

# chapter 25

## DNA METABOLISM

- 25.1 DNA Replication 950
- 25.2 DNA Repair 966
- 25.3 DNA Recombination 978

I have never known a dull enzyme.

—Arthur Kornberg, in the essay "For Love of Enzymes," 1975

As the repository of genetic information, DNA occupies a unique and central place among biological macromolecules. The nucleotide sequences of DNA encode the primary structures of all cellular RNAs and proteins and, through enzymes, indirectly affect the synthesis of all other cellular constituents. This passage of information from DNA to RNA and protein guides the size, shape, and functioning of every living thing.

DNA is a marvelous device for the stable storage of genetic information. The phrase "stable storage," however, conveys a static and misleading picture. It fails to capture the complexity of processes by which genetic information is preserved in an uncorrupted state and then transmitted from one generation of cells to the next. DNA metabolism comprises both the process by which copies of DNA molecules are faithfully made (replication) and the processes that affect the inherent structure of the information (repair and recombination). Together, these activities are the focus of this chapter.

The metabolism of DNA is shaped by the requirement for an exquisite degree of accuracy. The chemistry of joining one nucleotide to the next in DNA replication is elegant and simple, almost deceptively so. As we shall see, complexity arises in the form of enzymatic devices to ensure that the genetic information is transmitted intact. Uncorrected errors that arise during DNA synthe-

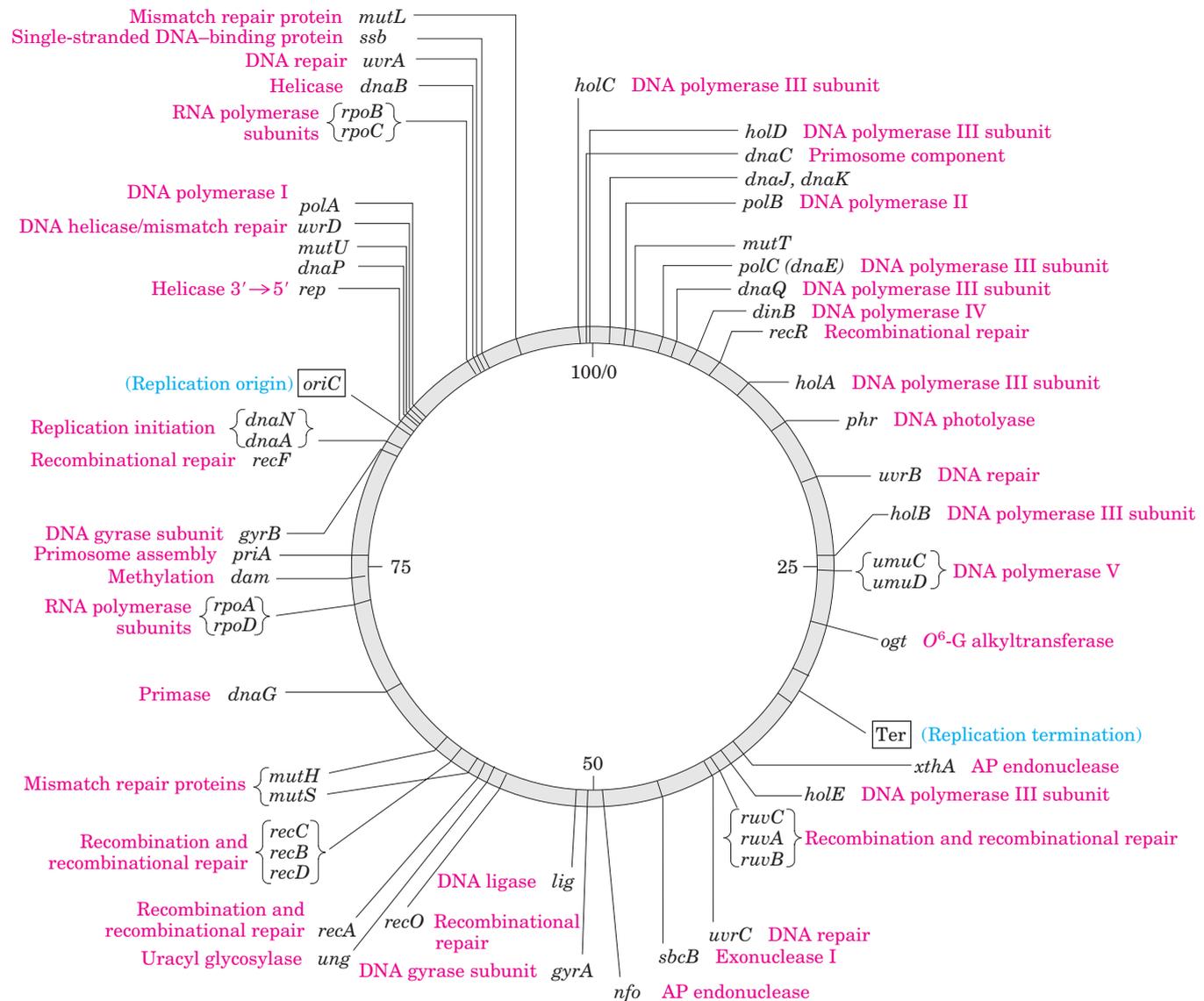
sis can have dire consequences, not only because they can permanently affect or eliminate the function of a gene but also because the change is inheritable.

The enzymes that synthesize DNA may copy DNA molecules that contain millions of bases. They do so with extraordinary fidelity and speed, even though the DNA substrate is highly compacted and bound with other proteins. Formation of phosphodiester bonds to link nucleotides in the backbone of a growing DNA strand is therefore only one part of an elaborate process that requires myriad proteins and enzymes.

