Chapter 15—
DNA II:
Repair, Synthesis, and Recombination

Stelios Aktipis

15.1— Overview

Although the processes of DNA repair, DNA synthesis, and DNA recombination are presented in a somewhat independent and self-contained manner, in reality DNA repair, synthesis, and recombination are intimately connected and interdependent. Furthermore, these DNA-directed processes are also closely associated with other DNA-dependent operations and more specifically DNA transcription reviewed in Chapter 16. Some of these interconnections are indicated in this chapter. The first area to be examined is the enzymatic repair of randomly induced changes in the chemical structure of the DNA bases. A review of the processes of DNA synthesis and DNA recombination completes the chapter.

Both the repair of DNA and particularly the replication of DNA are very complex processes. Although key similarities in the mechanisms of DNA repair are discernible among different organisms, a considerable amount of diversity exists in terms of individual detail. The same is true regarding the process of DNA synthesis. This diversity defeats any attempt to present a simplified and universally applicable model of these processes. To resolve this difficulty the basic elements of the substeps of each process are first described and subsequently integrated for prokaryotes, using as an example the *Escherichia coli* replication system. Eukaryotic replication is treated separately and its similarities and differences with prokaryotic replication are highlighted.

15.2— Formation of the Phosphodiester Bond *in Vivo*

DNA-Dependent DNA Polymerases of E. coli

An apparent common denominator between the processes of DNA replication and repair is the enzymatically catalyzed synthesis of DNA polynucleotide segments, which can be assembled with preexisting polynucleotides, leading to repair or replication. Synthesis of these polynucleotide segments is catalyzed by a family of enzymes, **DNA-dependent DNA polymerases.** In the case of *E. coli,* DNA polymerase has been isolated in three distinct forms, polymerases I, II, and III as listed in Table 15.1. All DNA polymerases have a 3 5 exonuclease activity in addition to the synthetic activities. Polymerase I also has a 5 3 exonuclease activity. Generally speaking, polymerase III is involved in DNA synthesis and polymerase I is involved in both synthesis and repair. Polymerase II is also involved in DNA repair but its function is highly specialized.

TABLE 15.1 Properties of DNA Polymerases I, II, and III of *E. coli*

	Pol I	*Pol II*	*Pol III (core)*
Function			
Polymerization: 5 3	Yes	Yes	Yes
Exonuclease: 3 5	Yes	Yes	Yes
Exonuclease: 5 3	Yes	No	No
Size (kDa)	103	90	(167, 130, 27.5, 10) [a]
Molecules per cell	400	—	10–20
Turnover number[b]	600	30	9000
Structural genes	*polA*	*polB*	*polC*[c]

Source: Adapted from Kornberg, A., and Baker, T. A. *DNA Replication,* 2nd ed. New York: Freeman, 1992.

[a] Sizes of the α, , and subunits.

[b] Nucleotides polymerized at 37°C/min/molecule of enzyme.

[c] Also known as dnaE, the gene for the large (α) subunit.

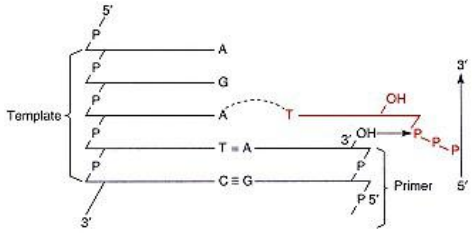


Figure 15.1
Synthetic activity of DNA polymerase.
DNA polymerase catalyzes polymerization of nucleotides in the 5 3 direction. A phosphodiester bond is formed between a free 3 -hydroxyl group of the strand undergoing elongation (the primer) and an incoming deoxyribonucleoside 5 -triphosphate. Pyrophosphate is eliminated.
Redrawn based on figure in Kornberg A. *Science* 163:1410, 1969.

Synthetic Activity

Figure 15.1 shows two complementary DNA strands of unequal length in which the shorter strand has a free 3 terminus. DNA polymerase catalyzes addition of 5 -deoxynucleoside triphosphates to the 3 terminus of the short strand, called the **primer.** The term primer applies to the terminus of a molecule, in this instance the 3 -polynucleotide end, onto which additional monomeric units can be added. The free portion of the longer complementary strand is the **template** that directs the condensation of selected 5 -deoxynucleotides onto the growing primer. The template is a single strand of nucleic acid providing the specific information necessary for the synthesis of a complementary strand. DNA polymerase requires both a primer and a template in order to function.

As seen from Table 15.2, polymerase III catalyzes the elongation of a primer with a much higher degree of efficiency than polymerase I. The enhanced catalytic efficiency of polymerase III is partially attributable to the higher **processivity** of this enzyme. After a polymerase has added a nucleotide residue on the 3 -OH terminus of the primer, it may dissociate from the primer and bind at random to another partially completed polynucleotide chain, or it may remain bound to the original template until many subsequent residues are added to it. Enzymes that tend to remain bound to their substrates through many rounds of polymerization are said to be processive. Polymerase I is less processive in that it tends to dissociate from the template after incorporating only a few nucleotides. Although processivity *per se* does not determine the catalytic rate, it is apparent that an enzyme with high catalytic activity, such as polymerase III, can achieve its optimal catalytic rate only if it is also highly processive.

DNA polymerases permit selection of 5 -deoxyribonucleoside triphosphates, one at a time, with a base complementary to that present in the corre-

TABLE 15.2 Major Subunits and Subassemblies of DNA Polymerase III

Subunit	*Mass (kDa)*	*Gene*			*Function*
α	130[a] 27.5[a]	*dnaE* *dnaQ (mutD)*	}	Pol III (core)	Polymerase 3 5 exonuclease
	10				, assembly?
γ	71[a] 47.5[a] 35 33 15 12	dnaX *dnaX* *holA* *holB* *holC* *holD*	}	γ Complex	Assembly of holoenzyme on DNA Part of the γ complex (Enhances processivity; assists in replisome assembly)
	40.6[a]	*dnaN*			Sliding clamp, processivity

Source: Adapted from Kornberg, A., and Baker, T. A. *DNA Replication*, 2nd ed. New York: Freeman, 1992.

[a] Subunits γ, , , , and form the so-called γ complex responsible for adding β subunits to DNA.

sponding position of the template. The specificity of the polymerase reaction with respect to the template is vested in the strong association of each of the bases of the template with their normal complementary partners present in the cell as free 5 -deoxyribonucleotides. Strong binding between complementary bases is apparently achieved because the bases become confined within custom-fitted cages created by appropriate hydrophobic regions of the DNA polymerase. As a result, the reading of the template is accurate but not completely free of error. Ionized forms of the bases apparently promote **mispairing** during DNA synthesis. As an example, 5-bromodeoxyuracil pairs with guanine when present in an ionized form, as shown in Figure 15.2, instead of its normal partner, adenine. In this instance, the hydroxyl group at C-4 upon loss of a proton acquires a negative charge and changes the hydrogen-bonding properties of 5-bromouracil. Similarly, 2-aminopurine, which normally pairs with thymine in its ionized form, may mispair with cytosine. The natural bases can also undergo ionizations, giving rise to a number of alternative base pairing schemes that produce atypical base pairs leading to misincorporation of bases.

Proofreading Activity

The presence of ionized bases accounts for the incorporation into DNA of inappropriate bases at a ratio of about 1 per 10^4 to 10^5 nucleotide incorporations. Yet, the experimentally measured misincorporation of nucleotides is lower and it does not exceed an error rate of 10^{-8}. The discrepancy is accounted for by the existence of a **"proofreading"** mechanism that allows removal, by the polymerase, of erroneously introduced nucleotides. The removal is carried out by the 3 5 exonuclease activity that characterizes almost every known polymerase, suggesting that proofreading is essential for accurate DNA synthesis. Because of this activity, polymerases can temporarily reverse their synthetic activities and function as exonucleases. The proofreading activity is triggered when a mismatch between the template base sequence and a newly introduced nucleotide at the 3 -OH terminus of the primer occurs. However, some polymerases ensure that a very large percentage of mismatched bases are removed by inadvertently removing a substantial percentage of correctly introduced bases as well. Overall, proofreading fails to remove less than 1 in 10^3 improperly incorporated nucleotides.

Structure of Polymerases

Recall that **polymerase I** has three distinct enzyme activities, namely, a 5 3 synthetic activity and 3 5 and 5 3 exonuclease activities. Chemical and mutation studies of the enzyme have shown that these activities originate from three distinct active sites on the enzyme. Cleavage of polymerase I by the protease subtilisin leads to the formation of a small fragment (30-kDa mass) with 5 3 activity and a larger fragment (70-kDa mass), known as the **Klenow fragment,** having the synthetic activity (5 3 polymerization) and 3 5 exonuclease activity, which is required for **proofreading** during DNA synthesis. X-ray diffraction studies, on cocrystals of DNA and polymerase I, suggest that DNA makes a sharp bend between the 3 5 exonuclease site and the synthetic

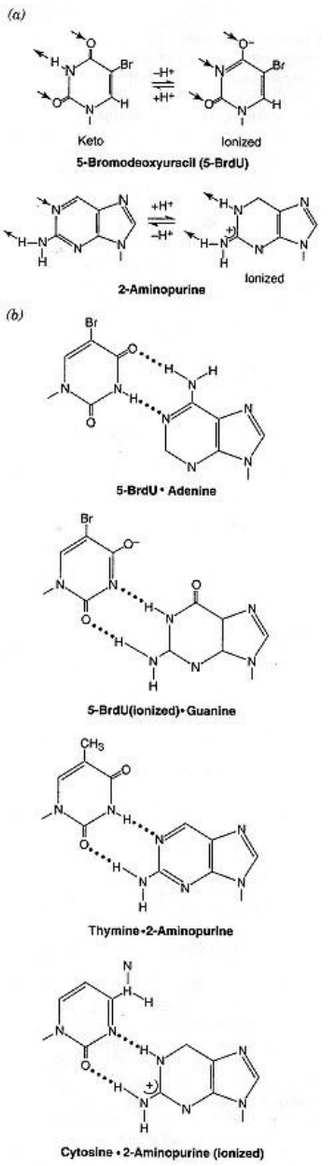


Figure 15.2
DNA base pairing of ionized forms of bases.
Ionization of 5 -bromodeoxyuracil (BrdU), a base analog of T, results in dissociation of a proton from the N-3 position of the pyrimidine ring whereas ionization of 2-aminopurine (2-AP), which is a base analog of A, involves dissociation of a proton from the N-1 position of the purine ring. Normal forms of these bases are in equilibrium with small amounts of the ionized forms. The ionized form of BrdU mispairs with G instead of the normal partner of T, which is A, and ionized 2-AP mispairs with C instead of the normal partner of A, which is T.

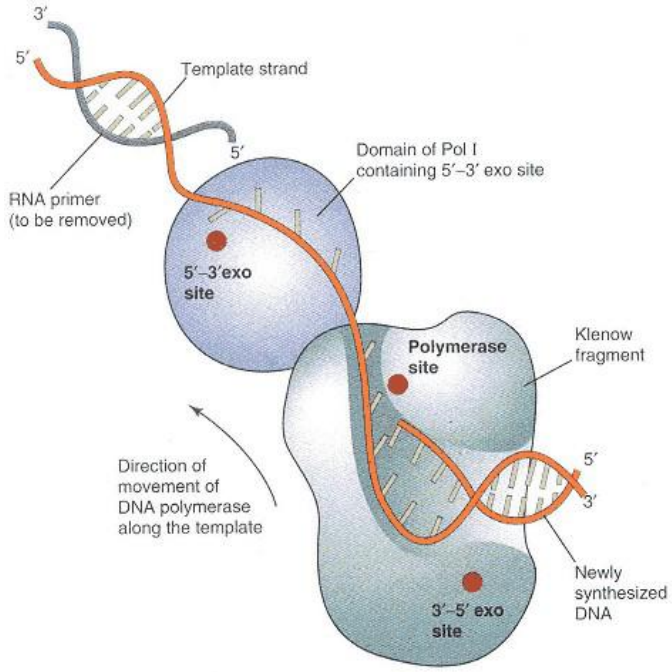


Figure 15.3
Model for the structure of DNA polymerase I-DNA complex.
Klenow fragment of DNA polymerase I includes the 5 3 polymerization site and the 3 5 proofreading site. The remaining segment of the enzyme contains the 5 3 exonuclease site, which is used for DNA repair and the removal of RNA primers from Okazaki segments. In this drawing the 3 growing end of a polynucleotide chain is in contact with the active site in the Klenow fragment, which is involved in elongation of the chain. The 3 end is shifted near the 3 5 exonuclease active site, probably by sliding of the enzyme along the DNA without dissociation from the template.
Adapted from Bease, L. S., Derbyshire, V., and Steitz, T. A. *Science* 260:352, 1993.

site located 3.5 nm away (Figure 15.3). When the polymerase active site detects a mismatch, the 3 terminus of the DNA primer is guided into the 3 5 exonuclease site for removal of the mismatched base and then guided back to the polymerization site for further elongation.

Polymerase III has the same 5 3 synthetic and 3 5 exonuclease activities as polymerase I except that the processivity and polymerase activity of the former are much higher than the corresponding properties of the latter. Polymerase III is a more complex enzyme than polymerase I, consisting of at least ten different protein subunits (Table 15.2). The catalytic core of the enzyme consists of subunits α, , and and has a composite mass of about 167 kDa. Polymerization activity is vested in subunit α and 3 5 exonuclease activity in subunit . The function of the subunit is not clear but it may contribute to the interaction between α and or α with other subunits of the polymerase. The γ subunit participates in initiation of DNA synthesis. Subunits , , , , and , appear to support the processivity properties of the enzyme. Formation of a complex of γ, , , , and during initiation of DNA synthesis catalyzes ATP-dependent transfer of a pair of β subunits to the DNA template. These two β subunits form a clamp around the template that allows the multisubunit assembly to slide along the DNA without dissociation from the template. The subsequent binding of the catalytic core to the clamp of the β subunits generates a molecule of template-bound polymerase III holoenzyme that is a fully functional assembly (Figure 15.4). This sliding clamp is responsible for the remarkable degree of processivity exhibited by DNA polymerase III.

Eukaryotic DNA Polymerases

Less is known about **eukaryotic DNA polymerases**, relative to the *E. coli* polymerase. Five main types of polymerases have been isolated from mammalian cells (Table 15.3). With the exception of polymerase γ, which occurs in mitochondria, the remaining polymerases are involved in chromosomal DNA synthesis and repair. As with the three polymerases of *E. coli*, all five eukaryotic polymerases are characterized by 5 3 synthetic activities, but unlike the prokaryotic polymerases not all eukaryotic polymerases are vested with 3 5 exonuclease (proofreading) activities. Among eukaryotic polymerases only polymerase , which is primarily a repair enzyme like its counterpart in *E. coli*,

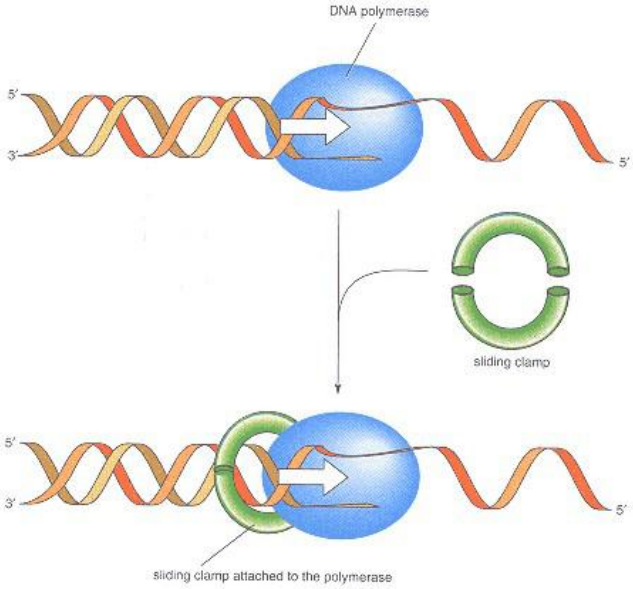


Figure 15.4
Model for the "clamp" that holds DNA polymerase III on the template.
The high processivity of DNA polymerase III is attributed to the formation of a sliding "clamp" that prevents the enzyme from dissociating from the template until DNA replication is completed. The sliding clamp is formed by the association of two β subunits of the polymerase that produces a donut-like structure having a hole with a diameter of about 3.5 nm. This hole easily accommodates B-DNA that has a diameter of no more than 2.5 nm. Upon completion of the synthesis, the two halves of the clamp dissociate and DNA polymerase is freed.

polymerase I, is vested with all three activities, namely, 5 3 synthetic, 3 5 exonuclease, and 5 3 exonuclease activities. Polymerase β is also a repair enzyme but, since it lacks 3 5 exonucleolytic activity necessary for proofreading, its fidelity is low. In analogy with polymerase III of *E. coli*, polymerases α and are the primary synthetic enzymes in eukaryotes and work in close association with each other. Of these two enzymes only polymerase α has a 3 5 exonuclease activity that is necessary for the proofreading function. It is not clear whether polymerase in fact lacks 3 5 activity or whether for some reason it is difficult to detect this activity *in vitro*. DNA polymerase is associated with a 37-kDa subunit, the **proliferating cell nuclear antigen (PCNA) protein**, that shows homology to the β subunit of polymerase III responsible for the high processivity of polymerase III. PCNA

TABLE 15.3 Biochemical Properties of Eukaryotic DNA Polymerases[a]

Property	**α**	**Δ**	**ε**	**β**	**γ**
Mass (kDa)					
Nativex	> 250	170	256	36–38	160–300
Catalytic core	165–180	125	215	36–38	125
Other subunits	70, 50, 60	48	55	None	35, 47
Activities					
3 5 Exonuclease	No	Yes	Yes	No	Yes
Processivity	Low	High	High	Low	High
Fidelity	High	High	High	Low	High

Source: Adapted from Kornberg, A., and Baker, T. A. *DNA Replication*, 2nd ed. New York: Freeman, 1992.

[a] With the exception of polymerase γ, which is a mitochondrial enzyme, all other polymerases are located in the cell nucleus.

endows polymerase with very high processivity and is also involved in eukaryotic DNA excision repair (see p. 636).

In an overall sense DNA polymerases operate at a high level of fidelity, which is required of their function as DNA replicating and repair enzymes. *Escherichia coli* polymerases have an overall error rate in base incorporation of 10^{-7} to 10^{-8}. The experimentally observed accuracy for DNA replication in *E. coli*, however, is substantially higher, with errors made at the rate of only one for every 10^9 to 10^{10} nucleotides incorporated. The discrepancy in these numbers is accounted for by the operation of a DNA repair system that removes mismatched bases that have escaped the scrutiny of the proofreading activity of the polymerases. This repair system, known as the **mismatch repair** system, is examined on page 638.

The necessity to maintain high fidelity in replication is probably also the reason why polymerases synthesize polynucleotides only in the 5 3 direction. If polynucleotide chains could be elongated in the 3 5 direction, the hypothetical growing 5 terminus, rather than the incoming nucleotide, would carry a triphosphate that is unsuitable for further elongation by the synthetic activity of the polymerase.

15.3— Mutation and Repair of DNA

Mutations Are Stable Changes in DNA Structure

One of the fundamental requirements for a structure that serves as a permanent depository of genetic information is high stability. Such stability is essential, at least in those parts that code for the genetic information. The structure of the DNA bases, however, is not totally exempt from gradual change. Normally, changes occur infrequently and they affect very few bases. Chemical and irradiation-induced reactions modify the structure of some bases, disrupt phosphodiester bonds, and sever strands. Extensive chemical changes of the bases occur spontaneously. Errors also occur during replication and DNA recombination, leading to incorporation of one or more erroneous bases. In almost every instance, however, a few cycles of DNA replication are required before a modification in the structure of a base can lead to irreversible damage. In effect, DNA polymerases must use the polynucleotide initially damaged as a template for the synthesis of a complementary strand for the initial change to become permanent. As Figure 15.5 suggests, use of the damaged strand as template extends the damage from a change of a single base to a change of a complete base pair and subsequent replication perpetuates the change. Other sources of permanent modifications of DNA include changes resulting from insertion to deletion from a DNA of short or longer nucleotide sequences during the process of DNA recombination (see p. 661). Intercalation of certain planar organic ring structures can also lead to insertion of nucleotides (see p. 631). Finally, deletions may occur as a result of chemical modification of the bases.

Figure 15.5
Mutation perpetuated by replication.
Mutations introduced on a DNA strand, such as the replacement of a cytosine by a uracil resulting from deamination of cytosine, extend to both strands when the damaged strand is used as a template during replication. In the first round of replication uracil selects adenine as omplementary base. In the second round of replication uracil is replaced by thymine. Similar events occur when the other bases are altered.

Irreversible alteration of a few DNA base pairs can cause drastic changes in the organism. These changes, referred to as **mutations,** may be hidden or visible, that is, **phenotypically silent** or **expressed.** Therefore a mutation is defined as a stable change in the DNA structure of a gene, which may be expressed as a phenotypic change in the organism. Mutations may be classified into two categories: **base substitutions** and **frameshift mutations.** Base substitutions include **transitions,** substitutions of one purine–pyrimidine pair by another, and **transversions,** substitutions of a purine–pyrimidine pair by a pyrimidine–purine pair. Frameshift mutations, which are the most radical, are the result of either the insertion of a new base pair or the deletion of a base pair or a block of base pairs from the DNA base sequence of the gene. These changes are illustrated in Figure 15.6.

Chemical Modification of Bases

Irradiation and certain chemical compounds are recognized as among the main mutagens. The incorporation of erroneous bases by DNA polymerase can also lead to mutations. Other mutations occur spontaneously. Bases in DNA are sensitive to the action of numerous chemicals including nitrous acid (HNO_2), hydroxylamine (NH_2OH), and various alkylating agents such as dimethyl sulfate and *N*-methyl-*N*^{8}-nitro-*N*-nitrosoguanidine. Chemical modifications of bases, brought about by these reagents, are shown in Figure 15.7.

Conversion of guanine to xanthine by nitrous acid has no effect on the hydrogen-bonding properties since xanthine, the new base, can pair with cytosine, the normal partner of guanine. However, the conversion of either adenine to hypoxanthine or the change from cytosine to uracil disrupts the normal hydrogen bonding of the double helix, because neither hypoxanthine nor uracil can form complementary pairs with the base present in the initial double helix (Figure 15.8). Subsequent replication of the DNA extends and perpetuates these base changes (Figure 15.5). Alkylating agents may affect the structure of the bases as well as disrupt phosphodiester bonds so as to lead to the fragmentation of the strands. In addition, certain alkylating agents can interact covalently with both strands, creating interstrand bridges.

DNA undergoes spontaneous changes as a result of various physical perturbations, such as thermal fluctuations or reactions with reactive forms of oxygen. Spontaneous **deamination of cytosine** in human DNA occurs at a rate of

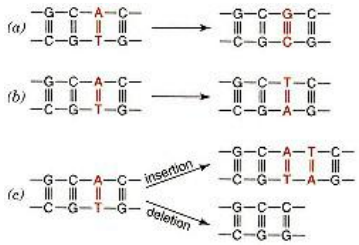


Figure 15.6
Mutations.
Mutations are classified as transition, transversion, and frameshift. Bases undergoing mutation are shown in color.
(*a*) Transition: A purine–pyrimidine base pair is replaced by another. This mutation occurs spontaneously or can be induced chemically by such compounds as 5-bromouracil or nitrous acid.
(*b*) Transversion: A purine–pyrimidine base pair is replaced by a pyrimidine–purine pair. This mutation occurs spontaneously and is common in humans. About one-half of the mutations in hemoglobin are of this type.
(*c*) Frameshift: This mutation results from insertion or deletion of a base pair. Some insertions can be caused by mutagens such as acridines, proflavin, and ethidium bromide. Deletions are often caused by deaminating agents. Alteration of bases by these agents prevents pairing.

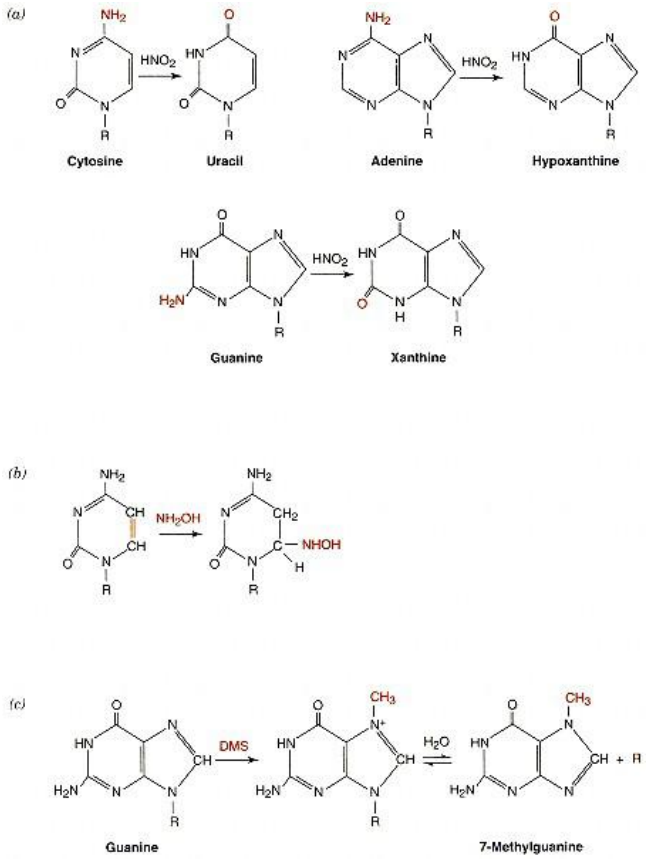


Figure 15.7
Reactions of various mutagens.
(a) Deamination by nitrous acid (HNO_2) converts cytosine to uracil, adenine to hypoxanthine, and guanine to xanthine.
(b) Reaction of bases with hydroxylamine (NH_2OH) as illustrated by the action of this reagent on cytosine.
(c) Alkylations of guanine by dimethyl sulfate (DMS). Formation of a quaternary nitrogen destabilizes the deoxyriboside bond and releases deoxyribose. Among the effective agents for methylation of bases are nitrosoguanidines such as *N*-methyl-N^8-nitro-*N*-nitrosoguanidine.

about 100 base pairs per genome per day and **DNA depurination** occurs at even higher rates of 5000 bases per genome per day (Figure 15.9) as a result of thermal disruption of the *N*-glycosyl bonds of the bases. Some other changes that occur in DNA (as shown in Figure 15.10) can lead to either deletion of one or more base pairs in the daughter DNA after DNA replication or to a base pair substitution.

Radiation Damage

Ultraviolet light, including sunlight, and X-ray irradiation are also effective means of producing mutations. **Radiation energy** absorbed by the DNA induces the formation of minor amounts of the **ionized forms of the bases.** These ionized forms cannot pair with the normal partners of the base, but, instead, they engage in atypical base pairing as shown in Figure 15.11. The presence of ionized base forms at the moment of DNA replication is therefore expected to increase the frequency of mutation in the newly synthesized DNA strands. UV irradiation of DNA causes formation of dimers between adjacent pyrimidine bases. Activation of the ethylene bond of these bases frequently leads to a **photochemical**

dimerization of two adjacent pyrimidines, as shown in Figure 15.12. Thymine residues are particularly susceptible to this reaction, although cytosine dimers and thymine–cytosine combinations are also produced.

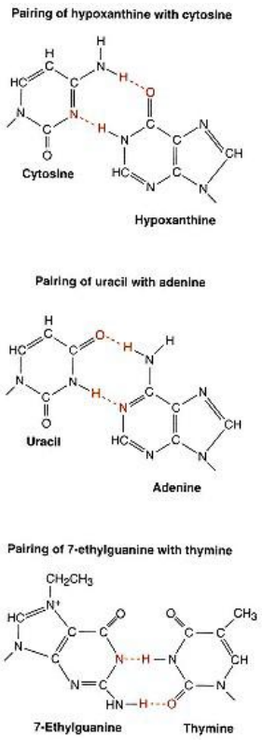


Figure 15.8
Chemical modifications that alter hydrogen-bonding properties of bases.
Hypoxanthine, obtained by deamination of adenine, has different hydrogen-bonding properties from adenine and pairs with cytosine. Similarly, uracil obtained from cytosine has a different hydrogen-bonding specificity than cytosine and pairs with adenine. Alkylation of guanine modifies hydrogen-bonding properties of the base.

High-energy radiation (X-rays or gamma rays) brings about direct modifications in the structure of the bases. Intermediates produced by electron expulsion can be rearranged, leading to the opening of the heterocyclic rings of the bases and the disruption of phosphodiester bonds. In the presence of oxygen additional reactions take place, yielding a variety of oxidation products.

DNA Polymerase Errors

With the appropriate deoxyribonucleoside triphosphates, DNA polymerases function with a high degree of fidelity. Some mutations do occur during DNA replication, but these changes are limited by the high synthetic fidelity of DNA polymerase and the "proofreading" exonuclease properties of this enzyme. The fidelity of DNA replication is further enhanced postreplicatively by an excision repair process known as the **mismatched repair system.** This system recognizes and corrects mismatches in newly replicated DNA by detecting distortions on the outside of the helix that are produced from poor fit between paired noncomplementary bases. Clearly, accurate correction of mismatched bases requires that the mismatched repair system discriminate between preexisting

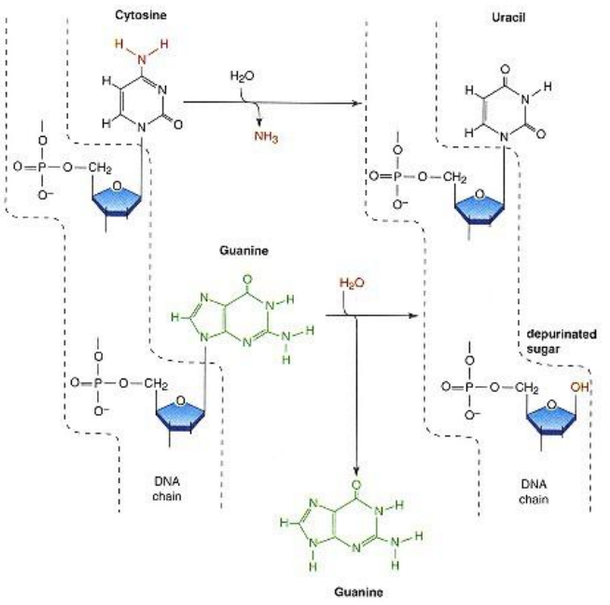


Figure 15.9
Spontaneous deamination of pyrimidines and depurination of polynucleotides.
DNA undergoes substantial structural modifications as a result of thermal perturbations that include
(1) extensive hydrolysis of the *N*-glycosyl bonds that connect purines to the deoxyribose residue and
(2) deamination of cytosine residues to uracil. In absence of repair mechanisms, these changes would have disastrous consequences for cell survival because of the high frequency of their occurrence.

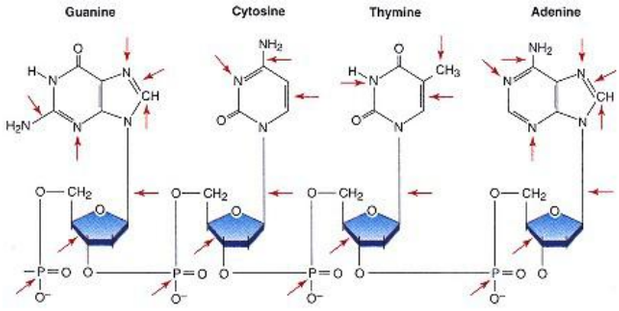


Figure 15.10
DNA sites subject to spontaneous chemical modifications.
Nucleotides are subject to various spontaneous chemical changes at sites indicated by arrows including (1) hydrolytic attack, (2) oxidative damage, and (3) methylation. The frequency and extent of chemical change vary from site to site.

and newly synthesized DNA strands. Such discrimination is feasible because certain adenine residues in DNA, which are part of a recurring GATC sequence, are subject to methylation that occurs posttranscriptionally, but with some delay. Mismatched proofreading is carried out by a multienzyme complex that excises mismatched nucleotides only from newly synthesized strands. The complex identifies these nucleotides by searching for unmethylated adenine residues in the GATC sequences of each strand. The mechanism of mismatched repair is described later.

DNA polymerases are unable to distinguish between the normal deoxyribonucleoside triphosphate substrates and other nucleotides with very similar structures, thus leading to their incorporation and a mutation. Classic examples of such analogs are deoxyribonucleotides of 5-bromouracil (5-BrdU) and 2-aminopurine (2-AP) that have been used experimentally for the introduction of mutations. Incorporation of 5-BrdU into DNA introduces, with a high frequency, a transition mutation in which a pu-py pair is transformed to another pu-py. Specifically, 5-BrdU paired with A is changed to a C-G pair, which amounts to a TA GC transition. The unusual pairing properties of 5-BrdU appear to relate to the higher tendency of this base to be transformed to an ionized form, relative to T for which it is a substitute. This occurs presumably because of the higher electronegative nature of the bromine atom in comparison to the corresponding methyl group in thymine.

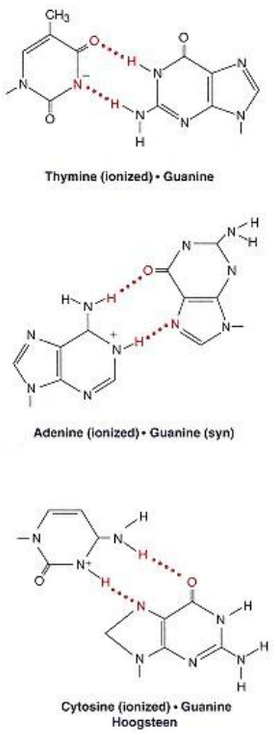


Figure 15.11
Base pairing between the ionized forms of the bases.
Adenine and cytosine are prone to protonation especially at lower pH. Also, an ionized form of thymine can be generated by loss of a proton. Reactions that give rise to ionized forms of bases occur readily at near-neutral pH, within certain nucleotide sequence contexts. Whereas some of the ionized complexes form with Watson–Crick hydrogen bonding, as, for instance, the T (ionized)–G pair, other ionized bases form more unusual types of H bonding. For example, the A (ionized)–G(syn) base pair involves H bonding between an A in the anti position and a G in the syn configuration.

Stretching of the Double Helix

Organic compounds characterized by planar aromatic ring structures of appropriate size and geometry can be inserted between base pairs in double-stranded DNA. This process is referred to as **intercalation.** During intercalation neighboring base pairs in DNA are separated to allow for the insertion of the intercalating ring system, causing an elongation of the double helix by stretching. In effect the double helix is locally unwound into a ladder-like structure in which the base pairs are transiently arranged at 0.68 nm apart. This localized arrangement doubles the 0.34-nm distance characteristic of the double helix and generates sufficient space between base pairs for the insertion of the intercalator. In effect, intercalation disrupts the continuity of the base sequences in DNA and the reading of the DNA template by the DNA polymerase, producing a daughter strand with an additional base incorporated into DNA. The resulting mutation is referred to as a **frameshift.** Acridines, ethidium bromide, and other intercalators are known to be effective frameshift mutagens (Figure 15.13). Clinical Correlation 15.1 discusses mutations and the etiology of cancer.

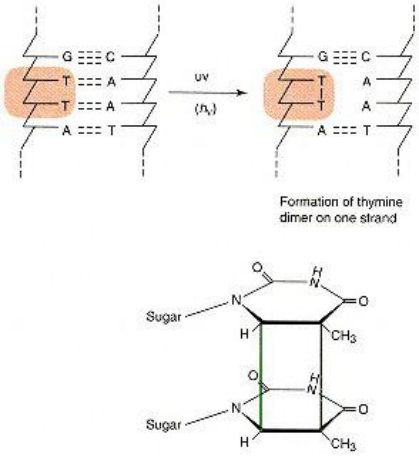


Figure 15.12
Dimerization of adjacent pyrimidines in irradiated DNA.
Thymine activated by absorption of UV light can react with a second neighboring thymine and form a thymine dimer.

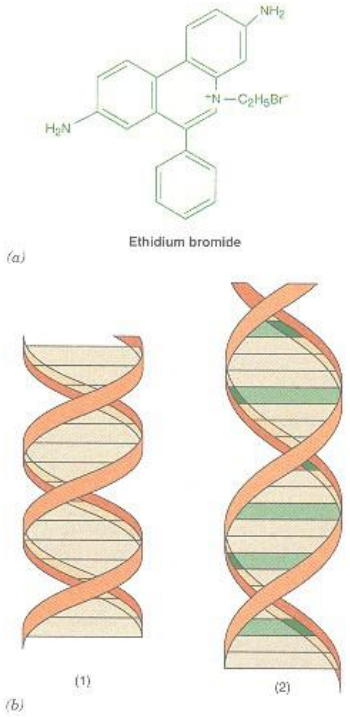


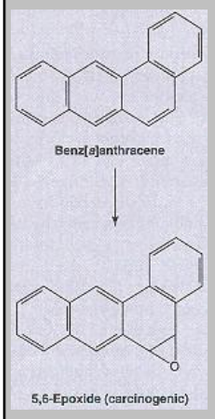
Figure 15.13
Intercalation between base pairs of the double helix.
(a) Insertion of planar ring system of intercalators between two adjacent base pairs requires stretching of the double helix
(*b*). During replication this stretching apparently changes the frame used by DNA polymerase for reading the sequence of nucleotides. Consequently, newly synthesized DNA is frameshifted.
(*b*-1) Original DNA helix;
(*b*-2) helix with intercalative binding of ligands.
Redrawn based on figure in Lippard, S. J. *Acct. Chem. Res*. 11:211, 1978.

CLINICAL CORRELATION 15.1

Mutations and the Etiology of Cancer

Considerable progress in understanding the etiology of cancer has been achieved in recent years by the realization that long-term exposure to certain chemicals leads to various forms of malignancy. It is now suggested that the great majority of cancers are triggered by agents in the environment that modify underlying genetic predisposition factors.

Carcinogenic (cancer-causing) compounds are not only introduced into the environment by the increasing use of new chemicals in industrial applications but are also present in the form of natural products. For instance, the aflatoxins, produced by certain molds, and benz*[a]*anthracene, present in cigarette smoke and charcoal-broiled foods, are carcinogenic. Some carcinogens act directly, while others, such as benz*[a]*anthracene, must undergo prior hydroxylation by arylhydroxylases, present mainly in the liver, before their carcinogenic potential can be expressed.



The reactivity of many carcinogenic compounds toward guanine residues results in modification of the guanine structure, usually by alkylation at the N-7 position and by cleavage of the phosphodiester bond, events that upon replication lead to permanent mutations. Chemical mutagens are generally carcinogenic and vice versa. Vulnerability of DNA to alkylating agents and other chemicals underscores the concerns expressed by many scientists about the ever-increasing exposure of our environment to new chemicals. What is of concern is that the carcinogenic potential of new chemicals released into the environment cannot be predicted with confidence even when they appear to be chemically innocuous toward DNA.

In the past, tests for carcinogenicity, that is, the ability of a substance to cause cancer, required the use of many experimental animals treated with high doses of suspected carcinogen over a long period of time. Such tests, which are time consuming as well as expensive, are the only approach still available for testing carcinogenicity directly. A much simpler and inexpensive indirect test for carcinogenicity is also available. This test, the Ames Test, is based on the premise that carcinogenicity and mutagenicity are essentially manifestations of the same underlying phenomenon—the structural modification of DNA. The test measures the rate of mutation that bacteria undergo when exposed to chemicals suspected to be carcinogens.

