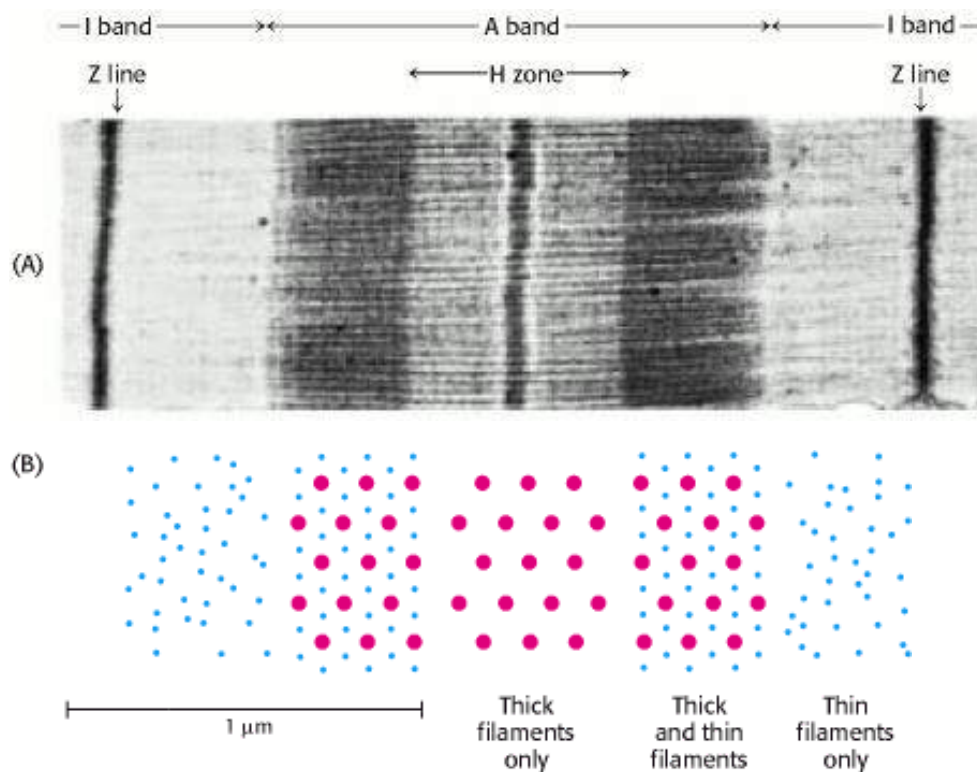


Figure 34.13. Sarcomere. (A) Electron micrograph of a longitudinal section of a skeletal muscle myofibril, showing a single sarcomere. (B) Schematic representations of cross sections correspond to the regions in the micrograph. [Courtesy of Dr. Hugh Huxley.]

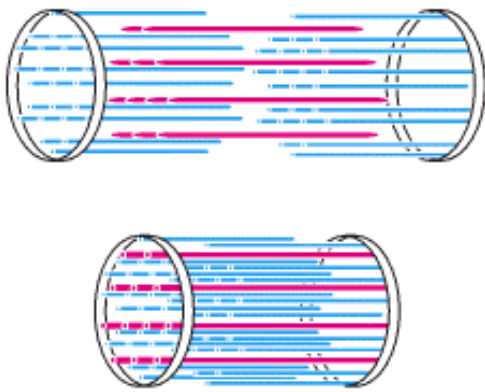


Figure 34.14. Sliding-Filament Model. Muscle contraction depends on the motion of thin filaments (blue) relative to thick filaments (red). [After H. E. Huxley. The mechanism of muscular contraction. Copyright © 1965 by Scientific American, Inc. All rights reserved.]

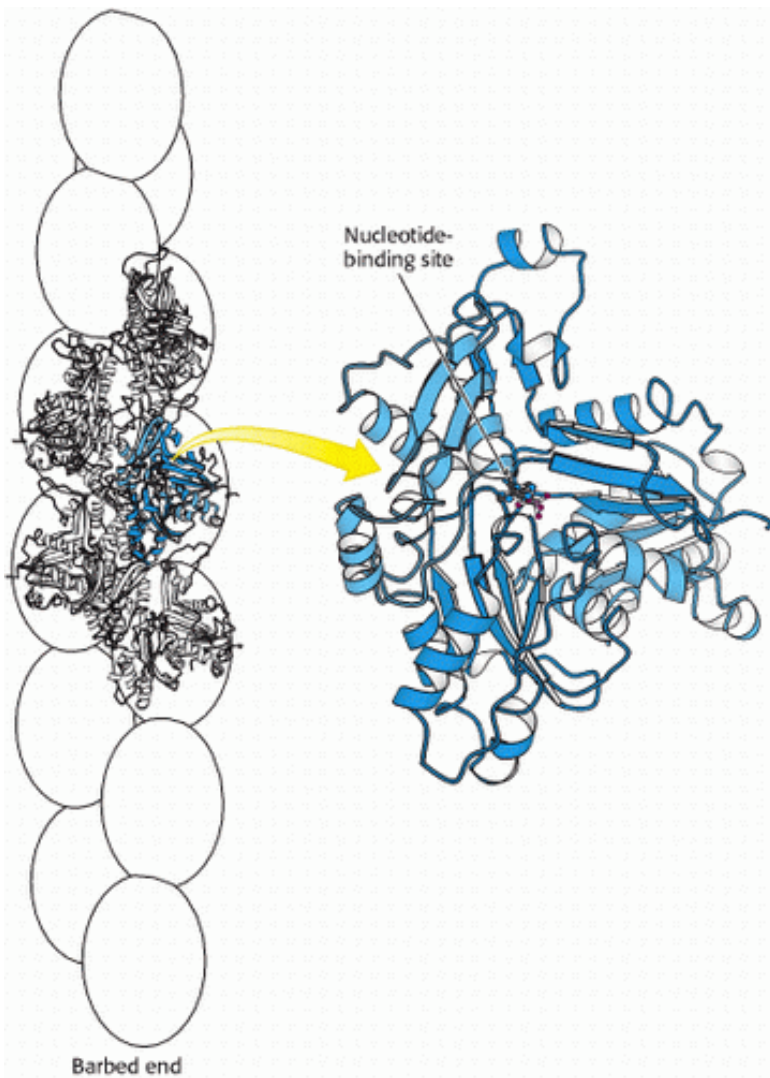


Figure 34.15. Actin Structure. (Left) Five actin monomers of an actin filament are shown explicitly, one in blue. (Right) The domains in the four-domain structure of an actin monomer are identified by different shades of blue.

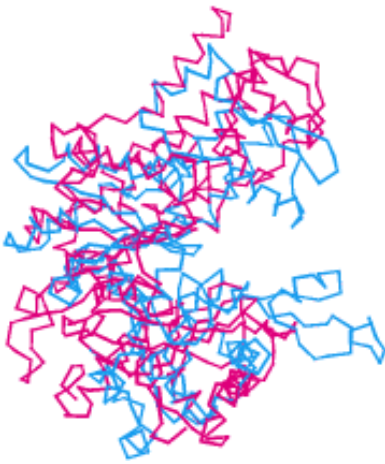


Figure 34.16. Actin and Hexokinase. A comparison of actin (blue) and hexokinase from yeast (red) reveals structural similarities indicative of homology. Both proteins have a deep cleft in which nucleotides bind.

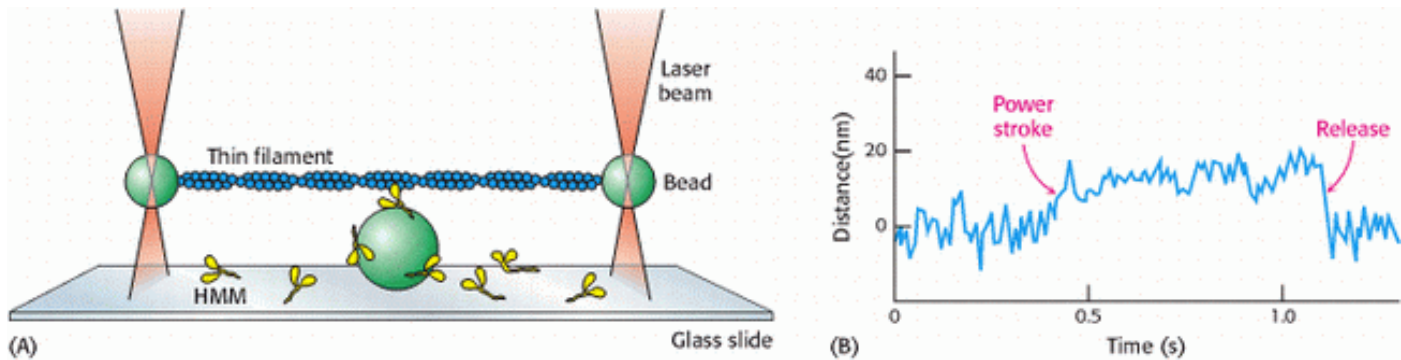


Figure 34.17. Watching a Single Motor Protein in Action. (A) An actin filament (blue) is placed above a heavy meromyosin (HMM) fragment (yellow) that projects from a bead on a glass slide. A bead attached to each end of the actin filament is held in an optical trap produced by a focused, intense infrared laser beam (orange). The position of these beads can be measured with nanometer precision. (B) Recording of the displacement of the actin filament induced by the addition of ATP. [After J. T. Finer, R. M. Simmons, and J. A. Spudich. *Nature* 368(1994):113.]