Maintaining the integrity of genetic information lies at the heart of DNA repair. As detailed in Chapter 8, DNA is susceptible to many types of damaging reactions. Such reactions are infrequent but significant nevertheless, because of the very low biological tolerance for changes in DNA sequence. DNA is the only macromolecule for which repair systems exist; the number, diversity, and complexity of DNA repair mechanisms reflect the wide range of insults that can harm DNA.

Cells can rearrange their genetic information by processes collectively called recombination—seemingly undermining the principle that the stability and integrity of genetic information are paramount. However, most DNA rearrangements in fact play constructive roles in maintaining genomic integrity, contributing in special ways to DNA replication, DNA repair, and chromosome segregation.

Special emphasis is given in this chapter to the enzymes of DNA metabolism. They merit careful study both for their increasing importance in medicine and for their everyday use as reagents in a wide range of modern biochemical technologies. Many of the seminal discoveries in DNA metabolism have been made with *Escherichia coli*, so its well-understood enzymes are generally used to illustrate the ground rules. A quick look at some relevant genes on the *E. coli* genetic map (Fig. 25-1) provides just a hint of the complexity of the enzymatic systems involved in DNA metabolism.



**FIGURE 25-1** Map of the *E. coli* chromosome. The map shows the relative positions of genes encoding many of the proteins important in DNA metabolism. The number of genes known to be involved provides a hint of the complexity of these processes. The numbers 0 to 100 inside the circular chromosome denote a genetic measurement called minutes. Each minute corresponds to ~40,000 bp along the

DNA molecule of *E. coli*. The three-letter names of genes and other elements generally reflect some aspect of their function. These include *mut*, mutagenesis; *dna*, DNA replication; *pol*, DNA polymerase; *rpo*, RNA polymerase; *uvr*, UV resistance; *rec*, recombination; *dam*, DNA adenine methylation; *lig*, DNA ligase; *Ter*, termination of replication; and *ori*, origin of replication.

**A Word about Terminology** Before beginning to look closely at replication, we must make a short digression into the use of abbreviations in naming genes and proteins. By convention, bacterial genes generally are named using three lowercase, italicized letters that often reflect their apparent function. For example, the *dna*, *uvr*, and *rec* genes affect DNA replication, resistance to the damaging effects of UV radiation, and recombination, respectively. Where several genes affect the same process, the letters A, B, C, and so forth, are added—as in *dnaA*, *dnaB*, *dnaQ*, for example—usually

reflecting their order of discovery rather than their order in a reaction sequence.

During genetic investigations, the protein product of each gene is usually isolated and characterized. Many bacterial genes have been identified and named before the roles of their protein products are understood in detail. Sometimes the gene product is found to be a previously isolated protein, and some renaming occurs. Often the product turns out to be an as yet unknown protein, with an activity not easily described by a simple enzyme name. In a practice that can be confusing,

these bacterial proteins often retain the name of their genes. When referring to the protein, roman type is used and the first letter is capitalized: for example, the *dnaA* and *recA* gene products are called the DnaA and RecA proteins, respectively. You will encounter many such examples in this chapter.

Similar conventions exist for the naming of eukaryotic genes, although the exact form of the abbreviations may vary with the species and no single convention applies to all eukaryotic systems.

## 25.1 DNA Replication

Long before the structure of DNA was known, scientists wondered at the ability of organisms to create faithful copies of themselves and, later, at the ability of cells to produce many identical copies of large and complex macromolecules. Speculation about these problems centered around the concept of a **template**, a structure that would allow molecules to be lined up in a specific order and joined, to create a macromolecule with a unique sequence and function. The 1940s brought the revelation that DNA was the genetic molecule, but not until James Watson and Francis Crick deduced its structure did the way in which DNA could act as a template for the replication and transmission of genetic information become clear: *one strand is the complement of the other*. The strict base-pairing rules mean that each strand provides the template for a sister strand with a predictable and complementary sequence (see Figs 8–16, 8–17).  **Nucleotides: Building Blocks of Nucleic Acids**

The fundamental properties of the DNA replication process and the mechanisms used by the enzymes that catalyze it have proved to be essentially identical in all species. This mechanistic unity is a major theme as we proceed from general properties of the replication process, to *E. coli* replication enzymes, and, finally, to replication in eukaryotes.

### DNA Replication Follows a Set of Fundamental Rules

Early research on bacterial DNA replication and its enzymes helped to establish several basic properties that have proven applicable to DNA synthesis in every organism.

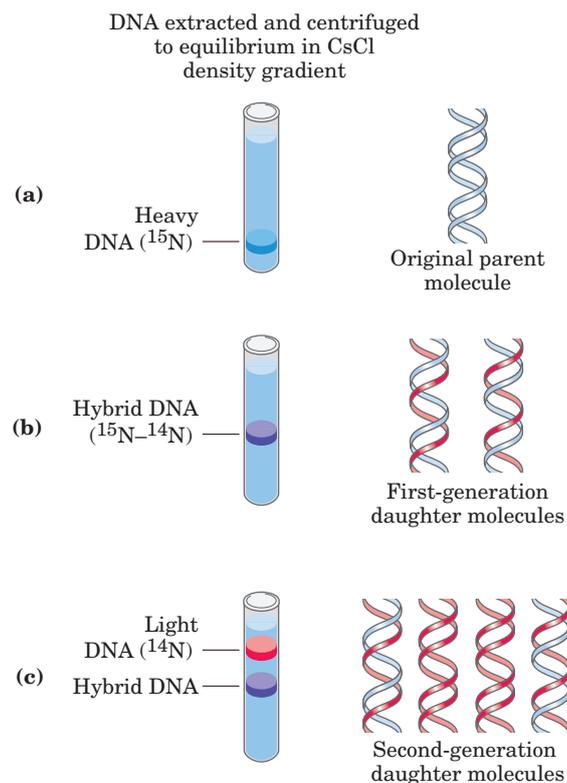
**DNA Replication Is Semiconservative** Each DNA strand serves as a template for the synthesis of a new strand, producing two new DNA molecules, each with one new strand and one old strand. This is **semiconservative replication**.