A major criticism of this test is that the assumption of an equivalence between mutagenicity and carcinogenicity is not always valid. Because of economic implications of labeling a chemical with widespread use as a potential carcinogen, the scrutiny often exercised in assessing the reliability of applicable tests for labeling a chemical as a carcinogen is understandable. Certain exceptions notwithstanding, the great majority of chemicals tested have shown that a good correlation exists between the tendency of a chemical to produce bacterial mutations and animal cancer. Even the direct and very costly tests for carcinogenicity have not completely escaped criticism. The reliability of such tests has been questioned because of the relatively large doses of chemicals employed, doses that are essential for shortening the long-term chemical exposure of the animals to a practically manageable period of time. Another criticism of direct tests is that they make projections from animals, usually rodents, to humans. This criticism has some merit. During the past few years it has became apparent that rodents are less efficient than humans at repairing certain types of damage in nontranscribed regions of their DNA. Damage in nontranscribed DNA regions is more slowly repaired than damage within transcribed genes, which have first priority for repair. Although damage in nontranscribed DNA regions has few immediate consequences, it appears with time that this damage leads to cancer. The relatively large doses of chemicals used for testing are likely to exceed the capacity of rodent DNA repair systems, making the extrapolation of the results obtained from rodents to humans unreliable.

The enzymes that activate carcinogens are often members of the cytochrome P450 family (Chapter 23) that can be induced by noncarcinogenic compounds such as ethanol; hence alcohol can increase the potential risk of cancer development after exposure to carcinogens.

Ames, B., Dursto, W. E., Yamasaki, E., and Lee, F. D. Carcinogens are mutagens: a simple test system combining liver homogenates for activation and bacteria for detection. *Proc. Natl. Acad. Sci. USA* 70:2281, 1973; and Culotta, E., and Koshland, D. E. Jr. DNA repair works its way to the top. *Science* 266:1926, 1994.

DNA Is Repaired Rather Than Degraded

DNA is the only macromolecule that is repaired rather than degraded. The repair processes are very efficient with fewer than 1 out of 1000 accidental changes resulting in mutations. The rest are corrected through various processes

of DNA repair. Mutation rates can be estimated using two entirely different approaches, that is, from the frequency with which new mutants arise either in populations, such as fruitflies, or in specific proteins in cells growing in tissue culture. These experiments provide estimates of mutation rates of 1 base pair change per 10^9 base pairs for each cell per each generation. On this basis, for an average-sized protein, which contains about 1000 coding base pairs, a mutation may occur once in 10^6 cell generations.

DNA repair is a high-priority process for maintaining cellular function. Germ cells must be protected against high rates of mutation to preserve the species, and somatic mutation must be controlled in order to avoid uncontrolled cell growth and disease. Unchecked accumulation of damage can lead to accumulation of nonfunctional proteins or unregulated growth characteristic of malignant cells. Commonly encountered DNA lesions are listed in Table 15.4.

There are multiple DNA repair pathways and each specializes in a certain type of damage, although some repair pathways have a wider versatility than others. Generally, repair mechanisms are applicable to both prokaryotic and eukaryotic DNA repair.

Repairs may be carried out under rare circumstances as a direct reversal of the damage or, far more commonly, by the replacement of the damaged DNA section. DNA repair depends on the existence of two complementary DNA strands except for **postreplication repair of rare lesions** and **postreplication SOS repair.** Damage or imperfection on one DNA strand can be corrected since the complementary strand provides the necessary information for accurate repairs. Postreplication repair is not a true repair mechanism but rather a stop-gap measure that allows for DNA replication to occur until damage can be repaired permanently. Postreplication repair cannot use the complementary DNA strand for repairs because this strand is also altered by the replication that precedes the repair. Postreplication repair depends, instead, on another process—DNA recombination. Recombination permits the use of homologous DNA strands, namely, DNA strands with the same or almost the same sequence as the damaged strand, for carrying out the repair of the damaged DNA section. An intriguing feature of DNA repair that has been appreciated recently is its apparent intimate coupling to other central processes in which DNA participates, such as recombination, transcription, and control of the cell cycle. Enzymes involved in DNA repair participate in DNA replication, DNA recombination, and particularly DNA transcription. DNA metabolism integrates important processes that are coordinated through the use of the same molecular tools to achieve different tasks.

TABLE 15.4 DNA Lesions that Require Repair

DNA Lesion	*Cause*
Missing base	Acid and heat remove purines (~10^4 purines per day per cell in mammals)
Altered base	Ionizing radiation; alkylating agents
Incorrect base	Spontaneous deaminations: C U, A hypoxanthine
Deletion–insertion	Intercalating agents (e.g., acridine dyes)
Cyclobutyl dimer	UV irradiation
Strand breaks	Ionizing radiation; chemicals (bleomycin)
Cross-linking of strands	Psoralin derivatives (light-activated); mitomycin C (antibiotic)

Source: From Kornberg, A. *DNA Replication*. San Francisco: Freeman, 1980, p. 608.

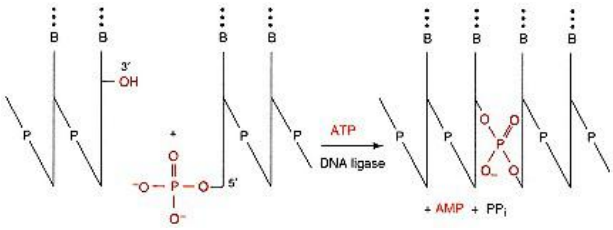


Figure 15.14
Action of DNA ligase.
The enzyme catalyzes the joining of polynucleotide strands that are part of a double- stranded DNA. A single phosphodiester bond is formed between 3 -OH and 5 -P ends of two strands. In *E. coli* cells, energy for formation of the bond is derived from cleavage of the pyrophosphate bond of NAD^+. In eukaryotic cells and bacteriophage-infected cells, energy is provided by hydrolysis of the α,β-pyrophosphate bond of ATP.

Excision Repair in E. coli

Excision repair is catalyzed by different enzymatic systems tailored to specific types of damage. This repair mechanism is universal, occurring in all organisms investigated. The mechanisms are characterized by four sequential steps: incision, excision, resynthesis, and ligation. Incision is the recognition step and is individualized for the specific type of damage present. It is also the rate-controlling step in the process. During excision the damaged DNA section is excised, leaving a gap in the DNA strand. In the resynthesis step the gap is filled by DNA polymerase I. This enzyme functions like DNA polymerase III in that it catalyzes the stepwise addition of nucleotide triphosphates on a 3 -OH generated by the preceding incision step. Polymerase I, however, differs from polymerase III in that it is less processive, tending to dissociate from the DNA after incorporation of 10–12 nucleotides. At this stage the gap is reduced to the size of a single phosphodiester bond. Because of the combined synthetic–nucleolytic action of polymerase I, the nick can move along the strand, undergoing repair until it is finally bridged during the ligation step by the action of DNA ligase (Figure 15.14). The ligation step appears to be very similar for all types of excision repair.

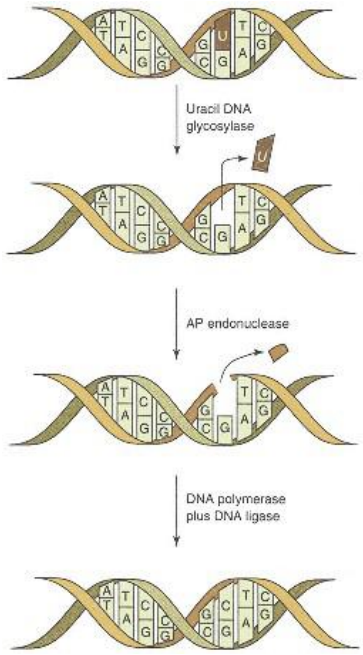


Figure 15.15
Uracil DNA glycosylase repair of DNA.
Uracil DNA glycosylase removes uracil, formed by accidental deamination of cytosine, by cutting the glycosidic bond, leaving DNA with a missing base. AP endonuclease subsequently cuts out the sugar–phosphate remnant. Repair is completed by DNA polymerase and ligase.

Base excision repair eliminates modified bases from DNA. The amino groups of cytosine, adenine, and guanine are susceptible to spontaneous elimination, and various chemicals lead to modifications in the structures of purines, including methylation and ring opening. In addition, ring opening may result from exposure to ionizing radiation. Bases that have been deaminated, methylated, or otherwise chemically modified are hydrolytically removed by enzymes referred to as **DNA glycosylases.** Removal of deaminated cytosine (i.e., uracil) by the enzyme **uracil DNA glycosylase** is illustrated in Figure 15.15. This enzyme removes the damaged cytosine, producing a deoxyribose residue with the base missing **[apurinic–apyrimidinic (AP) site].** AP sites are also generated without the involvement of DNA glycosylases, as in the case of spontaneous hydrolysis of purines **(depurination)** that occurs at very high rates in DNA. AP sites can also result from depyrimidination but the greater stability of the purine–glycoside bond makes this reaction almost insignificant. Once an AP site has been created, the enzyme **AP endonuclease** nicks the phosphodiester backbone at the depurinized site and excises the sugar–phosphate residue. The action of DNA polymerase I and ligase on this structure leads to the restoration of the damaged strand.

A second type of excision repair referred to as **nucleotide excision repair** is activated when DNA is damaged in a way that produces a "bulky" lesion. This occurs when DNA interacts with polycyclic aromatic hydrocarbons, such as benzo[*a*]pyrenes and dialkylbenzathracenes generated by smoking, thymine–psoralene adducts, and guanine–cisplatin adducts formed by chemotherapeutic drugs. UV light-induced dimerization of adjacent pyrimidines also causes bulky lesions. Nucleotide excision repair also corrects other lesions that do not distort the helix, such as the presence of methylated bases. Once the lesion has been located, an endonuclease activity cleaves the modified strand on both sides of

the distortion and the entire lesion is removed (Figure 15.16). Repair is initiated by recognition of the distortion of the DNA by an endonuclease system consisting of the products of three *E. coli* genes *uvrA, uvrB,* and *uvrC*. A tetramer consisting of two **UvrA** and two **UvrB** proteins, which is formed on DNA during a series of preincision steps, "melts" the DNA locally at the expense of ATP and locates the bulky lesion. The complex is subsequently subjected to incision at both sides of the bulky lesion. First, UvrB makes a 3 incision and then **UvrC** makes a 5 incision, leading to the release of an oligonucleotide consisting of 12 or 13 residues that includes the pyrimidine dimer. This nuclease activity, which is unique to DNA repair, has been christened **excision nuclease** or **excinuclease** to clearly distinguish it from other endonucleases. For the remainder of the repair, *E. coli* makes use of the protein **UvrD** which, acting as a **helicase,** unwinds and releases the oligonucleotide that was excised by UvrB and UvrC. The repair is completed by polymerase I and ligase.

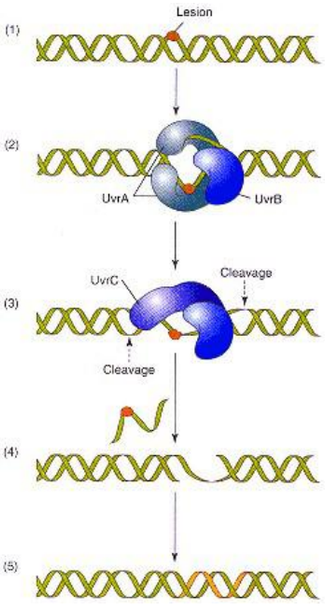


Figure 15.16
Nucleotide excision repair in *E. coli*.
Nucleotide excision repair in *E. coli* and in human DNA occurs in a series of analogous steps. Initial damage in *E. coli* is recognized by UvrA protein, which also serves as a "molecular matchmaker" by recruiting, at the damaged site, UvrB protein. UvrA binds to the lesion, unwinds and kinks DNA. UvrA also causes a conformational change in UvrB that promotes strong binding of UrvB at the site of the lesion. Subsequent dissociation of UvrA from UvrB–DNA complex makes the complex a target for UvrC. UvrB then makes a 3 cut that is followed by a 5 incision made by UvrC. Helicase II (UvrD) releases the excised oligonucleotide 12-mer and DNA polymerase displaces UvrB and fills the excision gap prior to ligation.
Redrawn based on figure in Moran, L. A., Scrimgeour, K. G., Horton, H. R., Ochs, R. S., and Rawn, J. D. *Biochemistry*. Englewood Cliffs, NJ: Neil Patterson/Prentice Hall, 1994.

Eukaryotic Excision Repair

Excision repair in prokaryotes and eukaryotes is remarkably similar with the following distinctions. The exonuclease activity of human cells consists of a much larger number of proteins (16–17 different polypeptides) as apposed to the four proteins (UvrA, B, C, and D) that constitute the exonuclease activity of *E. coli*. Some of the protein constituents of human excinucleases are listed in Table 15.5. Proteins XPA to XPG have been identified as seven different **genetic complementation groups** (A to G) of patients with **xeroderma pigmentosum (XP),** a condition characterized by UV sensitivity and corresponding deficiencies in DNA repair. The human nucleotide repair genes are therefore referred to by an XP or ERCC (excision repair component) designation. Nucleotide excision repair of human DNA begins with the binding of XPA to a dimer between XPF and ERCC1 (Figure 15.17). XPA recognizes and binds to the damaged site along with the replication protein HSSB. An intriguing aspect of human DNA repair is involvement of an additional enzymic complex con-

TABLE 15.5 Excinuclease Activity of Human DNA

Human Gene	*Protein Function*
XPA	Damage recognition protein (binds to damaged DNA)
XPB (*ERCC3*)	DNA helicase activity; subunit of transcription factor TFIIH
XPC	Interacts with general transcription factor TFIIH
XPD (*ERCC2*)	DNA helicase activity; subunit of transcription factor TFIIH
XPF	Nuclease activity
XPG	Nuclease activity
ERCC1	Part of nuclease activity (binds to XPF and to replication protein RPA)
HSSB (*RPA*)	Binds to the XPF–ERCC1 complex and together with XPA binds to the lesion site

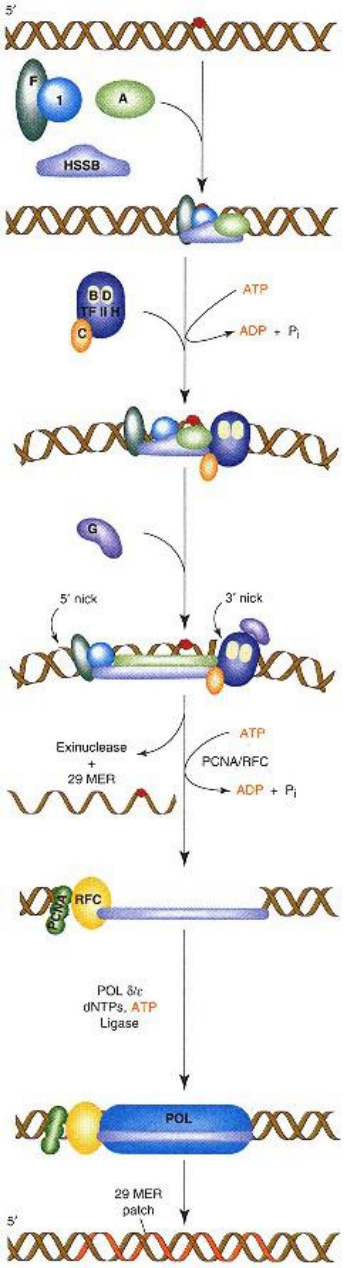


Figure 15.17
Nucleotide excision repair of human DNA.
In human DNA damage is recognized by the XPA factor (abbreviated in the figure as A) that recruits to the damaged site factors XPF and ERCC1 (abbreviated as F and 1, respectively) in the form of a dimer. XPF is an excinuclease that is recruited to the damaged site early on just as UvrB is recruited in the *E. coli* system. The replication protein (HSSB) binds to XPA and the lesion site. XPA also recruits to the damaged site the general transcription factor TFIIH, which, as it turns out, is also a repair protein since two of its protein subunits are repair factors XPB and XPD (abbreviated as B and D). In analogy with UvrA, TFIIH may be involved in kinking and unwinding of DNA at the damaged site and in recruiting XPC and XPG proteins, which are vested with helicase activity. Excinuclease cuts are made at the 3 site by XPG, whereas XPF nicks at the 5 site of the lesion, leading to the excision of a 23-mer oligonucleotide. Gap repair is carried out by polymerases and with PCNA and replication protein RFC, followed by ligation.
Redrawn based on figure in Sancar, A. *Science* 266: 1954, 1994. Copyright © 1994 American Association for the Advancement of Science.

sisting of eight different protein subunits and known as the **general transcription factor TFIIH.** This factor is essential for transcription initiation and for nucleotide excision repair. In fact, two of the eight subunits of TFIIH are the helicases XPB and XPD that evidently not only act in excision repair but also catalyze the opening of DNA to initiate transcription. This intimate involvement of a transcription factor suggests that DNA repair and transcription are not fully separable processes and may be coupled to each other. The TFIIH factor interacts with XPC and the entire complex is recruited to the damaged site by XPA, where it is joined by the endonuclease XPG. The two recruited endonucleases, XPF and XPG, complete the excinuclease systems with the XPG making the 3 nick and the XPF, in the form of a complex with ERCC1, making the 5 nick. The major XPG incision is made at the third phosphodiester bond 3 to the lesion, whereas the XPF–ERCC1 complex incises primarily at the 25th phosphodiester bond 5 to the lesion. The role of TFIIH is presumably to unwind the double helix at the damaged site so as to enable the endonucleases XPF and XPG to activate the excinuclease system. A protein associated with polymerase , PCNA **(proliferating cell nuclear antigen),** releases the excinuclease subunits and the excised oligomer, which is larger than the oligonucleotide released during *E. coli* repair (27–29 nucleotides versus 12–13 nucleotides in *E. coli*). The gap is filled by polymerases and and the DNA is ligated.

Excision repair also removes cross-links between complementary DNA strands, such as those introduced by the mustards and drugs used in cancer therapy (i.e., mitomycin D and platinum complexes). Error-free repair is not possible if the cross-link extends across directly opposing bases. Clinical Correlations 15.2 and 15.3 discuss defects in DNA repair that are associated with human disease; Clin. Corr. 15.4 examines the role of DNA repair in chemotherapy.

Mismatch Repair

Mismatch repair in both prokaryotic and eukaryotic cells deals with errors created during DNA replication. In effect, three serially operating mechanisms—base selection, exonucleolytic proofreading, and postreplicative mismatch re-

CLINICAL CORRELATION 15.2

Defects in Nucleotide Excision Repair and Hereditary Diseases

Defects in nucleotide excision repair are implicated in at least three rare hereditary disorders, xeroderma pigmentosum (XP), Cockayne's syndrome (CS), and trichothiodystrophy (TTD). XP patients exhibit sunlight-induced photodermatoses characterized by severe skin reactions that range initially from excessive freckling and skin ulcerations to the eventual development of skin cancers. Some forms are also accompanied by neurological abnormalities. The symptoms exhibited by CS and TTD patients are associated instead only with developmental abnormalities. CS syndrome is characterized by growth and mental retardation, neurological deficiencies, and photosensitivity but not an increased rate of cancer or skeletal abnormalities. TDD patients, on the other hand, have scaly skin, brittle hair, short stature, and neuroskeletal abnormalities.

Xeroderma pigmentosum is a group of closely related abnormalities in excision repair. About 80% of XP patients fall into one of seven complementation groups (different syndromes). Each group carries a mutation in a different gene and is characterized by varying levels of UV sensitivity caused by corresponding deficiencies in "excinuclease" repair activity. The remainder fall in the XPV (V for variant) group. In this variant UV irradiation produces different types of mutations compared to normal cells. During normal DNA synthesis, whenever the DNA polymerase bypasses a pyrimidine dimer in the template that has not yet been repaired, a purine (most often A) is incorporated into nascent DNA but this preference is not maintained by XPV cells. It appears that the mechanism of bypass by the DNA polymerase in XPV cells is altered possibly because of changes in one or more of the subunits of the polymerase or possibly some other protein factor that assists the polymerase to bypass the DNA lesions. The neurological abnormalities that frequently accompany XP appear to result from both abnormal gene expression and DNA deterioration caused by the accumulation of unrepaired DNA damage.

Cockayne's syndrome is associated with mutations in the *CSB/ERCC6, XPD,* and *XPB* genes. Trichothiodystrophy is caused by mutations in *XPB, XPD,* and *XPG* genes and perhaps in additional subunits of TFIIH or TFIIH-associated excision repair subunits. Obviously, different mutations in the *XPB* and *XPD* genes are responsible for each syndrome.

Tanaka, K., and Wood, R. D. Xeroderma pigmentosum and nucleotide excision repair of DNA. *TIBS* 9:83, 1994.

CLINICAL CORRELATION 15.3

DNA Ligase Activity and Bloom Syndrome

Bloom syndrome is a rare genetic disease that is characterized by chromosomal instability. Other chromosome breakage syndromes include Fanconi's anemia (FA), ataxia telangiestasia (AT), Werner's syndrome (WS), and Gardner's syndrome (GS). Deficiencies in the effective repair of DNA lesions, which can probably be attributed to defective DNA ligation, are presumably responsible for many of these syndromes. These repair deficiencies appear to increase the tendency to develop malignancies among those affected with the syndromes.

Bloom syndrome is a prototype of somatic mutation disease. The clinical features of Bloom syndrome are small body size, a sun-sensitive skin with well-defined hyper- and hypopigmented skin lesions, and increased sensitivity to bacterial infections due to immunodeficiency. Cancer, chronic lung disease, and diabetes are common complications. Cells from Bloom syndrome patients have high rates of mutation, and the excessive number of accumulated somatic mutations are responsible for many of the clinical features of this syndrome. In patients suffering from Bloom syndrome, hypermutability is responsible for the abolition of ligase I activity needed for completing DNA repair and (perhaps) DNA recombination.

German, J. *Bloom syndrome. Dermatol. Clin.* 13(1):7, 1995.

CLINICAL CORRELATION 15.4

DNA Repair and Chemotherapy

Many anticancer drugs cause DNA damage. For example, cisplatin, used for treatment of several forms of cancer and particularly effective against testicular tumors, forms two intrastrand adducts with DNA. The major one, the 1,2-intrastrand d(GpG) cross-link, is repaired by excision repair. DNA adducts are believed to be the primary cytotoxic lesion and cells deficient in excision repair are very sensitive to this drug. The high mobility group (HMG)-domain proteins "shield" and specifically inhibit DNA repair of this major cisplatin–DNA adduct, thus increasing the cytotoxicity of cisplatin. The types and levels of HMG-domain proteins in a given tumor may influence the responsiveness of that cancer to cisplatin chemotherapy. This information may provide a basis for the development of new platinum anticancer drugs that may have greater therapeutic potential.

Huang, J. C., Zamble, D. B., Reardon, J. T., Lippard, S. J., and Sancar, A. HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease. *Proc. Natl. Acad. Sci. USA* 91:10394, 1994.

pair-participate in ensuring fidelity of replication. The mismatch repair system recognizes and eliminates mispairing from newly synthesized DNA strands, improving the fidelity of the synthesis. Base selection and proofreading act more effectively against transversion than transitions, whereas mismatch repair does the opposite. DNA replication errors are difficult to recognize because mismatches consist of erroneous but unaltered base structures. The repair system relies on other signals within the helix to identify the newly synthesized strand, which by definition harbors the replication error. Such signals are provided in *E. coli* by a methylation reaction catalyzed by **Dam methylase** that modifies GATC sequences by introducing a methyl group at the *N*-6 position of adenines. Shortly after replication these GATC sequences exist in an unmethylated state that betrays the newly synthesized nature of the DNA strand and permits strand discrimination by the mismatch repair system (Figure 15.18).

The mismatch repair system in *E. coli* includes several different protein components, which repair mismatches in the vicinity of a GATC sequence according to complementary rules dictated by the base sequence of the methylated (i.e., preexisting) parental strand. Proteins that catalyze the process of mismatch repair have been named **MutS, MutH,** and **MutL.** Repair is initiated by binding of MutS to the mismatch followed by the addition of MutL. Formation of the MutS–MutL complex activates a latent GATC endonuclease activity, vested in the MutH protein, that nicks the unmodified strand at a hemimethylated GATC site. The strand break, which can occur on either side of the mismatch, will take place as long as the mismatched base is located within the general vicinity of the GATC site, which means within a few hundred base pairs from the GATC sequence. This nick marks the strand that will be excised. When the mismatch is located on the 5 side of the cleavage site the unmethylated strand is unwound, degraded, and replaced by new DNA synthesized in the 3 5 direction until the mismatch is reached and excised. This reaction requires a DNA helicase II, referred to also as the **MutU** protein, a 3 5 exonuclease (exonuclease I), DNA polymerase III, and finally DNA ligase to seal the repaired strand. If the mismatch is located on the 3 site of the cleavage, a series of completely analogous steps takes place, except that a 5 3 exonuclease **(RecJ)** replaces exonuclease I (an exonuclease with both 5 3 and 3 5 activity, exonuclease III can also substitute for RecJ in the latter repair). This unusual bidirectional excision activity of the mismatch repair system suggests that this system "keeps track" of the side on which the mispair of the GATC sequence signal is located.

Analogous mismatch repair systems have been identified in eukaryotes. Both yeast and human cells code for proteins homologous to the bacterial

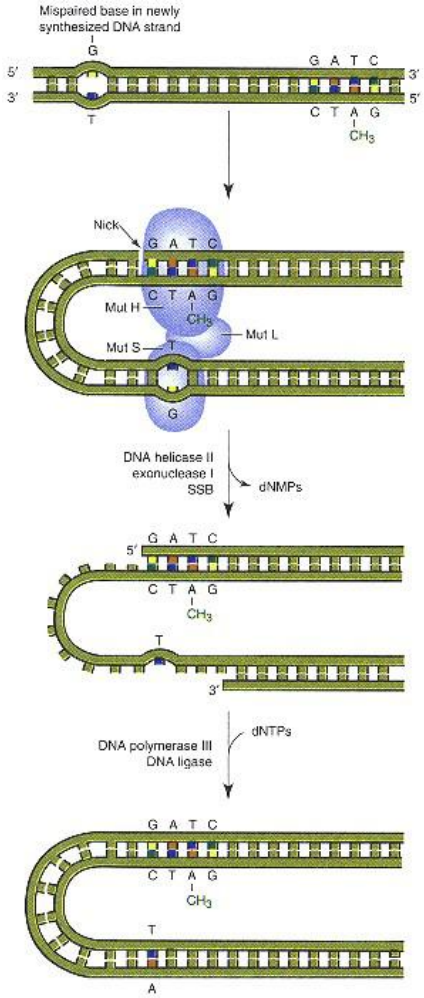


Figure 15.18
Mismatch DNA repair.
Methylation of adenine in palindromic 5 -GATC sequences serves to distinguish parental strands from newly synthesized strands that are methylated only after some delay. Methylation directs the mismatch repair system to repair mispaired bases. Methylated GATC sequences are recognized by MutH, which is also an endonuclease that cleaves the unmethylated strand on the 5 site of the G in the GATC sequence, whereas the mispaired site is recognized and bound by the MutS protein. MutL, which is a molecular matchmaker, links MutH and MutS together. The segment of the unmethylated strand, which represents newly synthesized DNA between the site cleaved by MutH and a point just past the mismatched base, is then removed by the action of helicase II, exonuclease I, and SSB protein. The gap is repaired by DNA polymerase III and ligase. A similar mechanism, but based on the presence of nicks to identify newly synthesized strands, is used by eukaryotes. The eukaryotic mismatch repair system does not use MutH and depends on MutL for the degradation of newly synthesized strands that contain base mismatches.

proteins MutS and MutL but lack the MutH protein. In eukaryotic mismatch repair the role of MutL is to scan nearby DNA for the presence of nicks. Upon finding a nick, MutL degrades the nicked strand starting at the nick site and extending just past the site of the mismatched base pair. Replication errors are thereby selectively removed. Clinical Correlation 15.5 describes the role of mismatch repair in the development of certain types of cancer.

Mechanisms That Reverse Damage

Formation of dimers can be directly reversed by the action of light. Photoreversal is catalyzed by deoxyribodipyrimidine **photolyase,** which disrupts the covalent

CLINICAL CORRELATION 15.5

Mismatch DNA Repair and Cancer

DNA is constantly being damaged. In the absence of efficient repair, this may be the cause of as much as 90% of all human cancers. The importance of defective mismatch repair in the development of certain types of human cancer has been demonstrated recently. Tumors associated with hereditary nonpolyposis colorectal cancer (HNPCC), which causes cancer predisposition and certain sporadic cancers, have been found to be prone to mutation by as much as two orders of magnitude higher than normal human cells. These high mutation rates have been found to be consistently associated with deficiencies in mismatch repair.

That loss of mismatch repair fidelity is a central step in the development of HNPCC tumors has been concluded from the finding that the majority of these tumors are attributable to defects at any one of four different human genome loci. These are the *hMSH2* gene, which codes for a protein homolog of bacterial MutS protein, and the *hMLH1*, *hPMS1*, and *hPMS2* genes, which specify three similar but distinct MutL analogs. These findings demonstrate that the primary event in the development of HNPCC tumors is the loss of critical mismatch repair activity. Inefficiencies in DNA repair presumably lead to mutations that circumvent the regulatory systems controlling cell proliferation. The link between mismatch repair and the development of colon cancer provides support for the hypothesis that cancers are initiated when cells accumulate a certain mutation load. A current emphasis in studies of cancer is the search for and study of particular genes, the mutations of which appear to lead to cancer. The new findings, which demonstrate the importance of mismatch repair defects in the development of cancers, may now expand the search from simply attempting to decipher the role of certain genes in carcinogenesis to also asking why and how some cells accumulate an excessive number of mutations.

Modrich, P. Mismatch repair, genetic stability and cancer. *Science* 266:1959, 1994.

bonds that hold together the pyrimidine molecules in the dimer. Photolyases are activated by light in the range of 300–600 nm. Photolyases are present in bacteria but are not essential for DNA repair; humans lack the enzymes.

Removal of a methyl or ethyl group from the 6 position of the enol form of a guanine residue reestablishes the normal structure of guanine. A specific protein accepts alkyl groups and becomes alkylated.

Postreplication Repair

The repair processes reviewed so far deal with damage of bases on one of the two DNA strands and use of the second complementary strand as a template for repair. Such repair occurs prior to replication of DNA that turns DNA damage into permanent mutation. For example, normal DNA replication with DNA polymerase III in *E. coli* cannot proceed past most types of DNA lesions until such lesions are first repaired. These lesions cannot be excised because excision would leave breaks in both strands that replication would perpetuate. Eventually, replication resumes past the site of the lesion with the polymerase skipping over a few of the damaged bases. After synthesis the daughter strand is found to be missing a base that would normally be present across the damaged base. The lesion itself is eventually repaired by borrowing template information from a homologous DNA strand. This type of repair is illustrated in Figure 15.19.

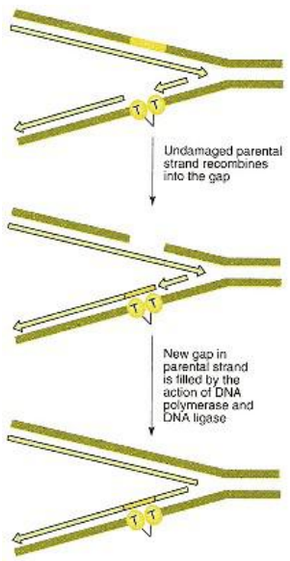


Figure 15.19
Postreplication repair.
Most DNA lesions in *E. coli* are repaired prior to replication. If an unrepaired lesion is encountered by the replication complex near the replication fork, replication is blocked at the site and resumes only beyond the unrepaired site. The gap, initially left behind in an unreplicated single-stranded segment of DNA, is eventually repaired by the process of recombination. Recombination allows the use of a complementary strand from another DNA as template.

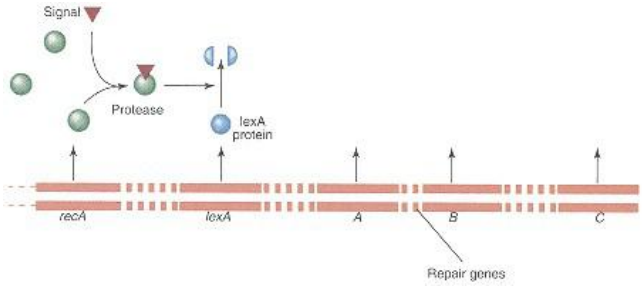


Figure 15.20
SOS DNA repair.
Under normal conditions the SOS repair proteins are not expressed. This is because a repressor protein, LexA, binds to promoter regions and inhibits the transcription of many genes required for DNA repair and DNA recombination. LexA also inhibits its own expression and the expression of another protein with multiple enzymatic roles, RecA. DNA damage, identified by the presence of single-stranded DNA, inactivates LexA. Inactivation of LexA is the result of proteolysis by the RecA protein, which when bound to single-stranded DNA functions as a specific protease. In the absence of LexA, genes that were previously inhibited by LexA can be expressed. After the damage of DNA is repaired, LexA begin to accumulate again, repressing the expression of SOS genes.

SOS Postreplication Repair

Many of the enzymes involved in DNA repair in *E. coli*, including the ABC excinuclease system, are inducible and regulated by proteins **LexA** and **RecA** that, together with the genes coding for the inducible proteins, form the **SOS repair** system.

Under normal conditions LexA binds tightly to the control region of genes that code for repair enzymes and several other proteins and prevents the expression. Genes in the SOS response also induce the *polB* gene encoding a polymerization subunit of DNA polymerase required for error-prone translesion replication. The SOS system is activated as a result of severe DNA damage. Activation can be described as the RecA-mediated cleavage and destruction of LexA in an autoproteolytic manner (Figure 15.20). The fragmented LexA dissociates from the DNA, allowing the efficient expression of the SOS response genes. Some of the products of the SOS response assemble at the lesion to form a specialized replication system that depends on DNA polymerase II for replicating past DNA lesions, which normally block DNA polymerase III. This **translesion replication** is made possible because of the distinct properties of polymerase II.

The signal that activates RecA is the binding of RecA onto exposed single-stranded DNA or damaged double-stranded DNA, when DNA replication is stalled because of extensive DNA damage. The SOS response to heavy DNA damage is a process that converts a lesion at a replication error-prone site and allows replication to be temporarily restored over the lesion.

15.4— DNA Replication

Complementary Strands Are Basic to the Mechanism of Replication

The double-stranded structure of DNA permits each strand to serve as a template for the synthesis of a new strand identical to the other strand, as suggested in Figure 15.21. The correctness of this overall scheme of replication has solidly

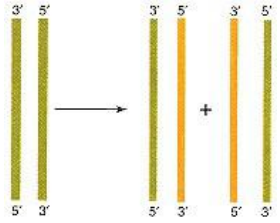


Figure 15.21
Each DNA strand serves as template for synthesis of a new complementary strand.
Replication of DNA proceeds by a mechanism in which a new DNA strand (indicated by a red line) is synthesized that matches each of the original strands (shown by green lines).

been established. Even some bacteriophages, which contain single-stranded instead of double-stranded DNA, have been shown to convert their DNA to a double-stranded form before replication. The simplicity of the basic scheme for replication conceals a rather complex set of coordinated intricate processes. A multiplicity of enzymes and protein factors participate in these processes. The enzymes involved in replication must also deal with a variety of topological problems. DNA-dependent DNA polymerase can synthesize new strands by operating only along the 5 3 direction, and therefore it is unable to elongate the two antiparallel strands of the helix in the same macroscopic direction. In addition, DNA polymerases are unable to start DNA synthesis in the absence of a preexisting primer and the replication cannot proceed unless the complementary strands are separated at an early stage of the synthesis. Separation requires the commitment of energy for disrupting the thermodynamically favorable double-helical arrangement and the unwinding of a highly twisted double helix at extremely rapid rates. Double-stranded DNA is normally a topologically closed domain, which, unless properly modified, will not tolerate strand unwinding to any appreciable degree. Obviously, these multiple difficulties must be dealt with before the replication of DNA can take place.

Replication Is Semiconservative

Three possibilities by which information transfer could take place during replication were initially visualized as indicated in Figure 15.22. Conservative replication could, in principle, yield a product consisting of a double helix of the original two strands and a daughter DNA consisting of completely newly synthesized chains. A second possibility, labeled dispersive, would have resulted if the nucleotides of the parental DNA were randomly scattered along the strands of the newly synthesized DNA. The synthesis of DNA eventually proved to be a **semiconservative process.** After each round of replication, the structure of parental DNA is found to preserve one of its own original strands combined with a newly synthesized complementary polynucleotide.

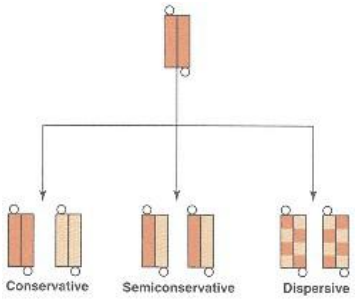


Figure 15.22
Three possible types of DNA replication.
Replication has been shown to occur exclusively according to the semiconservative model; that is, after each round of replication one of the parental strands is maintained intact, and it combines with one newly synthesized complementary strand. Circles represent the 5 terminals.

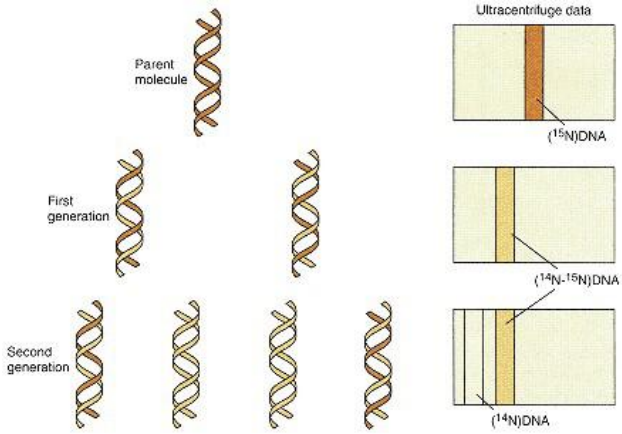


Figure 15.23
Semiconservative replication of DNA.
Schematic representation of the experiment of Meselson and Stahl that demonstrated semiconservative replication of DNA. This model of replication requires that, if the parent molecule (dark red) contains ^{15}N, each of the molecules produced during the first generation contain ^{15}N in one strand and ^{14}N in the other. Furthermore, in the second generation two molecules must contain only ^{14}N, and two molecules must contain equal amounts of ^{14}N and ^{15}N. The results of separating DNA molecules from successive generations, shown on the right, are consistent with this model.

The semiconservative nature of replication was elegantly suggested by a classic experiment that allowed the physical separation and identification of the parental and the newly synthesized strands. *Escherichia coli* was grown in a medium containing [^{15}N]-ammonium chloride as the exclusive source of nitrogen. Several cell divisions were allowed to occur, during which the naturally occurring ^{14}N in the DNA of *E. coli* was, for all practical purposes, replaced by the heavier ^{15}N isotope. The ^{14}N-containing nutrient was then added, and cells were removed at appropriate intervals. The DNA of these cells was extracted, and the ratios of ^{14}N to ^{15}N content were determined by equilibrium density gradient centrifugation. The separation between [^{14}N]DNA and [^{15}N]DNA was achieved based on the lower density of DNA, which contained the lighter isotope. In subsequent experiments, the newly synthesized DNA was thermally denatured and the individual strands were completely separated. The results, shown in Figure 15.23, demonstrated that daughter DNA molecules consisted of two strands with different densities, corresponding to the densities of single-stranded polynucleotides containing exclusively ^{14}N or ^{15}N. Conservative and dispersive replications are clearly inconsistent with these findings.