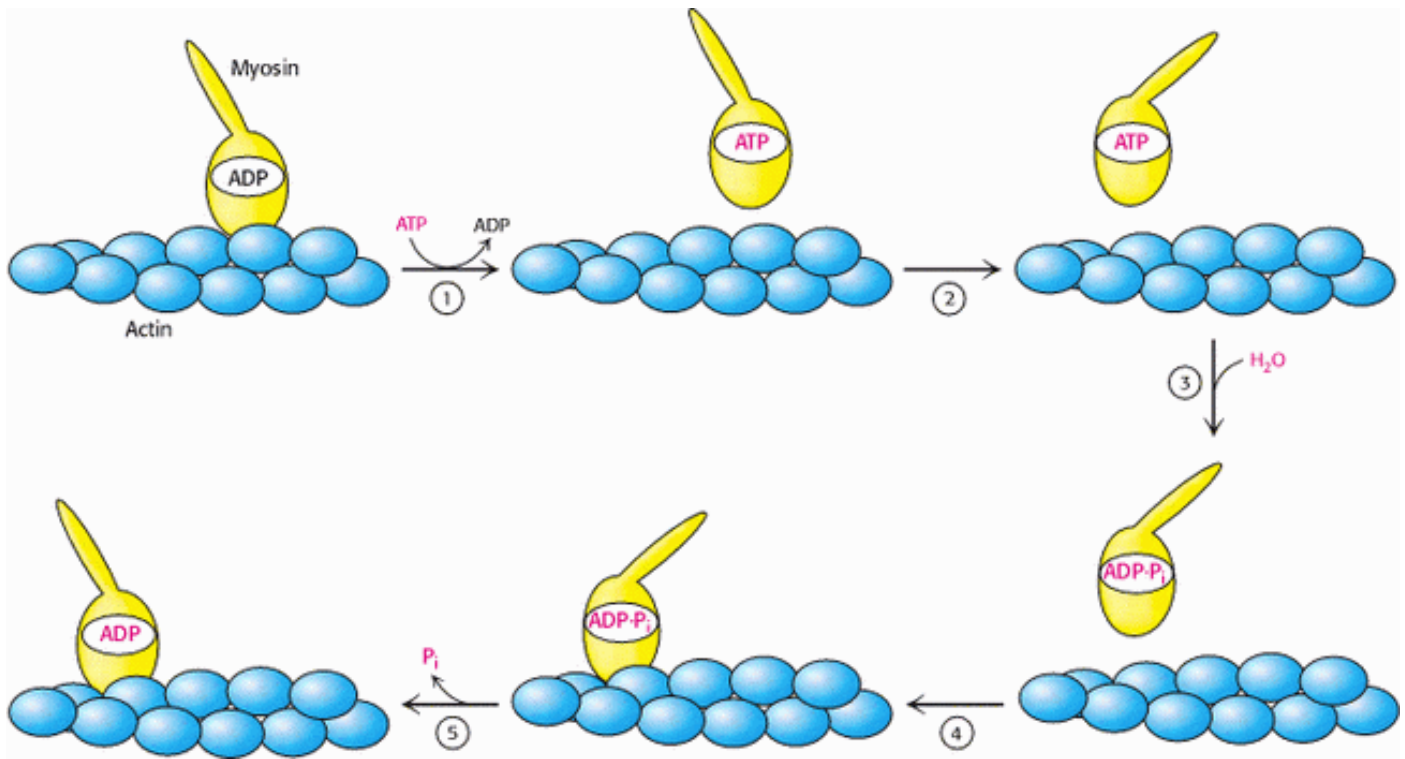


Figure 34.18. Myosin Motion Along Actin. A myosin head (yellow) in the ADP form is bound to an actin filament (blue). The exchange of ADP for ATP results in (1) the release of myosin from actin and (2) substantial reorientation of the lever arm of myosin. Hydrolysis of ATP (3) allows the myosin head to rebind at a site displaced along the actin filament (4). The release of P_i (5) accompanying this binding increases the strength of interaction between myosin and actin and resets the orientation of the lever arm.

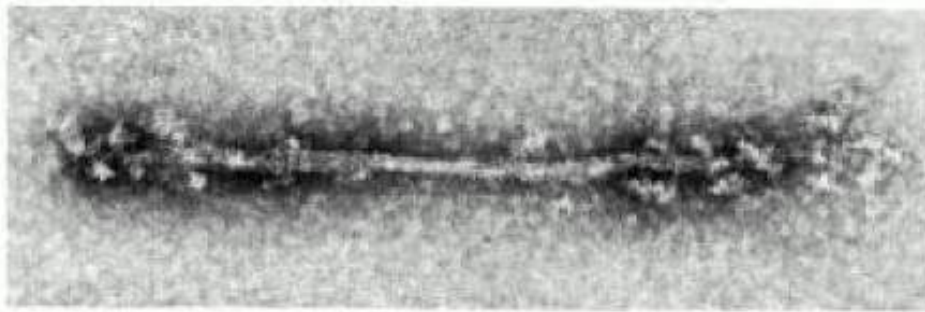


Figure 34.19. Thick Filament. (A) An electron micrograph of a reconstituted thick filament reveals the presence of myosin head domains at each end and a relatively narrow central region. (B) A schematic view shows how myosin molecules come together to form the thick filament. [Part A courtesy of Dr. Hugh Huxley.]

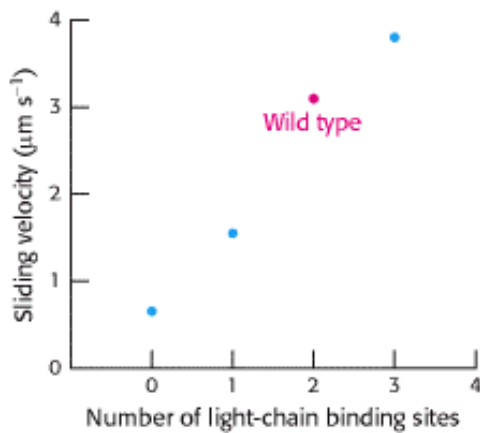



Figure 34.20. Myosin Lever Arm Length. Examination of the rates of actin movement supported by a set of myosin mutants with different numbers of light-chain binding sites revealed a linear relation; the greater the number of light-chain binding sites (and, hence, the longer the lever arm), the faster the sliding velocity. [After T. Q. P. Uyeda, P. D. Abramson, and J. A. Spudich. *Proc. Natl. Acad. Sci. USA* 93(1996):4459.]

34.3. Kinesin and Dynein Move Along Microtubules

In addition to actin, the cytoskeleton includes other components, notably intermediate filaments and microtubules. Microtubules serve as tracks for two classes of motor proteins — namely, kinesins and dyneins. Kinesins moving along microtubules usually carry cargo such as organelles and vesicles from the center of a cell to its periphery. Dyneins are important in sliding microtubules relative to one other during the beating of cilia and flagella on the surfaces of some eukaryotic cells.

 Some members of the kinesin family are crucial to the transport of organelles and other cargo to nerve endings at the periphery of neurons. It is not surprising, then, that mutations in these kinesins can lead to nervous system disorders. For example, mutations in a kinesin called KIF1B β can lead to the most common peripheral neuropathy (weakness and pain in the hands and feet), Charcot-Marie-Tooth disease, which affects 1 in 2500 people. A glutamine-to-leucine mutation in the P-loop of the motor domain of this kinesin has been found in some affected persons. Knockout mice with a disruption of the orthologous gene have been generated. Mice heterozygous for the disruption show symptoms similar to those observed in human beings; homozygotes die shortly after birth. Mutations in other kinesin genes have been tentatively linked to a predisposition to schizophrenia. In these disorders, defects in kinesin-linked transport may impair nerve function directly, and the decrease in the activity of specific neurons may lead to other degenerative processes.


34.3.1. Microtubules Are Hollow Cylindrical Polymers


Microtubules are built from two kinds of homologous 50-kd subunits, α - and β -tubulin, which assemble in an helical array of alternating tubulin types to form the wall of a hollow cylinder (Figure 34.21). Alternatively, a microtubule can be regarded as 13 protofilaments that run parallel to its long axis. The outer diameter of a microtubule is 30 nm, much larger than that of actin (5 nm). Like actin, microtubules are polar structures. One end, termed the minus end, is anchored near the center of a cell, whereas the plus end extends toward the cell surface.

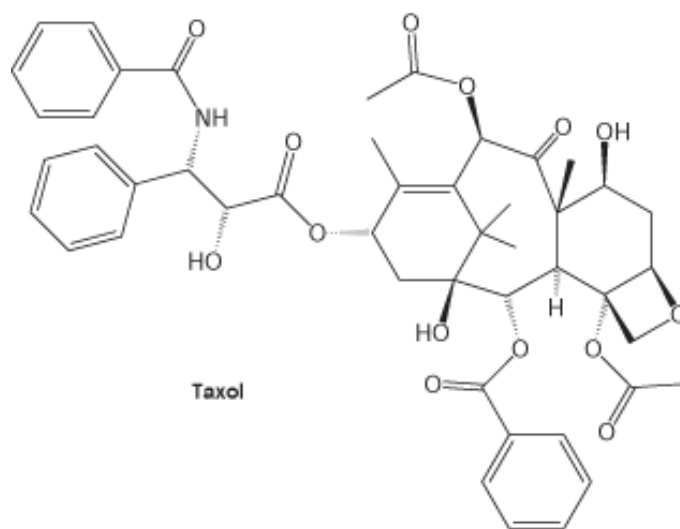
Microtubules are also key components of cilia and flagella present on some eukaryotic cells. For example, sperm propel themselves through the motion of flagella containing microtubules. The microtubules present in these structures adopt a common architecture (Figure 34.22). A bundle of microtubules called an *axoneme* is surrounded by a membrane contiguous with the plasma membrane. The axoneme is composed of a peripheral group of nine microtubule pairs surrounding two singlet microtubules. This recurring motif is often called a *9 + 2 array*. Dynein drives the motion of one

member of each outer pair relative to the other, causing the overall structure to bend.