Watson and Crick proposed the hypothesis of semiconservative replication soon after publication of their 1953 paper on the structure of DNA, and the hypothesis was proved by ingeniously designed experiments carried out by Matthew Meselson and Franklin Stahl in

1957. Meselson and Stahl grew *E. coli* cells for many generations in a medium in which the sole nitrogen source ( $\text{NH}_4\text{Cl}$ ) contained  $^{15}\text{N}$ , the “heavy” isotope of nitrogen, instead of the normal, more abundant “light” isotope,  $^{14}\text{N}$ . The DNA isolated from these cells had a density about 1% greater than that of normal [ $^{14}\text{N}$ ]DNA (Fig. 25–2a). Although this is only a small difference, a mixture of heavy [ $^{15}\text{N}$ ]DNA and light [ $^{14}\text{N}$ ]DNA can be separated by centrifugation to equilibrium in a cesium chloride density gradient.

The *E. coli* cells grown in the  $^{15}\text{N}$  medium were transferred to a fresh medium containing only the  $^{14}\text{N}$  isotope, where they were allowed to grow until the cell population had just doubled. The DNA isolated from these first-generation cells formed a *single* band in the CsCl gradient at a position indicating that the double-helical DNA molecules of the daughter cells were hybrids containing one new  $^{14}\text{N}$  strand and one parent  $^{15}\text{N}$  strand (Fig. 25–2b).

This result argued against conservative replication, an alternative hypothesis in which one progeny DNA



**FIGURE 25-2** The Meselson-Stahl experiment. (a) Cells were grown for many generations in a medium containing only heavy nitrogen,  $^{15}\text{N}$ , so that all the nitrogen in their DNA was  $^{15}\text{N}$ , as shown by a single band (blue) when centrifuged in a CsCl density gradient. (b) Once the cells had been transferred to a medium containing only light nitrogen,  $^{14}\text{N}$ , cellular DNA isolated after one generation equilibrated at a higher position in the density gradient (purple band). (c) Continuation of replication for a second generation yielded two hybrid DNAs and two light DNAs (red), confirming semiconservative replication.

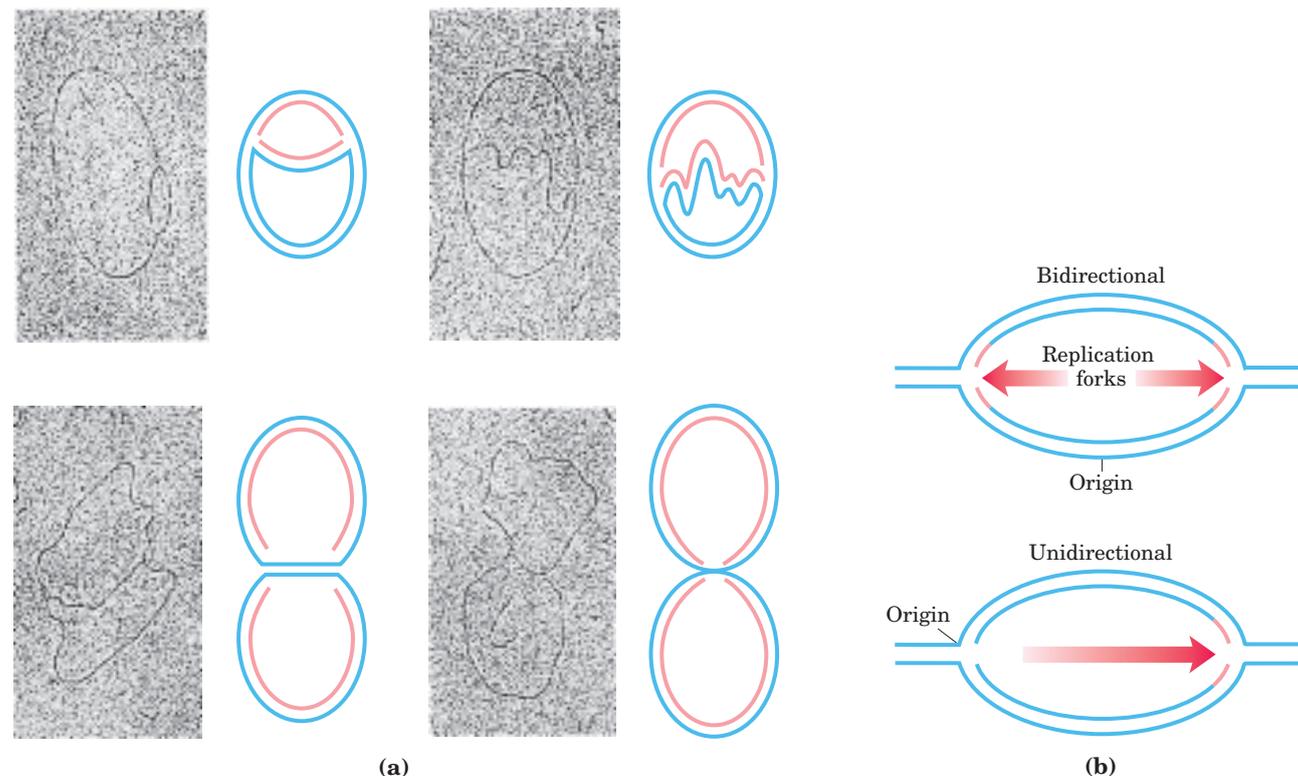
molecule would consist of two newly synthesized DNA strands and the other would contain the two parent strands; this would not yield hybrid DNA molecules in the Meselson-Stahl experiment. The semiconservative replication hypothesis was further supported in the next step of the experiment (Fig. 25-2c). Cells were again allowed to double in number in the  $^{14}\text{N}$  medium. The isolated DNA product of this second cycle of replication exhibited *two* bands in the density gradient, one with a density equal to that of light DNA and the other with the density of the hybrid DNA observed after the first cell doubling.