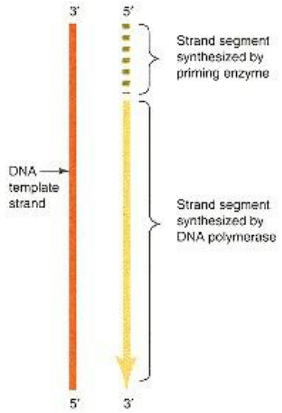


Figure 15.24
Synthesis of primer for DNA replication.
Primer (dashed line) is synthesized by primase. A primer permits new DNA (orange line) to be synthesized by DNA polymerases. The primer is excised at the completion of DNA synthesis.

A Primer Is Required

The semiconservative nature of replication requires that each strand serve as a DNA polymerase template for the synthesis of a new complementary strand. Elongation is catalyzed by polymerase III (Table 15.1), as distinguished from polymerase I, which is primarily involved in repair. Polymerase III, which is ATP-dependent, is unable to asemble the first few nucleotides of a new strand and requires a primer. In *E. coli* primers are segments 10–60 nucleotides long. With few exceptions, the primer is an oligonucleotide synthesized by other enzymes, as indicated in Figure 15.24. **Primers** are formed by **primases,** although in a few instances **RNA polymerases** are known to synthesize a primer. In some bacterial systems and phages, the priming enzyme has activity characteristic of an RNA polymerase because the ribonucleotides condense to form the primer. In other systems the primase does not discriminate between 5 -ribonucleotides and 5 -deoxyribonucleotides. As a general rule, however, primases use ribonucleotides for incorporation into primers. Some enzymes that catalyze the synthesis of primers act exclusively as primases, while others possess additional enzymatic activities. In mammalian cells primase activity is vested in **DNA polymerase α**, an enzyme that is also involved in DNA strand

elongation and in DNA repair. Once the primers have been synthesized, the DNA polymerase can move in and take over the process of synthesis. It is not clear what signal causes a switchover from primase to DNA polymerase, although it has been suggested that a specialized ribonuclease **(RNaseH)** is involved.

15.6 RNA Primers

Replicating System	*RNA Oligonucleotide*[a]
Bacteriophage T4	pppAC $(N)_3$
Bacteriophage T7	pppACCA
	pppACCC
Mouse polyoma virus	pppA $(N)_9$
	pppG $(N)_9$
Lymphoblastoid cells	pppA $(N)_8$
	pppG $(N)_8$

[a] N stands for any ribonucleotide. The primer lengths for the mouse polyoma virus and the animal cells are averages.

If DNA polymerase were the enzyme that would begin DNA synthesis by laying down the very first nucleotide complementary to the template, the efficiency of DNA synthesis would be severely reduced. Since the bases in a very short segment of a double helix have high configurational flexibility, the first nucleotide introduced into a newly synthesized DNA strand would likely be mispaired and would immediately activate the proofreading activity of DNA polymerase. The outcome would be a fruitless back-and-forth cycle of synthesis and proofreading by DNA polymerase with little net synthesis of new DNA. In contrast, primases, which have no proofreading ability, can quickly and efficiently position primers that can be elongated with DNA polymerases without appreciable backtracking. The primases ignore mismatches and produce an RNA chain long enough to allow the DNA polymerase to operate at the 3 end of a double-stranded structure that restricts newly introduced nucleotides on the basis of strict complementary rules. The mismatches introduced by the primase are irrelevant because the characteristic RNA-like structure of primers allows for their subsequent wholesale removal and replacement by DNA of an equivalent composition.

Although primers are almost invariably short RNA or RNA-like segments (Table 15.6), RNA priming is not used universally. In the "rolling circle" replication mechanism of DNA, a 3 -OH primer is generated by endonuclease digestion of parental DNA, and with **parvoviruses** a 3 -OH primer is generated by the folding back of an existing 3 terminus. A single deoxyribonucleotide can serve as primer in **adenovirus.** Such a nucleotide, with its 3 -OH terminus free, is attached to the end of a template strand through a virus-encoded specific protein (Figure 15.25).

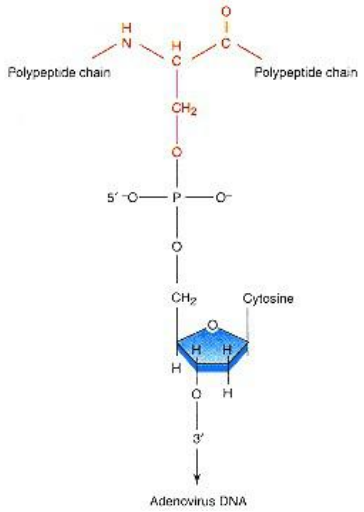


Figure 15.25
An unusual primer used in the replication of adenovirus DNA.
This primer is a single nucleotide attached, by its 5 -terminal phosphate, to a serine residue of a protein. Adenovirus DNA is synthesized by extension of the 3 terminus of this nucleotide.

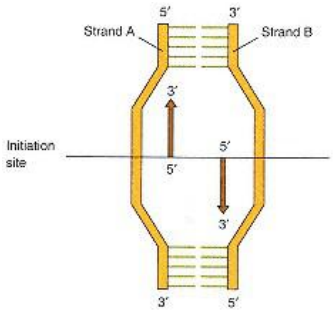


Figure 15.26
Both DNA strands serve as templates for DNA synthesis.
Each DNA strand must serve as a template for DNA synthesis. The new DNA can be synthesized only in the 5 3 direction. If only a single initiation origin were considered, the result of continuous synthesis would be the formation of two new nonidentical double-stranded DNA molecules (one above and one below the initiation origin). Also, the upper part of strand A and the lower part of strand B could not have been used as templates. In fact, the synthesis occurs both continuously and discontinuously.

Both Strands of DNA Serve As Templates Concurrently

In the preceding section, the events leading to the synthesis of DNA by DNA polymerase were examined and attention was directed to one of the two parental DNA strands used as template. In fact, synthetic events occur at both strands almost concurrently. This would appear to generate some problems of geometry. Specifically, if a single initiation site is considered, and the synthesis continued in the 5 3 direction until each template is completely copied, the result of the synthesis would be the creation of two new double-stranded molecules. Examination of Figure 15.26 indicates that, at least in the case of linear double-stranded DNA, neither of these two hypothetical DNA molecules would be identical to the parental DNA.

Such an outcome is not in agreement with the actual course of DNA replication. The discrepancy can be accounted for by recognizing that the microscopic synthesis of the new strands does not proceed uninterrupted. In fact, the synthesis occurs in a discontinuous fashion and in a manner that permits the assembly of the synthesized polynucleotide portions into appropriate complete DNA strands.

Synthesis Is Discontinuous

The overall process of DNA synthesis may now be considered past the immediate vicinity of initiation by examining a larger section of DNA. One of the two parts of DNA that would be generated if the macromolecule were divided at the site of chain initiation is shown in Figure 15.27. In almost every instance the synthesis is **bidirectional,** which means that the synthetic events occurring at the part of the molecule indicated by solid lines are of the same general nature as those occurring on the other site and indicated with dashed lines.

A prerequisite for the semiconservative mechanism of replication is that the two complementary strands of DNA gradually separate as the synthesis of new strands takes place. The mechanics of this separation are addressed later, but it may be apparent that as a result of separating the strands at an interior position, two topologically equivalent forks are created at the point of diversion of the two strands.

Various lines of evidence have indicated that DNA polymerase acts in a **discontinuous** manner; that is, along each DNA molecule there are numerous initiation points at which primers are formed. In eukaryotes primers may be formed at locations that are determined by nucleosome spacing. In the case of bacteriophage T7, primosomes appear to recognize TGGT and GGGT through prepriming proteins. Once a site for primer initiation has been recognized,

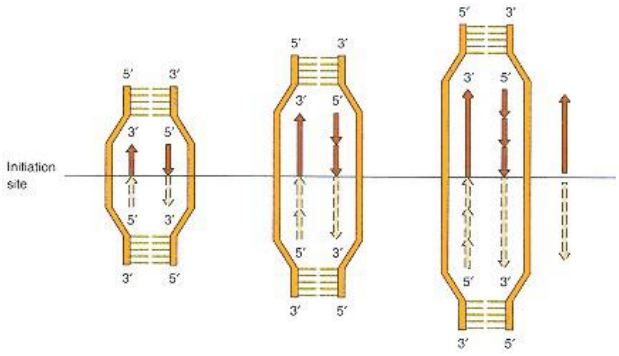


Figure 15.27
Discontinuous synthesis ofDNA.
This figure emphasizes the synthetic events occurring at only one side of the initiation site (dark red line). The two complementary strands of DNA separate as the discontinuous synthesis of small DNA segments takes place on both strands located at different sites on the DNA. After excision of the primers, the excised parts are repaired, and the segments are joined together. Although segments are clearly synthesized in opposite directions on the two strands, overall macroscopic impression is that DNA grows in the single direction suggested by the solid red arrow on the right.

single-strand binding proteins (SSB), which interact with single-stranded polynucleotides, are displaced and the primase lays down a primer. After promoting primer initiation at one point, prepriming proteins move along the template strand in order to synthesize the adjacent primer. At each one of these locations, DNA polymerase III makes use of the assembled primers for the synthesis of DNA. When DNA polymerase reaches the end of the single-stranded template, it comes upon the next primer annealed to the template. The polymerase, as indicated by its very high processivity, can overcome this hurdle by sliding over the intervening double-stranded DNA–RNA hybrid and resuming replication at the 3 end of this new primer.

The segments synthesized by DNA polymerase upon each primer, known as **precursor (Okazaki) fragments** or **nascent DNA,** vary in size from about 100 to 200 deoxyribonucleotides in eukaryotes to ten times as long in bacteria. Once these segments of the new DNA are synthesized on both strands of a fork (Figure 15.27), the fork opens up further, and the same process of synthesis is repeated. Shortly after synthesis, the primer portions of the Okazaki fragments are excised by the 5 3 exonuclease activity of DNA polymerase I, which also synthesizes short segments of DNA.

This discontinuous mechanism compensates for the inability of DNA polymerase to synthesize strands in the 3 5 direction. By synthesizing portions of DNA strands only in the 5 3 direction on both antiparallel strands of the parental DNA, the polymerase is able to create the illusion, when the synthesis is experimentally visualized by electron microscopy techniques, that both strands are concurrently elongated in the same macroscopic direction. In Figure 15.27 this direction is indicated by a large solid arrow. It should be noted that the first strand synthesized, often referred to as the **leading strand,** is synthesized continuously. It is the other strand, the **lagging strand,** that must be synthesized discontinuously.

Macroscopic Synthesis Is As a Rule Bidirectional

At the site of initiation of DNA synthesis two identical forks are created (Figure 15.27). Therefore two possibilities exist for the synthesis of DNA: the process may occur at only one fork and proceed in a single direction, as shown by the thick solid arrow, or alternatively it may occur at both forks and in both directions away from the starting point. The events occurring in the forks located below the starting line are simply a mirror image repetition of what occurs in the fork that is located above the line. **Bidirectional replication** is the mechanism of DNA synthesis. The only known exceptions are in a small number of phages and plasmids that replicate unidirectionally. In the case of a small linear chromosome (e.g., bacteriophage) each fork moves along, synthesizing new DNA, until the end of the chromosome is reached. In a circular chromosome (e.g., *E. coli*) the two forks proceed in opposite directions until they meet at a predetermined site on the other side of the chromosome, as depicted in Figure 15.28. As the two forks meet, a new copy of the parental DNA is completed and released. The average rate at which each fork moves during replication is of the order of 60,000 bases per minute at 37°C. Upon completion, new DNA is released by the action of a type II topoisomerase as illustrated in Figure 15.29.

Strands Must Unwind and Separate

Separation of the strands of the parental DNA prior to synthesis of new strands is a requirement because the bases of each template must be made accessible to the complementary deoxyribonucleotides from which the new strands are constructed. The overall process of separation consists of a number of enzymatically catalyzed, coordinated steps, including the local unwinding of the helix, and the nicking and rejoining of the strands necessary for continuation of the

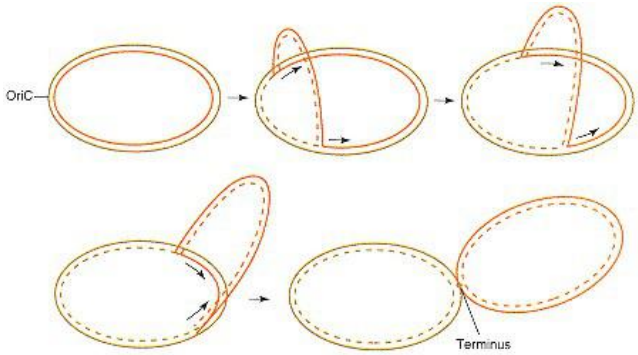


Figure 15.28
Bidirectional replication of a circular chromosome.
Replication starts at a fixed origin and proceeds at a constant rate in opposite directions until the two replication forks meet. Newly synthesized strands are indicated by dashed lines. After DNA synthesis is complete, two newly synthesized circular DNA molecules are separated by action of topoisomerases.

unwinding process. Once the strands are unwound, they must be kept separate so that they can operate freely as templates.

Specialized proteins accomplish rapid orderly unwinding of the strands. These proteins, **helicases,** separate DNA strands in advance of the moving replication fork and just in front of DNA polymerase. In *E. coli* they are referred to as **helicase II** and **rep protein.** Helicases move unidirectionally along DNA and separate the strands in advance of replication. They destabilize the interaction between complementary base pairs at the expense of ATP.

Once the strands have been separated, the single-stranded regions are stabilized by specific proteins, the **single-strand binding (SSB) proteins.** The DNA single strands are covered by the SSB proteins because of their high affinity for single-stranded DNA. As the helicase moves in advance of the replication fork, SSB proteins go on and off the DNA, with protein molecules that are displaced from one site reassociating with another (Figure 15.30). SSB proteins do not consume ATP and do not exhibit any enzymatic activities. Their role is only to keep the strands apart long enough for the priming process to occur.

In *E. coli* DNA, it is calculated that the parental double helix must unwind at a rate of about 6000 turns per minute. These high rates would generate insurmountable difficulties if strands were to separate over an appreciable length of DNA. The large free-energy requirements of bringing about the unwinding of large regions of DNA can, however, be reduced to manageable levels by the nicking of one or both of the DNA strands near the replicating fork. Since the fork is a moving entity, the nicking must be visualized as a reversible cut-and-rejoin process, which moves along with the fork. Nicking is indispensable for a topological reason as well. Unwinding at one of the two forks requires that the parental double helix rotate in the opposite direction to that necessary for the unwinding of the opposite fork. In the absence of a nick as the unwinding at one of the forks would progress, an increasing number of positive supercoils would have to be introduced into the double helix. Once the limit of the helix to accommodate the supercoils were reached, unwinding and replication would have to cease.

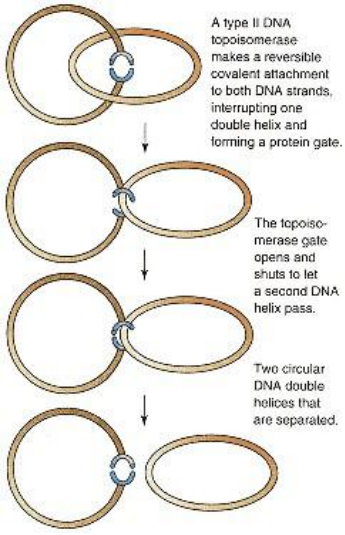


Figure 15.29
Function of topoisomerases II in separating interlocked DNA double helices.
Topoisomerase II attaches to both strands of DNA through reversible covalent bonds, thus forming an interrupted double helix with a topoisomerase "gate." A second DNA helix can pass through the portal using an "open-and-shut-the-gate" mechanism, leading to two separated DNA molecules. After separation of the molecules topoisomerase dissociates from DNA.

These topological restraints are overcome if DNA is maintained during replication in the **negative superhelical form.** This form could serve as a "sink" for the positive supercoils that could potentially be generated during replication. In *E. coli*, this is apparently achieved by the action of **gyrase,** a **topoisomerase** type II, which induces the formation of negative supercoils

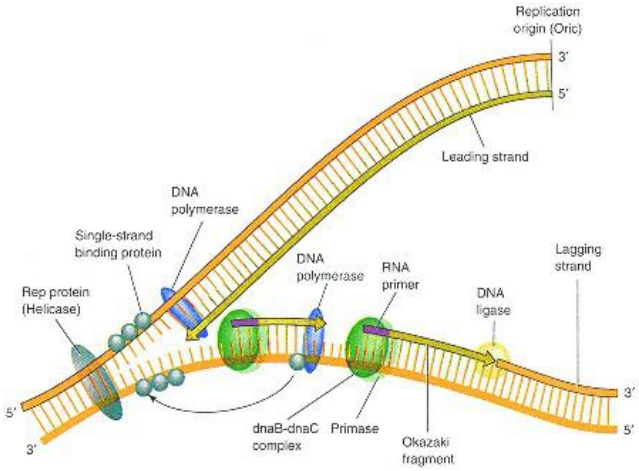


Figure 15.30
Model for DNA replication in *E. coli*.
The initial stages of replication are depicted. Primers are removed from newly synthesized segments of DNA at the lagging strand, and the segments are joined. Since replication is normally bidirectional, similar events take place concurrently at the other side of the initiation origin.

at the expense of ATP. **Topoisomerases type I** may also be involved. The superhelicity of DNA may be negatively regulated through a balance between topoisomerases of types I and II; that is, a diminishment of topoisomerase II activity may bring about a decrease in the amount of negative superhelicity that can be created, whereas an inhibition of topoisomerase I activity may increase it. During replication the linking number between parental strands decreases from a large value at the beginning of replication to zero at the end of a complete round of DNA synthesis.

TABLE 15.7 Components of the Replisome

Protein	*Function*
SSB	Single-strand binding
Protein i (dnaT) Protein n Protein n Protein n dnaG	Primosome assembly and function Primase
	(primer synthesis)
Pol III holoenzyme	Processive chain elongation
Pol I	Gap filling and primer excision
Ligase	Ligation
Gyrase	Supercoiling
gyrA	
gyrB	
rep	Helicase
Helicase II	
dnaB	Helicase
dnaA dnaC	Origin of replication

Escherichia coli *Provides Basic Model for Replication of DNA*

Extensive studies in *E. coli* and its phages have permitted the proposal of a replication model that depends on the action of a large number of proteins, some of which are listed in Table 15.7. With the specific exceptions noted in the sections that follow, this model may also be viewed as a basic scheme for DNA replication in most other cells.

Initiation and Progression of DNA Synthesis

Synthesis of DNA begins at a specific site of the chromosome referred to as the **replication origin,** which in *E. coli* is referred to as **OriC** (Figure 15.30). Initiation of DNA synthesis involves participation of as many as 20–30 different proteins, many of which are needed to be present at the origin of replication in multiple copies. OriC must be recognized by specific proteins, and the origin must unwind to allow helicase, primase, and DNA polymerase III to have access to each DNA strand. OriC is a sequence of 245 base pairs that contains four sites (nucleotide 9-mers with a similar nucleotide sequence) at which **dnaA,** a tetramer consisting of four identical subunits, can initiate the stepwise assembly of all the proteins and enzymes necessary to carry out replication (Figure 15.31). In addition, the origin contains 11 methylation sites recognized by **Dam methylase** and three AT-rich direct tandem repeats consisting of 13 base pairs each. This final assembly is called a **replisome.**

Formation of a replisome begins with the binding of one dnaA molecule

at each one of the 9-mers, provided that these binding sites are fully methylated. The dnaA apparently recognizes these 9-mers on the basis of their conformation, which appears to be slightly curved with the double helix somewhat elongated relative to typical B-DNA. Several more additional dnaA molecules are then added via a highly cooperative process to form a nucleosome-like structure. An additional factor, **HU protein,** participates in the formation of this complex.

The dnaA and HU protein interact with the OriC in a manner that promotes the opening of the DNA strands in the AT-rich regions adjacent to the origin. Finally, dnaA, with the aid of **dnaC,** adds **dnaB** in the complex. The dnaB, by virtue of its helicase activity, creates an initiation "bubble" consisting of a few hundred nucleotide pairs. The energy for the formation of the "bubble" is provided by ATP in a reaction catalyzed by topoisomerase II, and the "bubble" is stabilized by SSB proteins.

Synthesis of an RNA primer begins with the formation of a prepriming complex. The prepriming assembly consists of the dnaB–dnaC complex to which four other proteins (polypeptides n, n , n", and i) have been added. Addition of primase, **dnaG,** converts the prepriming complex to a **primosome**

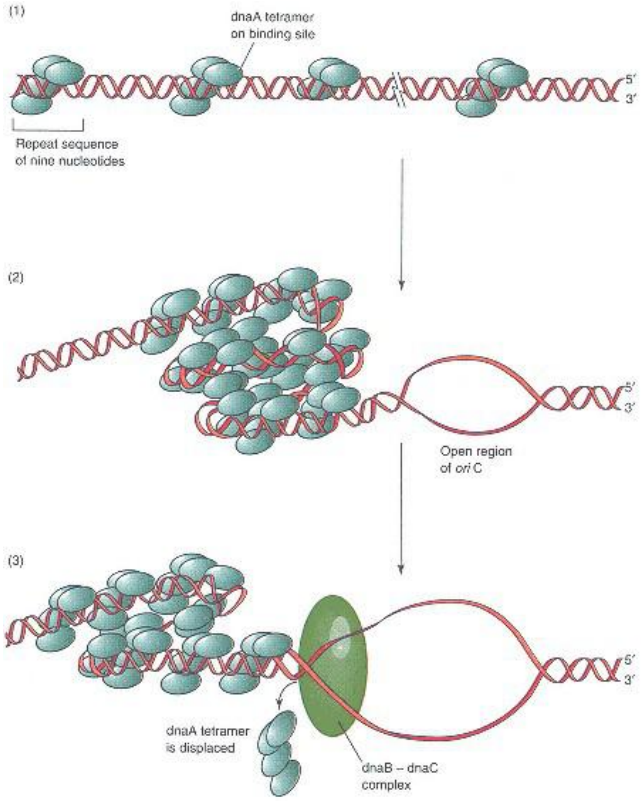


Figure 15.31
Model for initiation of replication in *E. coli*.
Step 1: Initiation of replication begins with binding of dnaA molecules to four sites consisting of nine-nucleotide long sequences each. These sequences are present at the origin of replication in *E. coli* (OriC).
Step 2: DNA-bound dnaA molecules subsequently coalesce and are joined by additional dnaA molecules to form a nucleosome-like DNA-protein complex, which promotes nearby "melting" of the double helix.
Step 3: The resulting opening of strands allows a dnaB–dnaC complex to become attached to DNA so that helicase activity of dnaB can further unwind the DNA. Unwinding is accompanied by a displacement of dnaA molecules.
Redrawn based on figure in Rawn, J. D., *Biochemistry*. Burlington, NC: Neil Patterson Publishers, 1989.

(Figure 15.32). The primosome interacts with a template, at each one of the two forks generated by the formation of a "bubble," and begins the synthesis of RNA primers on the two leading strands. Assembly of the replisome is completed by addition to the primosome of DNA polymerase III and **rep proteins.**

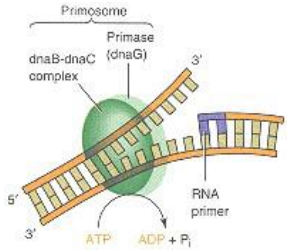


Figure 15.32
Primosome of *E. coli.*
The primosome is formed by binding of primase, together with a complex of dnaB and dnaC proteins, at specific sequences of DNA that serve as sites for formation of RNA primers. Additional factors, described as n proteins, are specific primosomal components that are responsible for placing the primosome at the appropriate sequences. In effect, the primosome "searches" the DNA for these sequences at the expense of ATP. Once the correct destination of the primosome is reached, RNA primer synthesis is initiated.

Initiation can be regulated by either restricting the availability of dnaA-binding sites at OriC or by limiting the concentration of dnaA. **Methylation** provides a switch for the availability of dnaA-binding sites. Once replication has been initiated, the dnaA near OriC binds to the plasma membrane and becomes unavailable to Dam methylase. In addition, binding of DNA in the vicinity of OriC to the cellular membrane sequesters the *dnaA* gene, which is situated near OriC (only 40 kb away). As a result, the synthesis of dnaA protein is inhibited and its cellular concentration is lowered.

Initiation of the leading DNA strand at OriC by the primosome is more complex than the subsequent initiation of synthesis of Okazaki fragments on the lagging strand initiated by primase at sites selected by the prepriming proteins. The initiation of the leading strand does not present the cell with serious topological problems, but for continuation of synthesis helicase II and rep protein are essential. These enzymes unwind and separate the strands in each of the two forks created by the initiation event. As the helicases move in advance of each fork, two single-stranded regions are generated on parental DNA. These regions are immediately covered by single-strand binding protein that keeps the fork open and allows DNA polymerase III to take over the elongation of primers. A signal for initiation of the lagging strand, uncovered on the template by the movement of helicase, leads to the binding of primase. Primase, the action of which is triggered by the prepriming proteins, synthesizes a brief complementary segment of the strand. This segment serves as a primer for covalent extension of the strand synthesized by DNA polymerase III and for formation of Okazaki fragments. DNA polymerase III complexes are endowed with similar but somewhat distinct properties, one tailored for the continuous synthesis of the leading strand and the other for the discontinuous synthesis of the lagging strand. This polymerase assembly, which appears to combine primase activity with nonidentical twin active sites for polynucleotide synthesis, allows for concurrent replication on both strands. In this scheme, looping of the lagging strand template by 180° brings it to the same orientation as the leading strand template (Figure 15.33). Thus a primer synthesized at the lagging strand is drawn past it. When a nascent (Okazaki) fragment reaches the 5 end of the previously synthesized Okazaki fragment, the lagging strand template is released and unlooped. Removal of the primer portions at the 5 end of the Okazaki fragments by DNA polymerase I, repair by the same enzyme, and joining of the repaired fragments by DNA ligase produces intact DNA strands.

Termination of DNA Synthesis

Termination occurs near the center of a 270-kb region across from OriC, the **ter** or **τ locus.** This region incorporates five ter sequences, that is, loci with the core sequence GTGTGTTGT that bind the **Tus protein (terminator utilization substance)** that promotes the termination of synthesis (Figure 15.34). Tus protein is a contrahelicase in that it functions by literally interfering with the ATP-dependent and dnaB helicase-promoted unwinding of DNA rather than simply impeding the propagation of this helicase along the double helix. The organization of the ter region is shown in Figure 15.34. Each Tus site has directional properties (asymmetry) and it arrests only those replisomes that reach the Tus site from one specific direction. Replisomes arriving from the opposite direction apparently force the dissociation of the Tus protein and thus

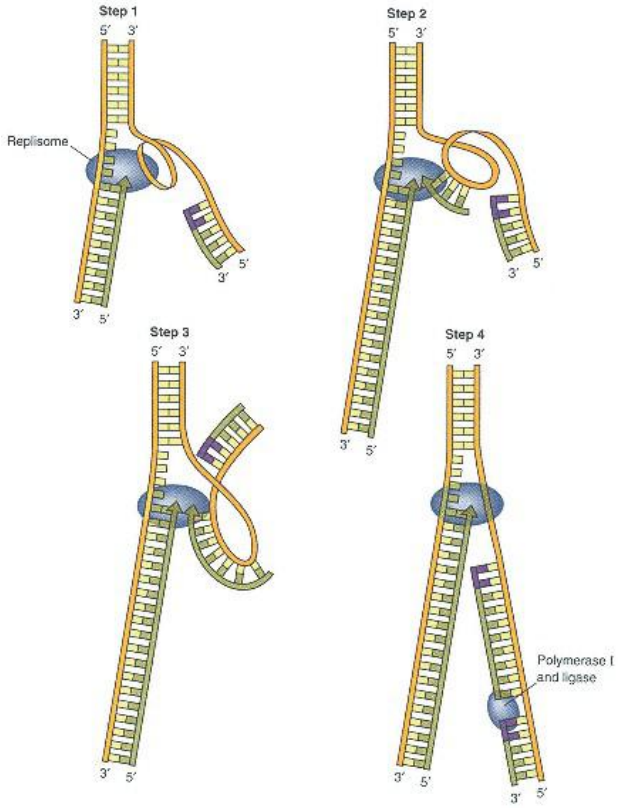


Figure 15.33
Model for the simultaneous synthesis of leading and lagging DNAstrands by DNA polymerase.
Two molecules of DNA polymerase operating in concert, and in the same rather than the opposite direction, may be participating in the simultaneous synthesis of DNA on both strands. In this model the replisome consists of a DNA polymerase dimer associated with the primosome and helicases. The primer made by the primosome is extended by the replisome as the lagging-strand template is looped through it. The primer continues to be extended until the previously completed Okazaki fragment is reached, at which point the loop is relaxed. The stretch of unpaired lagging-strand template then loops back again to participate in the formation of the next Okazaki fragment.
Redrawn based on figure in Kornberg, A. *DNA Replication*. San Francisco: Freeman, 1992.

can proceed unimpeded past the Ter–Tus site. Because of the distribution and orientation of sites in the ter region, each replisome must first pass over all sites that are oriented the opposite way before arriving at the Tus site that is oriented in a way that causes termination. This arrangement makes it inevitable that a replisome will not dissociate from DNA until it actually collides with the replisome entering the ter region from the opposite direction. This ensures the complete replication of the chromosome and prevents overreplication. The products of replication are two **concatenated** progeny chromosomes usually interwound by as many as 30 coils. The newly synthesized DNA is untangled from the parental DNA apparently by the action of a topoisomerase II.

Rolling Circle Model for Replication

DNA synthesis directed by circular mtDNA, and in some instances by bacteria and viruses, gives rise to linear daughter DNA molecules that contain the base sequence of parental DNA repeated numerous times. These repeated linear DNAs, which are known as **concatemers,** are essential for the bacterial mating and may be involved in gene amplification. The synthesis of concatemer DNA occurs by a mechanism known as **rolling circle replication.**

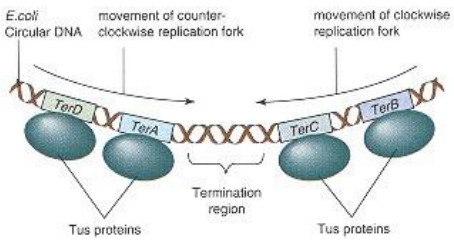


Figure 15.34
Termination of DNA replication in *E. coli*.
Termination region (ter) of *E. coli* incorporates five asymmetric ter sites. Each ter site can interact with Tus protein. TerB and terC are oriented in the same direction and the remaining three ter sites are oriented in the opposite direction. Because of the orientation of Tus-bound ter sites, each replisome that reaches the ter region must cross all the Tus–ter sites that are oriented the opposite way before arriving at a site that causes termination. A replisome moving in the direction shown by the arrow must first cross terE, terD, and terA before terminating replication at either the terC or terB site. This arrangement ensures that each replisome continues to synthesize DNA until it collides with a replisome entering the ter region from the opposite direction, leading to the dissociation of both replisomes from DNA.
Adapted from Hidaka, M., Kobayashi, T., and Horiuchi, T. *J. Bacteriol*. 173:381, 1991.

An example is the replication of certain circular single-stranded bacteriophages such as X174. When the virus enters a host bacterium the single-stranded genome is converted to a double-stranded DNA by action of primase and DNA polymerase III. The DNA strand complementary to the bacteriophage genome that is first synthesized [labeled the (–) strand] serves as the template for the genomic DNA [the (+) strand]. The atypical characteristic of this replication scheme is that the (+) strand is nicked at a specific site (by a phage-encoded endonuclease) so that it can serve as a primer for its own replication. The (+) strand is elongated from the 3 -hydroxyl end of the nick by DNA polymerase III by incrementally displacing segments of the (+) strand associated with the "helper" (–) strand (Figure 15.35).

A second characteristic is that the circular template does not dissociate from the complementary strand during the synthesis. Instead the replication of the leading strand goes on beyond the length of circle-generating linear concatemeric DNA. Appropriately sized DNA molecules are subsequently generated from concatemers by specific endonuclease cleavage.

Eukaryotic DNA Replication

The DNA synthesis in eukaryotes appears to be a process that is fundamentally similar to that occurring in prokaryotes. Formation of a replication fork, primer

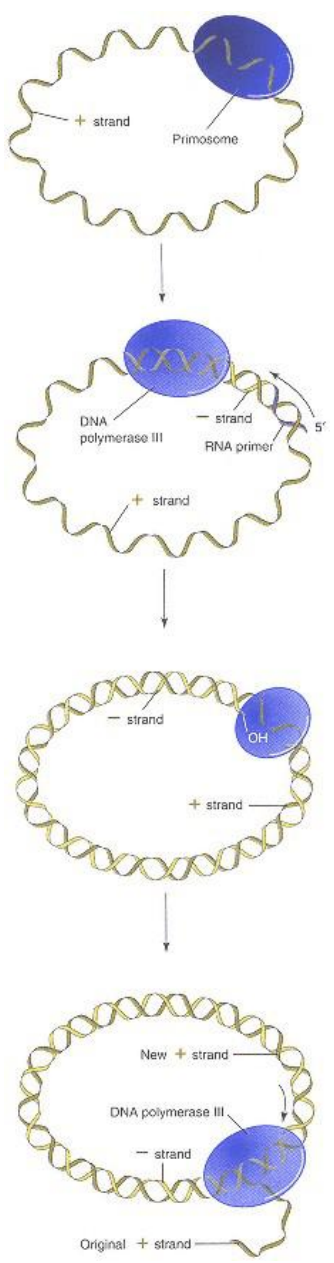


Figure 15.35
Replication by the rolling circle mechanism.
In ssDNA of certain bacteriophages, such as X174, the (+) strand is converted into dsDNA upon injection into a host bacterium. This transformation occurs by action of primase and polymerase III upon ssDNA that synthesizes a complementary (–) strand. Replication of (+) strands begins with nicking of (+) strand so that it can serve as a primer for its own replication. The (+) strand is elongated from the 3 -hydroxyl end of the nick, as the newly synthesized strand gradually displaces from the helper-strand the original (+) strand.
Redrawn based on figure in Moran, L. A., Scrimgeour, K. G., Horton, H. R. Achs, R. S., and Rawn, S. D. *Biochemistry*. Englewood Cliffs, NJ: Neil Patterson/Prentice Hall, 1994.

synthesis, Okazaki fragments, primer removal, and gap bridging between newly synthesized DNA segments, all parallel the corresponding steps that occur in prokaryotes, but the overall process is quite a bit more complex. Replication among eukaryotes, from yeasts to humans, shares similarities.

As expected, differences are more pronounced between prokaryotes and eukaryotes. In rapidly growing prokaryotes, DNA is replicated through much of the cell cycle and cell division occurs as soon as DNA synthesis has ceased. In contrast, eukaryotic DNA synthesis (and histone synthesis) is confined to only one part of the cell cycle, specifically the synthetic (S) phase of the interphase. This phase is preceded and followed by two periods during which DNA is not synthesized (gap periods G1 and G2). Cell division occurs at a different time within the interphase, referred to as the mitotic (M) period. Beyond this characteristic limitation of eukaryotic replication to a certain period of the cell cycle, important differences in replication between prokaryotes and eukaryotes arise primarily from the larger size of eukaryotic DNA (about 10^5–10^6 kb content) as compared to prokaryotic DNA (about 5×10^3 kb for *E. coli*), the distinct packaging of eukaryotic DNA in the form of chromatin, and the slower rates of fork movement in eukaryotes. For DNA to become available to DNA polymerases, nucleosomes must disassemble, a step that slows the rates of fork movement. DNA polymerase movement does not exceed 30,000 base pairs per minute, which is considerably slower than the rates observed for *E. coli*. Based on the higher DNA content of animal cells, and the lower activities of DNA polymerases in comparison to bacteria, the replication cycle of eukaryotic cells could be expected to take as long as a month to complete. In fact, however, the replication cycle is completed within hours, because compensating factors are in operation. Eukaryotic cells contain a large number of DNA polymerase molecules (often in excess of 20,000) as compared to a few dozen in each *E. coli* cell. DNA polymerase initiates bidirectional synthesis but at several origins of replication located anywhere between 5 and 300 kilobase pairs (kb) apart within the chromosome, depending on species and cell type (Figure 15.36). DNA segments between two origins of replication are termed **replicons.** An average human chromosome contains as many as 100 replicons and replication may proceed simultaneously at as many as 200 forks. More origins can be found in developmentally active cells that carry out DNA synthesis at very rapid rates. During early embryogenesis the largest chromosome of *Drosophila melanogaster* contains as many as 6000 replicating forks, or one for every 10 kb.

Role of Eukaryotic DNA Polymerases

In prokaryotes synthesis is catalyzed by two similar but distinct subunits of DNA polymerase III. In eukaryotes, synthesis of the leading and lagging strands is carried out by different enzymes (Table 15.2). **DNA polymerase δ,** a polymerase of high processivity, catalyzes the synthesis of the leading strand. This enzyme consists of a large subunit that is vested with 5 3 nucleotide polymerizing activity and a smaller subunit that has a 3 5 proofreading exonuclease activity. The high processivity of DNA polymerase is attributed to the presence of an accessory factor, the **proliferating cell nuclear antigen (PCNA),** that is found in large amounts in the nuclei of proliferating cells. PCNA (mol wt 25,000) is a multimeric protein that can act as a "clamp" to keep the enzyme from disassociating off the leading DNA strand. The "clamp" consists of three PCNA molecules, each containing two topologically identical domains that are tightly associated to form a closed ring. This suggests that in eukaryotes PCNA is the functional equivalent of the β subunit of *E. coli* polymerase III. Another accessory protein, the **replication factor C (RFC),** also binds to polymerase and probably assists with association between PCNA and DNA to form the "clamp." Alternatively, RFC may be involved in setting up a link

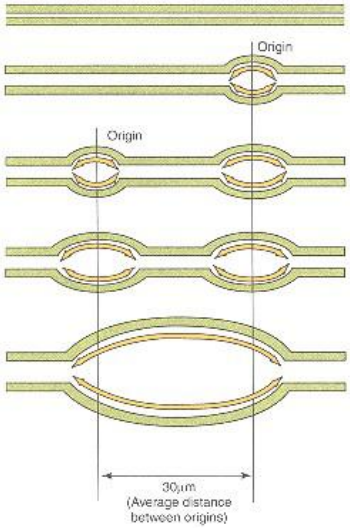


Figure 15.36
Replication of mammalian DNA.
Mammalian DNA replicates by using a very large number of replicating forks simultaneously. This mechanism accelerates the process of replication, which in mammalian systems is limited by rates of fork movement that are considerably slower than those characteristic of prokaryotes. Redrawn based on figure in Huberman, J. A., and Riggs, A. D. *J. Mol. Biol*. 32:327, 1968.

between polymerase and polymerase α. Therefore the role of RFC in DNA synthesis is analogous to the roles of the γ complex and the subunits of *E. coli* DNA polymerase III.