Microtubules are important in determining the shapes of cells and in separating daughter chromosomes in mitosis. They are highly dynamic structures that grow through the addition of α - and β -tubulin to the ends of existing structures. Like actin, *tubulins* also bind and hydrolyze nucleoside triphosphates, although for tubulin the nucleotide is GTP rather than ATP. The critical concentration for the polymerization of the GTP forms of tubulin is lower than that for the GDP forms. Thus, a newly formed microtubule consists primarily of GTP-tubulins. Through time, the GTP is hydrolyzed to GDP. The GDP-tubulin subunits in the interior length of a microtubule remain stably polymerized, whereas GDP subunits exposed at an end have a strong tendency to dissociate. Marc Kirschner and Tim Mitchison found that some microtubules in a population lengthen while others simultaneously shorten. This property, called *dynamic instability*, arises from random fluctuations in the number of GTP- or GDP-tubulin subunits at the plus end of the polymer. The dynamic character of microtubules is crucial for processes such as mitosis, which require the assembly and disassembly of elaborate microtubule-based structures.

 The structure of tubulin was determined at high resolution by electron crystallographic methods (Figure 34.23). As expected from their 40% sequence identity, α - and β -tubulin have very similar three-dimensional structures. Further analysis revealed that the tubulins are members of the P-loop NTPase family and contain a nucleotide-binding site adjacent to the P-loop. Tubulins are present only in eukaryotes, although a prokaryotic homolog has been found. Sequence analysis identified a prokaryotic protein called FtsZ (for filamentous temperature-sensitive mutant Z) that is quite similar to the tubulins. The homology was confirmed when the structure was determined by x-ray crystallography. Interestingly, this protein participates in bacterial cell division, forming ring-shaped structures at the constriction that arises when a cell divides. These observations suggest that tubulins may have evolved from an ancient cell-division protein.

 The continual lengthening and shortening of microtubules is essential to their role in cell division. *Taxol*, a compound isolated from the bark of the Pacific yew tree, was discovered through its ability to interfere with cell proliferation. Taxol binds to microtubules and stabilizes the polymerized form.



Taxol and its derivatives have been developed as anticancer agents because they preferentially affect rapidly dividing cells, such as those in tumors.

34.3.2. Kinesin Motion Is Highly Processive

Kinesins are motor proteins that move along microtubules. We have seen that myosin moves along actin filaments by a process in which actin is released in each cycle; a myosin head group acting independently dissociates from actin after every power stroke. In contrast, when a kinesin molecule moves along a microtubule, the two head groups of the kinesin molecule operate in tandem — one binds, and then the next one does. A kinesin molecule may take many steps before

both heads groups are dissociated at the same time. In other words, the motion of kinesin is highly processive. Single-molecule measurements allow processive motion to be observed (Figure 34.24). A single kinesin molecule will typically take 100 or more steps toward the plus end of a microtubule in a period of seconds before the molecule becomes detached from the microtubule. These measurements also revealed that the average step size is approximately 80 Å, a value that corresponds to the distance between consecutive α - or β -tubulin subunits along each protofilament.

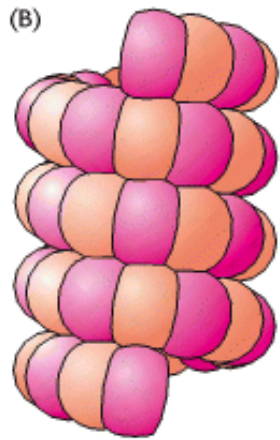
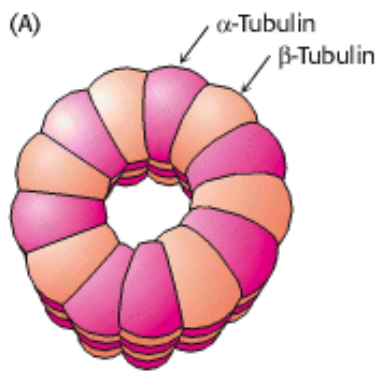
An additional fact is crucial to the development of a mechanism for kinesin motion — namely, that the addition of ATP strongly *increases* the affinity of kinesin for microtubules. This behavior stands in contrast with the behavior of myosin; ATP binding to myosin promotes its *dissociation* from actin. Do these differences imply that kinesin and myosin operate by completely different mechanisms? Indeed not. Kinesin-generated movement appears to proceed by a mechanism that is quite similar to that used by myosin (Figure 34.25). Let us begin with a two-headed kinesin molecule in its ADP form, dissociated from a microtubule. Recall that the neck linker binds the head domain when ATP is bound and is released when ADP is bound. The initial interaction of one of the head domains with a tubulin dimer on a microtubule stimulates the release of ADP from this head domain and the subsequent binding of ATP. The binding of ATP triggers a conformational change in the head domain that leads to two important events. First, the affinity of the head domain for the microtubule increases, essentially locking this head domain in place. Second, the neck linker binds to the head domain. This change repositions the other head domain acting through the coiled-coil domain that connects the two kinesin monomers. In its new position, the second head domain is close to a second tubulin dimer, 80 Å along the microtubule in the direction of the plus end. Meanwhile, the intrinsic ATPase activity of the first head domain hydrolyzes the ATP to ADP and P_i . When the second head domain binds to the microtubule, the first head releases ADP and binds ATP. Again, ATP binding favors a conformational change that pulls the first domain forward. This process can continue for many cycles until, by chance, both head domains are in the ADP form simultaneously and kinesin dissociates from the microtubule. Because of the relative rates of the component reactions, a simultaneous dissociation occurs approximately every 100 cycles.

Kinesin hydrolyzes ATP at a rate of approximately 80 molecules per second. Thus, given the step size of 80 Å per molecule of ATP, kinesin moves along a microtubule at a speed of 6400 Å per second. This rate is considerably slower than the maximum rate for myosin, which moves relative to actin at 80,000 Å per second. Recall, however, that myosin movement depends on the independent action of hundreds of different head domains working along the same actin filament, whereas the movement of kinesin is driven by the processive action of kinesin head groups working in pairs. Muscle myosin evolved to maximize the speed of the motion, whereas kinesin functions to achieve steady, but slower, transport in one direction along a filament.

34.3.3. Small Structural Changes Can Reverse Motor Polarity

Most members of the kinesin family move toward the plus end of microtubules. However, a small number, including the protein *ncd* (for *nonclaret disjunctional*, first identified in *Drosophila*), move toward the minus end. From an engineering perspective, there are many ways to change the polarity of a motor. How is polarity changed in this case?

Determination of the core structure of *ncd* revealed great similarity to other kinesins in the mechanical parts of the motor domain, including the structures of the switch regions, the relay helix, and the parts that bind microtubules. Significantly, however, the motor domain of *ncd* lies near the carboxyl terminus of the protein, whereas it lies near the amino terminus of conventional kinesin. Furthermore, when a larger fragment of *ncd* bound to ADP was analyzed, a short region just before the motor domain was seen to form an α helix that docks against the motor domain in a position similar to that occupied by the neck linker of conventional kinesin *in the ATP form* (Figure 34.26). This finding suggests that *ncd* moves toward the minus end of a microtubule by a mechanism only slightly different from that used by conventional kinesin to move in the opposite direction. Whereas ATP binding by conventional kinesin leads to the binding of the neck linker, ATP binding by *ncd* *releases* the helical region. Its release allows the second motor domain of the *ncd* dimer to bind to a site on the microtubule farther toward the minus end (Figure 34.27). We see once again the economical refinement of a protein by evolution — in this case, subtle adjustments have produced an opposite mechanical result in the activity of the protein assembly.



300 Å (30 nm)

Figure 34.21. Microtubule Structure. Schematic views of the helical structure of a microtubule. α -tubulin is shown in dark red and β -tubulin in light red. (A) Top view. (B) Side view.

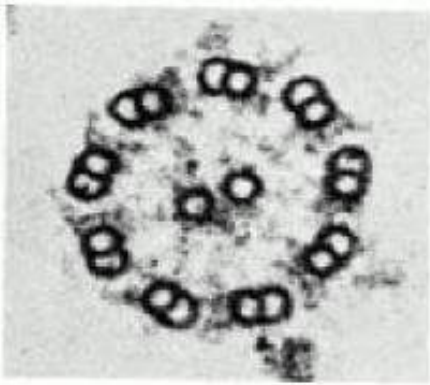


Figure 34.22. Microtubule Arrangement. Electron micrograph of a cross section of a flagellar axoneme shows nine microtubule doublets surrounding two singlets. [Courtesy of Dr. Joel Rosenbaum.]

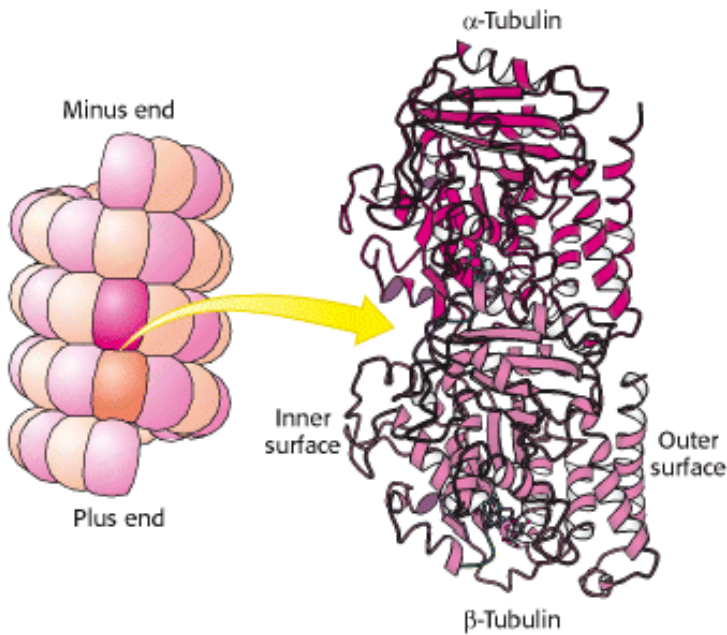


Figure 34.23. Tubulin. Microtubules can be viewed as an assembly of α -tubulin - β -tubulin dimers. The structures of α -tubulin and β -tubulin are quite similar; each includes a P-loop NTPase domain (purple shading) and a bound guanine nucleotide.