**Replication Begins at an Origin and Usually Proceeds Bidirectionally** Following the confirmation of a semiconservative mechanism of replication, a host of questions arose. Are the parent DNA strands completely unwound before each is replicated? Does replication begin at random places or at a unique point? After initiation at any point in the DNA, does replication proceed in one direction or both?

An early indication that replication is a highly coordinated process in which the parent strands are simultaneously unwound and replicated was provided by

John Cairns, using autoradiography. He made *E. coli* DNA radioactive by growing cells in a medium containing thymidine labeled with tritium ( $^3\text{H}$ ). When the DNA was carefully isolated, spread, and overlaid with a photographic emulsion for several weeks, the radioactive thymidine residues generated “tracks” of silver grains in the emulsion, producing an image of the DNA molecule. These tracks revealed that the intact chromosome of *E. coli* is a single huge circle, 1.7 mm long. Radioactive DNA isolated from cells during replication showed an extra loop (Fig. 25-3a). Cairns concluded that the loop resulted from the formation of two radioactive daughter strands, each complementary to a parent strand. One or both ends of the loop are dynamic points, termed **replication forks**, where parent DNA is being unwound and the separated strands quickly replicated. Cairns’s results demonstrated that both DNA strands are replicated simultaneously, and a variation on his experiment (Fig. 25-3b) indicated that replication of bacterial chromosomes is bidirectional: both ends of the loop have active replication forks.

The determination of whether the replication loops originate at a unique point in the DNA required landmarks along the DNA molecule. These were provided



**FIGURE 25-3 Visualization of bidirectional DNA replication.** Replication of a circular chromosome produces a structure resembling the Greek letter theta ( $\theta$ ). **(a)** Labeling with tritium ( $^3\text{H}$ ) shows that both strands are replicated at the same time (new strands shown in red). The electron micrographs illustrate the replication of a circular *E. coli* plasmid as visualized by autoradiography. **(b)** Addition of  $^3\text{H}$  for a

short period just before the reaction is stopped allows a distinction to be made between unidirectional and bidirectional replication, by determining whether label (red) is found at one or both replication forks in autoradiograms. This technique has revealed bidirectional replication in *E. coli*, *Bacillus subtilis*, and other bacteria.

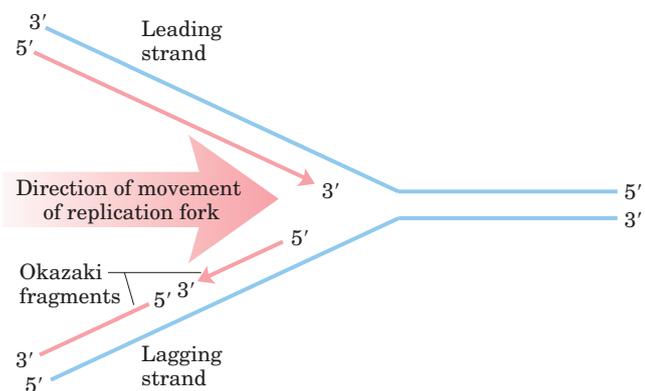
by a technique called **denaturation mapping**, developed by Ross Inman and colleagues. Using the 48,502 bp chromosome of bacteriophage  $\lambda$ , Inman showed that DNA could be selectively denatured at sequences unusually rich in A=T base pairs, generating a reproducible pattern of single-strand bubbles (see Fig. 8–31). Isolated DNA containing replication loops can be partially denatured in the same way. This allows the position and progress of the replication forks to be measured and mapped, using the denatured regions as points of reference. The technique revealed that in this system the replication loops always initiate at a unique point, which was termed an **origin**. It also confirmed the earlier observation that replication is usually bidirectional. For circular DNA molecules, the two replication forks meet at a point on the side of the circle opposite to the origin. Specific origins of replication have since been identified and characterized in bacteria and lower eukaryotes.

**DNA Synthesis Proceeds in a 5'→3' Direction and Is Semidiscontinuous** A new strand of DNA is always synthesized in the 5'→3' direction, with the free 3' OH as the point at which the DNA is elongated (the 5' and 3' ends of a DNA strand are defined in Fig. 8–7). Because the two DNA strands are antiparallel, the strand serving as the template is read from its 3' end toward its 5' end.

If synthesis always proceeds in the 5'→3' direction, how can both strands be synthesized simultaneously? If both strands were synthesized *continuously* while the replication fork moved, one strand would have to undergo 3'→5' synthesis. This problem was resolved by Reiji Okazaki and colleagues in the 1960s. Okazaki found that one of the new DNA strands is synthesized in short pieces, now called **Okazaki fragments**. This work ultimately led to the conclusion that one strand is synthesized continuously and the other discontinuously (Fig. 25–4). The continuous strand, or **leading strand**, is the one in which 5'→3' synthesis proceeds in the *same* direction as replication fork movement. The discontinuous strand, or **lagging strand**, is the one in which 5'→3' synthesis proceeds in the direction *opposite* to the direction of fork movement. Okazaki fragments range in length from a few hundred to a few thousand nucleotides, depending on the cell type. As we shall see later, leading and lagging strand syntheses are tightly coordinated.