Synthesis of the lagging strand is catalyzed by **DNA polymerase α.** This polymerase has similar structure and properties in all eukaryotes. The large subunit (mol wt ~ 180,000) of the tetrameric DNA polymerase α is vested with the usual 5 3 nucleotide polymerizing activity. Polymerase α, isolated from some but not most sources, also has a 3 5 exonuclease activity. Two of the other subunits of the enzyme are **primases.** The primary proofreading function in eukaryotes appears to be carried out by polymerase . Polymerase improves the fidelity of replication by a factor of 10^2 and contributes in limiting the rates of overall error to 10^{-9} to 10^{-12}.

The relatively low processivity of DNA polymerase α is typical for an enzyme involved in synthesis of the lagging strand that is assembled from segments of DNA that are no larger than 100–200 bp. The size of these Okazaki fragments is approximately equal to the length of DNA wrapped around a nucleosome. This observation suggests that eukaryotic DNA may be releasing one nucleosome at a time for priming of the lagging chain. The primase subunit of the enzyme synthesizes Okazaki segments as a closely coordinated priming–synthesizing activity, by laying down RNA primers containing 5–15 nucleotides that are subsequently extended by the synthetic activity of polymerase α. This polymerase catalyzes the synthesis of a polynucleotide chain at a rate of 50 nucleotides per second, which is about 1/20 the rate of *E. coli* DNA polymerase III synthesis. Looping of the lagging strand allows a combined polymerase α-polymerase asymmetric dimer to assemble and elongate both the leading and lagging strands in the same overall direction that corresponds to the direction of the fork movement. A third large monomeric protein, **polymerase ε,** is vested with a synthetic 5 3 polymerase activity and both a 3 5 proofreading exonuclease activity and a 5 3 exonuclease activity. Polymer-

ase is mainly required for DNA repair and for filling the gaps between Okazaki fragments on the lagging strand.

Eukaryotic DNA synthesis requires **replication protein A (RPA),** also known as **replication factor A (RFA).** This protein is the functional equivalent of prokaryotic single strand binding (SSB) protein. While helicase activities are part of the prokaryotic chromosome, eukaryotic helicases do not appear to be associated with primase activity. Eukaryotic helicase activity appears to be associated with DNA polymerase .

Initiation of Eukaryotic DNA Replication

Origins of replication in eukaryotic cells have been identified in yeast (*Saccharomyces*) and are termed ARS for **autonomously replicating sequence.** ARSs are about 100–120 bp long, each of which is characterized by an AT-rich central region. The 400 or so copies of the ARS in the yeast genome have highly conserved nucleotide sequences within the central region with variations in the flanking sequences. The core sequences of ARS contain 11-bp elements known as the **ARS consensus sequence** rich in AT pairs that appear to be analogous to the AT-rich 13-mers present in the OriC of *E. coli*. The flanking elements consist of overlapping sequences that include variants of the core sequence. Protein binding to form a so-called **origin of replication complex (ORC)** promotes DNA strand unwinding over the AT-rich sequences of the ARS cores. The unwound region is stabilized by single-strand-binding protein and RPA, and is extended by helicase. Polymerases α and , RFC, and PCNA are thus introduced into the origin of replication and begin DNA synthesis.

Weaker binding sites identified as B1, B2, and B3 are also present near the origin. B1 and B2 serve as sites for ORC formation, while B3 is associated with a protein that promotes initiation of transcription. This observation highlights the close association between eukaryotic DNA replication and transcription. Controlled activation of variant ARS-like subgroups, consisting of ARS-like sequences with different flanking elements, may determine the order of initiation of DNA synthesis in eukaryotes. Sequences completely comparable to yeast ARS have not been identified in higher eukaryotes. In mammals it appears that initiation depends more on chromosomal context than on specific sequences. Origins of initiation may be found within a broad section of the genome that also contains a small number of "hot spots," at which initiation is favored. In spite of these differences in the origins of replication between yeast and higher eukaryotes, the rest of the replication machinery appears to be remarkably analogous. Eukaryotic genomes replicate in a definite order, and at definite times within the S phase, with some DNA regions replicating early in the S phase and other DNA regions replicating later. Genes that replicate early are found in active segments of chromosomes, and genes that replicate later are located in the inactive areas of chromosomes. This pattern of activation changes with development. Differences in the rate of replication are regulated by variations in the duration of the S phase, which can be achieved either by controlling the number of replicons activated per unit length of chromosome or by slowing down the rate of DNA unwinding and replication. Sequence elements similar to the ARS subgroups in yeast may control replicon activation in other eukaryotes through the interaction of initiating proteins with these elements. Origins that are activated simultaneously are expected to share the same DNA sequences and bind to the same control proteins.

Since eukaryotic DNA is present in packaged form as chromatin, DNA replication is sandwiched between two additional steps, namely, a carefully ordered and incomplete dissociation of the chromatin and reassociation of DNA with the histone octamers to form nucleosomes. Methylation at the 5 position of cytosine residues by a **DNA methyltransferase** appears to function by loosening up the chromatin structure and allowing DNA access of proteins and enzymes needed for DNA replication. The synthesis of new histones occurs

mainly during the S phase simultaneously with DNA replication. Histone molecules appear to rarely leave the DNA to which they are bound. Instead transcription and replication forks are apparently able to move past the parental nucleosomes as they synthesize mRNA or new DNA. One possibility is that each nucleosome dissociates into two halves, thereby permitting DNA polymerase to replicate transiently uncoiled DNA. Newly synthesized DNA inherits some parental histones, which it combines with an equal amount of new histones to complete the structure of nascent nucleosomes that are formed behind the moving replication forks.

In coordinating the synthesis of DNA the eukaryotic cell copies millions of base pairs, distributed over numerous chromosomes, with remarkable accuracy and at just the right time in the cycle of cell division. Copying start at hundreds of different origins, some of which are triggered early in the S phase of the cell cycle while others are triggered late. Recent evidence indicates that the replication initiator, that is the ORC complex, does not act alone in controlling initiation. One or more additional proteins bind to the initiation origins late in mitosis and remain attached until the S phase begins. These proteins are known as **cyclin-dependent kinases** (CDKs) and operate in association with specific protein substrates **(cyclins).** Cyclins and CDKs may control the cell cycle; they push the cell to the S phase and initiation of DNA synthesis. Cyclin-CDK pair also prevents DNA synthesis from being initiated a second time, so that only one S phase occurs per cell cycle. Degradation of CDKs removes the signal that inhibits cell division and the cell cycle moves again to mitosis. This scheme suggests that DNA initiation depends upon the formation of a prereplication complex by adding to or removing from the ORC cyclins and CDKs in a cyclical manner. This scheme in which the same enzyme first activates DNA replication and then, once one round of DNA replication has begun, inhibits reformation of the prereplication complex provides an efficient arrangement for the coordination of the initiation of DNA synthesis.

DNA Replication at the End of Linear Chromosomes

Linear chromosomes cannot be fully replicated in the absence of additional steps that provide for the replication of their terminals. As a replisome falls off from the end of a linear chromosome, and the daughter DNA molecules separate, synthesis of DNA on the end of the lagging strand cannot be fully completed. A gap resulting from removal of a primer that was used to start replication is generated on the lagging strand (Figure 15.37). The exact size of this gap

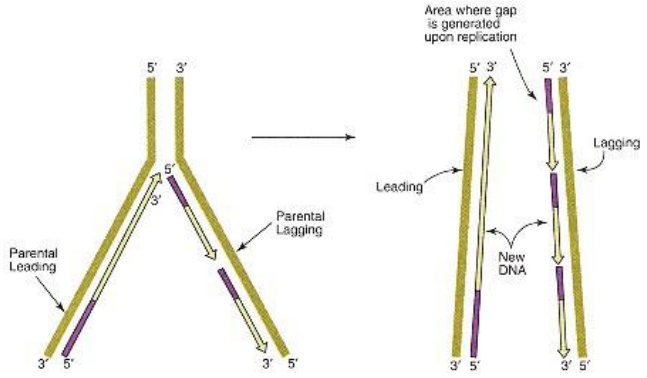


Figure 15.37
DNA replication at the ends of linear chromosomes.
In the absence of a special mechanism of replication operating at the ends of chromosomes, the completion of DNA synthesis of linear dsDNA would leave gaps at ends of newly synthesized strands. These gaps would result from removal of primers used to start replication. Upon each subsequent round of replication the gaps would be continuously expanded and accumulated because DNA polymerase requires a primer and therefore it cannot fill such gaps.

CLINICAL CORRELATION 15.6

Telomerase Activity in Cancer and Aging

Telomerase activity maintains appropriate length of the telomere sequences of chromosomes. Surprisingly, however, telomerase activity is absent from most somatic cells. In such cells telomere repeats gradually decrease in number with aging, as repeated cell divisions produce a substantial shortening of the telomere structure. Loss of telomerase activity in protozoans, such as *Tetrahymena*, is responsible for a gradual shortening of telomeres following each cell division, throughout the life of the cell. In human cultured fibroblast cells a linear inverse relationship exists between the length of telomeres and the age of the subject from which the cells are obtained. Eventual loss of telomeres leads to chromosomal instability and cell senescence and it may be an important factor that contributes to the process of aging. Specifically, telomere length appears to serve as a mitotic clock that limits the replication potential of mammalian cells. If it is true that the shortening of telomeres may be a contributing factor to the aging process, then the natural life span of an individual may be determined by the length of its telomere DNA. However, the possibility that telomere shortening may be the result, rather than the cause, of aging cannot be excluded. In any event, many other factors are also likely to contribute to the process of aging.

Since telomere length may serve as a mitotic clock, telomerase activity may stimulate cell division. The expression of telomerase may thus provide a selective advantage that allows tumor cells to divide indefinitely. Current understanding of telomere biology is still modest but as it improves telomerase may indeed become an important potential target for cancer chemotherapy.

Allsopp, R. C., Vaziri, H., Patterson, C. et al. Telomere length predicts replicative capacity of human fibroblasts. *Proc. Natl. Acad. Sci. USA* 89:10114, 1992; and Counter, C. M., Hirte, H. W., Bacchetti, S., and Harley, C. B. Telomerase activity in human ovarian carcinoma. *Proc. Natl. Acad. Sci. USA* 9:2900, 1994.

depends on the location of the last Okazaki fragment synthesized. As a minimum, the daughter DNA synthesized would have an 8–12 base gap generated by removal of the RNA primer for the Okazaki fragment. Without intervention this gap would be continuously regenerated and accumulated during each subsequent round of replication because it cannot be filled by DNA polymerase that requires a primer. The products of DNA replication would become shorter relative to parental DNA, leading to the gradual loss of DNA at the ends of human chromosomes. Cell senescence in humans and other mammals may be related to this chromosomal shortening as described in Clin. Corr. 15.6. In human cells that carry information to daughter cells (gamete cells) and in the linear chromosomes of bacteria and viruses, however, the integrity of DNA during replication cannot be compromised. Maintenance of intact chromosomal

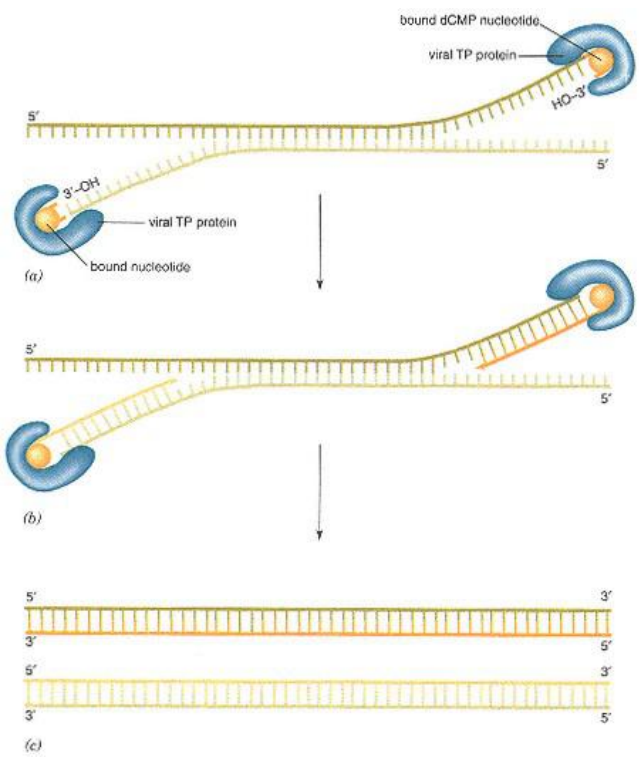


Figure 15.38
Replication of adenovirus DNA.
The adenovirus uses a protein as a primer, the terminal protein (TP), for synthesis of both strands of its DNA. TP, covalently associated with one dCMP, binds at the 3 end of each template chain and the dCMP residue provides a 3 -OH for DNA polymerase-catalyzed synthesis of a complementary strand. Since both strands of the viral DNA are synthesized continuously in the 5 3 direction, DNA synthesis is complete, leaving no gaps at the ends of the chromosome.
Redrawn based on figure in Wolfe, S. L. *Molecular and Cellular Biology*. Belmont, CA: Wadsworth, 1993.

structure requires a distinct mechanism for replication at the ends of DNA molecules.

Prokaryotic Replication

Different replication strategies have evolved to deal with the problem in viruses, plasmids, and organelle DNA. One approach is the use of a primer consisting of a protein, referred to as **terminal protein, TP,** that binds covalently to the 5 ends of viral DNA molecules via a phosphodiester bond with the hydroxyl group of a serine residue (Figure 15.38). Modified versions of TP that are distinct for different viruses also participate in replication. For instance, in the case of the mammalian adenovirus, the TP contains covalently bound dCMP. In bacteriophage 29, the bound nucleotide is dAMP. These nucleotides pair with the terminal nucleotides at the 3 end of each strand and serve as primers for replication. A special polymerase coded by each virus recognizes the TP and copies the strands unidirectionally from their 3 to 5 ends. With the priming limited to the ends of the parental DNA strands, both strands are replicated completely as if they both are leading strands. The TP molecule is cleaved from the primer nucleotide and it is released upon completion of the synthesis. Other viruses form circular intermediates that are copied by a rolling circle mechanism. Finally, some viruses, with identical sequences at the ends of their DNA, can hybridize their terminal sequences, forming linear repeats (linear concatenates). These concatenates are cleaved postreplicatively to generate progeny virus of the proper size (Figure 15.39).

Eukaryotic Replication:
Telomerases

Eukaryotes employ different strategies than prokaryotes and viruses for the replication of their chromosomal ends, known as **telomeres.** One approach that is used, albeit rarely, is the lengthening of chromosomal ends by the transposition of DNA segments known as **transposons.** This approach is apparently used for maintaining the chromosome ends in *Drosophila*. In most eukaryotes, however, telomere replication utilizes a specialized **reverse transcriptase enzyme** called **telomerase.** Telomerase activity depends on the presence of an RNA molecule that constitutes part of the telomerase structure and serves as an "internal" template. Maintenance of the chromosomal length depends on the action of telomerase on repetitive DNA sequences that constitute the telomeres of eukaryotic chromosomes (Figure 15.40). These telomeric tandem repeats can be several thousand nucleotides long and they consist of multiple copies of short G- and T-rich oligonucleotide sequences. Their size varies extensively from 20 bp in length for some protozoa to 150 kb in mouse telomers. For humans and other vertebrates the repetitive DNA is constructed with variants of the sequence TTAGGG. A short segment of single-stranded DNA ending in a 3 -OH group caps the end. Telomerase recognizes the G-rich single-strand at the 3 terminus and elongates it in the 5 3 direction, by adding telomere repeats at the end of the lagging chain. The RNA of telomerase, which has a sequence of about 150 nucleotides complementary to the telomer repeats, provides a movable template that substitutes for the absence of a normal DNA template. Telomerase provides in one package all that is needed for elongation of the strand that ends in a 3 terminus, namely, both template and enzymic activity. Extension of the telomeric sequence elongates the 3 end of DNA by about 100 nucleotides. This is then used as template for synthesis of the complementary strand by DNA polymerase α. Telomerase is then repositioned to repeat the process as illustrated in Figure 15.40. In this manner telomerase and polymerase α serve to maintain chromosomal length during repeated rounds of DNA replication. Maintenance is affected by such factors as telomerase processivity and its frequency of action on telomers as well as the rate of degradation of telomeric DNA. Telomeres may grow, shrink, or stay fairly stable depending

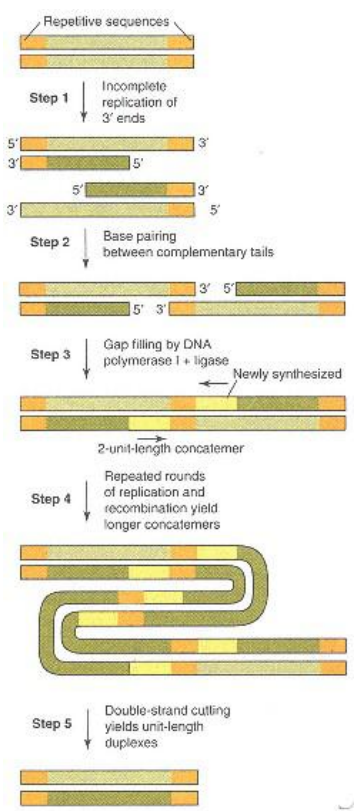


Figure 15.39
Replication of bacteriophage T7 DNA.
Bacteriophage T7 DNA has repetitive identical sequences at its chromosomal termini so that, following replication, the daughter molecules can hybridize end to end to form dimers. During subsequent rounds of replication the process is repeated until a large linear DNA, a concatenate, is formed. A specific nuclease then cleaves the large concatenate into fully replicated genome-size DNA segments.
Redrawn based on figure in
Mathews, C. K. and Van Holde, K. E. *Biochemistry*. Redwood City, CA: Benjamin/Cummings, 1990.

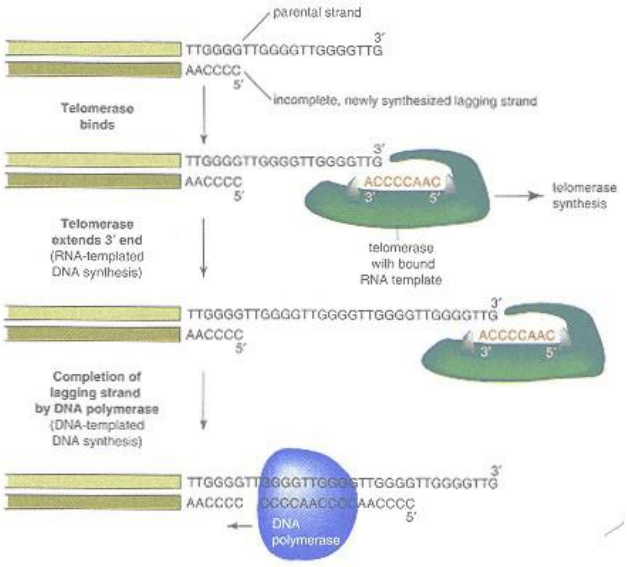


Figure 15.40
Telomere replication.
Telomerase contains an RNA template that codes for the extension of the ends of chromosomes and serves as a template for DNA polymerase. The DNA strand made on the lagging side of a replication fork of a linear chromosome is incomplete. For this strand to be completed, telomerase extends the 3 end on the complementary strand at the leading side of the fork. Telomerase first binds to a TG primer at the 3 end of this DNA strand. Binding is the result of base pairing between primer and RNA template that is part of the telomerase complex. The enzyme adds more T and G residues to the primer and repositions the RNA template so that more TG repeats can be added to the end of the primer. The extended primer is eventually recognized by DNA polymerase α, which proceeds to replicate the 5 end of the DNA using the single-stranded 3 end as template. Primase activity is vested in a subunit of DNA polymerase α.
Redrawn based on figure in Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. *Molecular Biology of the Cell*. New York: Garland, 1994.

on genetic or nutritional changes. For example, the size of yeast telomeres can vary from about 200 to 400 bp depending on conditions.

DNA Can Be Synthesized Using an RNA Template

For many years it had been assumed without reservation that the only direction in which genetic information can flow is from DNA to RNA. This dogma had to be revised, however, when it was discovered that the genomes of certain viruses, such as the **retroviruses,** consist of RNA instead of DNA and that during viral infection this genomic RNA is copied into DNA. The DNA that is obtained can either be transcribed to produce more viruses or it may be incorporated into the DNA of the host. In the latter case the viral genome is replicated along the DNA of the host and often remains latent for many host chromosome generations.

Enzymes that use RNA templates for DNA synthesis are called **reverse transcriptases.** Reverse transcriptases are often virally encoded but they are not limited to viruses. Enzymes with reverse transcriptase activities are also found in uninfected cells and are involved in the formation of pseudogenes and in the replication of transposable elements (see p. 669). Reverse transcriptases are the most error-prone type of DNA polymerases because they lack 3 5 exonuclease activities, thus lacking a proofreading function. Inhibitors of reverse transcriptase are used for the treatment of AIDS as described in Clin. Corr. 15.7.

DNA Replication, Repair, and Transcription Are Closely Coordinated

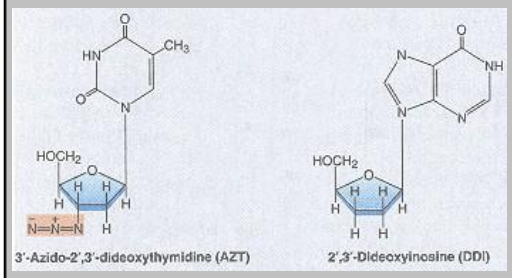
It has become increasingly clear that DNA replication, transcription, and repair are not separable, as most DNA lesions block both replication and transcription. Thus repair occurs with "expressed genes" as a priority, with the repair of dormant genes deferred. In addition, transcription and repair appear to cross paths at several points, with certain repair proteins participating in the activation

CLINICAL CORRELATION 15.7

Inhibitors of Reverse Transcriptase in Treatment of AIDS

AIDS is caused by a retrovirus, the human immunodeficiency virus (HIV). Treatment of AIDS is complicated by the high mutability of this virus, which reflects the low fidelity of the HIV reverse transcriptase responsible for the synthesis of the viral genome. This transcriptase is about one order of magnitude less accurate than other transcriptases and produces one or more mutations per generation, which means that any two HIV DNA molecules are almost never exactly the same in their nucleotide sequence.

The first drug that was used with some success, and continues in use, in controlling the rate of advancement of the disease is a structural analog of deoxythymidine, known as AZT.



This drug is converted to the triphosphate by cell kinases and the triphosphate is incorporated into the HIV genome in place of dTTP. AZT triphosphate competes successfully with dTTP for incorporation into the viral genome because of the higher binding affinity of AZT relative to dTTP toward the HIV reverse transcriptase. Since AZT has a lower affinity for cellular DNA polymerases than dTTP, it is not incorporated into cellular DNA. Incorporation of AZT triphosphate causes a premature termination of viral DNA synthesis because it lacks a 3 -OH site that is needed as the primer for incorporation of additional nucleotides.

Other nucleotide analogs, with similar reverse transcriptase-dependent mechanisms of actions, have been included in the treatment of AIDS. These include dideoxyinosine (ddI) dideoxycytidine (ddC), and azidothymidine (ZDV). Current approaches use ZDV or combination therapies of ZDV and ddI or ZDV and ddC. Other compounds that are not nucleotide analogs, referred to as nonnucleoside reverse transcriptase inhibitors (NNRTI), and a diverse group of other agents, such as protease inhibitors and HIV immune-based therapies, are currently under investigation for treatment of AIDS. A new class of drugs that inhibit proteases essential for HIV replication, when used in combination with reverse transcriptase inhibitors, is reported to reduce viral loads in AIDS patients to undetectable levels and in many instances reverse rather than simply arrest the symptoms of the disease.

Finkelstein, D. M., and Shoenfeld, D. A. (Eds.). *AIDS Clinical Trials*. New York: Wiley-Liss, 1995.

of initiation or elongation steps of transcription. For example, subunits of the TFIIH factor, which is essential for transcription, also participate in eukaryotic nucleotide excision repair. Repair and replication appear also to be coupled at the level of the protein factor, HSSB. This protein binds single-stranded DNA with high affinity during replication but it is also a repair protein required for the formation of the preincision complex. A protein induced as a result of DNA damage, the so-called Gadd45 protein, has regulatory effects on both DNA repair and replication. Gadd45 appears to both stimulate excision repair and inhibit DNA replication.

15.5— DNA Recombination

DNA recombination refers to a number of distinct processes during which genetic material is rearranged by breaking and joining portions of the same DNA molecule or portions of different DNA molecules. Recombination also takes place between the DNAs of different organisms to generate a new "composite" DNA. Both prokaryotic and eukaryotic DNAs undergo recombination. Three well-characterized processes listed in Table 15.8 fall under this general description of genetic recombination. Other DNA rearrangements have been noted whose mechanism and function are not well-understood and are referred to as illegitimate; these will not be reviewed in this chapter. Recombination creates new combinations of genes on the chromosome, which increase the chance of survival of a population. This increase of **genetic diversity** offers no advantage for individuals within a population. Individual survival partially

TABLE 15.8 Characteristics of Different Types of Genetic Recombination

Type	*Sequence Homology*	*Heteroduplex Sequences*	*Proteins Involved*	*DNA Synthesis*
Homologous	Extensive, but the homology is DNA sequence independent	Long	RecA, RecBCD, RuvAB, RuvC, and DNA repair enzymes[a]	Some
Site-specific	Short but specific DNA sequences are required on both DNAs	Short	Recombinases	Some
Transpositional	Homology is not required; specific sequences needed on one of the DNAs	None	Transposases	Minor (only to fill gaps)

[a] Several additional protein factors including RecE (exonuclease VIII), RecF, RecG, RecJ, RecN, RecOR, RecQ, RecT, SbcCD, DNA polymerase I, DNA gyrase, DNA topoisomerase I, DNA ligase, and DNA helicases participate in catalyzing homologous recombination.

depends, instead, on the operation of DNA repair. However, certain types of DNA repair depend on DNA recombination and therefore it is possible that recombination evolved as a mechanism of repair.

Homologous genetic recombination produces an exchange between a pair of distinct DNA molecules, often two slightly variant copies of the same chromosome, or two segments of DNA generated from the same DNA molecule. The main requirement for this process to occur is that the recombining DNAs are **homologous.** This means that the two DNAs share very similar base sequences over an extended region that may contain several thousand bases. An important example of homologous recombination in eukaryotes is the exchange of sections of homologous chromosomes during the early development of gametes (egg and sperm cells). In this manner slightly different versions of the same gene (alleles) can evolve during meiosis. Gene "mixing and reassortment" by general recombination is also widespread in bacteria. Homologous recombination is quite complex and involves a multistep mechanism catalyzed by a large number of different proteins. Prominent among them is the RecA protein, which also participates in SOS DNA repair.

Conservative site-specific recombination or **site-specific recombination** requires the presence of only short homologous DNA sequences. However, site-specific recombinations occur only in specific DNA sequences present in both the participating DNA molecules. The process is catalyzed by enzymes known as **recombinases.**

Transpositional site-specific recombination, or simply **transposition,** differs from conservative site-specific recombination in that it does not require a specific DNA sequence in the "target" chromosome. Transposition is catalyzed by **transposases**. Both transposases and recombinases recognize and act on specific DNA sequences. Recombination of either type is responsible for the insertion of viruses, plasmids, and **transposable elements (transposons)** into chromosomal DNA. Transposons are DNA elements that can move from location to location within a genome, in both bacteria and eukaryotes. Viruses are related to plasmids and transposons but also differ from these genetic elements in that viruses can synthesize a protein coat that allows them more host-independent existence. Plasmids and transposons are confined to replicate only within a specific cell and the progeny of that cell.

The most common recombination is the **homologous** type. **Site-specific recombination** and **transposition** are relatively rare, but important, events in that they may control replicative function in some viruses and certain aspects of development. Homologous recombination generates new combinations of genes that can lead to genetic diversity. DNA mutation and recombination are

the two principal approaches by which the cell creates variation that is required for evolution to occur. In addition, recombination events are involved in DNA repair. In those instances in which DNA damage occurs across complementary DNA sites, DNA repair can occur only through recombination. A large variety of protein structures used by the human immune system are produced by recombination as described in Clin. Corr. 15.8.

Homologous Recombination

Homologous recombination, which is accompanied by the formation of a **heteroduplex DNA** region, clearly requires breaking and rejoining of chromosomal DNA. Recombination occurs via a fairly complex multistep mechanism. A scheme that explains the outcome of recombination is shown in Figure 15.41. This scheme gives a minimal overview of recombination, in that each of the steps shown may represent more than one enzymatically catalyzed process. Numerous gene products are involved in homologous recombination.

Recombination may begin by introduction of a single-strand nick at a selected site of one of the DNA duplexes undergoing recombination. The resulting 3 -ended single-strand tail can then invade a homologous DNA duplex. Homologous DNA duplexes are chromosomes with the same linear arrangement of genes but with base sequences that may differ between the two duplexes. The variance is usually minor and may consist of no more than one different base among the millions of base pairs present in the chromosome. Single-strand invasion places the homologous DNA duplexes side by side in a process referred to as **synapsis.** Synapsis does not necessarily involve contacts between homologous sequences and further movement of the DNAs with respect to each other may be necessary until homologous sequences come into contact. This process is referred to as **homologous alignment. Strand invasion** is accompanied by **strand displacement** in the homologous DNA duplex resulting in the formation of a so-called **D-loop.** The "D-loop" strand that has been displaced by strand invasion is now nicked and it pairs with its complementary strand in the original duplex. The ends of exchanged strands are then ligated to form a stable **cross-stranded intermediate** known as **Holliday junction.** The junction can migrate in either direction by unwinding and rewinding of the two

CLINICAL CORRELATION 15.8

Immunoglobulin Genes Are Assembled by Recombination

Immunoglobulins (antibodies) are molecules that recognize and specifically bind to any substance that antibodies identify as foreign to the human body (see p. 88 for details). Because of the immense variety of infectious agents, including millions of microorganisms that are present in the environment, the human genome, which is equipped with only a limited pool of probably no more than 100,000 genes, does not have the capacity to directly produce an equivalent number of different antibodies necessary for specific recognition of all infectious agents. This inherent limitation in the gene-coding potential of the human genome is, however, overcome by recombination, which allows production, from a limited amount of gene-coding DNA, of an almost unlimited number of distinct antibodies.

Human immunoglobulins consist of two heavy and two light chains with each chain having a variable region, with a sequence that is characteristic for each immunoglobulin, and a chain with constant amino acid sequence (see p. 89). Recombination leads to diversity in the variable region of immunoglobulins. During the maturation of a bone marrow stem cell into a B lymphocyte, one V segment and one J segment are brought together by site-specific recombination. In the process the intervening DNA is deleted and a joint between the two regions is established by an RNA-splicing reaction that occurs following transcription. Since the V region consists of 300 segments and the J region of 4, at least 1200 different combinations can be generated by recombination.

Similar considerations apply to the light chains and the heavy chains, with the latter being assembled in as many as 5000 distinct combinations. Because individual light and individual heavy chains can subsequently be assembled in combination, at least 6×10^6 different IgG molecules can be produced. Furthermore, because some variations occur in the exact location of the V-J junction, the actual number of IgG molecules is two to three times higher than estimated above. Additional IgG diversity is produced during the process of maturation of B lymphocytes by mutational processes.

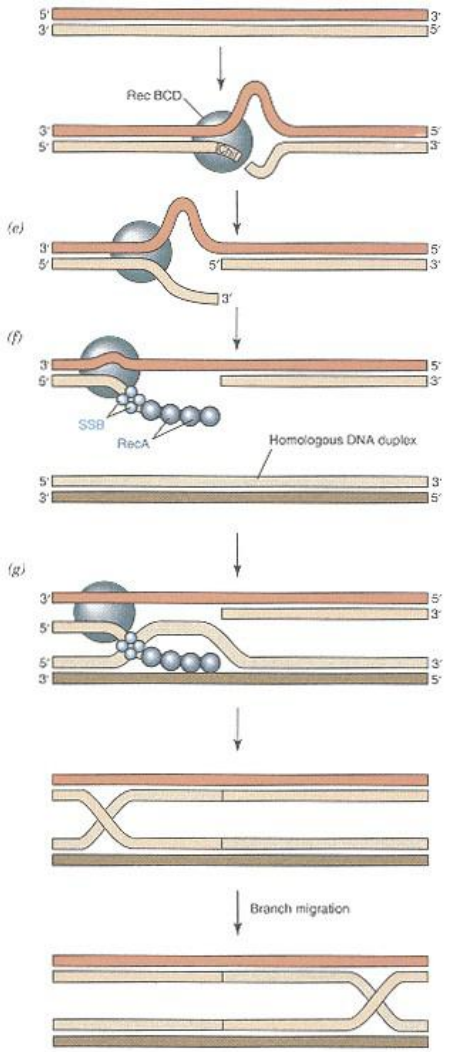


Figure 15.41
Overview of homologous recombination.
Transformations that can lead to formation of recombinant and nonre-combinant heteroduplexes, by participation of two homologous DNA molecules in homologous recombination are outlined. Each step indicated need not be the outcome of a single, enzymatically catalyzed or well-understood reaction. The sequence of steps shown is not necessarily universally applicable.

duplexes to produce a further exchange of single strands between interacting chromosomes. This process, known as **branch migration,** results in strand exchange and it produces **heteroduplex** regions of varying lengths. The resulting heteroduplex, shown in Figure 15.41, can also be presented in another form that is generated by merely pulling the ends of the heteroduplex together (Figure 15.42). A twist of this structure produces an isomeric heteroduplex, which is called the **Chi** form. In order to resolve the Chi form two additional single-strand nicks can be made, in either the horizontal direction or vertical

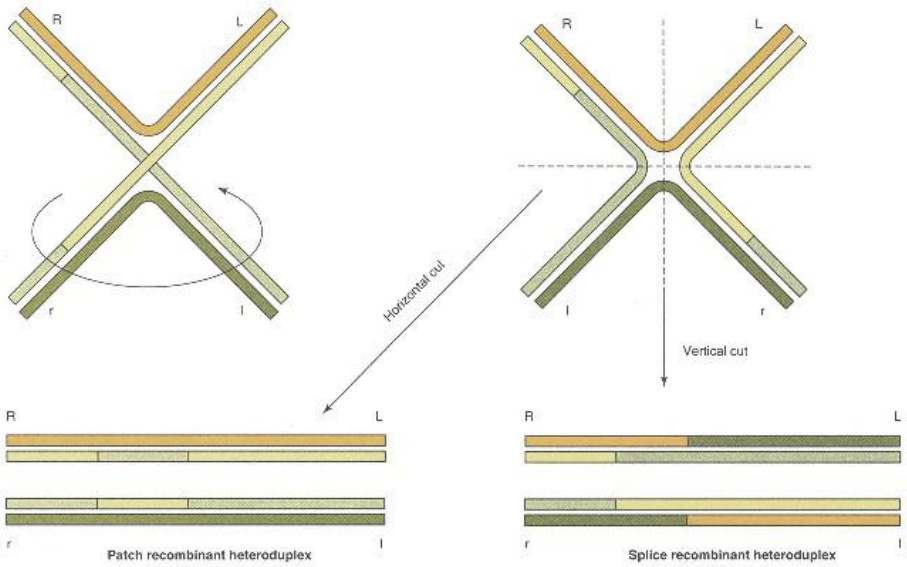


Figure 15.42
Patch and splice recombinant heteroduplexes.

direction, leading to two distinct products. Gaps present in these structures are repaired and ligated, leading to either one of the two products. The manner in which nicks are introduced in the horizontal and vertical directions is fundamentally different. In one case (horizontal direction) nicks are introduced again into the strands that were initially nicked, although at different sites, producing two duplexes in which one strand of each remains intact. These duplexes contain heteroduplex regions, generated by branch migration, that are misleadingly referred to as **Patch recombinant heteroduplexes.** These duplexes contain the same genes and in the same linear order as the initial duplexes. In vertical direction nicks, the complementary strands that previously were left intact are nicked again (though at different sites), producing two duplexes of true recombinant DNA, referred to as **splice recombinant heteroduplexes.** In these true recombinant heteroduplexes the linear order of DNA sequences contained in the original duplexes is clearly rearranged.

Support for this multistep recombination scheme has accumulated over the years based on genetic investigations, on electron microscopy of Holliday junctions, and by isolation of proteins and enzymes that can catalyze many of the transformations described in this recombination scheme.

Enzymes and Proteins That Catalyze Homologous Recombination

Homologous recombination in *E. coli* requires about 25 enzymes for recombination. A partial list includes **RecA** protein, **RecBCD** enzyme (which is the product of three distinct *E. coli* genes, *recB, recC*, and *recD*), **RuvAB** and **RuvC** proteins, DNA polymerase I, DNA gyrase, DNA topoisomerase I, DNA ligase, and DNA helicases (Table 15.8). Proteins homologous to RecA have also been isolated from yeast and human cells.

Homologous recombination in *E. coli* begins with RecBCD, which is a site-specific **endonuclease** and an ATP-dependent **helicase** (Figure 15.43).

RecBCD can initiate recombination by unwinding DNA and, on occasion, cleaving one strand. The enzyme binds to one end of linear DNA and travels along the helix at the expense of ATP, unwinding DNA as it moves and rewinding DNA behind it at a slower rate than unwinding. This produces a "bubble" consisting of two single-stranded loops that propagate on the DNA with the advance of the RecBCD. *Escherichia coli* DNA is characterized by the presence of about 1000 copies of the sequence 5 -GGTGGTGG-3 that, on average, occurs at intervals of 4–5 kb. These Chi sites are **"hot spots" for recombination** as they increase the frequency of recombination. When the advancing RecBCD encounters a Chi site within a "bubble," it cleaves the DNA strand that incorporates the 5 -GGTGGTGG-3 sequences 5–6 nucleotides to the 3 side of the Chi site. The helicase activity generates a 3 single-stranded tail of DNA that is progressively lengthened to several kilobases.

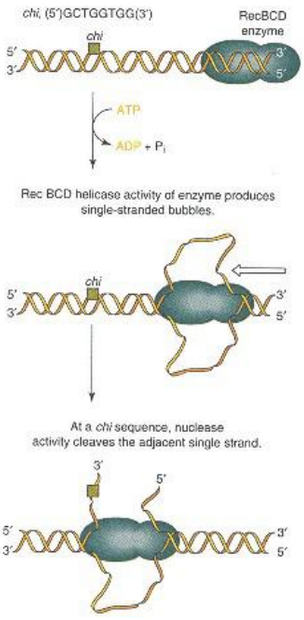


Figure 15.43
Activities of RecBCD protein.
RecBCD combines helicase and nuclease activities and appears to be involved in initiation of homologous genetic recombination in *E. coli*. RecBCD, using its helicase activity, enters the double helix and, using energy derived from ATP hydrolysis, travels along the helix until it encounters a Chi site, which consists of the sequence 5 -GCTGGTGG-3 . RecBCD introduces a cut, within the Chi site, that leads to displacement of a 3 -terminating single strand. This single strand initiates recombination by pairing with a homologous DNA double helix.
Redrawn based on figure in Liehninger, A. L., Nelson D. L., and Cox, M. M. *Principles of Biochemistry*. New York: Worth, 1993.