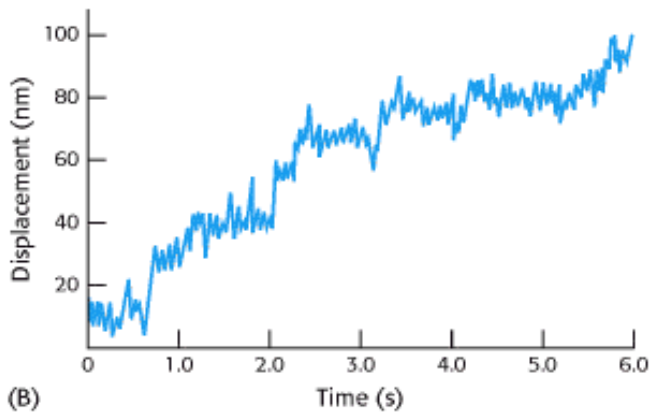
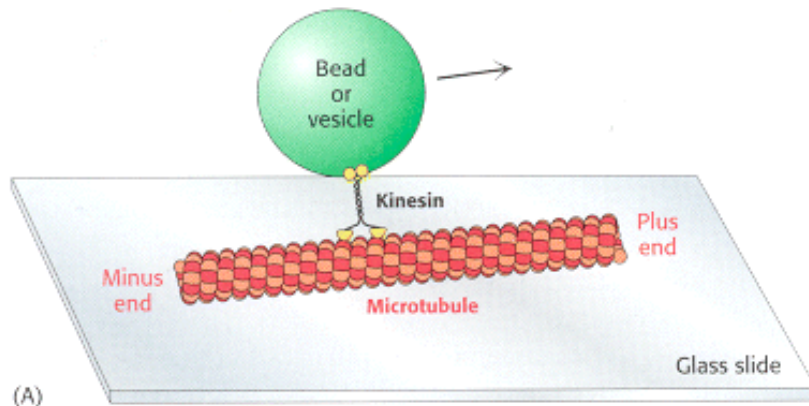


Figure 34.24. Monitoring Movements Mediated by Kinesin. (A) The movement of beads or vesicles, carried by individual kinesin dimers along a microtubule, can be directly observed. (B) A trace shows the displacement of a bead carried by a kinesin molecule. Multiple steps are taken in the 6-s interval. The average step size is about 8 nm (80 Å) [Part B after K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block. *Nature* 365(1993):721.]

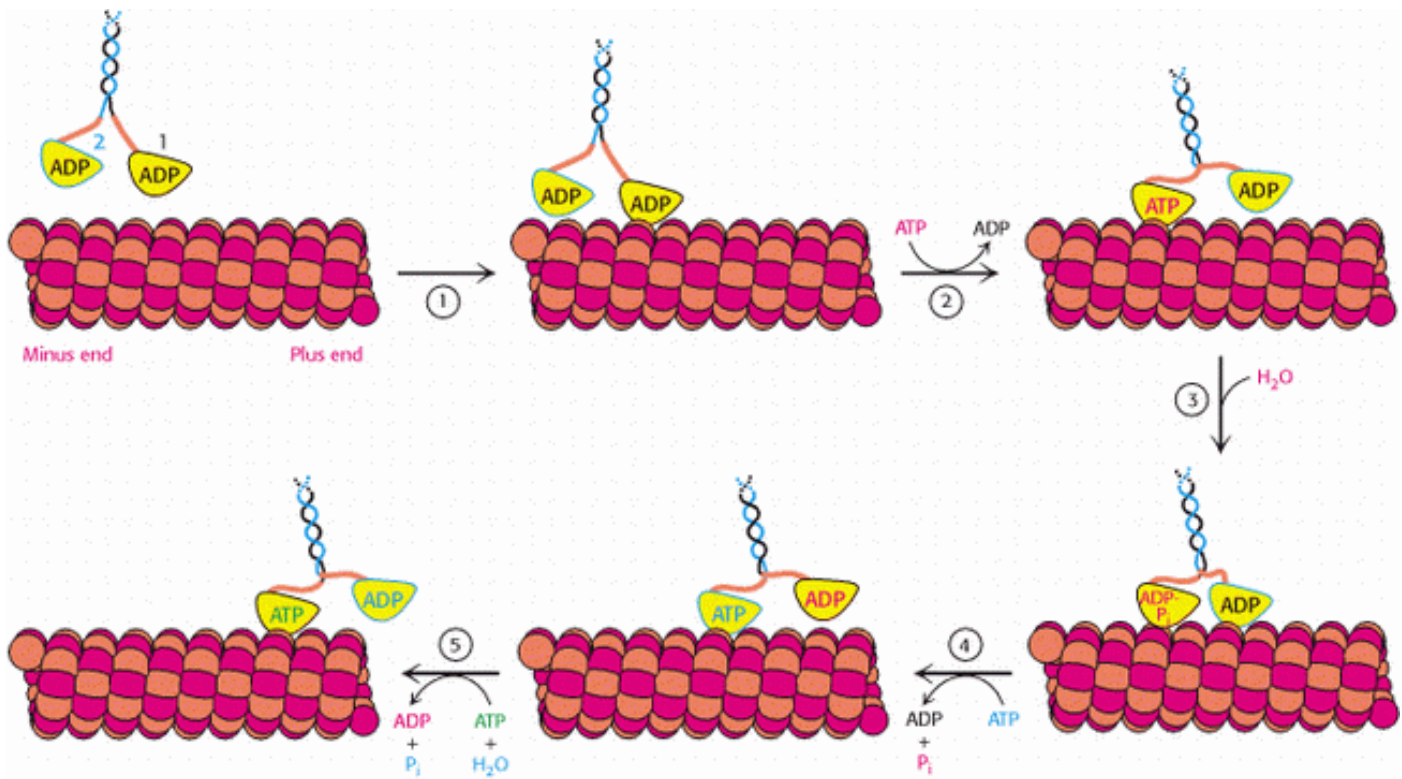


Figure 34.25. Kinesin Moving Along a Microtubule. (1) One head of a two-headed kinesin molecule, initially with both heads in the ADP form, binds to a microtubule. (2) Release of ADP and binding of ATP results in a conformational change that locks the head to the microtubule and pulls the neck linker (orange) to the head domain, throwing the second domain toward the plus end of the microtubule. (3) ATP hydrolysis occurs while the second head interacts with the microtubule. (4) The exchange of ATP for ADP in the second head pulls the first head off the microtubule, releasing P_i and moving the first domain along the microtubule. (5) The cycle repeats, moving the kinesin dimer farther down the microtubule.

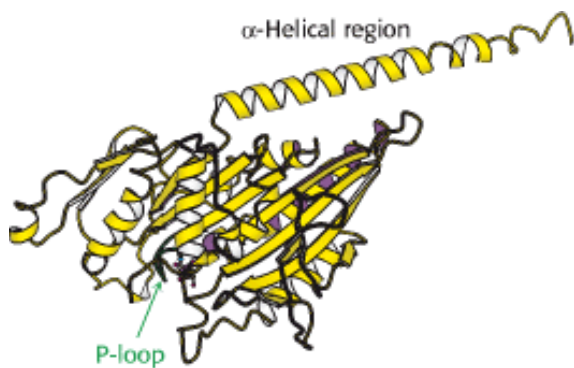


Figure 34.26. Structure of Ncd. The head domain of ncd is quite similar to that of conventional kinesin, including the presence of a P-loop NTPase domain (shaded in purple). In the ADP form of ncd (shown), the amino-terminal part of this fragment forms an α -helix that docks into the site occupied by the neck linker in the ATP form of conventional kinesin.

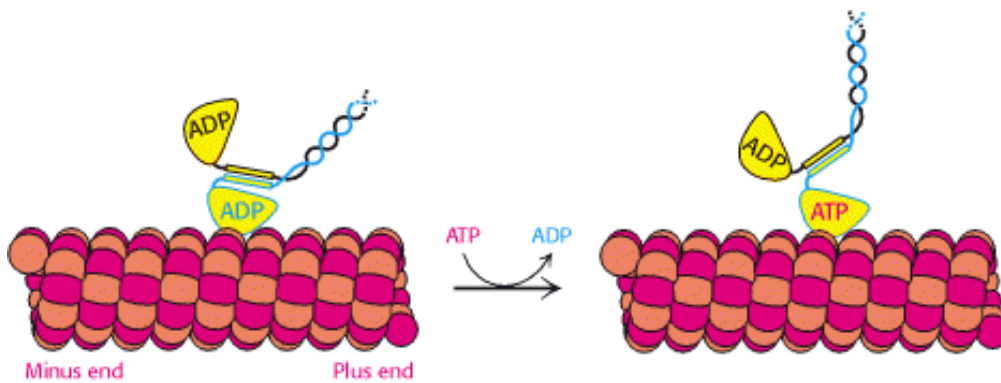


Figure 34.27. Motion in Ncd. The replacement of ATP for ADP releases the α -helical region from the head domain, moving the other head domain toward the minus end of the microtubule.