### DNA Is Degraded by Nucleases

To explain the enzymology of DNA replication, we first introduce the enzymes that degrade DNA rather than synthesize it. These enzymes are known as **nucleases**, or **DNases** if they are specific for DNA rather than RNA. Every cell contains several different nucleases, belonging to two broad classes: exonucleases and endonucleases. **Exonucleases** degrade nucleic acids from one



**FIGURE 25–4 Defining DNA strands at the replication fork.** A new DNA strand (red) is always synthesized in the 5'→3' direction. The template is read in the opposite direction, 3'→5'. The leading strand is continuously synthesized in the direction taken by the replication fork. The other strand, the lagging strand, is synthesized discontinuously in short pieces (Okazaki fragments) in a direction opposite to that in which the replication fork moves. The Okazaki fragments are spliced together by DNA ligase. In bacteria, Okazaki fragments are ~1,000 to 2,000 nucleotides long. In eukaryotic cells, they are 150 to 200 nucleotides long.

end of the molecule. Many operate in only the 5'→3' or the 3'→5' direction, removing nucleotides only from the 5' or the 3' end, respectively, of one strand of a double-stranded nucleic acid or of a single-stranded DNA. **Endonucleases** can begin to degrade at specific internal sites in a nucleic acid strand or molecule, reducing it to smaller and smaller fragments. A few exonucleases and endonucleases degrade only single-stranded DNA. There are a few important classes of endonucleases that cleave only at specific nucleotide sequences (such as the restriction endonucleases that are so important in biotechnology; see Chapter 9, Fig. 9–3). You will encounter many types of nucleases in this and subsequent chapters.

### DNA Is Synthesized by DNA Polymerases

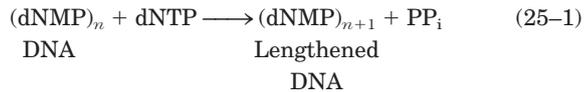
The search for an enzyme that could synthesize DNA began in 1955. Work by Arthur Kornberg and colleagues led to the purification and characterization of DNA polymerase from *E. coli* cells, a single-polypeptide enzyme now called **DNA polymerase I** ( $M_r$  103,000; encoded by the *polA* gene). Much later, investigators found that *E. coli* contains at least four other distinct DNA polymerases, described below.



Arthur Kornberg

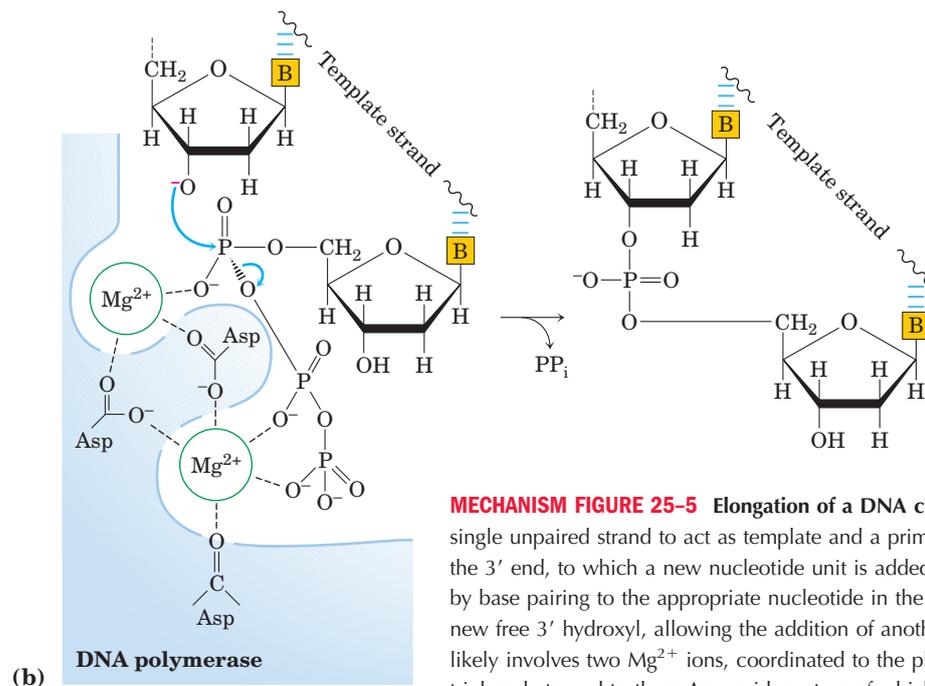
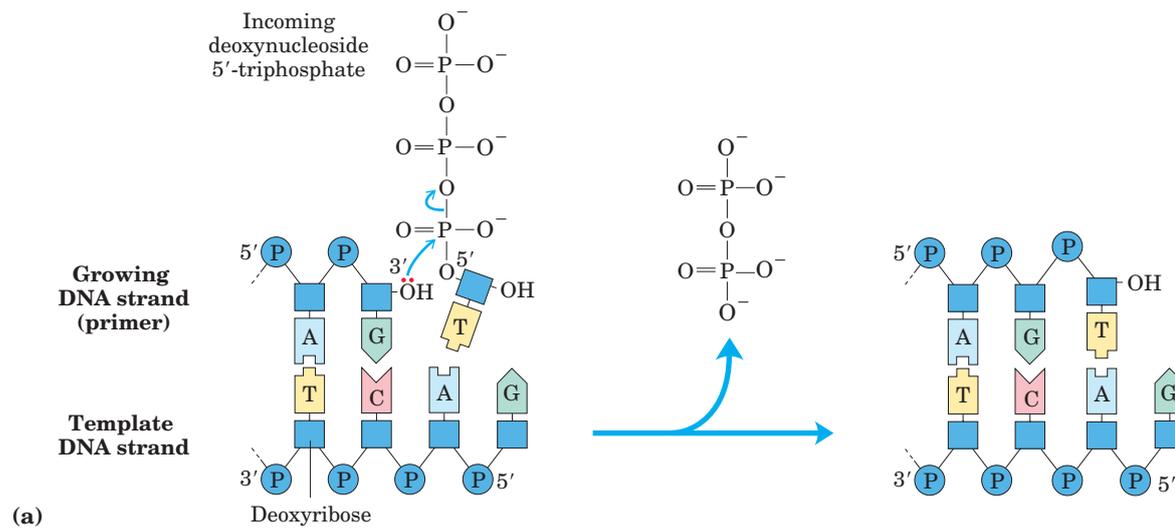
Detailed studies of DNA polymerase I revealed features of the DNA synthetic process that are now known to be common to all DNA polymerases. The fundamen-