This growing single-stranded tail can then initiate the **strand invasion** process with the assistance of RecA, which catalyzes a multiplicity of reactions in DNA recombination (Figure 15.41). RecA interacts with single-stranded (ss) and double-stranded (ds) DNA and catalyzes **pairing of homologous DNA sequences, invasion of ssDNA** into the homologous double helix, formation of the **Holliday junction,** and migration of this junction (branch migration). These activities of RecA depend on the presence of a RecA site that recognizes ssDNA and promotes the cooperative binding of the protein to ssDNA. Formation of a long and relatively stiff **nucleofilament** (Figure 15.44) prevents the

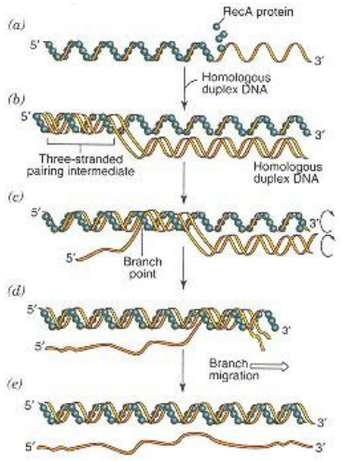


Figure 15.44
DNA strand exchange mediated by RecA.
Replacement of a complementary strand in a DNA duplex by a single-stranded DNA is catalyzed by RecA. RecA begins the exchange by coating both ssDNA and dsDNA by RecA (only coating of the single strand is shown). The coating modifies the conformation of both the single-stranded and double-stranded polynucleotides and catalyzes the invasion of the single-stranded intermediate. Switches in the base pairing between the strands, and the accompanying rotation of the DNA, move the three-stranded region from left to right as one strand of the DNA duplex is displaced by the identical, or nearly identical, invading ssDNA. Continuing branch migration leads to eventual separation of the displaced strand.

ssDNA tail from reassociating with the complementary strand within the DNA duplex, from which it originated, and prepares the single strand for invasion. In the resulting nucleofilament that binds one RecA molecule per 3 bases, the polynucleotide is positioned within a deep groove of the RecA protein. A second site on RecA recognizes and binds preferentially to dsDNA. In this nucleofilament each RecA monomer covers six nucleotides and each successive monomer binds to the opposite site of the DNA helix. For the sake of simplicity the dsDNA in Figure 15.44 is shown as free from RecA. The RecA–ssDNA and RecA–dsDNA nucleofilaments differ in their geometry from B-DNA, but both filaments represent partially unwound and unstacked helical structures that are extended lengthwise by 50% relative to B-DNA. DNA unwinding in the RecA–dsDNA nucleofilament (to about 18.6 bp per turn) exposes H-bond donors and acceptors in the major groove of the double helix, making them available for interaction with the ssDNA–RecA filament. Thus RecA contributes to the recognition of regions of homology between DNA strands. Once **homologous alignment** is established, a fairly stable **triple-stranded intermediate** can be formed (Figure 15.45). In this structure the third strand is in contact with the major groove of the duplex, aligned in a manner that permits RecA to flip the base pairing of the two identical strands.

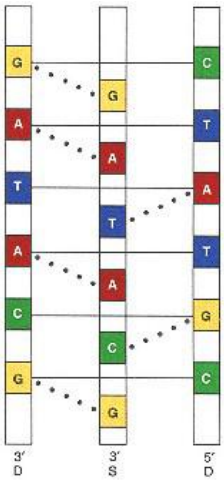


Figure 15.45
Model for the triple-stranded intermediate formed during DNA recombination.
RecA catalyzes the formation of a triple-stranded DNA intermediate as a result of the association of a dsDNA, the strands of which are marked D, and an invading ssDNA, marked S (shown in the middle). Both dsDNA and ssDNA are present in the form of complexes with RecA. This protein catalyzes unwinding of the strands of the double helix and makes the matrix of hydrogen-bond donors and acceptors in the major groove of the double helix available for pairing with ssDNA. The ssDNA is also unwound by RecA, providing for proper alignment between dsDNA and ssDNA.

The flipping of the base pairing and the resulting invasion of the RecA–ssDNA filament involve the exchange of two identical (or nearly identical) strands between helical structures, which therefore requires an ordered rotation of two aligned strands. The polynucleotides are prepared for this exchange by "the extended" conformation generated by RecA. Strand exchange can be extended by **branch migration,** which means that progression of the exchange requires both invasion and branch migration. Branch migration may be described as a process in which an unpaired region of a single DNA strand displaces a DNA strand from a region of homologous dsDNA and moves the branching point, without appreciably increasing the total number of disrupted base pairs. Migration is achieved by RecA-catalyzed rotation of RecA-bound DNA strands involved in the exchange (Figure 15.44). The resulting "spooling" action, in which topoisomerases may be involved, moves the branch as ATP is hydrolyzed.

Branch migration also occurs at the Holliday junction that is subsequently formed. In this intermediate homologous DNA helices that were initially paired are held together by mutual exchange of two of the four strands (Figure 15.46). Stereochemistry of the intermediate is determined by the juxtaposition of the grooves and the phosphate backbones of the participating helices, and the point of exchange or actual junction can be moved back and forth along the helices. Migration of the junction can proceed in the absence of RecA. This RecA-independent migration of the junction is catalyzed by a complex of **RuvA** and **RuvB.** RuvA binds to the junction and acts as a specificity factor that targets RuvB, which is an ATPase, to the junction. The RuvAB complex promotes migration and increases the length of the heteroduplex DNA at the expense of ATP. Finally, the Holliday junction is recognized and resolved into products by the **RuvC endonuclease,** a dimer of 19-kDa subunits related to each other by a dyad axis of symmetry. The catalytic center of this **resolvase** lies at the bottom of a cleft that fits a DNA duplex. Only strands with the same polarity are cleaved and produce two types of heteroduplex molecules, one type in which only single-strand segments are exchanged **(patch recombinants)** and another type, a true recombinant, in which the ends of molecules have been exchanged **(splice recombinants).** Resolution is completed by DNA polymerase I, DNA topoisomerase I, DNA gyrase, and DNA ligase.

RecA also exhibits a highly specific protease activity that is activated by unpaired DNA strands and is directed at specific regulatory proteins. Thus RecA has unique properties for coordinating regulation of a number of cellular

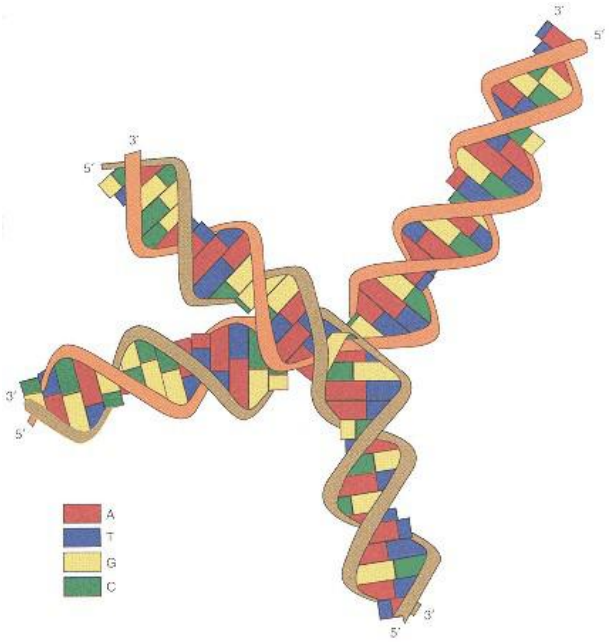


Figure 15.46
Structure of the Holliday junction.
The Holliday intermediate is a four-way junction that adopts a right-handed antiparallel X-shaped structure by pairwise coaxial stacking of the two double helices. The junction consists of fully stacked base pairs with the participating strands present as hydrogen-bonded DNA duplexes.
Redrawn based on figure in Moran, L. A., Scrimgeour, K. G., Horton, H. R., Ochs, R. S., and Rawn, J. D. *Biochemistry*. Englewood Cliffs, NJ: Neil Patterson/Prentice Hall, 1994.

functions that occur when DNA damage, or the interruption of DNA replication, leads to the production of ssDNA segments. An example is the postreplication repair of DNA damaged by UV light or other mutagens.

Site-Specific Recombination

This process separates and joins dsDNA molecules at specific sites. Site-specific recombination is limited to select regions of a genome and is driven by **recombinases** that recognize short (20–200 bp) specific sequences on both recombination sites. When recombinase binds to both recombination sites on DNA molecules it can produce an insertion of DNA. A well-studied example is provided by the integration of so-called temperate phages, such as *E. coli* bacteriophage , into the host chromosome of the corresponding host (Figure 15.47). The circular chromosome becomes integrated into a specific site in the *E. coli* chromosome consisting of about 20 nucleotides, the so-called attP site. Integration requires the alignment of the phage in a specific orientation with the *E. coli* chromosome. The alignment is achieved by a specific recombinase known as **integrase (Int)** and the participation of a protein known as the **integration host factor (IHF)** encoded by the bacterium. Integrase brings together the attB site of the bacterium with a corresponding specific site on the phage chromosome, which consists of 230 bp and is known as the attP site. Int generates a precise wrapping of DNA to juxtapose specific nucleotide sequences for the splicing reactions that follow. Functioning as a topoisomerase, Int unwinds the attP region and forms an Int–attP nucleoprotein. A corresponding nucleoprotein is also formed between Int and attB that brings the attP and attB

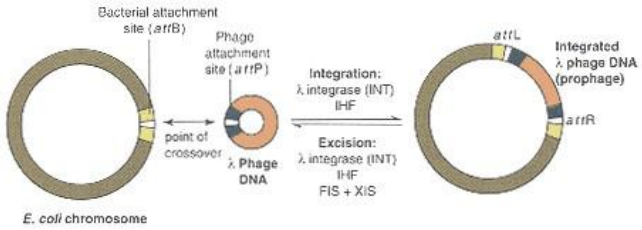


Figure 15.47
Site-specific recombination of λ phage.
Site-specific recombination is carried out by integrase. The phage chromosome undergoes recombination between the attP site and a corresponding site on the bacterium, attB. Integration of the phage chromosome generates two new attachment sites (attR and attL) that flank the integrated phage DNA. The reverse reaction, excision of the integrated phage chromosome, requires the participation of protein XIS produced by the bacteriophage and protein FIS encoded by the bacterium.

sites together. Integrase then generates a staggered cut, 7 base pairs apart within a core sequence of 15 bp, that is present in both the attP and attB sites and catalyzes the exchange of strands at the position of the cut to form a **Holliday intermediate.** To complete the exchange, cutting and rejoining must be repeated at a second point within each of the two recombination sites. Normally, limited branch migration is required prior to an Int-catalyzed second cleavage and strand exchange. Following ligation by Int, the original sequence of the recombination site is regenerated but the DNA on either side of the site is recombined. **Recombinases** often act in a reversible manner, restoring the sequences of original DNAs. **Integrase** also acts in a reversible manner so that the circular phage chromosome can be excised as conditions change. The forward and reverse steps of the integration reaction are separately regulated, with the reverse step being dependent on the presence of additional proteins: the XIS protein encoded by the phage and FIS encoded by the bacterium. Both reactions also require IHF.

Transposition

Transposition is a form of recombination catalyzed by recombinases called **transposases.** This type of recombination is best understood in bacteria but DNA of all cells, including eukaryotes such as *Drosophila,* maize, and yeast, contains segments that can move, generally with very low frequencies of 10^{-5}–10^{-7} per cell generation, from a **donor** site to another **target** site within a chromosome. These segments are known as **transposable elements (transposons).**

Transposition differs from homologous recombination in not requiring sequence homology between donor and target sites. Only the donor site, that is, the transposon, has specific nucleotide sequences located on both sides of the transposon that serve as binding sites for transposases. Most bacterial transposons have short repeats of about 15–25 bp at the two ends of the transposable DNA segment. In contrast, the target sites are not well defined and are not characterized by specific DNA sequences. **Heteroduplex** joints are not formed as a result of transposition.

Three classes, I, II, and III, of transposable elements are recognized. Class I transposons are called **insertion sequences (IS)** if they consist of a gene

CLINICAL CORRELATION 15.9

Transposons and Development of Antibiotic Resistance

Genes conferring to bacteria resistance to commonly used antibiotics such as pencillin or tetracycline are usually carried on plasmids. The DNA sequences of these plasmids do not have any homology with the chromosomal DNA sequences of the host. Yet, as a result of transposition, antibiotic resistance genes can be transferred to the chromosome of bacterial hosts. The existence of genes that can move from one chromosome to another is of course of great importance in understanding the factors that produce changes in the organization of genomes. From the clinical standpoint these "transposable" genes are of critical significance for understanding how populations of antibiotic-resistant bacteria arise with use of antibiotics in the treatment of bacterial infections in humans and animals.

coding for transposase together, of course, with the repeats that normally flank the transposable element. IS elements vary in size between 800 and 1300 bp. When Class I transposons also contain an additional gene, such as a gene conferring antibiotic resistance to bacteria, they are called **composite transposons (Tn).** Class II transposons differ from Class I in that, in addition, they code for the gene of a second enzyme, **resolvase.** Typically, **composite transposons** and Class II transposons are several thousand base pairs long. Finally, a small group of bacteriophages, such as bacteriophage Mu, that insert their chromosome into a host chromosome are classified as Class III transposable elements.

Transposition begins by a transposase-catalyzed introduction of a staggered cut at the target DNA sequence. Cuts are also made on each side of the transposon so that it can be moved onto the target site. The relocation leaves a double-stranded break at the site from which the transposon is excised. At the target site the transposon is spliced into the staggered cut as shown in Figure 15.48. Specifically, 3–12 bp at the target site are duplicated by DNA polymerase I, to form an additional short repeat at each end of the inserted transposon, and the "tailored" transposon then is ligated within the target site. In Class II and III transposition, in addition to duplication of the short repeats, the transposon itself is replicated and one copy of it remains at the donor site while the other copy is transferred to the target site. This type of transposition, referred to as **replicative transposition,** requires the enzyme resolvase and therefore does not occur in Class I transposition. Replicative transposition can reshape the structure of a chromosome beyond the simple act of relocating a transportable element from one site to another. Because this type of transposition places two homologous sequences within the same chromosome, homologous recombination between these two sequences can produce either a deletion or an insertion, depending on whether these sequences are oriented in the same or in opposite directions, as shown in Figure 15.49.

Finally, transposition may inactivate a gene by mutation if a transposon is inserted into a coding sequence and interrupts it. Alternatively, insertion by transposition of a promoter or a transcriptional activator next to a gene may activate the gene. Clinical Correlation 15.9 reviews the role of transposition and Clin. Corr. 15.10 the role of DNA amplification in the development of drug resistance.

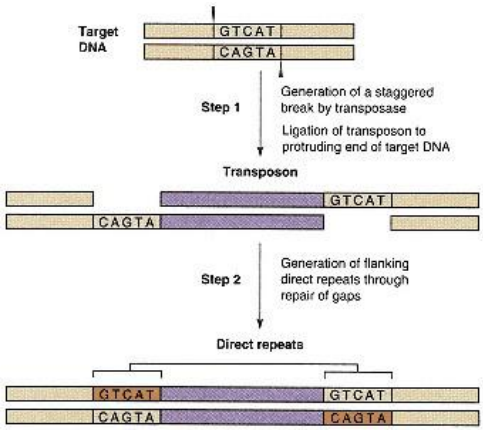


Figure 15.48
Direct repeats at the ends of transposons.
Transposons are inserted into gaps generated at a target sequence by introduction of a staggered cut by a transposase. Ligation of transposon to the protruding ends of target DNA leaves gaps at both sides of the transposon. Repair of these gaps is responsible for the presence of direct repeats that flank transposons.
Redrawn based on figure from Mathews, C. K. and Van Holde, K. E. *Biochemistry*. Redwood City, CA: Benjamin/Cummings, 1990.

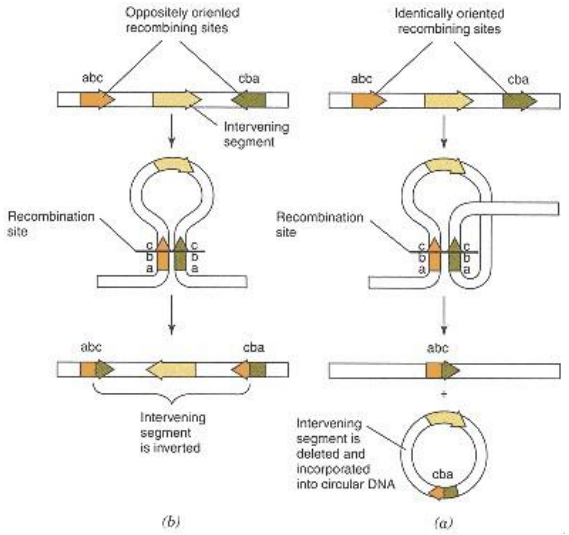


Figure 15.49
Genomic rearrangements promoted by transposons.
In replicative transposition a transposable element is replicated, with one copy of it remaining at the donor site and a new copy becoming inserted within a different location. This produces two homologous sequences within the same chromosome that can subsequently enter into homologous recombination.
(a) When these homologous sequences are of the same polarity, recombination can yield a deletion of DNA by a process that is superficially analogous, but certainly not similar, to the reverse reaction occurring in site-specific recombination.
(b) Inversion of DNA flanked by these transposons can result when transposable elements are present in the chromosome oriented in opposite direction.
Redrawn based on figure from Mathews, C. K. and Van Holde, K. E. *Biochemistry*. Redwood City, CA: Benjamin/Cummings, 1990.

15.6— Sequencing of Nucleotides in DNA

Restriction Maps Give the Sequence of Segments of DNA

The sequences of many genes and adjoining DNA segments have been determined for bacteria, viruses, plants, and humans. The determination of the sequence of a large DNA molecule begins by cutting the DNA into pieces of a more manageable size with appropriate restriction endonucleases. **Restriction digests** permit the construction of a characteristic **restriction map** for each DNA. One protocol depends on the generation of partial restriction digests of end-labeled DNA. Partial digests are obtained by setting the conditions so that

CLINICAL CORRELATION 15.10

DNA Amplification and Development of Drug Resistance

An important limitation in the effectiveness of chemotoxic drugs in the treatment of cancer is the development of drug resistance. Thus cancer cells become resistant to methotrexate (see p. 520), an inhibitor of dihydrofolate reductase (DHFR). Drug resistance in cultured cells results from the specific amplification of a large DNA segment that incorporates the *DHFR* gene but the exact mechanism by which amplification occurs is not clear. It appears likely that amplification results from recombination of identically oriented homologous sequences that flank the amplified DNA. Amplification can occur by tandem duplication of DNA that contains the *DHFR* gene or alternatively the DHFR-containing segment can be excised (apparently by a recombination process), producing extrachromosomal DNA (minichromosomes). The two mechanisms of *DHFR* gene amplification are not mutually exclusive and, in fact, some resistant cells contain both types of amplified DNA.

Gene amplification is gradually reversed in the absence of methotrexate, first with the disappearance of the extrachromosomal copies. Chromosomally amplified genes, however, persist for several generations after removal of the drug. The amplification of genes is a general phenomenon not limited to methotrexate or the development of cell resistance toward other drugs. In fact, gene amplification and the accompanying resistance extend to areas well beyond clinical medicine, as, for instance, in agriculture with the development of pesticide-resistant insects.

CLINICAL CORRELATION 15.11

Nucleotide Sequence of the Human Genome

The purpose of the Human Genome Project is to provide a detailed map of the human genome and establish what DNA sequences determine human phenotypic characteristics and guide human development. A corollary to this goal is to identify genes responsible for human disease so that new approaches can be developed for diagnosis, prevention, and therapy.

The human genome is believed to consist of 70,000–100,000 different genes that determine the genetic characteristics of every cell in the human individual. The human genome consists of about three billion base-paired nucleotides that are assembled in the form of 23 pairs of chromosomes. The availability of restriction endonucleases and the development of effective physical mapping procedures for DNA, combined with the increasing rapidity of contemporary nucleotide sequencing methods, have provided strong impetus for the very ambitious undertaking of determining the nucleotide sequence of the entire human genome.

Extensive physical mapping has been completed. In addition, genetic mapping seeks to locate over 500 known genetic markers on the human chromosomes. Cumulatively over 150 million base pair sequences, representing parts of the chromosome sequences of both human DNA and that of other organisms, have been determined. Also, the sequences of certain continuous stretches of DNA, ranging from one million to several million base pairs in length, are being determined. Considering that the size of different human chromosomes varies from 263 million to less than 50 million base pairs, the determination to date of a total of about 150 million base pairs represents an important accomplishment. It is conservatively estimated that complete sequencing of the genome will take more than a decade and a half.

Because of the routine nature of determining the nucleotide sequences involved, many scientists have questioned the wisdom of diverting resources from perhaps more creative scientific endeavor, to the effort required to sequence the human genome. Others have pointed out that the project is fraught with technical uncertainties. Proponents point out the great potential benefits of determining the imprint that controls the genetic properties of the human cell at the highest possible level of resolution. Presently, as many as 4000 genetic diseases have been identified and many of them, namely, those inherited in Mendelian fashion, are caused by a single mutant gene. Searching for the imprint of human disease at the level of nucleotide sequences may permit understanding of all disease states at the genomic level. Determination of the complete sequence appears to be one of the prerequisites for understanding human disease at the molecular level. There is little doubt that the sequencing of the human genome will present us with many new challenges and opportunities in medicine.

Grant Cooper, N. (Ed.). *The Human Genome Project*. Mill Valley, CA: University Science Books, 1994.

the restriction endonuclease will not recognize all sites in every DNA molecule but will instead produce a digest that includes a collection of partial fragments. Double-stranded DNA is end-labeled by treatment with alkaline phosphatase, which removes the phosphate residue at the 5 end, and then γ-labeled with [^{32}P]ATP and a polynucleotide kinase, which incorporates the ^{32}P into the two 5 termini of the DNA strands. Alternately, the ^{32}P-label can be introduced at the 3 termini by the incorporation of ^{32}P-labeled deoxyribonucleotide triphosphates using DNA polymerase. End-labeling allows for each fragment to be identified on an electrophoresis gel. The details of this procedure are presented on page 762. Thus, with a series of different site cuts, the fragments can be mapped directly relative to the labeled end. Restriction maps are used for characterization of various DNAs and for ordering of smaller DNA fragments within a particular DNA sequence. Such ordering is essential before the nucleotide sequence of large DNA molecules can be determined.

Several methods have been developed for rapid sequencing of large poly-deoxyribonucleotides. They are impressively accurate. Digests obtained using different restriction enzymes produce segments with overlapping lengths of nucleotide sequences. The accuracy of sequencing methods are increased by sequencing the complementary strand. These procedures can also be used for sequencing of RNA molecules by prior conversion of the RNA sequence to a complementary DNA by use of reverse transcriptase. Sequences up to 500 bp can be determined in a single automated operation and stretches of 10,000 bp, which correspond to the average length of a gene, are now routinely determined. Clinical Correlation 15.11 discusses the application of these procedures for obtaining the sequence of the human genome.

Bibliography

Adams, R. L. P. *DNA Replication*. Oxford, England: Oxford University Press; New York: IRL Press, 1991.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. *Molecular Biology of the Cell*, 3rd ed. New York: Garland, 1994.

Bardwell, A. J., Bardwell, L., Wang, Z., Siede, W., et al. Recent insights on DNA repair: the mechanism of damaged nucleotide excision in eukaryotes and its relationship to other cellular processes. *Ann. N.Y. Acad. Sci.* 726:281, 1994.

Bell, S. P., and Stillman, B. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* 357:128, 1992.

Bernstein, C. *Aging, Sex and DNA Repair*. San Diego, CA: Academic Press, 1991.

Bhattacharyya, A., Murchie, A. I., von Kitzing, E., Diekmann, S., Kemper, B., and Lilley, D. M. J. A model for the interaction of DNA junctions and resolving enzymes. *J. Mol. Biol.* 221:1191, 1991.

Binden, R. R. *DNA Structure and Function*. San Diego, CA: Academic Press, 1994.

Blackburn, E. H. Telomerases. *Annu. Rev. Biochem.* 61:113, 1992.

Bootsma, D., Weeda, G., Vermeulen, W., van Vuuren, H., et al. Nucleotide excision repair syndromes: molecular basis and clinical symptoms. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 347:75, 1995.

Calladine, C., and Drew, H. *Understanding DNA—The Molecule and How It Works*. San Diego, CA: Academic Press, 1992.

Calsou, P., and Salles, B. Properties of damage-dependent DNA incision by nucleotide excision repair in human cell-free extracts. *Nucleic Acids Res.* 22:4937, 1994.

Chiu, S. K., Rao, B. J., Story, R. M., and Radding, C. M. Interactions of three strands in joints made by RecA protein. *Biochemistry* 32:13146, 1993.

Dianov, G., and Lindahl, T. Reconstitution of the DNA base excision-repair pathway. *Curr. Biol.* 4:1069, 1994.

Duckett, D. R., Murchie, A. I., Giraud-Panis, M. J., Pohler, J. R., and Lilley, D. M. Structure of the four-way DNA junction and its interaction with proteins. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 347:27, 1995.

Echols, H., and Goodman, M. F. Fidelity mechanisms in DNA replication *Annu. Rev. Biochem.* 60:477, 1991.

Fangman, W. L., and Brewer, B. J. Activation of replication origins in yeast chromosomes. *Annu. Rev. Cell Biol.* 7:375, 1991.

Friedberg, E. C., Walker, G. C., and Siede, W. *DNA Repair and Mutagenesis*. Materials Park, OH: ASM Press, 1995.

Hanawalt, P. C. Transcription-coupled repair and human disease. *Science* 266:1957, 1994.

Hiasa, H., and Marians, K. J. Tus prevents overreplication of oriC plasmid DNA. *J. Biol. Chem.* 269:26959, 1994.

Horiuchi, T., Fujimura, Y., Nishitani, H., Kobayashi, T., and Hidaka, M. The DNA replication fork blocked at the Ter site may be an entrance for the RecBCD enzyme into duplex DNA. *J. Bacteriol.* 176:4656, 1994.

Huang, J.-C., Svoboda, D. L., Reardon, J. T., and Sancar, A. Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5 and the 6th phosphodiester bond 3 to the photodimer. *Proc. Natl. Acad. Sci. USA* 89:3664, 1992.

Joyce, C. M., and Steitz, T. A. Function and structure relationships in DNA polymerases. *Annu. Rev. Biochem.* 63:777, 1994.

Kamenetskii, F. *Unraveling DNA*. New York: VCH Publishers, 1993.

Kim, M. G., Zhurkin, V. B., Jernigan, R. L., and Camerini-Otero, R. D. Probing the structure of a putative intermediate in homologous recombination: the third strand in the parallel DNA triplex is in contact with the major groove of the duplex. *J. Mol. Biol.* 247:874, 1995.

Kim, S. T., and Sancar, A. Photochemistry, photophysics, and mechanism of pyrimidine dimer repair by DNA photolyase. *Photochem. Photobiol.* 57:895, 1993.

Kornberg, A. *DNA Replication*, 4th ed. New York: Wiley, 1992.

Landy, A. Dynamic, structural and regulating aspects of site-specific recombination. *Annu. Rev. Biochem.* 58:913, 1989.

Lewin, B. *Genes V*. New York: Oxford University Press, 1994.

Lindsten, J., and Pettersson, U. *Etiology of Human Disease at the DNA Level*. New York: Raven Press, 1991.

Luzio, J. P., and Thompson, R. J. In: *Molecular Medical Biochemistry*. New York: Cambridge University Press, 1990.

Marians, K.J. Prokaryotic DNA replication. *Annu. Rev. Biochem.* 61:673, 1992.

Matsunaga, T., Mu, D., Park, C.-H., Reardon, J. T., and Sancar, A. Human DNA repair excision nuclease. *J. Biol. Chem.* 270:20862, 1995.

Modrich, P. Mismatch repair, genetic stability, and cancer. *Science* 266:1959, 1994.

Parsons, C. A., Stasiak, A., Bennett, R. J., and West, B. C. Structure of a multisubunit complex that promotes DNA branch migration. *Nature* 374:375, 1995.

Pederson, D. S. *Transcription Factors and DNA Replication*. Austin, TX: R. B. Landes, 1994.

Richardson, C. C., and Lehman, I. R. *Molecular Mechanisms in DNA Replication and Recombination*. New York: Wiley, 1990.

Sancar, A. Mechanisms of DNA excision repair. Science 266:1954, 1994.

Schlotterer, C., and Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20:211, 1992.

Takahashi, M., and Nordien, B. Structure of RecA–DNA complex and mechanism of DNA strand exchange reaction in homologous recombination. *Adv. Biophys.* 30:1, 1994.

Vos, J.-M. H. *DNA Repair Mechanism: Impact on Human Disease and Cancer*. New York: Demos Vermande, 1995.

Wang, T. S. F. Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* 60:513, 1991.

Wolfe, S. L. *Molecular and Cellular Biology*. Belmont, CA: Wadsworth, 1993.

Yang, S. S., Fliakas-Boltz, V., Bader, J. P., and Buckheit, R. W. Jr. Characteristics of a group of nonnucleoside reverse transcriptase inhibitors with structural diversity and potent anti-human immunodeficiency virus activity. *Leukemia* 9:575, 1995.

Questions

C. N. Angstadt and J. Baggott

1. Which of the following statements about *E. coli* DNA polymerases is correct?

   A. All polymerases have both 3 5 and 5 3 exonuclease activity.

   B. The primary role of polymerase III is in DNA repair.

   C. Polymerases I and III require both a primer and a template.

   D. Polymerase I tends to remain bound to the template until a large number of nucleotides have been added.

   E. The specificity of the polymerase reaction is inherent in the nature of polymerases.

2. Proofreading activity to maintain the fidelity of DNA synthesis:

   A. occurs after the synthesis has been completed.

   B. is a function of the 3 5 exonuclease activity of the DNA polymerases.

   C. requires the presence of an enzyme separate from the DNA polymerases.

   D. occurs in prokaryotes but not eukaryotes.

   E. is independent of the polymerase activity in prokaryotes.

3. Which of the following would result in a frameshift mutation?

A. formation of ionized bases by radiation.

B. substitution of a purine–pyrimidine pair by a pyrimidine–purine pair.

C. intercalation of ethidium bromide into the nucleotide chain.

D. deamination of cytosine to uracil.

E. conversion of guanine to xanthine.

4. One way of introducing a transition mutation into DNA is by:

A. using a structural analog of a base during synthesis.

B. the action of an acridine dye.

C. introducing a methyl group on the adenine of a GATC sequence.

D. blocking the proofreading action of DNA polymerase.

E. stretching the DNA helix.

5. Which of the following is (are) step(s) in excision repair mechanisms?

A. excision.

B. incision.

C. ligation.

D. all of the above.

E. none of the above.

6. Base excision repair:

A. is used only for bases that have been deaminated.

B. uses enzymes called DNA glycosylases to generate an abasic sugar site.

C. removes about 10–15 nucleotides.

D. requires the action of DNA polymerase III (*E. coli*).

E. recognizes a bulky lesion.

7. All of the following are true about nucleotide excision repair EXCEPT:

A. it is deficient in the disease xeroderma pigmentosum.

B. it removes thymine dimers generated by UV light.

C. it involves the activity of excision nuclease, which is an endonuclease.

D. it requires polymerase I (*E. coli*) and ligase.

E. it occurs in prokaryotes but not in eukaryotes.

8. Mismatch repair:

A. recognizes and removes mismatched bases during the process of replication.

B. occurs only if the mismatch is on a strand containing methylated bases.

C. in *E. coli*, recognizes mismatches within a few hundred base pairs of a GATC sequence.

D. looks for a distortion where the base structure has been altered.

E. is characterized by all of the above being correct.

9. Both strands of DNA serve as templates concurrently in:

A. replication.

B. excision repair.

C. mismatch repair.

D. repair catalyzed by photolyase.

E. all of the above.

10. Replication:

A. is semiconservative.

B. requires only proteins with DNA polymerase activity.

C. uses 5 3 polymerase activity to synthesize one strand and 3 5 polymerase activity to synthesize the complementary strand.

D. requires a primer in eukaryotes but not in prokaryotes.

E. must begin with an incision step.

11. The discontinuous nature of DNA synthesis:

A. requires that DNA polymerase III dissociate from the template when it reaches the end of each single-stranded region.

B. is necessary only because synthesis is bidirectional from the initiation point.

C. leads to the formation of Okazaki fragments.

D. means that synthesis occurs on the second strand of DNA only after synthesis on the first strand is completed.

E. means that both 3 5 and 5 3 polymerases are used.

12. All of the following are factors in the unwinding and separation of DNA strands for replication EXCEPT:

A. the tendency of negative superhelices to partially unwind.

B. destabilization of complementary base pairs by helicases.

C. the action of topoisomerases.

D. the enzymatic activity of SSB proteins.

E. energy in the form of ATP.

13. Initiation of replication in *E. coli*:

A. begins with dnaA binding at the OriC site if certain bases are methylated.

B. results in the formation of several "bubbles," each consisting of a few nucleotide pairs.

C. forms a primosome, which then uses a topoisomerase to open a replication fork.

D. requires the action of helicase to initiate synthesis on the leading strands.

E. begins with the formation of the replisome, followed by the formation of the primosome to begin replication.

14. In eukaryotic DNA replication:

A. only one replisome forms because there is a single origin of replication.

B. the leading and lagging strands are synthesized by the same enzyme.

C. helicase dissociates from DNA as soon as the initiation bubble forms.

D. at least one DNA polymerase has a 3 5 exonuclease activity.

E. the process occurs throughout the cell cycle.

15. All of the following statements about telomerase are correct EXCEPT:

A. the RNA component acts as a template for the synthesis of a segment of DNA.

B. it adds telomeres to the 5 ends of the DNA strands.

C. it provides a mechanism for replicating the ends of linear chromosomes in most eukaryotes.

D. telomerase recognizes a G-rich single strand of DNA.

E. it is a reverse transcriptase.

16. Homologous recombination:

A. occurs only between two segments from the same DNA molecule.

B. requires that a specific DNA sequence be present.

C. requires that one of the duplexes undergoing recombination be nicked in both strands.

D. may result in strand exchange by branch migration.

E. is catalyzed by transposases.

17. All of the following are true about transpositions EXCEPT:

A. transposons move from one location to a different one within a chromosome.

B. both the donor and target sites must be homologous.

C. composite transposons contain an additional gene that is not present in an insertion sequence (IS).

D. transposase introduces a staggered cut in the target DNA sequence.

E. transposition may either activate or inactivate a gene.

Answers

1. C The primer is the initial 3 terminus of an existing strand and the template is the free portion of the complementary stand. B: Polymerase III functions in synthesis. D: Polymerase I has low processivity because it dissociates after only a few nucleotides add. E: Specificity is a function of complementary hydrogen-bonding between the base being added and the template (pp. 622–623).

2. B This activity removes a newly added base if there is a mismatch with the template. A: This is called repair. C: The polymerases are multifunctional enzymes. D: Not all eukaryotic polymerases have 3 5 exonuclease activity but some do. E: The polymerase active site seems to be the one that detects the mismatch and directs the 3 terminus to the proofreading site (p. 624).

3. C Since the bases are read in groups of three, insertion of an additional base would shift the reading frame (p. 628). Intercalation stretches the DNA so when DNA is replicated an additional base is inserted near the intercalation site (p. 633). A: Ionized bases show atypical base pairing. B and D are both examples of base substitution type of mutations. E: This change probably wouldn't make any difference (p. 628).

4. A 5-Bromouracil and 2-aminopurine are used to deliberately introduce mutations for research. B and E: Intercalating agents stretch the helix, allowing the insertion of an extra base—a frameshift mutation. C: The mismatched repair system uses methylated adenine to distinguish between the old and newly synthesized strands. D: This could lead to a mutation but not necessarily a transition (p. 628).

5. D The other step in the process is resynthesis to fill in the gap left by the actions of A and B (p. 635).

6. B These catalyze the first step of the process. A: Methylated and other chemically modified bases can also be removed. C and E: These are characteristics of a different repair system. D: Polymerase I is the repair enzyme (p. 635).

7. E It is common to both systems. A: This is a genetic disease which requires more proteins than the prokaryotic system. B: Thymine dimers are only one cause of bulky lesions. C and D: The excision nuclease is a complex of proteins needed to unwind the DNA and remove the lesion. The polymerase and ligase fill in the gap (pp. 635–636).

8. C Methylated adenine in this sequence is a postreplicative event and signals the correct strand (i.e., unmethylated strand is newly synthesized). A: This is the function of proofreading; mismatch repair is postreplicative. B: Unmethylated sequences shortly after replication denote the newly synthesized strand. D: The bases are unaltered (p. 639).

9. A This allows for the synthesis of two identical DNA molecules. B and C: In both of these the damaged segment is removed so both strands are not available. D: This simply disrupts the inappropriate covalent bond of thymine dimers; no synthesis is involved (pp. 642 and 646).

10. A B and D: Replication requires a primer, usually synthesized by a primase. Ligases, helicases, and other proteins are required as well. C: Replication involves Okazaki fragments because synthesis occurs only in the 5 3 direction. E: Incision is the recognition step for DNA repair (p. 643).

11. C These are the segments of DNA built upon the primer. A: DNA polymerase remains bound to the template and slides over the next primer to continue synthesis. B and E: This mechanism compensates for the inability to synthesize 3 5 and would be necessary even if synthesis were unidirectional. D: Both strands are synthesized concurrently (pp. 646–647).

12. D SSB proteins stabilize the single strands after separation but have no enzymatic activity. A: This is especially true in regions of high AT pairs. B and E: This helps in the original unwinding at the expense of ATP. C: Topoisomerases nick and reseal one of the strands to prevent the introduction of an increasing number of positive supercoils (p. 648).

13. A Methylation seems to be a key in recognition of the OriC site. B: *Escherichia coli* forms only one bubble, a few hundred nucleotide pairs in size. C: The forks form and are stabilized before primase adds. D: The negative superhelicity favors initiation but helicase is necessary for the continuation of synthesis. E: The replisome is the final assembly and includes DNA polymerase III and rep proteins (pp. 649–651).

14. D Polymerase α shows this activity that provides proofreading during synthesis. and have this for proofreading. A: There are multiple initiation sites. The DNA segments between two initiation points are called replicons (p. 654). B: In prokaryotes, the DNA polymerase III does both; in eukaryotes synthesizes the leading strand and α the lagging strand, at least for the initiation process. C: Helicase activity is also necessary for the continuation of synthesis, that is, the opening of the forks (p. 655). E: Replication is confined to the S phase (p. 654).

15. B It is the 3 end of each strand that cannot be conventionally replicated. A and C: Telomerase both positions itself at the 3 ends of the DNA and provides the template for extending that end (p. 660, Figure 15.40). D: This is a characteristic of the 3 -end. E: It is using an RNA template to synthesize DNA (p. 659).

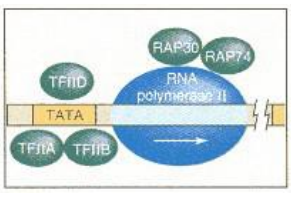
16. D This is just one of the events in this complex process. A and B: It may occur between two distinct DNA molecules; the

requirement is that the two sequences be homologous but not that they be specific sequences. C: The nicks are usually on a single strand. E: These are the enzymes of transpositional site-specific recombination (p. 663).