34.4. A Rotary Motor Drives Bacterial Motion

In one second, a motile bacterium can move approximately 25 μm , or about 10 body lengths. A human being sprinting at a proportional rate would complete the 100-meter dash in slightly more than 5 seconds. The motors that power this impressive motion are strikingly different from the eukaryotic motors that we have seen so far. In the bacterial motor, an element spins around a central axis rather than moving along a polymeric track. The direction of rotation can change rapidly, a feature that is central to chemotaxis, the process by which bacteria swim preferentially toward an increasing concentration of certain useful compounds and away from potentially harmful ones.

34.4.1. Bacteria Swim by Rotating Their Flagella

Bacteria such as *Escherichia coli* and *Salmonella typhimurium* swim by rotating flagella that lie on their surfaces (Figure 34.28). When the flagella rotate in a counterclockwise direction (viewed from outside the bacterium), the separate flagella form a bundle that very efficiently propels the bacterium through solution.

Bacterial flagella are polymers approximately 15 nm in diameter and as much as 15 μm in length, composed of 53-kd subunits of a protein called *flagellin* (Figure 34.29). These subunits associate into a helical structure that has 5.5 subunits per turn, giving the appearance of 11 protofilaments. Each flagellum has a hollow core. Remarkably, flagella form not by growing at the base adjacent to the cell body but, instead, by the addition of new subunits that pass through the hollow core and add to the free end. Each flagellum is intrinsically twisted in a left-handed sense. At its base, each flagellum has a rotary motor.

34.4.2. Proton Flow Drives Bacterial Flagellar Rotation

Early experiments by Julius Adler demonstrated that ATP is *not* required for flagellar motion. What powers these rotary motors? The necessary free energy is derived from the proton gradient that exists across the plasma membrane. The flagellar motor is quite complex, containing as many as 40 distinct proteins (Figure 34.30). Five components particularly crucial to motor function have been identified through genetic studies. MotA is a membrane protein that appears to have four transmembrane helices as well as a cytoplasmic domain. MotB is another membrane protein with a single transmembrane helix and a large periplasmic domain. Approximately 11 *MotA* - *MotB* pairs form a ring around the base of the flagellum. The proteins *FliG*, *FliM*, and *FliN* are part of a disc-like structure called the MS (*m*embrane and *s*upramembrane) ring, with approximately 30 *FliG* subunits coming together to form the ring. The three-dimensional structure of the carboxyl-terminal half of *FliG* reveals a wedge-shaped domain with a set of charged amino acids, conserved among many species, lying along the thick edge of the wedge (Figure 34.31).

The MotA - MotB pair and FliG combine to create a proton channel that drives rotation of the flagellum. How can proton flow across a membrane drive mechanical rotation? We have seen such a process earlier in regard to ATP synthase (Section 18.4.4). Recall that the key to driving the rotation of the γ subunit of ATP synthase is the **a** subunit of the F_0 fragment. This subunit appears to have two half-channels; protons can move across the membrane only by moving into the half-channel from the side of the membrane with the higher local proton concentration, binding to a disc-like structure formed by the **c** subunits, riding on this structure as it rotates to the opening of the other half-channel, and exiting to the side with the lower local proton concentration. Could a similar mechanism apply to flagellar rotation? Indeed, such a mechanism was first proposed by Howard Berg to explain flagellar rotation before the rotary mechanism of ATP synthase was elucidated. Each MotA - MotB pair is conjectured to form a structure that has two half-channels; FliG serves as the rotating proton carrier, perhaps with the participation of some of the charged residues identified in crystallographic studies (Figure 34.32). In this scenario, a proton from the periplasmic space passes into the outer half-channel and is transferred to an FliG subunit. The MS ring rotates, rotating the flagellum with it and allowing the proton to pass into the inner half-channel and into the cell. Ongoing structural and mutagenesis studies are testing and refining this hypothesis.

34.4.3. Bacterial Chemotaxis Depends on Reversal of the Direction of Flagellar Rotation

Many species of bacteria respond to changes in their environments by adjusting their swimming behavior. Examination of the paths taken is highly revealing (Figure 34.33). The bacteria swim in one direction for some length of time (typically about a second), tumble briefly, and then set off in a new direction. The tumbling is caused by a brief reversal in the direction of the flagellar motor. When the flagella rotate counterclockwise, the helical filaments form a coherent bundle favored by the intrinsic shape of each filament, and the bacterium swims smoothly. When the rotation reverses, the bundle flies apart because the screw sense of the helical flagella does not match the direction of rotation (Figure 34.34). Each flagellum then pulls in a different direction and the cell tumbles.

In the presence of a gradient of certain substances such as glucose, bacteria swim preferentially toward the direction of the higher concentration of the substance. Such compounds are referred to as *chemoattractants*. Bacteria also swim preferentially away from potentially harmful compounds such as phenol, a *chemorepellant*. The process of moving in specific directions in response to environmental cues is called *chemotaxis*. In the presence of a gradient of a chemoattractant, bacteria swim for longer periods of time without tumbling when moving toward higher concentrations of chemoattractant. In contrast, they tumble more frequently when moving toward lower concentrations of chemoattractant. This behavior is reversed for chemorepellants. The result of these actions is a *biased random walk* that facilitates net motion toward conditions more favorable to the bacterium.

Chemotaxis depends on a signaling pathway that terminates at the flagellar motor. The signaling pathway begins with the binding of molecules to receptors in the plasma membrane (Figure 34.35). In their *unoccupied* forms, these receptors initiate a pathway leading eventually to the phosphorylation of a specific aspartate residue on a soluble protein called *CheY*. In its phosphorylated form, CheY binds to the base on the flagellar motor. When bound to phosphorylated CheY, the flagellar motor rotates in a clockwise rather than a counterclockwise direction, causing tumbling.

The binding of a chemoattractant to a surface receptor blocks the signaling pathway leading to CheY phosphorylation. Phosphorylated CheY spontaneously hydrolyzes and releases its phosphate group in a process accelerated by another protein, CheZ. The concentration of phosphorylated CheY drops, and the flagella are less likely to rotate in a clockwise direction. Under these conditions, bacteria swim smoothly without tumbling. Thus, the reversible rotary flagellar motor and a phosphorylation-based signaling pathway work together to generate an effective means for responding to environmental conditions.

Bacteria sense spatial gradients of chemoattractants by measurements separated in time. A bacterium sets off in a random direction and, if the concentration of the chemoattractant has increased after the bacterium has been swimming for a period of time, the likelihood of tumbling decreases and the bacterium continues in roughly the same direction. If the concentration has decreased, the tumbling frequency increases and the bacterium tests other random directions. The

success of this mechanism once again reveals the power of evolutionary problem solving — many possible solutions are tried at random, and those that are beneficial are selected and exploited.

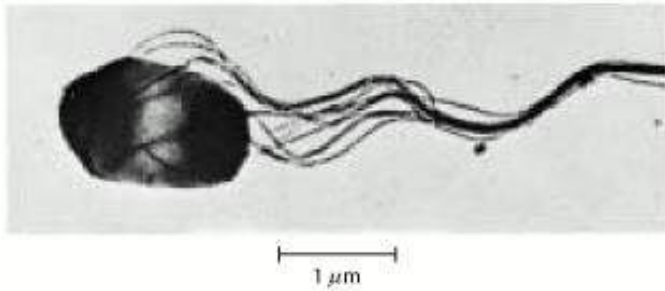


Figure 34.28. Bacterial Flagella. Electron micrograph of *S. typhimurium* shows flagella in a bundle. [Courtesy of Dr. Daniel Koshland, Jr.]

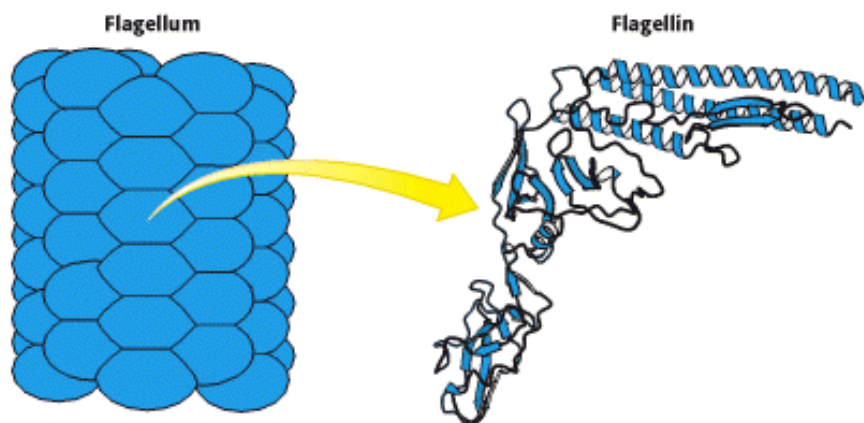


Figure 34.29. Structure of Flagellin. A bacterial flagellum is a helical polymer of the protein flagellin.

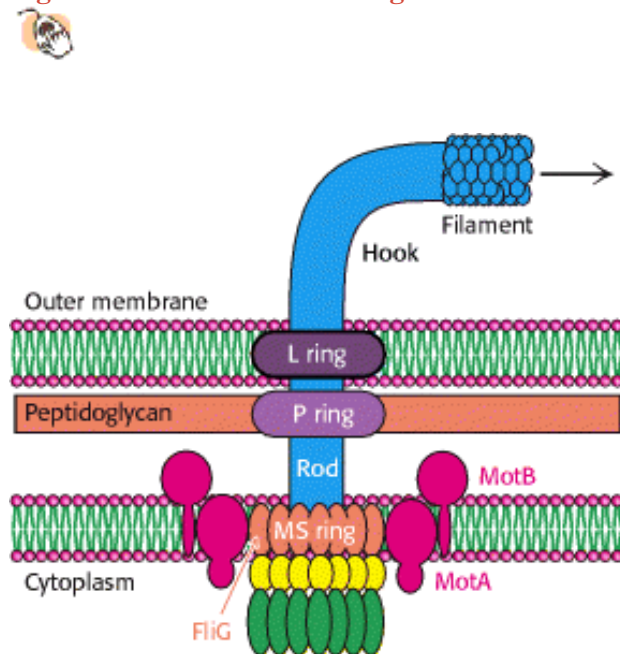


Figure 34.30. Flagellar Motor. A schematic view of the flagellar motor, a complex structure containing as many as 40 distinct types of protein. The approximate positions of the proteins MotA and MotB (red), FliG (orange), FliN (yellow), and FliM (green) are shown.