tal reaction is a phosphoryl group transfer. The nucleophile is the 3'-hydroxyl group of the nucleotide at the 3' end of the growing strand. Nucleophilic attack occurs at the  $\alpha$  phosphorus of the incoming deoxynucleoside 5'-triphosphate (Fig. 25-5). Inorganic pyrophosphate is released in the reaction. The general reaction is



where dNMP and dNTP are deoxynucleoside 5'-mono-

phosphate and 5'-triphosphate, respectively. The reaction appears to proceed with only a minimal change in free energy, given that one phosphodiester bond is formed at the expense of a somewhat less stable phosphate anhydride. However, noncovalent base-stacking and base-pairing interactions provide additional stabilization to the lengthened DNA product relative to the free nucleotide. Also, the formation of products is facilitated in the cell by the 19 kJ/mol generated in the subsequent hydrolysis of the pyrophosphate product by the enzyme pyrophosphatase.



**MECHANISM FIGURE 25-5** Elongation of a DNA chain. (a) DNA polymerase I activity requires a single unpaired strand to act as template and a primer strand to provide a free hydroxyl group at the 3' end, to which a new nucleotide unit is added. Each incoming nucleotide is selected in part by base pairing to the appropriate nucleotide in the template strand. The reaction product has a new free 3' hydroxyl, allowing the addition of another nucleotide. (b) The catalytic mechanism likely involves two  $\text{Mg}^{2+}$  ions, coordinated to the phosphate groups of the incoming nucleotide triphosphate and to three Asp residues, two of which are highly conserved in all DNA polymerases. The top  $\text{Mg}^{2+}$  ion in the figure facilitates attack of the 3'-hydroxyl group of the primer on the  $\alpha$  phosphate of the nucleotide triphosphate; the lower  $\text{Mg}^{2+}$  ion facilitates displacement of the pyrophosphate. Both ions stabilize the structure of the pentacovalent transition state. RNA polymerases use a similar mechanism (See Fig. 26-1b).  **Nucleic Acid Synthesis**

Early work on DNA polymerase I led to the definition of two central requirements for DNA polymerization. First, all DNA polymerases require a **template**. The polymerization reaction is guided by a template DNA strand according to the base-pairing rules predicted by Watson and Crick: where a guanine is present in the template, a cytosine deoxynucleotide is added to the new strand, and so on. This was a particularly important discovery, not only because it provided a chemical basis for accurate semiconservative DNA replication but also because it represented the first example of the use of a template to guide a biosynthetic reaction.

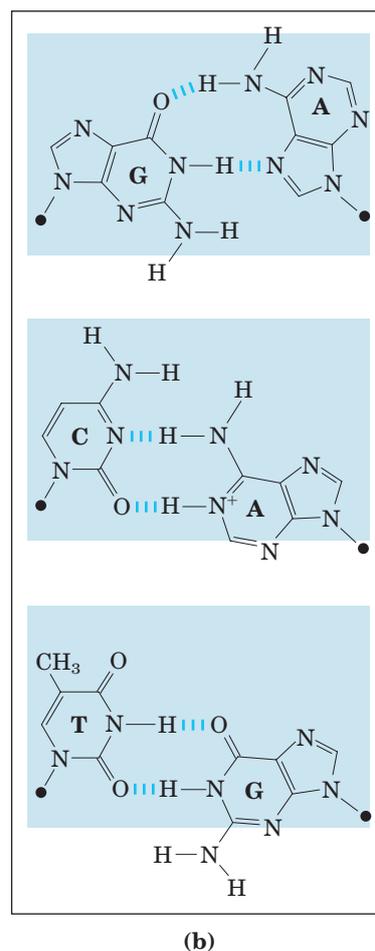
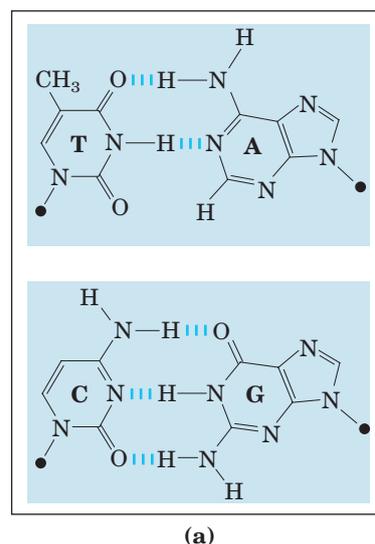
Second, the polymerases require a **primer**. A primer is a strand segment (complementary to the template) with a free 3'-hydroxyl group to which a nucleotide can be added; the free 3' end of the primer is called the **primer terminus**. In other words, part of the new strand must already be in place: all DNA polymerases can only add nucleotides to a preexisting strand. Most primers are oligonucleotides of RNA rather than DNA, and specialized enzymes synthesize primers when and where they are required.

After adding a nucleotide to a growing DNA strand, a DNA polymerase either dissociates or moves along the template and adds another nucleotide. Dissociation and reassociation of the polymerase can limit the overall polymerization rate—the process is generally faster when a polymerase adds more nucleotides without dissociating from the template. The average number of nucleotides added before a polymerase dissociates defines its **processivity**. DNA polymerases vary greatly in processivity; some add just a few nucleotides before dissociating, others add many thousands.  **Nucleotide Polymerization by DNA Polymerase**

### Replication Is Very Accurate

Replication proceeds with an extraordinary degree of fidelity. In *E. coli*, a mistake is made only once for every  $10^9$  to  $10^{10}$  nucleotides added. For the *E. coli* chromosome of  $\sim 4.6 \times 10^6$  bp, this means that an error occurs only once per 1,000 to 10,000 replications. During polymerization, discrimination between correct and incorrect nucleotides relies not just on the hydrogen bonds that specify the correct pairing between complementary bases but also on the common geometry of the standard A=T and G=C base pairs (Fig. 25-6). The active site of DNA polymerase I accommodates only base pairs with this geometry. An incorrect nucleotide may be able to hydrogen-bond with a base in the template, but it generally will not fit into the active site. Incorrect bases can be rejected before the phosphodiester bond is formed.