17. B Only the donor site requires a specific nucleotide sequence; homology is not required. A: This is the definition. C: The IS contains the gene for transposon plus the flanking sequences; composite transposons have an additional gene—for example, one that confers antibiotic resistance in bacteria. D: This permits the transposon to be inserted. E: Insertion into the middle of a gene would inactivate it; insertion of a promoter next to a gene may activate it (pp. 670–671).

Chapter 16—
RNA:
Structure, Transcription, and Processing

Francis J. Schmidt



16.1— Overview

The primary information store of a cell is its genetic complement, that is, its DNA. DNA information is exactly analogous to the master copies of a computer program or any database: It is the source of cellular information and therefore must be kept as error-free as possible. Chapter 14 has detailed some of the elaborate mechanisms that are employed to keep DNA information intact from one cell generation to the next. This chapter describes another type of information transfer that helps to ensure the integrity of genomic information. Just as a careful computer programmer makes working copies of a program or data set, a cell makes macromolecular copies of the information in DNA. These macromolecules, ribonucleic acids (RNAs), are linear polymers of ribonucleoside monophosphates. The sequence of DNA is copied exactly into RNA. The process by which RNA copies of selected DNA sequences are made is termed transcription. The primary role of RNA within the cell is its involvement in protein synthesis, that is, translation.

The overall process of information transfer in the cell is therefore given by the so-called **central dogma** of molecular biology:

$$\text{DNA} \xrightarrow{\text{transcription}} \text{RNA} \xrightarrow{\text{translation}} \text{protein}$$

TABLE 16.1 Characteristics of Cellular RNAs

Type of RNA	*Abbreviation*	*Function*	*Size and Sedimentation Coefficient*	*Site of Synthesis*	*Structural Features*
Messenger RNA Cytoplasmic	mRNA	Transfer of genetic information from nucleus to cytoplasm, or from gene to ribosome	Depends on size of protein 1000–10,000 nucleotides	Nucleoplasm	Blocked 5 end; poly(A) tail on 3 end; nontranslated sequences before and after coding regions; few base pairs and methylations
Mitochondrial	mt mRNA		9S–40S	Mitochondria	
Transfer RNA Cytoplasmic	tRNA	Transfer of amino acids to mRNA–ribosome complex and correct sequence insertion	65–110 nucleotides 4S	Nucleoplasm	Highly base paired; many modified nucleotides; common specific structure
Mitochondrial	mt tRNA		3.2S–4S	Mitochondria	
Ribosomal RNA Cytoplasmic	rRNA	Structural framework for ribosomes	28S, 5400 nucleotides 18S, 2100 nucleotides 5.8S, 158 nucleotides 5S, 120 nucleotides	Nucleolus Nucleolus Nucleolus Nucleoplasm	5.8S and 5S highly base paired; 28S and 18S have some base paired regions and some methylated nucleotides
Mitochondrial	mt rRNA		16S, 1650 nucleotides 12S, 1100 nucleotides	Mitochondria	
Heterogeneous nuclear RNA	hnRNA	Some are precursors to mRNA and other RNAs	Extremely variable 30S–100S	Nucleoplasm	mRNA precursors may have blocked 5 ends and 3 -poly (A) tails; many have base paired loops
Small nuclear RNA	snRNA	Structural and regulatory RNAs in chromatin	100–300 nucleotides	Nucleoplasm	
Small cytoplasmic RNA [7S(L) RNA]	scRNA	Selection of proteins for export	129 nucleotides	Cytosol and rough endoplasmic reticulum	Associated with proteins as part of signal recognition particle

RNA information is occasionally **reverse transcribed** into DNA, a process important in the life cycle of infectious retroviruses such as the human immunodeficiency virus (HIV), which causes the acquired immunodeficiency syndrome (AIDS). Reverse translation of protein sequence into nucleic acid sequence information, however, does not occur in nature.

RNA molecules are classified according to the roles they play in information transfer processes (Table 16.1). In prokaryotes, transcription and translation occur close together; in fact, ribosomes can begin translating a mRNA while it is still being synthesized. In eukaryotes, these processes are spatially separated: transcription occurs in the nucleus and translation in the cytoplasmic portions of the cell. **Messenger RNAs (mRNA)** serve as templates for the synthesis of protein; they carry information from the DNA to the cellular protein synthetic machinery. Here a number of other RNA species contribute to the synthesis of the peptide bond.

The molecules that transfer specific amino acids from soluble amino acid pools to ribosomes, and ensure the alignment of these amino acids in the proper sequence prior to peptide bond formation, are **transfer RNAs (tRNA).** All tRNA molecules are approximately the same size and shape. The assembly site, or factory, for peptide synthesis involves ribosomes. These complex subcellular particles contain three or four **ribosomal RNA (rRNA)** molecules and 70–80 ribosomal proteins.

Protein synthesis requires a close interdependent relationship between mRNA, the informational template, tRNA, the amino acid adaptor molecule, and rRNA, part of the synthetic machinery. In order for protein synthesis to occur at the correct time in a cell's life, the syntheses of mRNA, tRNA, and rRNA must be coordinated with the cell's response to the intra- and extracellular environments.

All cellular RNA is synthesized on a DNA template and reflects a portion of the DNA base sequence. Therefore all RNA is associated with DNA at some time. Although DNA is the more prevalent genetic store of information, RNA can also carry genetic information. Genomic RNA is found in the RNA tumor viruses and the other small RNA viruses, such as poliovirus and reovirus.

16.2— Structure of RNA

RNA Is a Polymer of Ribonucleoside 5′-Monophosphates

Chemically, RNA is similar to DNA. Although RNA is one of the more stable components within a cell, it is not as stable as DNA. The presence of the adjacent 2 -hydroxyl group makes the RNA phosphodiester bond more susceptible to chemical and enzymatic hydrolysis than its DNA counterpart. Some RNAs, such as bacterial mRNA, are synthesized, used, and degraded within minutes, whereas others, such as rRNA, are more stable metabolically.

RNA is an unbranched linear polymer of ribonucleoside monophosphates. The purines found in RNA are *adenine* and *guanine*; the pyrimidines are *cytosine* and *uracil*. Except for uracil, which replaces thymine, these are the same bases found in DNA.

A, C, G, and U nucleotides are incorporated into RNA during transcription. Many RNAs also contain **modified nucleotides**, which are synthesized after transcription. Modified nucleotides are especially characteristic of stable RNA species (i.e., tRNA and rRNA); however, some methylated nucleotides are also present in eukaryotic mRNA. For the most part, the functions of the modified nucleotides in RNA have not been identified. Where known, the function of nucleotide modification seems to involve "fine tuning" rather than an indispensable role in the cell.

The 3 ,5 -phosphodiester bonds of RNA form a chain or backbone from

which the bases extend (Figure 16.1). Eukaryotic RNAs vary from approximately 65 nucleotides long to more than 200,000 nucleotides long. RNA sequences are complementary to the base sequences of specific portions of only one strand of DNA. Thus, unlike the base composition of DNA, molar ratios of A + U and G + C in RNA are not equal. All cellular RNA so far examined is linear and single stranded, but double-stranded RNA is present in some viral genomes.

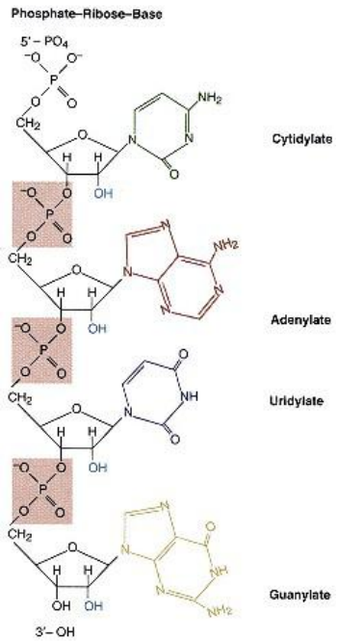


Figure 16.1
Structure of the 3′,5′-phosphodiester bonds between ribonucleotides forming a single strand of RNA.
The phosphate joins the 3 -OH group of one ribose with the 5 -OH group of the next ribose. This linkage produces a polyribonucleotide having a sugar–phosphate "backbone." The purine and pyrimidine bases extend away from the axis of the backbone and may pair with complementary bases to form double helical base paired regions.

Secondary Structure of RNA Involves Intramolecular Base Pairing

RNA, being single stranded rather than double stranded, does not usually form an extensive double helix. Rather, the structure in an RNA molecule arises from relatively short regions of **intramolecular base pairing.** Considerable helical structure exists in RNA even in the absence of extensive base pairing, for example, in the portions of an RNA that do not form intramolecular Watson–Crick base pairs. This helical structure is due to the strong base-stacking forces between A, G, and C residues. Base stacking is more important than simple hydrogen bonding in determining inter- and intramolecular interactions. These forces act to restrict the possible conformations of an RNA molecule (Figure 16.2). **RNA helical structures** generally are of the "A type" with 11 nucleotides per turn in a double helix.

Double helical regions in RNA are often called "hairpins." There are considerable variations in the fine structural details of **"hairpin"** structures, including the length of base paired regions and the size and number of unpaired loops (Figure 16.3). Transfer RNAs are excellent examples of base stacking and hydrogen bonding in a single-stranded molecule (Figure 16.4*a*). About 60% of the bases are paired in four double helical stems. In addition, the unpaired regions have the capability to form base pairs with free bases in the same or other looped regions, thereby contributing to the molecule's tertiary structure. The anticodon region in tRNA is an unpaired, base-stacked, loop of seven nucleotides. The partial helix caused by base stacking in this loop binds, by specific base pairing, to a complementary codon in mRNA so that translation (peptide bond formation) can occur.

RNA Molecules Have Tertiary Structures

The actual functioning structures of RNA molecules are more complex than the base-stacked and hydrogen-bonded helices mentioned above. RNAs *in vivo* are

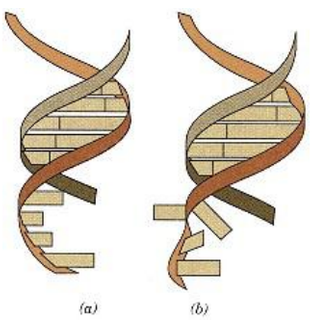


Figure 16.2
Helical structure of tRNA.
Models indicating a helical structure due to
(a) base stacking in the CCA terminus of tRNA and
(b) the lack of an ordered helix when no stacking occurs in this non-base paired region.
Redrawn from Sprinzl, M., and Cramer, F. *Prog. Nucl. Res. Mol. Biol.* 22:9, 1979.

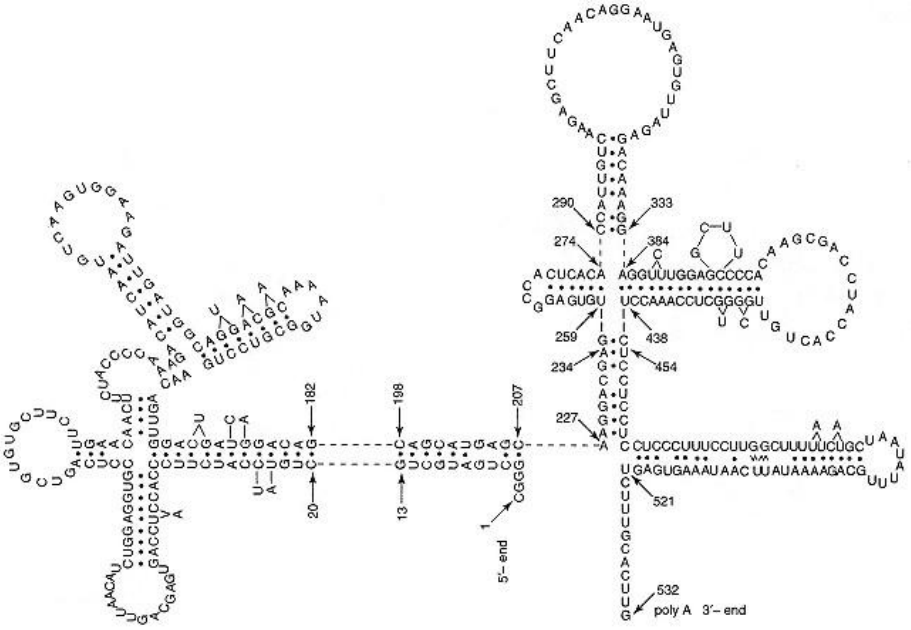


Figure 16.3
Proposed base pairing regions in the mRNA for mouse immunoglobulin light chain.
Base paired structures shown have free energies of at least −5 kcal. Note the variance in loop size and length of paired regions.
Redrawn from Hamlyn, P. H., Browniee, G. G., Cheng, C. C., Gait, M. J., and Milstein, C. *Cell* 15:1067, 1978.

dynamic molecules that undergo changes in conformation during synthesis, processing, and functioning. Proteins associated with RNA molecules often lend stability to the RNA structure; in fact, it is perhaps more correct to think of RNA–protein complexes rather than naked RNA molecules as functioning components of the cell. In addition to the secondary, base paired structure, RNA molecules also form other hydrogen bonds to form the **tertiary structure** of the molecule. Again, the structure of tRNA provides a number of examples. In solution, tRNA is folded into a compact "**L**-shaped" conformation (Figure 16.4*b*). The arms and loops are folded in specific conformations held in position not only by Watson–Crick base pairing, but also by base interactions involving more than two nucleotides. Bases can donate hydrogen atoms to bond with the phosphodiester backbone. The 2 -OH of the ribose is an important donor and acceptor of hydrogens. All these interactions contribute to the folded shape of an RNA molecule.

16.3— Types of RNA

RNA molecules are traditionally classified as transfer, ribosomal, and messenger RNAs according to their usual function; however, we now know that RNA molecules perform or facilitate a variety of other functions in a cell.

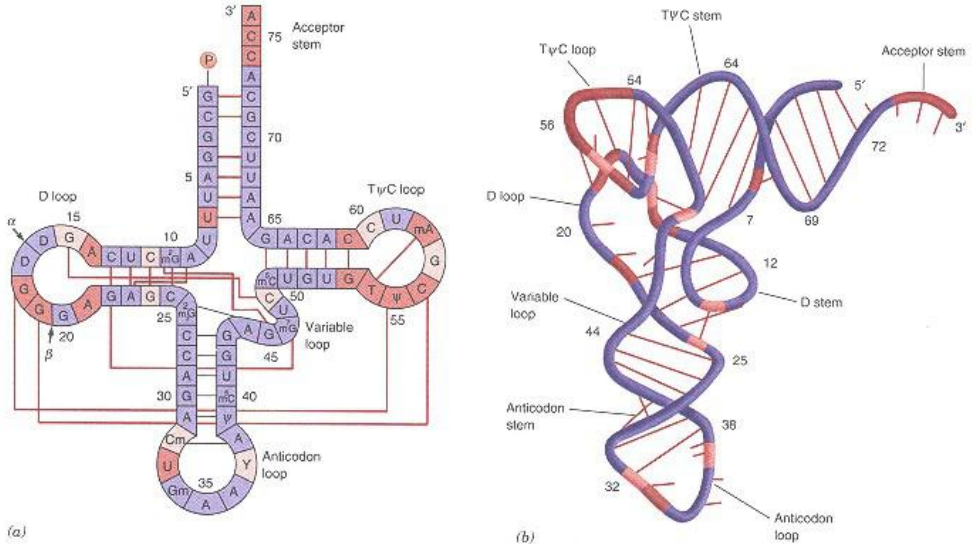


Figure 16.4
Coverleaf structure of tRNA
(a) Cloverleaf diagram of the two-dimensional structure and nucleotide sequence of yeast $tRNA^{Phe}$. Red lines connecting nucleotides indicate hydrogen-bonded residues. Insertion of nucleotides in the D loop occurs at positions α and β for different tRNAs.
(b) Tertiary folding of the cloverleaf structure in (*a*). Hydrogen bonds are indicated by cross rungs.
Redrawn with permission from Quigley, G. J., and Rich, A. *Science* 194:797, 1976.

Transfer RNA Has Two Roles: Activating Amino Acids and Recognizing Codons in mRNA

About 15% of the total cellular RNA is **tRNA**. Transfer RNA has two functions that are essential for its cellular role as an "adapter" of nucleic acid to protein information. First, tRNA molecules **activate amino acids** for protein synthesis so that formation of peptide bonds is energetically favored. The activated amino acid is transported to the polyribosome where it is transferred to the growing peptide chain (hence tRNA's name). The second function of tRNA is to recognize codons in mRNA to ensure that the correct amino acid is incorporated into the growing peptide chain. These two functions are reflected in the fact that tRNAs have two primary active sites, the 3 -OH terminal CCA, to which specific amino acids are attached enzymatically, and the **anticodon triplet**, which base pairs with mRNA codons.

Each tRNA can transfer only a single amino acid. Although there are only 20 amino acids used in protein synthesis, free-living organisms synthesize a larger set of tRNAs. For example, analysis of the recently determined genomic sequence of *Haemophilus influenzae* identified genes for 54 tRNA species. Mitochondria synthesize a much smaller number of tRNAs. Transfer RNAs that accept the same amino acid are called **isoacceptors**. A tRNA that accepts phenylalanine would be written as $tRNA^{Phe}$, whereas one accepting tyrosine would be written $tRNA^{Tyr}$.

CLINICAL CORRELATION 16.1

Staphylococcal Resistance to Erythromycin

Bacteria exposed to antibiotics in a clinical or agricultural setting often develop resistance to the drugs. This resistance can arise from a mutation in the target cell's DNA, which gives rise to resistant descendants. An alternative and clinically more serious mode of resistance arises when plasmids coding for antibiotic resistance proliferate through the bacterial population. These plasmids may carry multiple resistance determinants and render several antibiotics useless at the same time.

Erythromycin inhibits protein synthesis by binding to the large ribosomal subunit. *Staphylococcus aureus* can become resistant to erythromycin and similar antibiotics as a result of a plasmid-borne RNA methylase that converts a single adenosine in 23S rRNA to N^6-dimethyladenosine. Since the same ribosomal site binds lincomycin and clindamycin, the plasmid causes cross-resistance to these antibiotics as well. Synthesis of the methylase is induced by erythromycin.

The microorganism that produces an antibiotic must also be immune to it or else it would be inhibited by its own toxic product. The producer of erythromycin, *Streptomyces erythreus*, itself possesses an rRNA methylase that acts at the same ribosomal site as the one from *S. aureus*.

Which came first? It is likely that many of the resistance genes in target organisms evolved from those of producer organisms. In several cases, DNA sequences from resistance genes of the same specificity are conserved between producer and target organisms. We may therefore look on plasmid-borne antibiotic resistance as a case of "natural genetic engineering," whereby DNA from one organism (e.g., the *Streptomyces* producer) is appropriated and expressed in another (e.g., the *Staphylococcus* target).

Cundliffe, E. How antibiotic-producing microorganisms avoid suicide. *Annu. Rev. Microbiol.* 43:207, 1989.

Transfer RNAs range from 65 to 110 nucleotides in length, corresponding to a molecular weight range of 22,000–37,000. The sequences of all tRNA molecules (over 1000 are known) can be arranged into a common secondary structure that has the appearance of a cloverleaf. The cloverleaf structure is determined by complementary Watson–Crick base pairs forming three stem and loop or hairpin structures. The anticodon triplet sequence is at one "leaf" of the **cloverleaf** while the CCA acceptor stem is at the "stem" (see Figure 16.4). This arrangement where the two active sites of a tRNA are spatially separated is preserved in the tertiary structure of $tRNA^{Phe}$ shown in Figure 16.4. Additional, non-Watson–Crick, hydrogen bonds form in the **L**-shaped molecule.

The nucleotide sequence and structure of the $tRNA^{Phe}$ molecule depicted in Figure 16.4 show that tRNAs have several modified nucleotides. The modified nucleotides affect tRNA structure and stability but are not required for the formation or maintenance of tertiary conformation. For example, a modified base in the anticodon loop makes codon recognition more efficient but a tRNA without this modification can still be read correctly by the ribosome.

Many structural features are common to all tRNA molecules. Seven base pairs are present in the amino acid acceptor stem, which terminates with the nucleotide triplet CCA. This CCA triplet is not base paired. The dihydrouracil or "D" stem has three or four base pairs, while the anticodon and T stems have five base pairs each. Both the anticodon loop and T loop contain seven nucleotides. Differences in the number of nucleotides in different tRNAs are accounted for by the variable loop. Thus 80% of tRNAs have small variable loops of 4–5 nucleotides, while others have larger loops of 13–21 nucleotides. The positions of some nucleotides are constant in all tRNAs (see Figure 16.4*a*).

Ribosomal RNA Is Part of the Protein Synthesis Apparatus

Protein synthesis takes place on ribosomes. These complex assemblies are composed in eukaryotes of four RNA molecules, representing about two-thirds of the particle mass, and 82 proteins. The smaller subunit, the **40S particle,** contains one **18S RNA** and 33 proteins. The larger subunit, the **60S particle,** contains the **28S,** the **5.8S,** and the **5S rRNAs** and 49 proteins. The total assembly is called the **80S ribosome.** Prokaryotic ribosomes are somewhat smaller: the **30S subunit** contains a single **16S rRNA** and 21 proteins, while the larger subunit **(70S)** contains **5S** and **23S rRNAs** as well as 34 ribosomal proteins.

The rRNAs account for 80% of the total cellular RNA and are metabolically stable. This stability, required for repeated functioning of the ribosome, is enhanced by close association with the ribosomal proteins. The 28S (4718 nucleotides), 18S (1874 nucleotides), and 5.8S (160 nucleotides) rRNAs are synthesized in the nucleolar region of the nucleus. The 5S rRNA (120 nucleotides) is not transcribed in the nucleolus but rather from separate genes within the nucleoplasm (Figure 16.5). Processing of the rRNAs (see Section 16.5) includes cleavage to the functional size, internal base pairing, modification of particular nucleotides, and association with ribosomal proteins to form a stable tertiary conformation.

The larger rRNAs contain most of the altered nucleotides found in rRNA. These are primarily **methylations** on the 2 position of the ribose, yielding 2 -*O*-methylribose. Methylation of rRNA has been directly related to bacterial antibiotic resistance in a pathogenic species (see Clin. Corr. 16.1). A small number of N^6-dimethyladenines are present in 18S rRNA. The 28S rRNA has about 45 methyl groups and the 18S rRNA has 30 methyl groups.

Biochemical studies of ribosome function indicate that rRNA molecules are more than macromolecular scaffolds for enzymatic proteins. The exact extent to which rRNA participates in protein biosynthetic reactions is the subject of current investigation. Several lines of evidence indicate that the actual formation of a peptide bond may be catalyzed by the large RNA subunit of the ribosome.

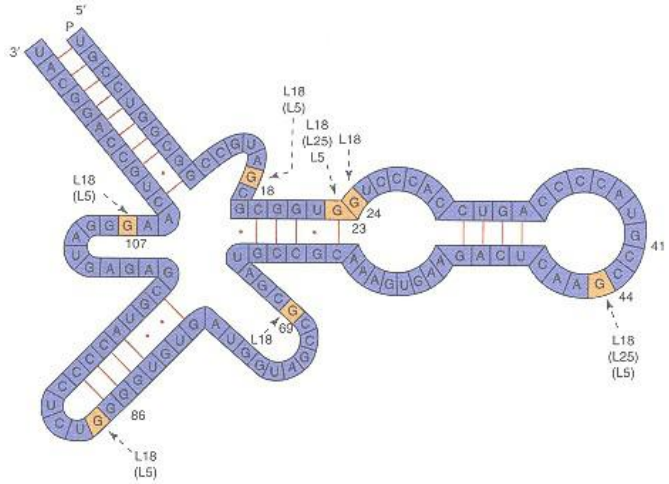


Figure 16.5
Secondary, base paired, structure proposed for 5S rRNA.
Arrows indicate regions protected by proteins in the large ribosomal subunit.
Combined information from Fox, G. E., and Woese, C. R.
Nature 256:505, 1975; and R. A. Garrett and P. N. Gray.

Messenger RNAs Carry the Information for the Primary Structure of Proteins

The **mRNAs** are the direct carriers of genetic information from genomes to the ribosomes. Each eukaryotic mRNA is **monocistronic;** that is, it contains information for only one polypeptide chain. In prokaryotes, mRNA species often encode more than one protein in a **polycistronic** molecule. A cell's phenotype and functional state are related directly to its mRNA content.

In the cytoplasm mRNAs have relatively short life spans. Some mRNAs are known to be synthesized and stored in an inactive or dormant state in the cytoplasm, ready for a quick protein synthetic response. An example of this is the unfertilized egg of the African clawed toad, *Xenopus laevis*. Immediately upon fertilization the egg undergoes rapid protein synthesis in the absence of transcription, indicating the presence of preformed mRNA.

Eukaryotic mRNAs have unique structural features not found in rRNA or tRNA (see Figure 16.6). Since the information within mRNA lies in the linear sequence of the nucleotides, the integrity of this sequence is extremely im-

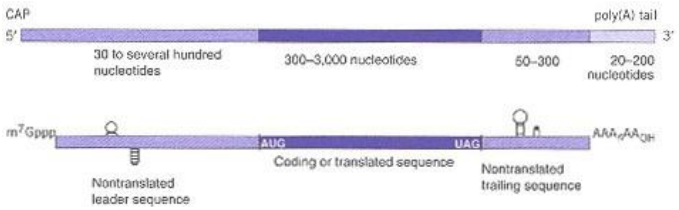


Figure 16.6
General structure for a eukaryotic mRNA.
There is a "blocked" 5 terminus (cap) followed by the nontranslated leader containing a promoter sequence. The coding region usually begins with the initiator codon AUG and continues to the translation termination sequence UAG, UAA, or UGA. This is followed by the nontranslated trailer and a poly(A) tail on the 3 end.

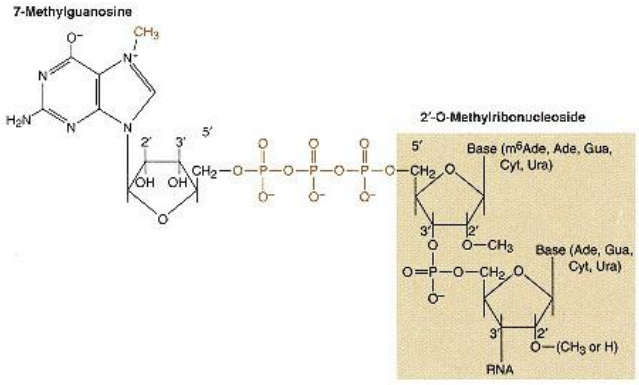


Figure 16.7
Diagram of the "cap" structure or blocked 5′ terminus in mRNA.
The 7-methylguanosine is inverted to form a 5 -phosphate to 5 -phosphate linkage with the first nucleotide of the mRNA. This nucleotide is often a methylated purine.

portant. Any loss or change of nucleotides could alter the protein being translated. The translation of mRNA on the ribosomes must also begin and end at specific sequences. Structurally, starting from the 5 terminus, eukaryotic mRNA is capped with an inverted methylated base attached via **5′-phosphate–5′-phosphate bonds** rather than the usual 3 ,5 -phosphodiester linkages. The **cap** is attached to the first transcribed nucleotide, usually a purine, methylated on the 2 -OH of the ribose (see Figure 16.7). The cap is followed by a nontranslated or **"leader"** sequence to the 5 side of the coding region. Following the leader sequence are the **initiation sequence or codon,** most often AUG, and the translatable coding region of the molecule. At the end of the coding sequence is a **termination sequence** signaling termination of polypeptide formation and release from the ribosome. A second nontranslated or **"trailer"** sequence follows, terminated by a string of 20–200 adenine nucleotides, called a **poly(A) tail,** which makes up the 3 terminus of the mRNA.

The 5 cap has a positive effect on the initiation of message translation. In the initiation of translation of a mRNA, the cap structure is recognized by a single ribosomal protein, an initiation factor (see Chapter 17). The poly(A) sequence is correlated with the stability of the mRNA molecule; for example, **histone mRNA** molecules lack a poly(A) tail and are also present in the cell only transiently.

Mitochondria Contain Unique RNA Species

Mitochondria (mt) have their own protein-synthesizing apparatus, including ribosomes, tRNAs, and mRNAs. The mt rRNAs, 12S and 16S, are transcribed from the mitochondrial DNA (mt DNA), as are 22 specific tRNAs and 13 mRNAs, most of which encode proteins of the electron transport chain and ATP synthetase. Note that there are fewer mt tRNAs than prokaryotic or cytoplasmic tRNA species; there is only one mt tRNA species per amino acid. The mt RNAs account for 4% of the total cellular RNA. They are transcribed by a mitochondrial-specific RNA polymerase and are processed from a pair of mt RNA precursors. Each precursor is an exact copy of the entire mitochondrial genome, complementary to either the heavy (H) or light (L) strand of mt DNA. Genes for 12 tRNAs are located on the heavy mt DNA strand and 7 on the light strand. Some of the mRNAs have eukaryotic characteristics, such as 3 -poly(A) tails. A large degree of coordination exists between the nuclear and mitochondrial genomes. Most of the aminoacylating enzymes for the mt tRNAs and all of the mitochondrial ribosomal proteins are specified by nuclear genes, translated in the cytoplasm

and transported into the mitochondria. The modified bases in mt tRNA species are synthesized by enzymes encoded in nuclear DNA.

RNA in Ribonucleoprotein Particles

Besides tRNA, rRNA, and mRNA, small, stable RNA species can be found in the nucleus, cytoplasm, and mitochondria. These small RNA species function as ribonucleoprotein particles (RNPs), with one or more protein subunits attached. Different RNP species have been implicated in RNA processing, splicing, transport, and control of translation, as well as in the recognition of proteins due to be exported. The actual roles of these species, where known, are described more fully in the discussion of specific metabolic events.

Some RNAs Have Catalytic Activity

RNA can be an enzyme. In several cases the RNA component of a ribonucleoprotein particle has been shown to be the catalytically active subunit of the enzyme. In other cases, *in vitro* catalytic reactions can be carried out by RNA in the absence of any protein. Enzymes whose RNA subunits carry out catalytic reactions are called **ribozymes.** There are four classes of ribozyme. Three of these RNA species carry out self-processing reactions while the fourth, **ribonuclease P (RNase P),** is a true catalyst.

In the ciliated protozoan *Tetrahymena thermophila*, an intron in the rRNA precursor is removed by a multistep reaction (Figure 16.8). A guanosine nucleoside or nucleotide reacts with the intron–exon phosphodiester linkage to displace the donor exon from the intron. This reaction, a transesterification, is promoted by the folded intron itself. The free donor exon then similarly attacks the intron–exon phosphodiester bond at the acceptor end of the intron. Introns of this type **(Group I introns)** have been found in a variety of genes in fungal mitochondria and in the bacteriophage T4. Although these introns are not true enzymes *in vivo* because they only work for one reaction cycle, they can be made to carry out catalytic reactions under specialized conditions.

Group II self-splicing introns are found in the mitochondrial RNA precursors of yeasts and other fungi. The self-splicing of these introns proceeds through a lariat intermediate similar to the lariat intermediate in the splicing of nuclear mRNA precursors (see below). Since this reaction is carried out by a ribozyme the catalytic activity of the small nucleus ribonucleoproteins (snRNPs) involved in nuclear mRNA splicing may also reside in the RNA component.

A third class of **self-cleaving RNAs** is found in the genomic RNAs of several plant viruses. These RNAs self-cleave during the generation of single genomic RNA molecules from large multimeric precursors. The three-dimensional structure of the **hammerhead ribozyme,** a member of this third class, has recently been determined (Figure 16.9). Catalysis is carried out by a bound Mg^{2+} ion positioned near the bond to be cleaved in the folded ribozyme structure. The phosphate of the cleaved bond is left at the 3 hydroxyl position of the RNA product. A self-cleaving RNA is found in a small satellite virus, hepatitis delta virus, that is implicated in severe cases of human infectious hepatitis. All of the above self-processing RNAs can be made to act as true catalysts (i.e., exhibiting multiple turnover) *in vitro* and *in vivo*.

Ribonuclease P contains both a protein and an RNA component. It acts as a true enzyme in the cell, cleaving tRNA precursors to generate the mature 5 end of the tRNA molecule. RNase P recognizes constant structures associated with tRNA precursors (e.g., the acceptor stem and CCA sequence) rather than using extensive base pairing to bind the substrate RNA to the ribozyme. The product of cleavage contains a 5 phosphate in contrast to the products of hammerhead and similar RNAs. In all of these events the structure of the catalytic RNA is essential for intramolecular or enzyme catalysis.

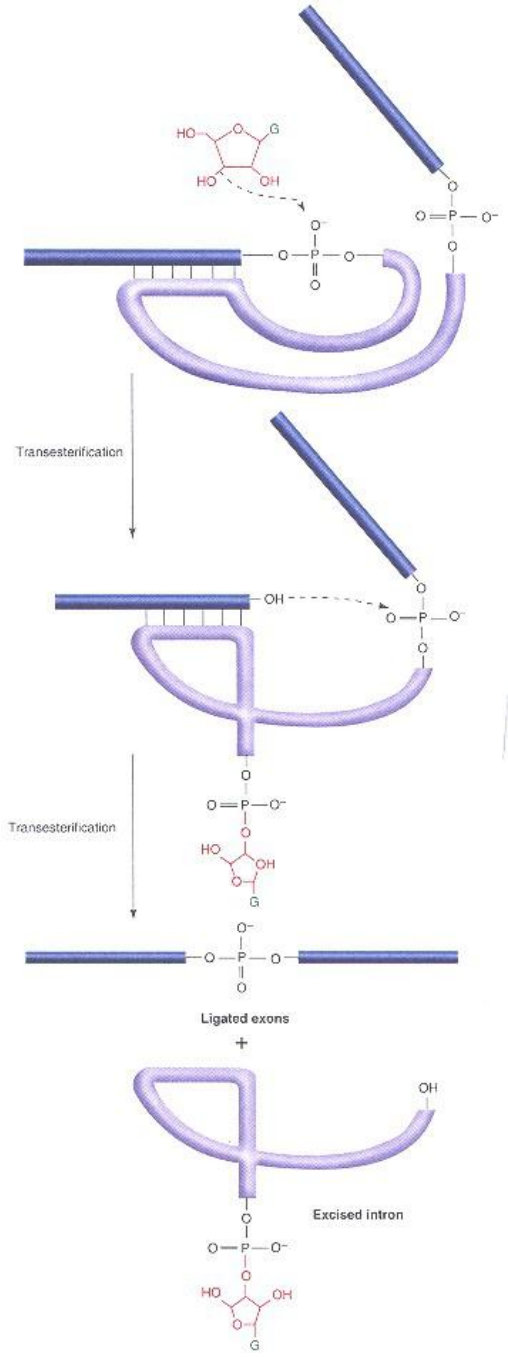


Figure 16.8
Mechanism of self-splicing of the rRNA precursor of *Tetrahymena*.
The two exons of the rRNA are denoted by dark blue. Catalytic functions reside in the intron, which is purple. This splicing function requires an added guanosine nucleoside or nucleotide.
Reproduced from Cech, T. R. *JAMA* 260:308, 1988.

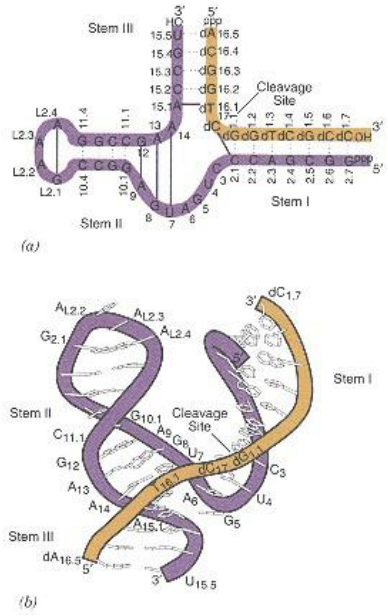


Figure 16.9
"Hammerhead" structure of viral RNA
(a) The hammerhead structure of a self-cleaving viral RNA. This artificial molecule is formed by the base pairing of two separate RNAs. The cleavage of the RNA sequence at the site indicated by the arrow in the top strand requires its base pairing with the sequence at the bottom of the molecule. The boxed nucleotides are a consensus sequence found in self-cleaving viral RNAs. (b) The three-dimensional folding of the hammerhead catalytic RNA. The star indicates the position of the cleaved bond while M indicates a binding site for a metal ion. Helices II and III stack to form an apparently continuous helix while non-Watson–Crick interactions position the noncomplementary bases in the hammerhead into a "uridine turn" structure identical to that found in tRNA.
Part (a) redrawn from Sampson, J. R., Sullivan, F. X., Behlen, L. S., DiRenzo, A. B., and Uhlenbeck, O. C. *Cold Spring Harbor Symp. Quant. Biol.* 52:267, 1987; part (b) redrawn from Pley, H. W., Flaherty, K. M., and McKay, D. B. *Nature* 372:68 1994.

The discovery of RNA catalysis has greatly altered our concepts of biochemical evolution and the range of allowable cellular chemistry. First, we now recognize that RNA can serve as both a catalyst and a carrier of genetic information. This has raised the possibility that the earliest living organisms were based entirely on RNA and that DNA and proteins evolved later. This model is sometimes referred to as the **"RNA world."** Second, we know that many viruses, including human pathogens, use RNA genetic information; some of these RNAs have been shown to be catalytic. Thus catalytic RNA presents opportunities for the discovery of RNA-based pharmaceuticals. Third, many of the information processing events in protein synthesis and mRNA splicing require RNA components. These RNAs may also be fulfilling a catalytic function.

RNAs Can Form Binding Sites for Other Molecules

Consideration of the RNA world has led to a new type of biological chemistry based on the large number of potential sequences (4^N) that would be made if A, C, G, or U were inserted randomly in each of N positions in a nucleic acid. A set of chemically synthesized, randomized, nucleic acid molecules 25 nucleotides long would contain $4^{25} = 10^{15}$ potential members. Individual molecules within this large collection of RNAs would be expected to fold into a similarly large collection of shapes. The large number of molecular shapes implies that some member of this collection will be capable of strong, specific binding to any ligand, much as group I introns bind guanosine nucleotides specifically. Though a single molecule would be too rare to study within the original population, the RNA capable of binding can be selected and preferentially replicated *in vitro*. In one case, for example, an RNA capable of distinguishing **theophylline** from **caffeine** was selected from a complex population (see

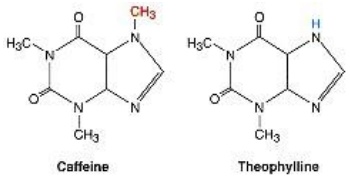


Figure 16.10
Structures of theophylline and caffeine.
Although these compounds differ only by a single methyl group, a specific synthetic RNA can bind to theophylline 10,000-fold more tightly than to caffeine.

Figure 16.10). Theophylline is used in the treatment of chronic asthma but the level must be carefully controlled to avoid side effects. The monitoring of theophylline by conventional antibody-based clinical chemistry is difficult because caffeine and theophylline differ only by a single methyl group. Therefore anti-theophylline antibodies show considerable cross-reaction with caffeine. RNA molecules have been found that bind theophylline 10,000-fold more tightly than caffeine.

Other extensions of the technology have used selection procedures to identify new, synthetic ribozymes and potential therapeutic RNAs.