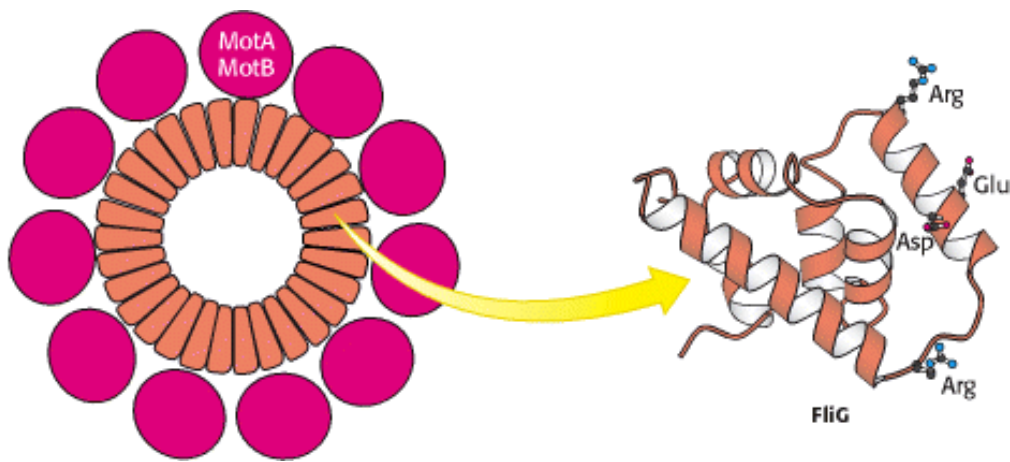


Figure 34.31. Flagellar Motor Components. Approximately 30 subunits of FliG assemble to form part of the MS ring. The ring is surrounded by approximately 11 structures consisting of MotA and MotB. The carboxyl-terminal domain of FliG includes a ridge lined with charged residues that may participate in proton transport.

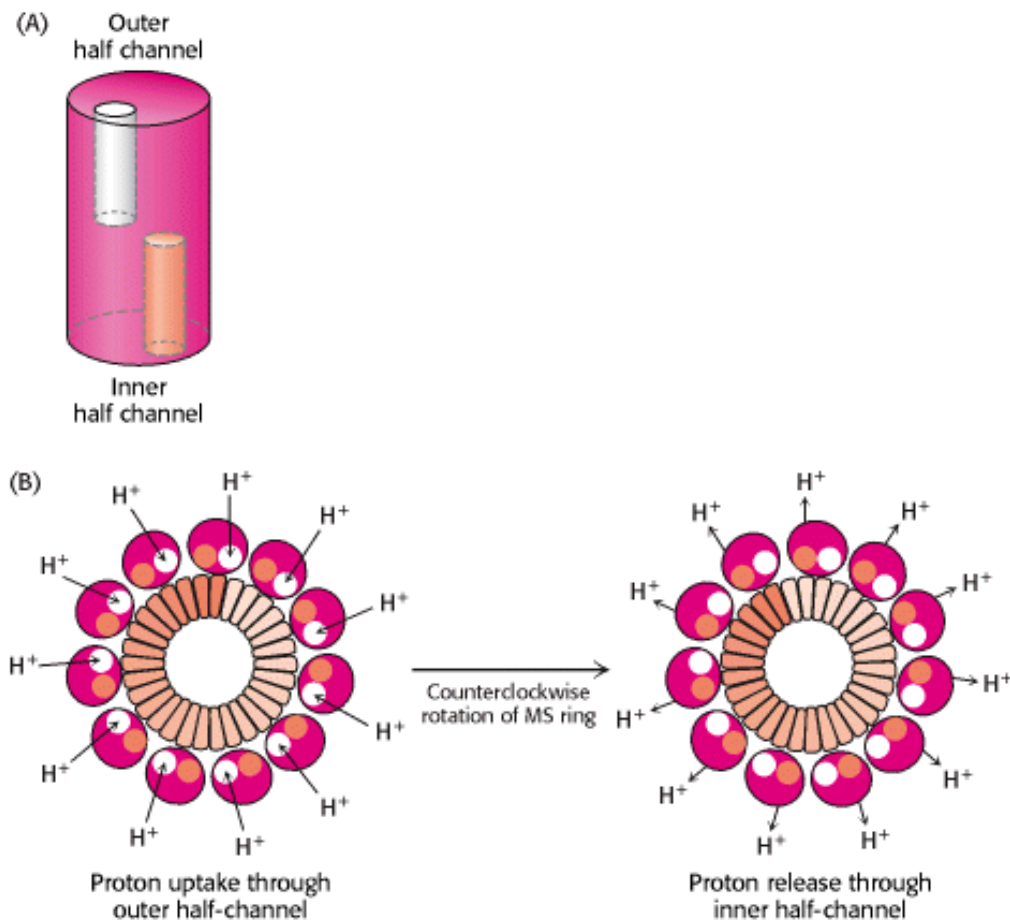


Figure 34.32. Proton Transport-Coupled Rotation of the Flagellum. (A) MotA-MotB may form a structure having two half-channels. (B) One model for the mechanism of coupling rotation to a proton gradient requires protons to be taken up into the outer half-channel and transferred to the MS ring. The MS ring rotates in a counterclockwise direction, and the protons are released into the inner half-channel. The flagellum is linked to the MS ring and so the flagellum rotates as well.



Figure 34.33. Charting a Course. This projection of the track of an *E. coli* bacterium was obtained with a microscope that automatically follows bacterial motion in three dimensions. The points show the locations of the bacterium at 80-ms intervals. [After H. C. Berg. *Nature* 254(1975):390.]

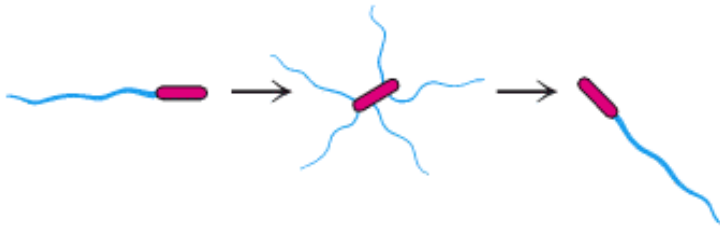


Figure 34.34. Changing Direction. Tumbling is caused by an abrupt reversal of the flagellar motor, which disperses the flagellar bundle. A second reversal of the motor restores smooth swimming, almost always in a different direction. [After a drawing kindly provided by Dr. Daniel Koshland, Jr.]

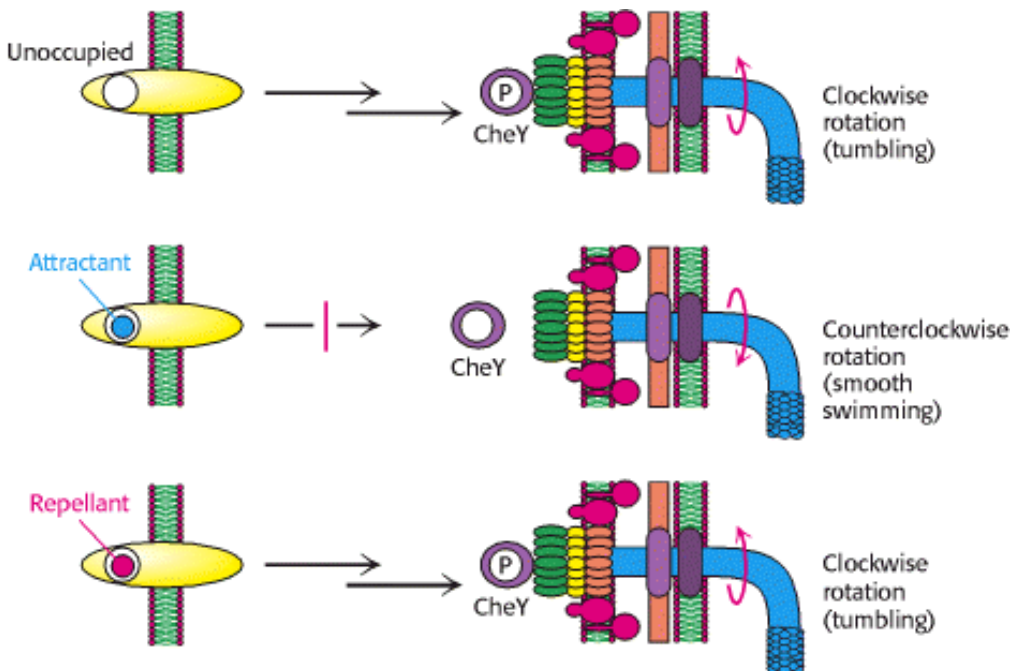


Figure 34.35. Chemotaxis Signaling Pathway. Receptors in the plasma membrane initiate a signaling pathway leading to the phosphorylation of the CheY protein. Phosphorylated CheY binds to the flagellar motor and favors clockwise rotation. When an attractant binds to the receptor, this pathway is blocked, and counterclockwise flagellar rotation and, hence, smooth swimming results. When a repellent binds, the pathway is stimulated, leading to an increased concentration of phosphorylated CheY and, hence, more frequent clockwise rotation and tumbling.