The accuracy of the polymerization reaction itself, however, is insufficient to account for the high degree of fidelity in replication. Careful measurements in vitro have shown that DNA polymerases insert one incorrect nucleotide for every  $10^4$  to  $10^5$  correct ones. These



**FIGURE 25-6** Contribution of base-pair geometry to the fidelity of DNA replication. (a) The standard A=T and G=C base pairs have very similar geometries, and an active site sized to fit one (blue shading) will generally accommodate the other. (b) The geometry of incorrectly paired bases can exclude them from the active site, as occurs on DNA polymerase I.

mistakes sometimes occur because a base is briefly in an unusual tautomeric form (see Fig. 8–9), allowing it to hydrogen-bond with an incorrect partner. In vivo, the error rate is reduced by additional enzymatic mechanisms.

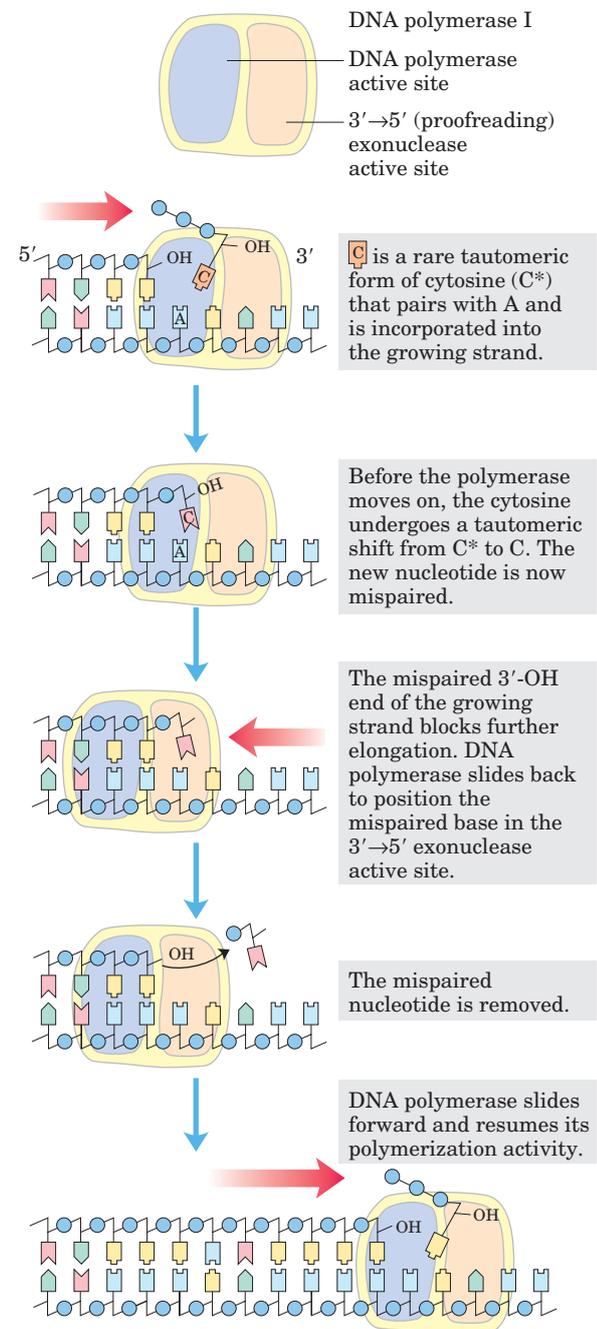
One mechanism intrinsic to virtually all DNA polymerases is a separate 3'→5' exonuclease activity that double-checks each nucleotide after it is added. This nuclease activity permits the enzyme to remove a newly added nucleotide and is highly specific for mismatched base pairs (Fig. 25–7). If the polymerase has added the wrong nucleotide, translocation of the enzyme to the position where the next nucleotide is to be added is inhibited. This kinetic pause provides the opportunity for a correction. The 3'→5' exonuclease activity removes the mispaired nucleotide, and the polymerase begins again. This activity, known as **proofreading**, is not simply the reverse of the polymerization reaction (Eqn 25–1), because pyrophosphate is not involved. The polymerizing and proofreading activities of a DNA polymerase can be measured separately. Proofreading improves the inherent accuracy of the polymerization reaction  $10^2$ - to  $10^3$ -fold. In the monomeric DNA polymerase I, the polymerizing and proofreading activities have separate active sites within the same polypeptide.

When base selection and proofreading are combined, DNA polymerase leaves behind one net error for every  $10^6$  to  $10^8$  bases added. Yet the measured accuracy of replication in *E. coli* is higher still. The additional accuracy is provided by a separate enzyme system that repairs the mismatched base pairs remaining after replication. We describe this mismatch repair, along with other DNA repair processes, in Section 25.2.

### *E. coli* Has at Least Five DNA Polymerases

More than 90% of the DNA polymerase activity observed in *E. coli* extracts can be accounted for by DNA polymerase I. Soon after the isolation of this enzyme in 1955, however, evidence began to accumulate that it is not suited for replication of the large *E. coli* chromosome. First, the rate at which it adds nucleotides (600 nucleotides/min) is too slow (by a factor of 100 or more) to account for the rates at which the replication fork moves in the bacterial cell. Second, DNA polymerase I has a relatively low processivity. Third, genetic studies have demonstrated that many genes, and therefore many proteins, are involved in replication: DNA polymerase I clearly does not act alone. Fourth, and most important, in 1969 John Cairns isolated a bacterial strain with an altered gene for DNA polymerase I that produced an inactive enzyme. Although this strain was abnormally sensitive to agents that damaged DNA, it was nevertheless viable!