16.4— Mechanisms of Transcription

The Initial Process of RNA Synthesis Is Transcription

The process by which RNA chains are made from DNA templates is called **transcription.** All known transcription reactions take the following form:

$$\text{DNA template} + n(\text{NTP}) \rightarrow \text{pppN(pN)}_{n-1} + (n-1)\text{PP}_i + \text{DNA template}$$

Enzymes that catalyze this reaction are designated RNA polymerases; it is important to recognize that they are absolutely template dependent. In contrast to DNA polymerases, however, **RNA polymerases** do not require a primer molecule. The energetics favoring the RNA polymerase reaction are twofold: first, the 5 α-nucleotide phosphate of the ribonucleoside triphosphate is converted from a phosphate anhydride to a phosphodiester bond with a change in free energy (G) of approximately 3 kcal (12.5 kJ) mol^{-1} under standard conditions; second, the released pyrophosphate, PP_i, can be cleaved into two phosphates by pyrophosphatase so that its concentration is low and phosphodiester bond formation is more favored relative to standard conditions (see Chapter 6 for a fuller discussion of metabolic coupling).

Since a DNA template is required for RNA synthesis, eukaryotic transcription takes place in the cell nucleus or mitochondrial matrix. Within the nucleus, the *nucleolus* is the site of rRNA synthesis, whereas mRNA and tRNA are synthesized in the nucleoplasm. Prokaryotic transcription is accomplished on the cell's DNA, which is located in a relatively small region of the cell. In the case of prokaryotic plasmids, the DNA template need not be associated with the chromosome.

Structural changes occur in DNA during its transcription. In the polytene chromosomes of *Drosophila*, transcriptionally active genes are visualized in the light microscope as puffs distinct from the condensed, inactive chromatin. Furthermore, the nucleosome patterns of active genes are disrupted so that active chromatin is more accessible to, for example, DNase attack. In prokaryotes and eukaryotes, the DNA double helix is transiently opened (unwound) as the transcription complex proceeds down the DNA.

These openings and **unwindings** are a manifestation of a topological necessity. If the RNA chain were copied off DNA without this unwinding, the transcription complex and growing end of the RNA chain would have to wind around the double helix once every 10 base pairs as they travel from the beginning of the gene to its end. Such a process would wrap the newly synthesized RNA chain around the DNA double helix. Local opening and unwinding of the DNA solves this problem before it occurs by allowing transcription to proceed on a single face or side of the DNA. In addition, the opening of DNA base pairs during transcription allows Watson–Crick base pairing between template DNA and the bases in the newly synthesized RNA.

The process of transcription is divided into three parts: **initiation** refers to the recognition of an active gene starting point by RNA polymerase and the beginning of the bond formation process; **Elongation** is the actual synthesis of the RNA chain and is followed by chain **termination** and release.

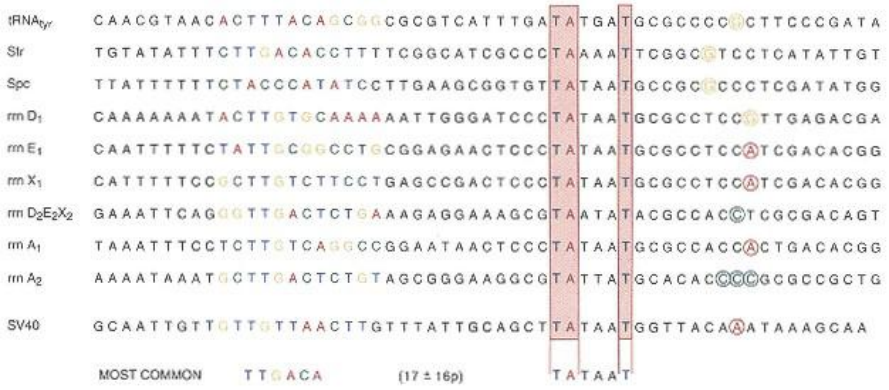


Figure 16.11
Determination of a consensus sequence for prokaryotic promoters.
A portion of the data set used for the identification of the consensus sequence for E. coli promoter activity. The –10 region (sometimes called the Pribnow box) is shaded in red and the –35 region nucleotides are colored. Note that none of the individual promoters has the entire consensus sequence.
Modified from Rosenberg, M., and Court, D. *Ann. Rev. Genet.* 13:319, 1979.

The Template for RNA Synthesis Is DNA

Each cycle of transcription begins and ends with the recognition of specific sites in the DNA template. The DNA sequencing of a large number of transcription start regions, called **promoters,** has shown that certain **conserved sequences** occur in promoters with great regularity.

An example is shown in Figure 16.11. Similar considerations demonstrate that termination occurs at different conserved sequences. In addition, sites within a transcript may allow premature termination of transcription. These sites can act as molecular switches affecting the continuation of synthesis of an RNA molecule.

Conserved sequences near the transcription start are found for both prokaryotic and some eukaryotic promoters. In addition, eukaryotic transcription has been shown in some cases to be affected by **internal promoter** elements and other sequences called enhancers. **Enhancers** are gene-specific sequences that positively affect transcription. Enhancer sequences can stimulate transcription whether they are located at the beginning, in the middle, or at the end of a gene. An enhancer sequence must be on the same DNA strand as the transcribed gene (genetically in a cis position) but can function in either orientation. Cellular protein factors are known that specifically bind different enhancers. The most likely hypothesis is that protein factors bound to enhancers cause a structural change in the DNA template, allowing protein–protein interaction with other factors or with RNA polymerase itself. This interaction facilitates transcription.

RNA Polymerase Catalyzes the Transcription Process

RNA polymerases all synthesize RNA in the 5 3 direction using a DNA template; in this respect, they are similar to template-dependent DNA polymerases discussed in Chapter 15. Unlike DNA polymerases, however, RNA polymerases initiate polymerization at a promoter sequence without the need of a DNA

or RNA primer. Cellular RNA polymerases, both prokaryotic and eukaryotic, are large multisubunit enzymes whose mechanisms are only partially understood.

The most intensely studied prokaryotic RNA polymerase is that from *Escherichia coli*, which consists of five subunits having an aggregate molecular weight of over 500,000 (Table 16.2). Two α subunits, one β subunit, and one subunit constitute the **core enzyme,** which is capable of faithful transcription but not of specific (i.e., promoter-initiated) RNA synthesis. The addition of a fifth protein subunit, designated **σ,** results in the **holoenzyme** that is capable of specific RNA synthesis *in vitro* and *in vivo*. The logical conclusion, that is involved in the specific recognition of promoters, has been borne out by a variety of biochemical studies and is discussed below. Specific **σ factors** can recognize different classes of genes. For example, a specific factor recognizes promoters for genes that are induced as a result of heat shock. In sporulating bacteria, specific factors recognize genes induced during sporulation. Some bacteriophage synthesize factors that allow the appropriation of the cell's RNA polymerase for transcription of the viral DNA.

The common prokaryotic RNA polymerases are inhibited by the antibiotic **rifampicin** (used in treating tuberculosis), which binds to the β subunit (see Clin. Corr. 16.2). Eukaryotic nuclear RNA polymerases are inhibited differentially by the compound **α-amanitin,** which is synthesized by the poisonous mushroom *Amanita phalloides*. Three nuclear RNA polymerase classes can be distinguished by these experiments. Very low concentrations of α-amanitin inhibit the synthesis of mRNA and some small nuclear RNAs (snRNAs); higher concentrations inhibit the synthesis of tRNA and other snRNAs, whereas rRNA synthesis is not inhibited at these concentrations of drug. Messenger RNA synthesis is the function of **RNA polymerase II.** Synthesis of transfer RNA, 5sRNA, and some snRNAs are carried out by **RNA polymerase III.** Ribosomal RNA genes are transcribed by **RNA polymerase I,** which is concentrated in the nucleolus. (The numbers refer to the order of elution of the enzymes from a chromatography column.) Each enzyme is highly complex structurally (Table 16.2).

In addition, a mitochondrial RNA polymerase is responsible for the synthesis of this organelle's mRNA, tRNA, and rRNA species. This enzyme, like bacterial RNA polymerase, is inhibited by rifampicin.

TABLE 16.2 Comparative Properties of Some RNA Polymerases

	Nuclear				
	I (A)	***II (B)***	***III (C)***	***Mitochondrial***	***E. coli***
High MW subunits[a]	195–197	240–214	155	65	160 ( )
	117–126	140	138		150 ( )
Low MW subunits	61–51	41–34	89		86 ( )
	49–44	29–25	70		40 ( )
	29–25	27–20	53		10 ( )
	19–16.5	19.5	49		
		19	41		
		16.5	32		
			29		
			19		
Variable forms	2–3 types	3–4 types	2–4 types	1	1
Specialization	Nucleolar; rRNA	mRNA	tRNA	All mtRNA	None
		Viral RNA	5S rRNA		
Inhibition by α-amanitin	Insensitive (>1 mg mL^{-1})	Very sensitive (10^{-9}–10^{-8} M)	Sensitive (10^{-5}–10^{-4} M)	Insensitive, but sensitive to rifampicin	Rifampicin sensitive

[a] Molecular weight × 10^{-3}.

CLINICAL CORRELATION 16.2

Antibiotics and Toxins That Target RNA Polymerase

RNA polymerase is obviously an essential enzyme for life since transcription is the first step of gene expression. No RNA polymerase means no enzymes. Two natural products point out this principle; in both cases inhibition of RNA polymerase leads to death of the organism.

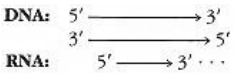
The "death cap" or "destroying angel" mushroom, *Amanita phalloides*, is highly poisonous and still causes several deaths each year despite widespread warnings to amateur mushroom hunters (it is reputed to taste delicious, incidentally). The most lethal toxin, α-amanitin, inhibits the largest subunit of eukaryotic RNA polymerase II, thereby inhibiting mRNA synthesis. The course of the poisoning is twofold: initial, relatively mild, gastrointestinal symptoms are followed about 48 h later by massive liver failure as essential mRNAs and their proteins are degraded but not replaced by newly synthesized molecules. The only therapy is supportive, including liver transplantation; but this latter course is clearly a desperate measure of unproven efficacy.

More benign (at least from the point of view of our own species) is the action of the antibiotic rifampicin to inhibit the RNA polymerases of a variety of bacteria, most notably in the treatment of tuberculosis. *Mycobacterium tuberculosis*, the causative agent, is insensitive to many commonly used antibiotics, but it is sensitive to rifampicin, the product of a soil streptomycetes. Since mammalian RNA polymerase is so different from the prokaryotic variety, inhibition of the latter enzyme is possible without great toxicity to the host. This consideration implies a good therapeutic index for the drug, that is, the ability to treat a disease without causing undue harm to the patient. Together with improved public health measures, antibiotic therapy with rifampicin and isoniazid (an anti-metabolite) has greatly reduced the morbidity due to tuberculosis in industrialized countries. Unfortunately, the disease is still endemic in impoverished populations in the United States and in other countries. Furthermore, in increasing numbers, immunocompromised individuals, especially AIDS patients, have active tuberculosis.

Mitchel, D. H. *Amanita* mushroom poisoning. *Annu. Rev. Med.* 31:51, 1980; Gilman, A. G., Rall, T. W., Nies, A. S., and Taylor, P. (Eds.). *The Pharmacological Basis of Therapeutics*, 8th ed. New York: Pergamon Press, 1990, pp. 129–130; DeCock, K. M., Soro, B., Colibaly, I. M., Lucas, S. B. Tuberculosis and HIV infection in sub-Saharan Africa. *JAMA* 268:1581, 1992.

The Steps of Transcription in Prokaryotes Have Been Determined

Transcription is a strand-selective process; most double helical DNA is transcribed in only one direction. This is illustrated as follows:



The DNA strand that serves as the template for RNA synthesis is sometimes called the sense strand because it is complementary to the RNA transcript. Conventionally, the *sense strand* is usually the "bottom" strand of a double-stranded DNA as written. The other strand, the "top" strand, has the same direction as the transcript when read in the 5 3 direction; this strand is sometimes (confusingly) called the *antisense strand*. When only a single DNA sequence is given in this book, the antisense strand is represented. Its sequence can be converted to the RNA transcript of a gene by simply substituting U (uracil) for T (thymine) bases. Prokaryotic transcription begins with the binding of RNA polymerase to a gene's promoter (Figures 16.11 and 16.12). RNA polymerase holoenzyme binds to one face of the DNA extending 45 bp or so upstream and 10 bp downstream from the RNA initiation site. Two short oligonucleotide sequences in this region are highly conserved. One sequence that is located about 10 bp upstream from the transcription start is the consensus sequence (sometimes called a Pribnow box):

T*A*TAAT*

The positions marked with an asterisk are the most conserved; indeed, the last T residue is always found in *E. coli* promoters.

A second consensus sequence is located upstream from the **Pribnow** or "–10" box. This "–35 sequence"

T*T*G*ACA

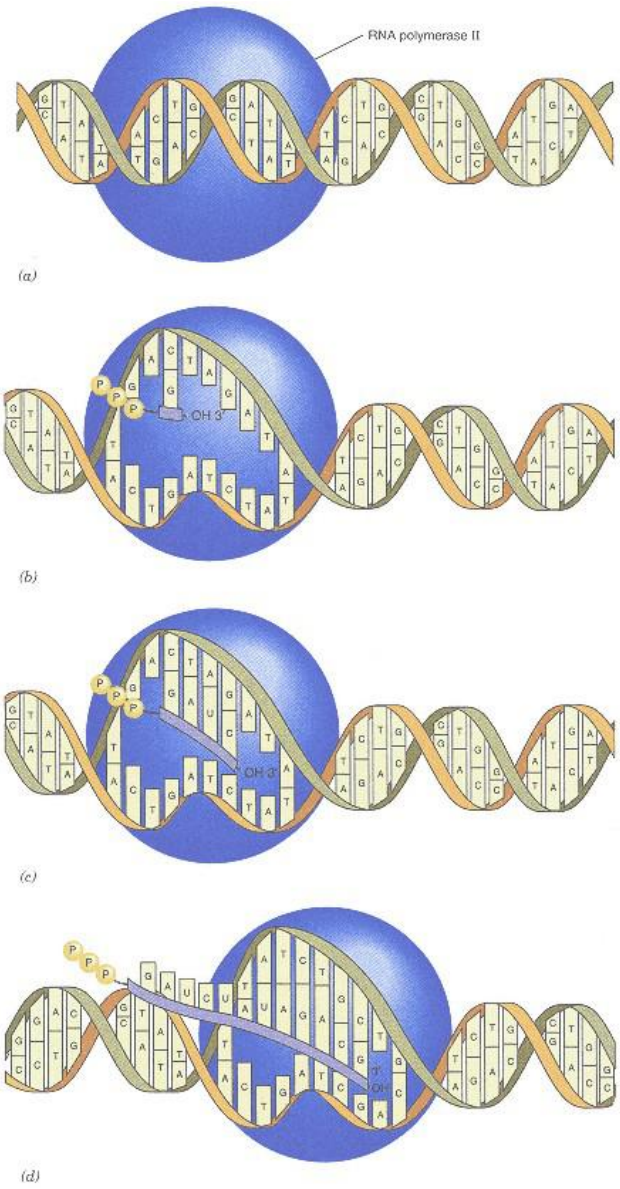


Figure 16.12
Early events in prokaryotic transcription.
(a) Recognition: RNA polymerase (drawn smaller than scale) with "sigma" factor binds to a DNA promoter region in a "closed" conformation.
(b) Initiation: The complex is converted to an "open" conformation and the first nucleoside triphosphate aligns with the DNA.
(c) Bond formation: The first phosphodiester bond is formed and the "sigma" factor released.
(d) Elongation: Synthesis of nascent RNA proceeds with movement of the RNA polymerase along the DNA. The double helix reforms.

is centered about 35 bp upstream from the transcription start; the nucleotides with asterisks are most conserved. The spacing between the –35 and –10 sequences is crucial with 17 bp being highly conserved. As shown in Figure 16.13, the TTGACA and TATAAT sequences are asymmetrical, that is, they do not have the same sequence if the complementary sequence is read. Thus the promoter sequence itself determines that transcription will proceed in only one direction. What difference do the consensus sequences make to a gene? Measurements of RNA polymerase binding affinity and initiation efficiency to

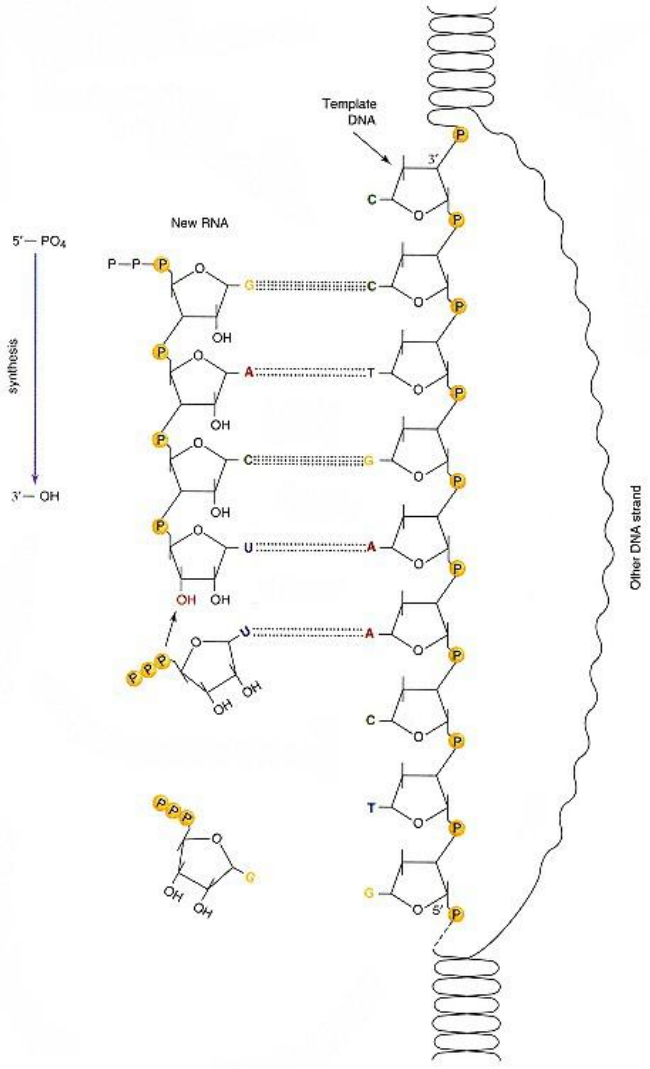


Figure 16.13
Biosynthesis of RNA showing asymmetry in transcription.
Nucleoside 5 -triphosphates align with complementary bases on one DNA strand, the template. RNA polymerase catalyzes the formation of the 3 ,5 -phosphodiester links by attaching the 5 -phosphate of the incoming nucleotide to the 3 -OH group of the growing nascent RNA releasing P_i. The new RNA is synthesized from its 5 end toward the 3 end.

various promoter sequences have shown that the most active promoters fit the consensus sequences most closely. Statistical measurements of promoter homology conform closely to the measured "strength" of a promoter, that is, its kinetic ability to initiate transcription with –35 purified RNA polymerase.

Bases flanking the –35 and –10 sequences, bases near the transcription start, and bases located near the –16 position are weakly conserved. In some of these weakly conserved regions, RNA polymerase may require that a particular nucleotide not be present or that local variations in DNA helical structure be present.

Promoters for *E. coli* heat shock genes have different consensus sequences at the –35 and –10 homologies. This is consistent with their being recognized by a different factor.

An RNA transcript usually starts with a purine riboside triphosphate; that is, pppG · · · or pppA · · ·, but pyrimidine starts are also known (Figures 16.11 and 16.12). The position of transcription initiation differs slightly among various promoters but usually is from five to eight base pairs downstream from the invariant T of the Pribnow box.

Initiation

Two kinetically distinct steps are required for RNA polymerase to initiate the synthesis of an RNA transcript. In the first step, RNA polymerase holoenzyme binds to the promoter DNA to form a **"closed complex."** In the second step, the holoenzyme forms a more tightly bound **"open complex,"** which is characterized by a local opening of about 10 bp of the DNA double helix. Since the consensus Pribnow box is A-T rich, it can facilitate this local unwinding. As discussed in Chapter 14, opening 10 bp of DNA is topologically equivalent to the relaxation of a single negative supercoil. As might be predicted from this observation, the activity of some promoters depends on the superhelical state of the DNA template; some promoters are more active on highly supercoiled DNA while others are more active when the superhelical density of the template is lower. The unwound DNA binds the initiating triphosphate and RNA polymerase then forms the first phosphodiester bond. The enzyme translocates to the next position (this is the rifampicin-inhibited step) and continues synthesis. At or a short time after the initial bond formation, factor is released and the enzyme is considered to be in an elongation mode. Other RNA polymerase molecules can now bind to the promoter so that a gene can be transcribed many times (Figure 16.14).

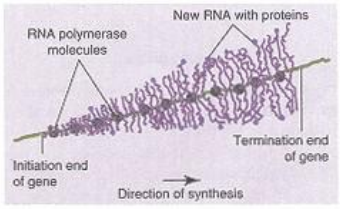


Figure 16.14
Simultaneous transcription of a gene by many RNA polymerases, depicting the increasing length of nascent RNA molecules. Courtesy of Dr. O. L. Miller, University of Virginia. Reproduced with permission from Miller, O. L., and Beatty, B. R. *J. Cell Physiol.* 74:225, 1969.

Elongation

RNA polymerase continues the binding–bond formation–translocation cycle at a rate of about 40 nucleotides per second. This rate is only an average, however, and there are many examples known for which RNA polymerase pauses or slows down at particular sequences, usually inverted repeats (palindrome sequence of nucleotides). As will be discussed below, these pauses can bring about transcription termination.

As RNA polymerase continues down the double helix, it continues to separate the two strands of the DNA template. As seen in Figure 16.12, this process allows the template (sense) strand of the DNA to base pair with the growing RNA chain. Thus a single mechanism of information transfer (Watson–Crick base pairing) serves several processes: DNA replication, DNA repair, and transcription of genetic information into RNA. (As will be seen in Chapter 17, base pairing is essential for translation as well.) The process of unwinding and restoring the DNA double helix is aided by DNA topoisomerases I and II, which are components of the transcription complex.

Changes in the transcription complex during the elongation phase can affect subsequent termination events. These changes depend on the binding of another cellular protein (nusA protein) to core RNA polymerase. Failure to

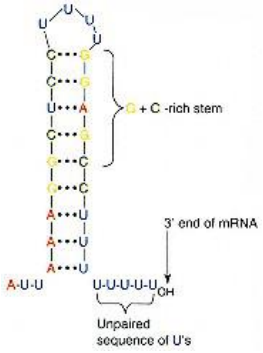


Figure 16.15
The stem–loop structure of the RNA transcript that determines rho-independent transcriptional termination.
Note the two components of the structure: the G + C-rich stem and loop,followed by a sequence of U residues.

bind sometimes results in an increased frequency of termination and, consequently, a reduced level of gene expression.

Termination

The RNA polymerase complex also recognizes the ends of genes (Figure 16.15). Transcription termination can occur in either of two modes, depending on whether or not it is dependent on the protein factor rho. Terminators are thus classified as rho independent or rho dependent.

Rho-independent terminators are better characterized (Figure 16.15). A consensus-type sequence is involved here: a G-C rich palindrome (inverted repeat) precedes a sequence of 6–7 U residues in the RNA chain. As a result the RNA chain forms a stem and loop structure preceding the U residues. The secondary structure of the stem and loop is crucial for termination; base change mutations in the stem and loop that disrupt pairing also reduce termination. Furthermore, the most efficient terminators are the most G-C rich and therefore most stable. The terminator stem and loop stabilize prokaryotic mRNA against nucleolytic degradation.

Rho-dependent terminators are less well defined. Rho factor is a hexameric protein possessing an essential RNA-dependent ATPase activity. The sequences of rho-dependent termination sites feature regularly spaced C residues within a relatively unstructured length of the transcript. The nascent RNA is thought to wrap around rho factor while ATP hydrolysis leads to dissociation of the transcript from the template.

Prokaryotic ribosomes usually attach to the nascent mRNA while it is being transcribed. This coupling between transcription and translation is important in gene control by *attenuation*, which is discussed in Chapter 19.

Transcription in Eukaryotes Involves Many Additional Molecular Events

Eukaryotic transcription is considerably more complex than the process in prokaryotes. While the information specifying a promoter is still carried in a DNA sequence, several molecular events besides RNA polymerase binding are required for transcription initiation. First, chromatin containing the promoter sequence must be spatially accessible to the transcription machinery. Second, protein **transcription factors** distinct from RNA polymerase must bind to sequences in the promoter region for a gene to be active. Third, other sequences located some distance away from the promoter affect transcription; these sequences are termed **enhancers** and they, too, bind protein factors to stimulate transcription. Finally, recall that the eukaryotic RNA polymerase consists of three distinct enzyme forms, each specific form capable of transcribing only a

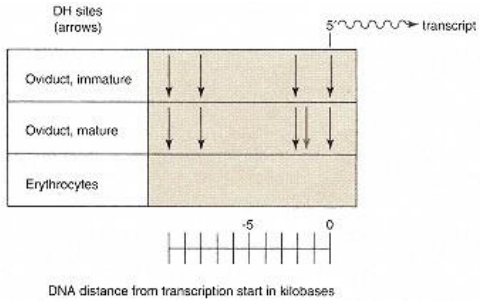


Figure 16.16
DNase-hypersensitive (DH) sites upstream of the promoter for the chick lysozyme gene, a typical eukaryotic transcriptional unit.
Hypersensitive sites, that is, sequences around the lysozyme gene where nucleosomes are not bound to the DNA, are indicated by arrows. Note that some hypersensitive sites are found in the lysozyme promoter whether the oviduct is synthesizing or not synthesizing lysozyme; the synthesis of lysozyme is accompanied by the opening up of a new hypersensitive site in mature oviduct. In contrast, no hypersensitive sites are present in nucleated erythrocytes that never synthesize lysozyme.
Adapted from Elgin, S. C. R. *J. Biol. Chem.* 263:1925, 1988.

CLINICAL CORRELATION 16.3

Fragile X Syndrome: A Chromatin Disease?

Fragile X syndrome is the single most common form of inherited mental retardation, affecting 1/1250 males and 1/2000 females. A variety of anatomical and neurological symptoms result from the inactivation of the FMR1 gene, located on the X chromosome. The genetics of the syndrome are complex due to the molecular mechanism of the Fragile X mutation.

The Fragile X condition results from the expansion of a trinucleotide repeat sequence, CGG, found at the 5 -untranslated region of the FRM1 gene. Normally, this repeat is present in 30 copies, although normal individuals can have up to 200 copies of the repeat. In individuals with Fragile X syndrome, the FMR1 gene contains many more copies, from 200 to thousands, of the CGG repeat. The complex genetics of the disease result from the potential of the CGG repeat sequence to expand from generation to generation.

The presence of an abnormally high number of CGG repeats induces extensive DNA methylation of the entire promoter region of FMR1. Methylated DNA is transcriptionally inactive, so FMR1 mRNA is not synthesized. The absence of FMR1 protein leads to the pathology of the disease.

FMR1 protein normally is located in the cytoplasm in all tissues of the early fetus and, later, especially in the fetal brain. Its sequence has some characteristics of an RNA-binding protein. One hypothesis is that the protein aids in the translation of brain-specific mRNAs during development.

Warren, S. L., and Nelson, D. L. Advances in molecular analysis of Fragile X syndrome. JAMA 271:536, 1994; and Caskey, C. T. Triple repeat mutations in human disease. *Science* 256:784, 1992.

single class of cellular RNA. By contrast, transcription in prokaryotes requires, in the simplest case, only an appropriate sequence of DNA, RNA polymerase holoenzyme, and nucleoside triphosphate substrates.

The Nature of Active Chromatin

The structural organization of eukaryotic chromosomes was discussed in Chapter 14. Although chromatin is organized into **nucleosomes** whether or not it is capable of being transcribed, an active gene has a generally "looser" configuration than does transcriptionally inactive chromatin. This difference is most striking in the promoter sequences, parts of which are not organized into nucleosomes at all (Figure 16.16). The lack of nucleosomes is manifested experimentally by the enhanced sensitivity of promoter sequences to external reagents that cleave DNA, such as the enzyme DNase I. This enhanced accessibility of promoter sequences (termed **DNase I hypersensitivity**) ensures that transcriptional factors will be able to bind to appropriate regulatory sequences. In addition, although the transcribed parts of a gene may be organized into nucleosomes, the nucleosomes are less tightly bound than those in an inactive gene. Finally, DNA may be transcriptionally inactivated by methylation (see Clin. Corr. 16.3). The overall theme is one of partially unfolded chromatin being necessary but not sufficient for transcription.

Enhancers

Enhancer sequences increase (enhance) the expression of a gene about 100-fold, hence the name. They function only when located on the same DNA molecule (chromosome) as the promoter whose activity they affect. They can function when located in either the 5 or 3 direction and as much as 1000 bp away from the relevant promoter. Protein factors bind to enhancer DNA and are necessary for enhancer function.

Transcription of Ribosomal RNA Genes

Recall that rRNA genes are located in a specialized nuclear structure, the nucleolus. There are several hundred copies of each rRNA gene in a eukaryotic cell, tandemly repeated in the DNA of a specific region of one chromosome, the **nucleolar organizer.** The repeat units contain a copy of each RNA sequence (28S, 5.8S, and 18S) and are separated from each other by **nontranscribed spacer regions.** Figure 16.17 is a diagram of this arrangement. Each repeat unit is transcribed as a unit, yielding a primary transcript containing one copy each of the 28S, 5.8S, and 18S sequences, ensuring synthesis of equimolar amounts of these three RNAs. The primary transcript is then processed by ribonucleases and modifying enzymes to the three mature rRNA species (see Section 16.5). Termination of transcription occurs within the nontranscribed spacer region before RNA polymerase I reaches the promoter of the next repeat unit.

The promoter recognized by RNA polymerase I is located within the non-transcribed spacer, from about positions –40 to +10 and from –150 to –110. A transcription factor binds to the promoter and thereby directs RNA polymerase

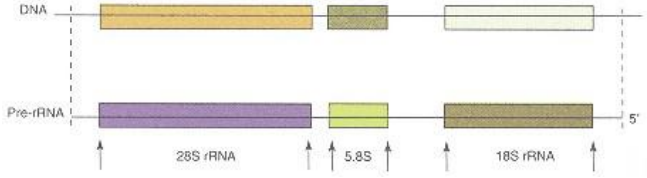


Figure 16.17
Structure of a rRNA transcription unit.
Ribosomal RNA genes are arranged with many copies one after another. Each copy is transcribed separately and each transcript is processed into three separate RNA species. Promoter and enhancer sequences are located in the nontranscribed regions of the tandemly repeated sequences.

recognition of the promoter sequence. In addition, an enhancer element is located about 250 bp upstream from the promoter in human ribosomal DNA. The size of the nontranscribed spacer varies considerably from one organism to the next, as does the position of the enhancer element.

Transcription of rRNA can be very rapid; this reflects the fact that synthesis of ribosomes is rate-limiting for cell growth. Phosphorylation of RNA polymerase I may activate especially rapid transcription of rRNA, for example, during embryonic growth or liver regeneration.

Transcription by RNA Polymerase II

RNA polymerase II is responsible for the synthesis of mRNA in the nucleus. Three common themes have emerged from research on a large number of genes (Figure 16.18). (1) The DNA sequences controlling transcription are complex; a single gene may be controlled by as many as six or eight DNA sequence elements in addition to the promoter (RNA polymerase binding region) itself. The controlling sequence elements function in combination to give a finely tuned pattern of control. (2) The effect of the controlling sequences on transcription is mediated by the binding of protein molecules to each sequence element. These transcription factors recognize the nucleotide sequence of the appropriate controlling sequence element. (3) Bound transcription factors bind with each other and with RNA polymerase to activate transcription. The DNA binding and activation activities of the factors reside in separate domains of the proteins.

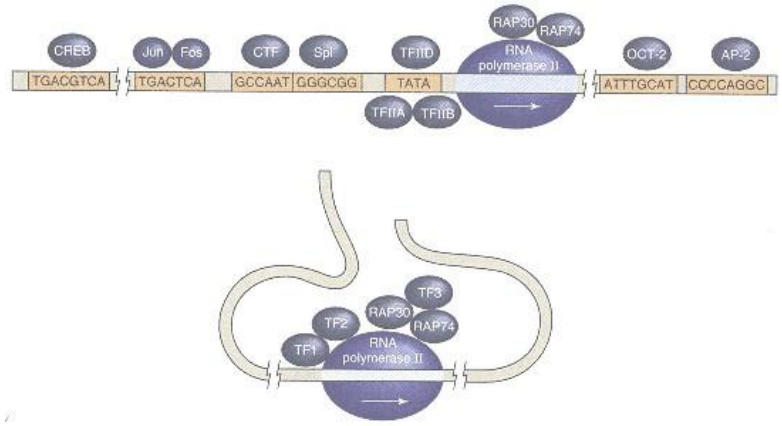


Figure 16.18
Interaction of transcription factors with promoters.
A large number of transcriptional factors interact with eukaryotic promoter regions.
(a) A hypothetical array of factors that interact with specific DNA sequences near the promoter. This includes a factor, TFIID, which binds to the TATA box and the Jun and Fos proteins, which are proto-oncogenes (Clin. Corr. 16.4). The figure is not meant to imply that all of the DNA binding factors bind to the promoter simultaneously.
(b) One way in which the DNA binding factors are hypothesized to bind to each other and to RNA polymerase. Although this model is not completely proved, it is known that proteins that bind to distant DNA sequences make protein–protein contacts with each other.
Reprinted with permission from Mitchell, P. J., and Tjian, R. *Science* 245:371, 1989.

Promoters for mRNA Synthesis

In contrast to prokaryotic RNA polymerase which recognizes only a single promoter sequence, RNA polymerase II can initiate transcription by recognizing several classes of consensus sequences upstream from the mRNA start site. The first and most prominent of these, sometimes called the **TATA box,** has the sequence

```
   A A
TATA A
   T T
```

The TATA box is centered about 25 bp upstream from the transcription unit. Experiments in which it was deleted suggest that it is required for efficient transcription, although some promoters may lack it entirely.

A second region of homology is located further upstream, in which the **CAAT box** sequence

```
  T
GG CAATCT
  C
```

is found. This sequence is not as highly conserved as the TATA box, and some active promoters may not possess it. Other sequences, described in Figure 16.18, may also promote transcription. The CAAT and TATA boxes, as well as the other sequences shown in Figure 16.15, do not contact RNA polymerase II directly. Rather, they require the binding of specific transcription factors to function. The current model for the **activation of genes** in this manner is shown in Figure 16.18. Note how protein factors bind not only to their recognition sequences but also to each other and to RNA polymerase, itself a very large and complex enzyme. Despite the complexities of the detailed interactions, the three principles elaborated above account for the known mechanisms of all class II transcription factors. Mutated forms of several of these transcription factors function as nuclear oncogenes (see Clin. Corr. 16.4).

Transcription by RNA Polymerase III

The themes elaborated above for the transcription of class I and class II promoters hold for the transcription of 5S RNA and tRNA by RNA polymerase III. Transcription factors bind to DNA and direct the action of RNA polymerase. One unusual feature of RNA polymerase III action in the transcription of 5S RNA is the location of the factor-binding sequence; it can be located within the DNA sequence encoding the RNA. The DNA in the region that would normally be thought of as a promoter, that is, the sequence immediately 5 to the transcribed region of the gene, has no specific sequence and can be substituted by other sequences without a substantial effect on transcription. Figure 16.19 diagrams this unusual sequence arrangement. In other cases, for example, tRNA transcription, the factor-binding sequence is located more conventionally at the 5 region of the gene, that is, preceding the transcribed sequences.

16.5— Posttranscriptional Processing

The immediate product of transcription is a **precursor RNA molecule,** called the **primary transcript,** which is modified to a mature, functional molecule. The reactions of RNA processing can include removal of extra nucleotides, base modification, addition of nucleotides, and separation of different RNA sequences by the action of specific nucleases. Finally, in eukaryotes, RNAs must be exported from the nucleus.

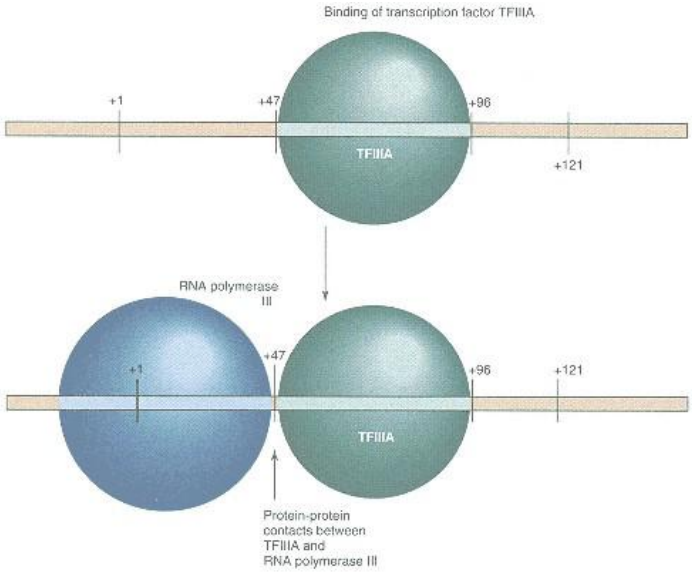


Figure 16.19
Transcription factor for a class III eukaryotic gene.
The transcription factor TFIIIA binds to a sequence located within the *Xenopus* gene for 5S rRNA. The RNA polymerase III then binds to the factor and initiates transcription of the 5S sequence. No specific sequence in the DNA is required other than the factor binding sequence.

Transfer RNA Precursors Are Modified by Cleavage, Additions, and Base Modification

Cleavage

The primary transcript of a tRNA gene contains extra nucleotide sequences both 5 and 3 to the tRNA sequence. In some cases these primary transcripts contain introns in the anticodon region of the tRNA also. Processing reactions occur in a closely defined but not necessarily rigid temporal order. First, the primary transcript is trimmed in a relatively nonspecific manner to yield a precursor molecule with shorter 5 and 3 extensions. Then ribonuclease P, a ribozyme (see above), removes the 5 extension by endonucleolytic cleavage. The 3 end is trimmed exonucleolytically, followed by synthesis of the CCA terminus. Synthesis of the modified nucleotides occurs in any order relative to the nucleolytic trimming. **Intron removal** is dictated by the secondary structure of the precursor (see Figure 16.20, p. 702) and is carried out by a soluble, two-component enzyme system; one enzyme removes the intron and the other reseals the nucleotide chain.

Additions

Each functional tRNA has the sequence CCA at its 3 terminus. In most instances this sequence is added sequentially by the enzyme **tRNA nucleotidyltransfer-**

CLINICAL CORRELATION 16.4

Involvement of Transcriptional Factors in Carcinogenesis

The conversion of a normally well-regulated cell into a cancerous one requires a number of independent steps whose end result is a transformed cell capable of uncontrolled growth and metastasis. Insights into this process have come from recombinant DNA studies of the genes whose mutated or overexpressed products contribute to carcinogenesis. These genes are termed oncogenes. Oncogenes were first identified as products of DNA or RNA tumor viruses but normal cells have copies of these genes as well. The normal, nonmutated cellular analogs of oncogenes are termed proto-oncogenes. The products of proto-oncogenes are components of the many pathways that regulate growth and differentiation of a normal cell; mutation into an oncogenic form involves a change that makes the regulatory product less responsive to normal control.