Summary

Most Molecular-Motor Proteins Are Members of the P-Loop NTPase Superfamily

Eukaryotic cells contain three families of molecular-motor proteins: myosins, kinesins, and dyneins. These proteins move along tracks defined by the actin and microtubule cytoskeletons of eukaryotic cells, contributing to cell and organismal movement and to the intracellular transport of proteins, vesicles, and organelles. Despite considerable differences in size and a lack of similarity detectable at the level of amino acid sequence, these proteins are homologous, containing core structures of the P-loop NTPase family. The ability of these core structures to change conformations in response to nucleoside triphosphate binding and hydrolysis is key to molecular-motor function. Motor proteins consist of motor domains attached to extended structures that serve to amplify the conformational changes in the core domains and to link the core domains to one another or to other structures.

Myosins Move Along Actin Filaments

The motile structure of muscle consists of a complex of myosin and actin, along with accessory proteins. Actin, a highly abundant 42-kd protein, polymerizes to form long filaments. Each actin monomer can bind either ATP or ADP. Muscle contraction entails the rapid sliding of thin filaments, based on actin, relative to thick filaments, composed of myosin. A myosin motor domain moves along actin filaments in a cyclic manner: (1) myosin complexed to ADP and P_i binds actin; (2) P_i is released; (3) a conformational change leads to a large motion of a lever arm that extends from the motor domain, moving the actin relative to myosin; (4) ATP replaces ADP, resetting the position of the lever arm and releasing actin; (5) the hydrolysis of ATP returns the motor domain to its initial state. The length of the lever arm determines the size of the step taken along actin in each cycle. The ability to monitor single molecular-motor proteins has provided key tests for hypotheses concerning motor function.

Kinesin and Dynein Move Along Microtubules

Kinesin and dynein move along microtubules rather than actin. Microtubules are polymeric structures composed of α - and β -tubulin, two very similar guanine-nucleotide-binding proteins. Each microtubule comprises 13 protofilaments with alternating α - and β -tubulin subunits. Kinesins move along microtubules by a mechanism quite similar to that used by myosin to move along actin, but with several important differences. First, ATP binding to kinesin favors motor-domain binding rather than dissociation. Second, the power stroke is triggered by the binding of ATP rather than the release of P_i . Finally, kinesin motion is processive. The two heads of a kinesin dimer work together, taking turns binding and releasing the microtubule, and many steps are taken along a microtubule before both heads dissociate. Whereas most kinesins move toward the plus end of a microtubule, some such as *Drosophila ncd* move in the opposite direction. Movement in the direction of the minus end is accomplished through a mechanism very similar to that used for plus-end-directed motion, but with critical structural differences accounting for the difference in direction.

A Rotary Motor Drives Bacterial Motion

Many motile bacteria use rotating flagella to propel themselves. When rotating counterclockwise, multiple flagella on the surface of a bacterium come together to form a bundle that effectively propels the cell through solution. When rotating clockwise, the flagella fly apart and the cell tumbles. In a homogeneous environment, bacteria swim smoothly for approximately 1 second and then reorient themselves by tumbling. Bacteria swim preferentially toward chemoattractants

in a process called chemotaxis. When bacteria are swimming in the direction of an increasing concentration of a chemoattractant, clockwise flagellar motion and tumbling is suppressed, leading to a biased random walk in the direction of increasing chemoattractant concentration. A proton gradient across the plasma membrane, rather than ATP hydrolysis, powers the flagellar motor. The mechanism for coupling transmembrane proton transport to macromolecular rotation appears to be similar to that used by ATP synthase.

Key Terms

myosin

kinesin

dynein

S1 fragment

conventional kinesin

level arm

relay helix

neck linker

actin

myofibril

sarcomere

tropomyosin

troponin complex

G-actin

F-actin

critical concentration

optical trap

power stroke

microtubule

tubulin

dynamic instability

ncd (nonclaret disjunctional)

flagellin

MotA - MotB pair

FliG

chemoattractant

chemorepellant

chemotaxis

CheY

Problems

1. *Diverse motors.* Skeletal muscle, eukaryotic cilia, and bacterial flagella use different strategies for the conversion of free energy into coherent motion. Compare and contrast these motility systems with respect to (a) the free-energy source and (b) the number of essential components and their identity.

See answer

2. *You call that slow?* At maximum speed, a kinesin molecule moves at a rate of 6400 Å per second. Given the dimensions of the motor region of a kinesin dimer of approximately 80 Å, calculate its speed in "body lengths" per second. What speed does this body-length speed correspond to for an automobile 10 feet long?

See answer

3. *Heavy lifting.* A single myosin motor domain can generate a force of approximately 4 piconewtons (4 pN). How many times its "bodyweight" can a myosin motor domain lift? Note that 1 newton = 0.22 pounds. Assume a molecular mass of 100 kd for the motor domain.

See answer

4. *Rigor mortis.* Why does the body stiffen after death?

See answer

5. *Now you see it, now you don't.* Under certain stable concentration conditions, actin monomers in their ATP form will polymerize to form filaments that disperse again into free actin monomers over time. Provide an explanation.

See answer

6. *Open and Schutt case?* In our consideration of muscle contraction, we have assumed that actin filaments play an entirely passive role, being pulled by myosin. What property of actin suggests that it might play a more active role?

See answer

7. *Designer kinesins.* Hybrids of kinesin have been prepared that consist of either (a) the sequence of conventional kinesin with the motor domain replaced by that of ncd or (b) the sequence of ncd with the motor domain replaced by that of conventional kinesin. In which direction along microtubules would you predict these hybrid kinesins to move?

See answer

8. *Helicases as motors.* Helicases such as PcrA (Section 27.2.5) can use single-stranded DNA as tracks. In each cycle, the helicase moves one base in the $3' \rightarrow 5'$ direction. Given that PcrA can hydrolyze ATP at a rate of 50 molecules per second in the presence of a single-stranded DNA template, calculate the velocity of the helicase in micrometers per second. How does this velocity compare with that of kinesin?

See answer

9. *New moves.* When bacteria such as *E. coli* are starved to a sufficient extent, they become nonmotile. However, when such bacteria are placed in an acidic solution, they resume swimming. Explain.

See answer

10. *Hauling a load.* Consider the action of a single kinesin molecule in moving a vesicle along a microtubule track. The force required to drag a spherical particle of radius a at a velocity v in a medium having a viscosity η is

$$F = 6\pi\eta av$$

Suppose that a 2- μm bead is carried at a velocity of $0.6 \mu\text{m s}^{-1}$ in an aqueous medium ($\eta = 0.01 \text{ poise} = 0.01 \text{ g cm}^{-1} \text{ s}^{-1}$).

(a) What is the magnitude of the force exerted by the kinesin molecule? Express the value in dynes (1 dyne = 1 g cm s^{-2}).

(b) How much work is performed in 1 second? Express the value in ergs (1 erg = 1 dyne cm).

(c) A kinesin motor hydrolyzes approximately 80 molecules of ATP per second. What is the energy associated with the hydrolysis of this much ATP in ergs? Compare this value with the actual work performed.

See answer

11. *Unusual strides.* A publication describes a kinesin molecule that is claimed to move along microtubules with a step size of 6 nm. You are skeptical. Why?

See answer

12. *The sound of one hand clapping.* KIF1A is a processive motor protein that moves toward the plus end of microtubules *as a monomer*. KIF1A has only a single motor domain. What additional structural elements would you expect to find in the KIF1A structure.

See answer

Mechanism Problem

13. *Backward rotation.* On the basis of the proposed structure in [Figure 34.32](#) for the bacterial flagellar motor, suggest a pathway for transmembrane proton flow when the flagellar motor is rotating clockwise rather than counterclockwise.

See answer

Chapter Integration Problem

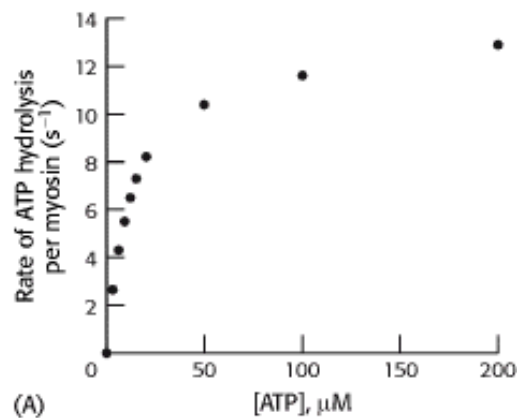
14. *Smooth muscle.* Smooth muscle, in contrast with skeletal muscle, is not regulated by a tropomyosin - troponin mechanism. Instead, vertebrate smooth muscle contraction is controlled by the degree of phosphorylation of its light chains. Phosphorylation induces contraction, and dephosphorylation leads to relaxation. Like that of skeletal muscle, smooth muscle contraction is triggered by an increase in the cytoplasmic calcium ion level. Propose a mechanism for this action of calcium ion on the basis of your knowledge of other signal-transduction processes.

See answer

Data Interpretation Problem

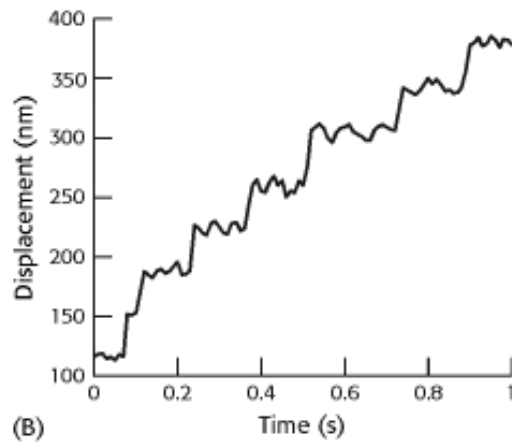
15. *Myosin V.* An abundant myosin-family member, myosin V is isolated from brain tissue. This myosin has a number of unusual properties. First, on the basis of its amino acid sequence, each heavy chain has six tandem binding sites for calmodulin-like light chains. Second, it forms dimers but not higher-order oligomers. Finally, unlike almost all other myosin-family members, myosin V is highly processive.

The rate of ATP hydrolysis by myosin has been examined as a function of ATP concentration, as shown in graph A.



- (a) Estimate the values of k_{cat} and K_{M} for ATP.

With the use of optical-trap measurements, the motion of single myosin V dimers could be followed, as shown in graph B.



(b) Estimate the step size for myosin V.

The rate of ADP release from myosin V is found to be approximately $13 \text{ molecules s}^{-1}$.

(c) Combine the observations about the amino acid sequence of myosin, the observed step size, and the kinetics results to propose a mechanism for the processive motion of myosin V.

[Based on M. Rief, R. S. Rock, A. D. Mehta, M. S. Mooseker, R. E. Cheney, and J. A. Spudich. *Proc. Natl. Acad. Sci. USA* 97(2000):9482.]

See answer

Selected Readings

Where to start

R.D. Vale and R.A. Milligan. 2000. The way things move: Looking under the hood of molecular motor proteins *Science* 288: 88-95. ([PubMed](#))

R.D. Vale. 1996. Switches, latches, and amplifiers: Common themes of G proteins and molecular motors *J. Cell Biol.* 135: 291-302. ([PubMed](#))

A.D. Mehta, M. Rief, J.A. Spudich, D.A. Smith, and R.M. Simmons. 1999. Single-molecule biomechanics with optical methods *Science* 283: 1689-1695. ([PubMed](#))

S.C. Schuster and S. Khan. 1994. The bacterial flagellar motor *Annu. Rev. Biophys. Biomol. Struct.* 23: 509-539. ([PubMed](#))

Books

Howard, J., 2001. *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer.

Squire, J. M., 1986. *Muscle Design, Diversity, and Disease*. Benjamin Cummings.

Pollack, G. H., and Sugi, H. (Eds.), 1984. *Contractile Mechanisms in Muscle*. Plenum.

Myosin and actin

K.C. Holmes. 1997. The swinging lever-arm hypothesis of muscle contraction *Curr. Biol.* 7: R112-R118. ([PubMed](#))

J.S. Berg, B.C. Powell, and R.E. Cheney. 2001. A millennial myosin census *Mol. Biol. Cell* 12: 780-794. ([PubMed](#)) ([Full Text in PMC](#))

A. Houdusse, V.N. Kalabokis, D. Himmel, A.G. Szent-Györgyi, and C. Cohen. 1999. Atomic structure of scallop myosin subfragment S1 complexed with MgADP: A novel conformation of the myosin head *Cell* 97: 459-470. ([PubMed](#))

A. Houdusse, A.G. Szent-Györgyi, and C. Cohen. 2000. Three conformational states of scallop myosin S1 *Proc. Natl. Acad. Sci. USA* 97: 11238-11243. ([PubMed](#)) ([Full Text in PMC](#))

T.Q. Uyeda, P.D. Abramson, and J.A. Spudich. 1996. The neck region of the myosin motor domain acts as a lever arm to generate movement *Proc. Natl. Acad. Sci. USA* 93: 4459-4464. ([PubMed](#)) ([Full Text in PMC](#))

A.D. Mehta, R.S. Rock, M. Rief, J.A. Spudich, M.S. Mooseker, and R.E. Cheney. 1999. Myosin-V is a processive actin-based motor *Nature* 400: 590-593. ([PubMed](#))

L.R. Otterbein, P. Graceffa, and R. Dominguez. 2001. The crystal structure of uncomplexed actin in the ADP state *Science* 293: 708-711. ([PubMed](#))

K.C. Holmes, D. Popp, W. Gebhard, and W. Kabsch. 1990. Atomic model of the actin filament *Nature* 347: 44-49. ([PubMed](#))

C.E. Schutt, J.C. Myslik, M.D. Rozycki, N.C. Goonesekere, and U. Lindberg. 1993. The structure of crystalline profilin-beta-actin *Nature* 365: 810-816. ([PubMed](#))

F. van den Ent, L.A. Amos, and J. Lowe. 2001. Prokaryotic origin of the actin cytoskeleton *Nature* 413: 39-44. ([PubMed](#))

C.E. Schutt and U. Lindberg. 1998. Muscle contraction as a Markov process I: Energetics of the process *Acta Physiol. Scand.* 163: 307-323. ([PubMed](#))

M. Rief, R.S. Rock, A.D. Mehta, M.S. Mooseker, R.E. Cheney, and J.A. Spudich. 2000. Myosin-V stepping kinetics: A molecular model for processivity *Proc. Natl. Acad. Sci. USA* 97: 9482-9486. ([PubMed](#)) ([Full Text in PMC](#))

T.B. Friedman, J.R. Sellers, and K.B. Avraham. 1999. Unconventional myosins and the genetics of hearing loss *Am. J. Med. Genet.* 89: 147-157. ([PubMed](#))

Kinesin, dynein, and microtubules

R.D. Vale and R.J. Fletterick. 1997. The design plan of kinesin motors *Annu. Rev. Cell. Dev. Biol.* 13: 745-777. ([PubMed](#))

F.J. Kull, E.P. Sablin, R. Lau, R.J. Fletterick, and R.D. Vale. 1996. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin *Nature* 380: 550-555. ([PubMed](#))

M. Kikkawa, E.P. Sablin, Y. Okada, H. Yajima, R.J. Fletterick, and N. Hirokawa. 2001. Switch-based mechanism of kinesin motors *Nature* 411: 439-445. ([PubMed](#))

R.H. Wade and F. Kozielski. 2000. Structural links to kinesin directionality and movement *Nat. Struct. Biol.* 7: 456-460. ([PubMed](#))

M. Yun, X. Zhang, C.G. Park, H.W. Park, and S.A. Endow. 2001. A structural pathway for activation of the kinesin motor ATPase *EMBO J.* 20: 2611-2618. ([PubMed](#)) ([Full Text in PMC](#))

F. Kozielski, S. De Bonis, W.P. Burmeister, C. Cohen-Addad, and R.H. Wade. 1999. The crystal structure of the minus-end-directed microtubule motor protein ncd reveals variable dimer conformations *Structure Fold Des.* 7: 1407-1416. ([PubMed](#))

J. Lowe, H. Li, K.H. Downing, and E. Nogales. 2001. Refined structure of α β -tubulin at 3.5 Å resolution *J. Mol. Biol.* 313: 1045-1057. ([PubMed](#))

E. Nogales, K.H. Downing, L.A. Amos, and J. Lowe. 1998. Tubulin and FtsZ form a distinct family of GTPases *Nat. Struct. Biol.* 5: 451-458. ([PubMed](#))

C. Zhao, J. Takita, Y. Tanaka, M. Setou, T. Nakagawa, S. Takeda, H.W. Yang, S. Terada, T. Nakata, Y. Takei, M. Saito, S. Tsuji, Y. Hayashi, and N. Hirokawa. 2001. Charcot-Marie-Tooth disease type 2A caused by mutation in a microtubule motor KIF1Bbeta *Cell* 105: 587-597. ([PubMed](#))

D.J. Asai and M.P. Koonce. 2001. The dynein heavy chain: Structure, mechanics and evolution *Trends Cell Biol.* 11: 196-202. ([PubMed](#))

G. Mocz and I.R. Gibbons. 2001. Model for the motor component of dynein heavy chain based on homology to the AAA family of oligomeric ATPases *Structure* 9: 93-103. ([PubMed](#))

Bacterial motion and chemotaxis

H.C. Berg. 2000. Constraints on models for the flagellar rotary motor *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355: 491-501. ([PubMed](#))

D.J. DeRosier. 1998. The turn of the screw: The bacterial flagellar motor *Cell* 93: 17-20. ([PubMed](#))

W.S. Ryu, R.M. Berry, and H.C. Berg. 2000. Torque-generating units of the flagellar motor of *Escherichia coli* have a high duty ratio *Nature* 403: 444-447. ([PubMed](#))

S.A. Lloyd, F.G. Whitby, D.F. Blair, and C.P. Hill. 1999. Structure of the C-terminal domain of FliG, a component of the rotor in the bacterial flagellar motor *Nature* 400: 472-475. ([PubMed](#))

E.M. Purcell. 1977. Life at low Reynolds number *Am. J. Physiol.* 45: 3-11.

R.M. Macnab and J.S. Parkinson. 1991. Genetic analysis of the bacterial flagellum *Trends Genet.* 7: 196-200. ([PubMed](#))

Historical aspects

H.E. Huxley. 1965. The mechanism of muscular contraction *Sci. Am.* 213: (6) 18-27. ([PubMed](#))

K.E. Summers and I.R. Gibbons. 1971. ATP-induced sliding of tubules in trypsin-treated flagella of sea-urchin sperm *Proc. Natl. Acad. Sci. USA* 68: 3092-3096. ([PubMed](#))

R.M. Macnab and D.E. Koshland Jr. 1972. The gradient-sensing mechanism in bacterial chemotaxis *Proc. Natl. Acad. Sci. USA* 69: 2509-2512. ([PubMed](#))

E.W. Taylor. 2001. 1999 E. B. Wilson lecture: The cell as molecular machine *Mol. Biol. Cell* 12: 251-254. ([PubMed](#)) ([Full Text in PMC](#))

Appendix A: Physical Constants and Conversion of Units