Some proto-oncogenic products are involved in the transduction of hormonal signals or the recognition of cellular growth factors and act cytoplasmically. Other proto-oncogenes have a nuclear site of action; their gene products are often associated with the transcriptional apparatus and they are synthesized in response to growth stimuli. It is easy to visualize how the overproduction or permanent activation of such a positive transcription factor could aid the transformation of a cell to malignancy: genes normally transcribed at a low or controlled level would be overexpressed by such a deranged control mechanism.

A more subtle genetic effect predisposing to cancer is exemplified by the human tumor suppressor protein p53. This protein is the product of a dominant oncogene. A single copy of the mutant gene causes Li–Fraumeni syndrome, an inherited condition predisposing to carcinomas of the breast and adrenal cortex, sarcomas, leukemia, and brain tumors.

Somatic mutations in p53 can be identified in about half of all human cancers. Mutations represent a loss of function, affecting either the stability or DNA-binding ability of p53. Thus wild-type p53 functions as a tumor suppressor. The wild-type protein helps to control the checkpoint between the G1 and S phases of the cell cycle, activates DNA repair, and, in other circumstances, leads to programmed cell death (apoptosis). Thus the biochemical actions of p53 serve to keep cell growth regulated, maintain the information content of the genome, and, finally, eliminate damaged cells. All of these functions would counteract neoplastic transformation of a cell.

These varied roles are a function of p53's action as a transcription factor, inhibiting some genes and activating others. For example, p53 inhibits transcription of genes with TATA sequences, perhaps by binding to the complex formed between transcription factors and the TATA sequence. Alternatively, p53 is a site-specific DNA-binding protein and promotes transcription of some other genes, for example, those for DNA repair.

The three-dimensional structure of p53 has been determined. Mutations found in p53 from tumors affect the DNA-binding domain of the protein. For example, nearly 20% of all mutated residues involve mutations at two positions in p53. The crystal structure of the protein–DNA complex shows that these two amino acids, both arginines, form hydrogen bonds with DNA. Arginine 248 forms hydrogen bonds in the minor groove of the DNA helix with a thymine oxygen and with a ring nitrogen of adenine. Mutation disrupts this H-bonded network and therefore the ability of p53 to regulate transcription.

Weinberg, R. A. Oncogenes, antioncogenes, and the molecular basis of multistep carcinogenesis. *Cancer Res.* 49:3713, 1989; Cho, Y., Gorina, S., Jeffrey, P. D., and Pavletich, N. P. Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. *Science* 265:346, 1994; Friend, S. p53: A glimpse at the puppet behind the shadow play. *Science* 265:334, 1994; and Harris, C. C., and Hollstein, M. Clinical implications of the p53 tumor-suppressor gene. *N. Engl. J. Med.* 329:1318, 1993.

ase. Nucleotidyltransferase uses ATP and CTP as substrates and always incorporates them into tRNA at a ratio of 2C/1A. The CCA ends are found on both cytoplasmic and mitochondrial tRNAs.

Modified Nucleosides

Transfer RNA nucleotides are the most highly modified of all nucleic acids. More than 60 different modifications to the bases and ribose, requiring well over 100 different enzymatic reactions, have been found in tRNA. Many are simple, one-step methylations, but others involve multistep synthesis. Two derivatives, **pseudouridine** and **queuosine** (7–4, 5-*cis*-dihydroxy-1-cyclopenten-3-ylamino methyl-7-deazaguanosine), actually require severing of the β-glycosidic bond of the altered nucleotide. One enzyme or set of enzymes produces a single site-specific modification in more than one species of tRNA molecule. Separate enzymes or sets of enzymes produce the same modifications at more than one location in tRNA. In other words, most modification enzymes are site or nucleotide sequence specific, not tRNA specific. Most modifications are completed before the tRNA precursors have been cleaved to mature tRNA size.

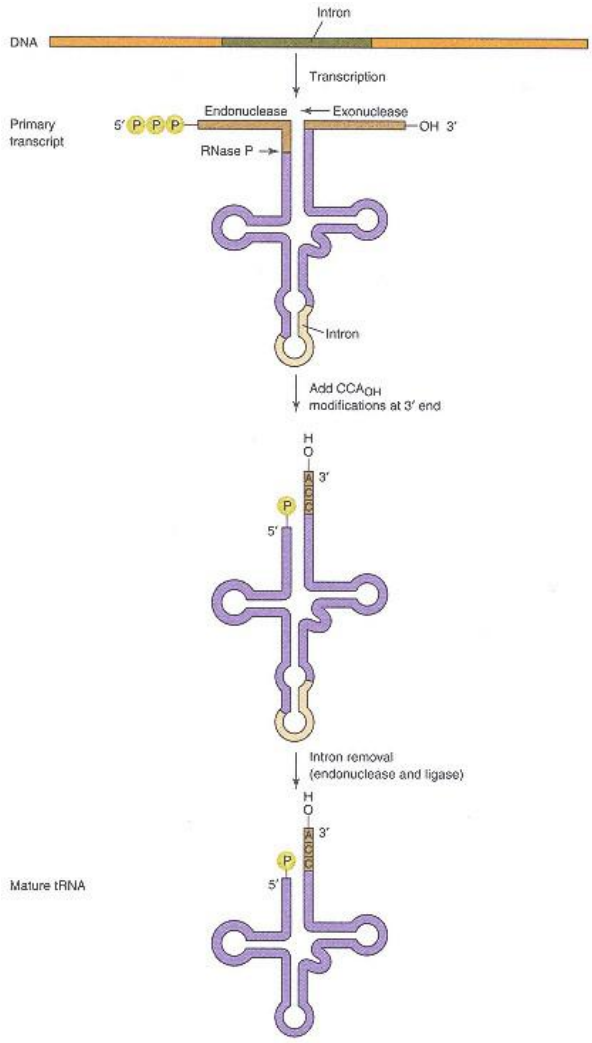


Figure 16.20
Scheme for processing a eukaryotic tRNA.
The primary transcript is cleaved by RNase P and a 3 -exonuclease, and the terminal CCA is synthesized by tRNA nucleotidyltransferase before the intron is removed, if necessary.

Ribosomal RNA Processing Releases the Various RNAs from a Longer Precursor

The primary product of rRNA transcription is a long RNA, termed 45S RNA, which contains the sequences of 28S, 5.8S, and 18S rRNAs. Processing of 45S RNA occurs in the nucleolus. Like the processing of mRNA precursors (see below), processing of the rRNA precursors is carried out by large multisubunit ribonucleoprotein assemblies. At least three RNA species are required for processing. These all function as **small nucleolar ribonucleoprotein complexes** (snoRNPs). Processing of the rRNAs follows a sequential order (Figure 16.21).

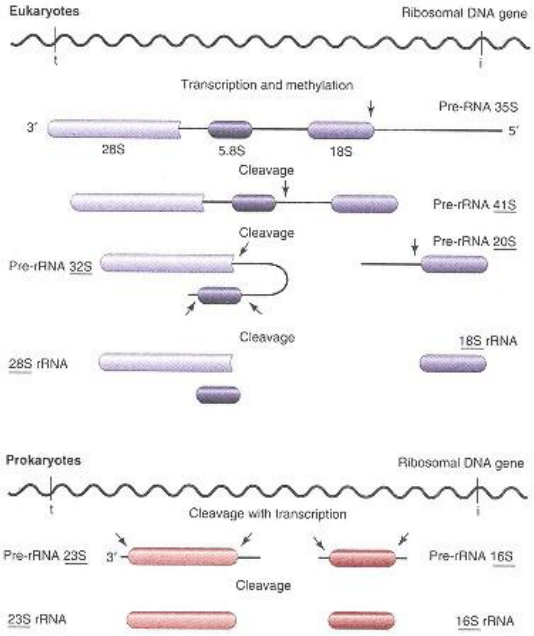


Figure 16.21
Schemes for transcription and processing of rRNAs.
Redrawn from Perry, R. *Annu. Rev. Biochem.* 45:611, 1976.

Processing of pre-rRNA in prokaryotes also involves cleavage of high molecular weight precursors to smaller molecules (see Figure 16.21). Some of the bases are modified by methylation on the ring nitrogens of the bases rather than the ribose and by the formation of pseudouridine. The *E. coli* genome has seven rRNA transcriptional units dispersed throughout the DNA. Each contains one 16S, one 23S, and one 5S rRNA or tRNA sequence. Processing of the rRNA is coupled directly to transcription, so that cleavage of a large precursor primary transcript rapidly yields pre-16S, pre-23S, pre-5S, and pre-tRNAs. These precursors are slightly larger than the functional molecules and only require trimming for maturation.

Messenger RNA Processing Requires Maintenance of the Coding Sequence

Most eukaryotic mRNAs have distinctive structural features added in the nucleus by enzyme systems other than RNA polymerase. These include the 3 -terminal poly (A) tail, methylated internal nucleotides, and the cap 5 terminus. Cytoplasmic mRNAs are shorter than their primary transcripts, which can contain additional terminal and internal sequences. Noncoding sequences present within pre-mRNA molecules, but not present in mature mRNAs, are called **intervening sequences** or **introns.** The **expressed** or **retained sequences** are called **exons.** The general pattern for mRNA processing is depicted in Figure 16.22. Incompletely processed mRNAs make up a large part of the heterogeneous nuclear RNA (hnRNA).

Processing of eukaryotic pre-mRNA involves a number of molecular reactions, all of which must be carried out with exact fidelity. This principle is most clear in the removal of introns from an mRNA transcript. An extra nucleotide in the coding sequence of mature mRNA would cause the reading frame of

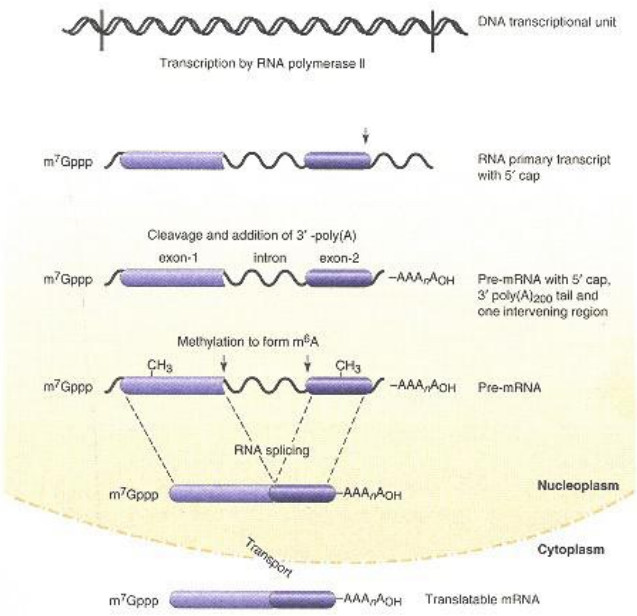


Figure 16.22
Scheme for processing mRNA.
The points for initiation and termination of transcription are indicated on the DNA. Arrows indicate cleavage points. The many proteins associated with the RNA and tertiary conformations are not shown.

that message to be shifted and the resulting protein will almost certainly be nonfunctional. Indeed, mutations in the β-globin gene that interfere with intron removal are a major cause of the genetic disease ***β*-thalassemia** (see Clin. Corr. 16.5). The task for the cell becomes even more daunting when seen in the light of the structure of some important human genes that consist of over 90% intron sequences. The complex reactions to remove introns are accomplished by multicomponent enzyme systems that act in the nucleus; after these reactions are completed the mRNA is exported to the cytoplasm where it interacts with ribosomes to initiate translation.

Blocking of the 5′ Terminus and Poly(A) Synthesis

Addition of the **cap structures** occurs during transcription by RNA polymerase II (Figure 16.22). As the transcription complex moves along the DNA, the capping enzyme complex modifies the 5 end of the nascent mRNA. This is the only eukaryotic pre-mRNA processing event that is known to occur cotranscriptionally, that is, while RNA polymerase is still transcribing the downstream portions of the gene.

After initiation and cap synthesis, RNA polymerase continues transcribing the gene until a **polyadenylation signal sequence** is reached (Figure 16.23). This sequence, which has the consensus AAUAAA, appears in the mature mRNA but usually does not form part of its coding region. Rather, it signals cleavage of the nascent mRNA precursor about 20 or so nucleotides downstream. The poly(A) sequence is then added by a soluble polymerase to the free 3 end produced by this cleavage. Note that polyadenylation does not require a template. Somewhat paradoxically, RNA polymerase II continues transcription for as many as 1000 nucleotides beyond the point at which the transcript is released from chromatin. Nucleotides incorporated into RNA by this process are apparently turned over and never appear in any cytoplasmic RNA species.

CLINICAL CORRELATION 16.5

Thalassemia Due to Defects in Messenger RNA Synthesis

The thalassemias are genetic defects in the coordinated synthesis of α- and β-globin peptide chains; a deficiency of β chains is termed β-thalassemia while a deficiency of α chains is termed α-thalassemia. Patients suffering from either of these conditions present with anemia at about 6 months of age as HbF synthesis ceases and HbA synthesis would become predominant. The severity of symptoms leads to the classification of the disease into either thalassemia major, where a severe deficiency of globin synthesis occurs, or thalassemia minor, representing a less severe imbalance. Occasionally, an intermediate form is seen. Therapy for thalassemia major involves frequent transfusions, leading to a risk of complications from iron overload. Unless chelation therapy is successful, the deposition of iron in peripheral tissues, termed hemosiderosis, can lead to death before adulthood. Carriers of the disease usually have thalassemia minor, involving mild anemia. Ethnographically, the disease is common in persons of Mediterranean, Arabian, and East Asian descent. As is the case for sickle cell anemia (HbS) and glucose 6-phosphate dehydrogenase deficiency, the abnormality of the carriers' erythrocytes affords some protection from malaria. Maps of the regions where one or another of these diseases is frequent in the native population superimpose over the areas of the world where malaria is endemic.

α-Thalassemia is usually due to a genetic deletion, which can occur because the α-globin genes are duplicated; unequal crossing over between adjacent α alleles apparently has led to the loss of one or more loci. In contrast, β-thalassemia can result from a wide variety of mutations. Known events include mutations leading to frameshifts in the β-globin coding sequence, as well as mutations leading to premature termination of peptide synthesis. Many β-thalassemias result from mutations affecting the biosynthesis of β-globin mRNA. Genetic defects are known that affect the promoter of the gene, leading to inefficient transcription. Other mutations result in aberrant processing of the nascent transcript, either during splicing out of the two introns from the transcript or during polyadenylation of the mRNA precursor. Examples where the molecular defect illustrates a general principle of mRNA synthesis are discussed in the text.

Orkin, S. H. Disorders of hemoglobin synthesis: the thalassemias. In: G. Stamatoyannopoulis, A. W. Nienhuis, P. Leder, and P. W. Majerus (Eds.). *The Molecular Basis of Blood Diseases* Philadelphia: Saunders, 1987; and Weatherall, D. J., Clegg, J. B., Higgs, D. R., and Wood, W. G. The hemoglobinopathies. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.). *The Metabolic and Molecular Bases of Inherited Disease,* 7th ed. New York: McGraw-Hill, 1995.

Removal of Introns from mRNA Precursors

As pre-RNA is extruded from the RNA polymerase complex, it is rapidly bound by **small nuclear ribonucleoproteins, snRNPs** (snurps), which carry out the dual steps of **RNA splicing:** (1) breakage of the intron at the 5 donor site and (2) joining the upstream and downstream exon sequences together. All introns begin with a GU sequence and end with AG; these are termed the donor and acceptor intron–exon junctions, respectively. Not all GU or AG sequences are spliced out of RNA, however. How does the cell know which GU sequences are in introns (and therefore must be removed) and which are destined to remain in mature mRNA? This discrimination is accomplished by the formation of base pairs between **U1 RNA** and the sequence of the mRNA precursor surrounding the donor GU sequence (see Clin. Corr. 16.6). See Figure 16.24 for an illustration of this process. Another snRNP, containing **U2 RNA,** recognizes

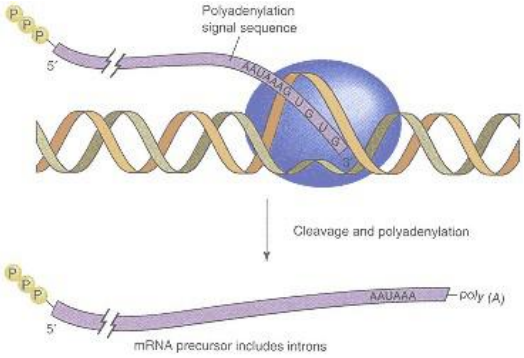


Figure 16.23
Cleavage and polyadenylation of eukaryotic mRNA precursors.
The 3 termini of eukaryotic mRNA species are derived by processing. The sequence AAUAAA in the mRNA specifies the cleavage of the mRNA precursor. The free 3 -OH end of the mRNA is a primer for poly(A) synthesis.
Adapted from Proudfoot, N. J. *Trends Biochem. Sci*. 14:105, 1989

CLINICAL CORRELATION 16.6

Autoimmunity in Connective Tissue Disease

Humoral antibodies in sera of patients with various connective tissue diseases recognize a variety of ribonucleoprotein complexes. Patients with systemic lupus erythematosus exhibit a serum antibody activity designated Sm, and those with mixed connective tissue disease exhibit an antibody designated RNP. Each antibody recognizes a distinct site on the same RNA–protein complex, U1 RNP, that is involved in mRNA processing in mammalian cells. The U1–RNP complex contains U1 RNA, a 165-nucleotide sequence highly conserved among eukaryotes, that at its 5′ terminus includes a sequence complementary to intron–exon splice junctions. Addition of this antibody to *in vitro* splicing assays inhibits splicing, presumably by removal of the U1 RNP from the reaction. Sera from patients with other connective tissue diseases recognize different nuclear antigens, nucleolar proteins, and/or chromosomal centromeres. Sera of patients with myositis have been shown to recognize cytoplasmic antigens such as aminoacyl-tRNA synthetases. Although humoral antibodies have been reported to enter cells via Fc receptors, there is no evidence that this is part of the mechanism of autoimmune disease.

important sequences at the 3′ acceptor end of the intron. Still other snRNP species, among them U5 and U6, then bind to the RNA precursor, forming a large complex termed a **spliceosome** (by analogy with the large ribonucleoprotein assembly involved in protein synthesis, the ribosome). The spliceosome uses ATP energy to carry out the accurate removal of the intron. First, the phosphodiester bond between the exon and the donor GU sequence is broken, leaving a free 3′-OH group at the end of the first exon and a 5′ phosphate on the donor G of the intron. This pG is then used to form an unusual linkage with the 2′-OH group of an adenosine within the intron to form a branched or **lariat RNA** structure, as shown in Figure 16.25. After the lariat is formed, the second step of splicing occurs. The phosphodiester bond immediately following the AG is cleaved and the two exon sequences are ligated together. In pre-mRNAs containing a large number of introns, splicing occurs roughly in order from the 5′ to the 3′ end of the mRNA precursor. However, this is not a hard and fast rule as there is no singly preferred order for removal. The end result of processing is a fully functional coding mRNA, all introns removed, and ready to direct protein synthesis.

Mutations in Splicing Signals Cause Human Diseases

Messenger RNA splicing is an intricate process dependent on many molecular events. If these events are not carried out with precision, functional mRNA is not produced. This principle is illustrated in the human thalassemias, which affect the balanced synthesis of α and β-globin chains (see Clin. Corr. 16.5). Some of the mutations leading to **β-thalassemia** interfere with the splicing of β-globin mRNA precursors. For example, we know that all intron sequences begin with the dinucleotide GU. Mutation of the G in this sequence to an A means that the splicing machinery will no longer recognize this dinucleotide as a donor site. Splicing will "pass by" the correct exon–intron junction. This could lead to two results: extra sequences that would normally be spliced out will appear in the β-globin mRNA, or, alternatively, sequences could be deleted from the mRNA product (Figure 16.26). In either event, functional β-globin will be made in reduced amounts and the anemia characteristic of the disease will result.

Alternate pre-mRNA Splicing Can Lead to Multiple Proteins Being Made from a Single DNA Coding Sequence

The existence of intron sequences is paradoxical. Introns must be removed precisely so that the mRNA can accurately encode a protein. As we have seen above, a single base mutation can drastically interfere with splicing and cause a serious disease. Furthermore, the presence of intron sequences in a gene means that its overall sequence is much larger than is required to encode its

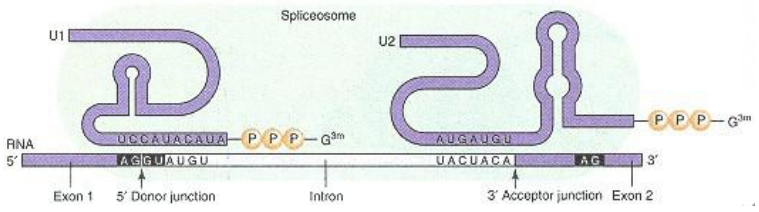


Figure 16.24
Mechanism of splice junction recognition.
The recognition of the 5′ splice junction involves base pairing between the intron–exon junction and the U1 RNA snRNP. This base pairing targets the intron for removal.
Adapted from Sharp, P. A. *JAMA* 260:3035, 1988.

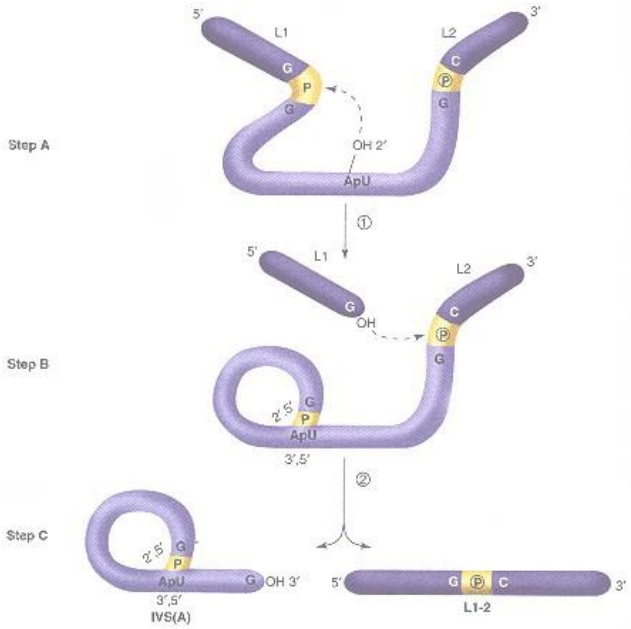


Figure 16.25
Proposed scheme for mRNA splicing to include the lariat structure.
A messenger RNA is depicted with two exons (in dark blue) and an intervening intron (in light blue). A 2 -OH group of the intron sequence reacts with the 5 -phosphate of the intron's 5 -terminal nucleotide producing a 2 –5 linkage and the lariat structure. Simultaneously, the exon 1–intron phosphodiester bond is broken, leaving a 3 -OH terminus on this exon free to react with the 5 -phosphate of the exon 2, displacing the intron and creating the spliced mRNA. The released intron lariat is subsequently digested by cellular nucleases.

protein product. A large gene is a target for more mutagenic events than is a small one. Indeed, common human genetic diseases like Duchenne muscular dystrophy occur in genes that encompass millions of base pairs of DNA information. Why has nature not removed introns completely over the long time scale of eukaryotic evolution? There are no clear answers to questions of this type but some introns do have beneficial effects.

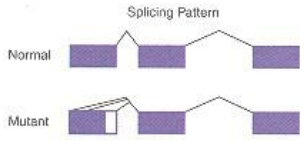


Figure 16.26
Nucleotide change at an intron–exon junction of the human β-globin gene, which leads to aberrant splicing and β-thalassemia.
This figure shows the splicing pattern of a mutated transcript containing a change of G-U to A-U at the first two nucleotides of the first intron. Loss of this invariant sequence means that the correct splice junction cannot be used; therefore transcript sequences that base pair with the U1 snRNA less well than the correct sequence junction are used as splice donors. The diagonal lines indicate the portions spliced together in mutant transcripts. Note that some of the mutant mRNA precursor molecules are spliced so that portions of the first intron (denoted as a white box) appear in the processed product. In other instances the donor junction lies within the first exon and portions of the first exon are deleted. In no case is wild-type globin mRNA produced.
Adapted from Orkin, S. H. In: G. Stamatoyannopoulis et al. (Eds.). *The Molecular Basis of Blood Diseases*. Philadelphia: Saunders, 1987.

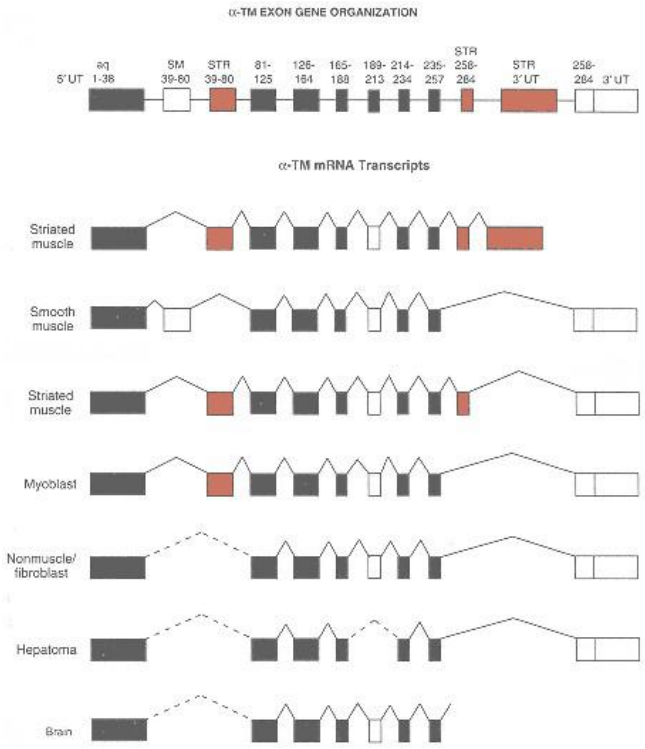


Figure 16.27
Alternate splicing of tropomyosin gene transcripts results in a family of tissue-specific tropomyosin proteins.
Redrawn from Breitbart, R. E., Andreadis, A., and Nadal-Ginard, B. *Annu. Rev. Biochem.* 56:467, 1986.

Tropomyosin proteins are essential components of the contractile apparatus in the three types of muscle (see, p. 948) and each contractile cell type contains a specific tropomyosin type. This diversity arises from a single gene that is transcribed into a primary transcript. The transcript is then processed as diagramed in Figure 16.27. All cells containing tropomyosin make the same primary transcript but each cell type processes this transcript in a characteristic fashion. The resulting mRNA species then are translated to yield the tropomyosins characteristic of each cell type. About 40 examples are well documented of tissue-specific splicing. Thus the existence of introns supplies the organism with still another method of generating protein diversity.

16.6— Nucleases and RNA Turnover

The different roles of RNA and DNA in genetic expression are reflected in their metabolic fates. A cell's information store (DNA) must be preserved, thus the myriad DNA repair and editing systems in the nucleus. Although individual stretches of nucleotides in DNA may turn over, the molecule as a whole is metabolically inert when not replicating. The various RNA molecules, on the other hand, are individually dispensable and can be replaced by newly synthe-

sized species of the same specificity. It is therefore no surprise that RNA repair systems are not known. Instead, defective RNAs are removed from the cell by degradation into nucleotides, which then are repolymerized into new RNA species.

This principle is clearest for mRNA species, which are classified as unstable. However, even the so-called stable RNAs turn over; for example, the half-life of tRNA species in liver is on the order of 5 days. A fairly long half-life for a mammalian mRNA would be 30 h. Removal of RNAs from the cytoplasm is accomplished by cellular **ribonucleases.** Messenger RNAs are initially degraded in the cytoplasm. The rates vary for different mRNA species, raising the possibility of control by differential degradation.

Two examples of the role of RNA stability in gene control illustrate how the stability of mRNA influences gene expression. **Tubulin** is the major component of the microtubules found in many cell types as part of the cytoskeleton. When there is an excess of tubulin in the cell, the monomeric protein binds to and promotes the degradation of tubulin mRNA, thereby reducing tubulin synthesis. A second example is provided by **herpes simplex viruses** (HSV), the agent causing cold sores and some genital infections. An early event in the establishment of HSV infection is the ability of the virus to destabilize all the cellular mRNA molecules, thereby reducing the competition for free ribosomes. Thus the viral proteins are more efficiently translated.

Nucleases are of several types and specificities. The most useful distinction is between **exonucleases,** which degrade RNA from either the 5 or 3 end, and **endonucleases,** which cleave phosphodiester bonds within a molecule. The products of RNase action contain either 3 or 5 terminal phosphates, and both endo- and exonucleases can be further characterized by the position (5 or 3) at which the monophosphate created by the cleavage is located.

The structure of RNA also affects nuclease action. Most ribonucleases are less efficient on regions of highly ordered RNA structure. Thus tRNAs are preferentially cleaved in unpaired regions of the sequence. On the other hand, many RNases involved in RNA processing require a defined three-dimensional structure for enzyme activity. These enzymes are discussed more fully above in the consideration of RNA processing pathways.

Bibliography

Bradshaw, R. A. (Ed.). Transcription. Special issue of *Trends Biochem. Sci.* 16, November, 1991.

Breitbart, R. E., Andreadis, A., and Nadal-Ginard, B. Alternative splicing: A ubiquitous mechanism for the generation of multiple protein isoforms from single genes. *Annu. Rev. Biochem.* 56:467, 1987.

Caskey, C.T. Triple repeat mutations in human disease. *Science* 256:784, 1992.

Cundliffe, E. How antibiotic-producing microorganisms avoid suicide. *Annu. Rev. Microbiol.* 43:207, 1989.

Das, A. Control of transcription by RNA-binding proteins. *Annu. Rev. Biochem.* 62:893, 1993.

Friend, S. p53: A glimpse at the puppet behind the shadow play. *Science* 265:334, 1994.

Gesteland, R. F., and Atkins, J. F. (Eds.). *The RNA World*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1993.

Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* 64:763, 1995.

Koleske, A. J., and Young, R. A. The RNA polymerase II holoenzyme and its implications for gene regulation. *Trends Biochem. Sci.* 20:113, 1995.

Lai, M. M. C. The molecular biology of hepatitis delta virus. *Annu. Rev. Biochem.* 64:259, 1995.

Larson, D. E., Zahradka, P., and Sells, B. H. Control points in eucaryotic ribosome biogenesis. *Biochem. Cell. Biol.* 69:5, 1991.

Orkin, S. H. Disorders of hemoglobin synthesis: The thalassemias. In: G. Stamatoyannopoulis, A. W. Nienhuis, P. Leder, and P. W. Majerus (Eds.), *The Molecular Basis of Blood Diseases*. Philadelphia: Saunders, 1987, pp. 106–126.

Pace, N. R., and Brown, J. W. Evolutionary perspective on the structure and function of ribonuclease P, a ribozyme. *J. Bacteriol.* 177:1919, 1995.

Rosenberg, M., and Court, D. Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* 12:319, 1979.

Soll, D., and Rajbhandary, U. L. (Eds.). *tRNA: Structure, Biosynthesis and Function.* Washington DC: American Society for Microbiology, 1994.

Weatherall, D.J., Clegg, J.B., Higgs, D.R., and Wood, W.G. The hemoglobinopathies. In: C. R. Scriver, A. L. Beaudet, W. S. Sly, and D. Valle (Eds.), *The Metabolic and Molecular Bases of Inherited Disease*, 7th ed. New York: McGraw-Hill, 1995.

Weinberg, R. A. Oncogenes, antioncogenes and the molecular basis of multistep carcinogenesis. *Cancer Res*. 49:3713, 1989.

Wise, J. A. Guides to the heart of the spliceosome. *Science* 262:1978, 1993.

Questions

C. N. Angstadt and J. Baggott

1. RNA:

A. incorporates both modified and unmodified purine and pyrimidine bases during transcription.

B. does not exhibit any double helical structure.

C. structures exhibit base stacking and hydrogen-bonded base pairing.

D. usually contains about 65–100 nucleotides.

E. does not exhibit Watson–Crick base pairing.

Refer to the following for Questions 2–4.

A. HnRNA

B. mRNA

C. rRNA

D. snRNA

E. tRNA

2. Has the highest percentage of modified bases of any RNA.

3. Stable RNA representing the largest percentage by weight of cellular RNA.

4. Contains both a 7-methylguanosine triphosphate cap and a poly-adenylate segment.

5. Ribozymes:

A. are any ribonucleoprotein particles.

B. are enzymes whose catalytic function resides in RNA subunits.

C. carry out self-processing reactions but cannot be considered true catalysts.

D. bind to the mRNA precursor to recognize the 5 -splice site for intron removal.

E. function only in the processing of mRNA.

6. In eukaryotic transcription:

A. RNA polymerase does not require a template.

B. all RNA is synthesized in the nucleolus.

C. consensus sequences are the only known promoter elements.

D. phosphodiester bond formation is favored, in part, because it is followed by pyrophosphate hydrolysis.

E. RNA polymerase requires a primer.

7. An enhancer:

A. is a consensus sequence in DNA located where RNA polymerase first binds.

B. may be located in various places in different genes.

C. may be located on a separate chromosome from the gene it regulates.

D. functions by binding RNA polymerase.

E. stimulates transcription in both prokaryotes and eukaryotes.

8. The sigma ( ) subunit of prokaryotic RNA polymerase:

A. is part of the core enzyme.

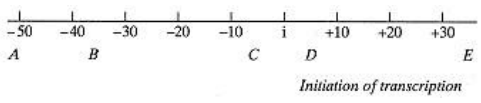
B. binds the antibiotic rifampicin.

C. is inhibited by αamanitin.

D. must be present for transcription to occur.

E. specifically recognizes promoter sites.

Use this schematic representation of a prokaryotic gene to answer Questions 9–11. Numbers refer to positions of base pairs relative to the beginning of transcription.



9. Sigma ( ) factor might be released from RNA polymerase.

10. An "open complex" should form in this region.

11. Events beyond this region should be catalyzed by core enzyme.

12. Termination of a prokaryotic transcript:

A. is a random process.

B. requires the presence of the rho subunit of the holoenzyme.

C. does not require rho factor if the end of the gene contains a G-C rich palindrome.

D. is most efficient if there is an A-T rich segment at the end of the gene.

E. requires an ATPase in addition to rho factor.

13. Eukaryotic transcription:

A. is independent of the presence of consensus sequences upstream from the start of transcription.

B. may involve a promoter located within the region transcribed rather than upstream.

C. requires a separate promoter region for each of the three ribosomal RNAs transcribed.

D. requires that the entire gene be in the nucleosome form of chromatin.

E. is affected by enhancer sequences only if they are adjacent to the promoter.

14. All of the following are correct about a primary transcript in eukaryotes EXCEPT it:

A. is usually longer than the functional RNA.

B. may contain nucleotide sequences that are not present in functional RNA.

C. will contain no modified bases.

D. usually contains information for more than one RNA molecule.

E. contains a TATA box.

15. The processing of transfer RNA involves all of the following EXCEPT:

A. addition of a methylated guanosine at the 5 end.

B. cleavage of extra bases from both the 3 and 5 ends.

C. nucleotide sequence-specific methylation of bases.

D. addition of the sequence CCA by a nucleotidyl transferase.

E. sometimes, removal of intron from the anticodon region.

16. Cleavage and splicing:

A. are features of ribosomal RNA processing.

B. always occur in the same way for a given primary transcript.

C. remove noninformational sequences occurring anywhere within a primary transcript.

D. are usually the first events in mRNA processing.

E. are catalyzed by enzymes that recognize and remove specific introns.

17. In the cellular degradation of RNA:

A. any of the nucleotides released may be recycled.

B. regions of extensive base pairing are more susceptible to cleavage.

C. endonucleases may cleave the molecule starting at either the 5 or 3 end.

D. the products are nucleotides with a phosphate at either the 3 - or 5 -OH group.

E. all species except rRNA are cleaved.

Answers

1. C Stacking stabilizes the single-stranded helix. A: Only the four bases A, G, U, and C are incorporated during transcription. B and C: Although single stranded, RNA exhibits considerable secondary and tertiary structure. D: Only tRNA would be this small; sizes can range to more than 6000 nucleotides. E: This occurs in the intrachain helical regions (pp. 679–680).

2. E Modified bases seem to be very important in the three-dimensional structure of tRNA (p. 683).

3. C Stability of rRNA is necessary for repeated functioning of ribosomes (p. 683).

4. B These are important additions during processing that yield a functional eukaryotic mRNA (p. 685, Table 16.1).

5. B A: Ribozymes are a very specific type of particle. C: One of the four classes, RNase P, catalyzes a cleavage reaction. D: This is the function of one of the snRNPs, several of which binding to mRNA result in a spliceosome. E: Ribozymes have been implicated in the processing of ribosomal and tRNAs (p. 686).

6. D This is an important mechanism for driving reactions. A and B: Transcription is directed by the genetic code, generating rRNA precursors in the nucleolus and mRNA and tRNA precursors in nucleoplasm. C: Eukaryotic transcription may have internal promoter regions as well as enhancers. E: This is a difference from DNA polymerase (p. 689).

7. B B and C: Enhancer sequences seem to work whether they are at the beginning or end of the gene, but they must be on the same DNA strand as the transcribed gene. D: They seem to function by binding proteins which themselves bind RNA polymerase (p. 697).

8. E A, D, and E: Sigma factor is required for correct initiation and dissociates from the core enzyme after the first bonds have been formed. Core enzyme can transcribe but cannot correctly initiate transcription. B and C: Rifampicin binds to the β subunit, and α-amanitin is an inhibitor of eukaryotic polymerases (p. 691).

9. D Sigma factor is released when, or a short time after, the initial bond is formed.

10. C The high A-T content of the Pribnow box is believed to facilitate initial unwinding.

11. E Elongation, which requires only the core enzyme, is well underway in this region (p. 695, Figure 16.2).

12. C A, B, and E: There is a rho-dependent as well as a rho-independent process. Rho is a separate protein from RNA polymerase and appears to possess ATPase activity (p. 696). C and D: Rho-independent termination involves secondary structure, which is stabilized by high G-C content.

13. B RNA polymerase III uses an internal promoter. A: RNA polymerase II activity involves the TATA and CAAT boxes. C: RNA polymerase I produces one transcript, which is later processed to yield three rRNAs. D: Parts of the promoter are not in a nucleosome. E: Enhancers may be as much as 1000 bp away (pp. 696–697).

14. E The TATA box is part of the promoter, which is not transcribed. A–D: Modification of bases, cleavage, and splicing are all important events in posttranscriptional processing to form functional molecules (pp. 700–707).

15. A Capping is a feature of mRNA. B: The primary transcript is longer than the functional molecule. C: The same modifications, catalyzed by a certain (set of) enzyme(s), occurs at more than one location. D: This is a posttranscriptional modification (pp. 700–701).

16. C A: Cleavage occurs, but splicing does not. B: Alternate splicing leads to different proteins from a single gene. D: Splicing occurs after other events. E: Specificity of cleavage is related to specific sequences at the intron-exon junctions, not to the sequence of the intron itself (pp. 702–707).

17. D A: Modified bases cannot be recycled. B: Although some enzymes of maturation may require an ordered structure, degradative enzymes are less efficient on an ordered structure. C: An endonuclease cleaves an interior phosphodiester bond. E: Even rRNA turns over although it is more stable than the other species (p. 709).