A search for other DNA polymerases led to the discovery of *E. coli* **DNA polymerase II** and **DNA polymerase III** in the early 1970s. DNA polymerase II



**FIGURE 25–7** An example of error correction by the 3'→5' exonuclease activity of DNA polymerase I. Structural analysis has located the exonuclease activity ahead of the polymerase activity as the enzyme is oriented in its movement along the DNA. A mismatched base (here, a C–A mismatch) impedes translocation of DNA polymerase I to the next site. Sliding backward, the enzyme corrects the mistake with its 3'→5' exonuclease activity, then resumes its polymerase activity in the 5'→3' direction.

is an enzyme involved in one type of DNA repair (Section 25.3). DNA polymerase III is the principal replication enzyme in *E. coli*. The properties of these three DNA polymerases are compared in Table 25–1. DNA

**TABLE 25-1** Comparison of DNA Polymerases of *E. coli*

	DNA polymerase		
	I	II	III
Structural gene*	<i>polA</i>	<i>polB</i>	<i>polC (dnaE)</i>
Subunits (number of different types)	1	7	≥10
$M_r$	103,000	88,000 <sup>†</sup>	791,500
3'→5' Exonuclease (proofreading)	Yes	Yes	Yes
5'→3' Exonuclease	Yes	No	No
Polymerization rate (nucleotides/s)	16–20	40	250–1,000
Processivity (nucleotides added before polymerase dissociates)	3–200	1,500	≥500,000

\*For enzymes with more than one subunit, the gene listed here encodes the subunit with polymerization activity. Note that *dnaE* is an earlier designation for the gene now referred to as *polC*.

<sup>†</sup>Polymerization subunit only. DNA polymerase II shares several subunits with DNA polymerase III, including the  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\delta'$ ,  $\chi$ , and  $\psi$  subunits (see Table 25-2).

polymerases IV and V, identified in 1999, are involved in an unusual form of DNA repair (Section 25.2).

DNA polymerase I, then, is not the primary enzyme of replication; instead it performs a host of clean-up functions during replication, recombination, and repair. The polymerase's special functions are enhanced by its 5'→3' exonuclease activity. This activity, distinct from the 3'→5' proofreading exonuclease (Fig. 25-7), is located in a structural domain that can be separated from the enzyme by mild protease treatment. When the 5'→3' exonuclease domain is removed, the remaining fragment ( $M_r$  68,000), the **large fragment** or **Klenow fragment** (Fig. 25-8), retains the polymerization and

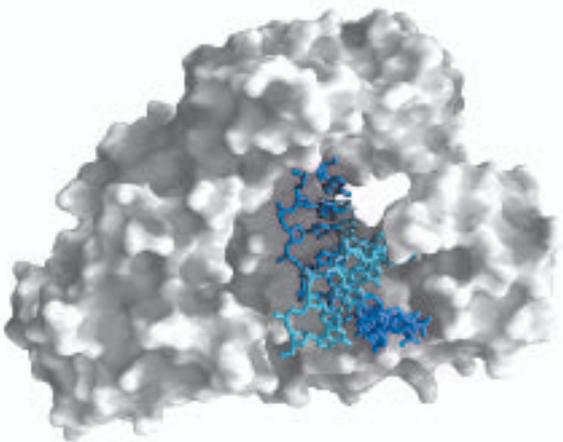
proofreading activities. The 5'→3' exonuclease activity of intact DNA polymerase I can replace a segment of DNA (or RNA) paired to the template strand, in a process known as nick translation (Fig. 25-9). Most other DNA polymerases lack a 5'→3' exonuclease activity.

DNA polymerase III is much more complex than DNA polymerase I, having ten types of subunits (Table 25-2). Its polymerization and proofreading activities reside in its  $\alpha$  and  $\epsilon$  (epsilon) subunits, respectively. The  $\theta$  subunit associates with  $\alpha$  and  $\epsilon$  to form a core polymerase, which can polymerize DNA but with limited processivity. Two core polymerases can be linked by

**TABLE 25-2** Subunits of DNA Polymerase III of *E. coli*

Subunit	Number of subunits per holoenzyme	$M_r$ of subunit	Gene	Function of subunit	
$\alpha$	2	129,900	<i>polC (dnaE)</i>	Polymerization activity	} Core polymerase
$\epsilon$	2	27,500	<i>dnaQ (mutD)</i>	3'→5' Proofreading exonuclease	
$\theta$	2	8,600	<i>holE</i>		
$\tau$	2	71,100	<i>dnaX</i>	Stable template binding; core enzyme dimerization	} Clamp-loading ( $\gamma$ ) complex that loads $\beta$ subunits on lagging strand at each Okazaki fragment
$\gamma$	1	47,500	<i>dnaX*</i>	Clamp loader	
$\delta$	1	38,700	<i>holA</i>	Clamp opener	
$\delta'$	1	36,900	<i>holB</i>	Clamp loader	
$\chi$	1	16,600	<i>holC</i>	Interaction with SSB	
$\psi$	1	15,200	<i>holD</i>	Interaction with $\gamma$ and $\chi$	
$\beta$	4	40,600	<i>dnaN</i>	DNA clamp required for optimal processivity	

\*The  $\gamma$  subunit is encoded by a portion of the gene for the  $\tau$  subunit, such that the amino-terminal 66% of the  $\tau$  subunit has the same amino acid sequence as the  $\gamma$  subunit. The  $\gamma$  subunit is generated by a translational frameshifting mechanism (see Box 27-1) that leads to premature translational termination.



**FIGURE 25-8** Large (Klenow) fragment of DNA polymerase I. This polymerase is widely distributed in bacteria. The Klenow fragment, produced by proteolytic treatment of the polymerase, retains the polymerization and proofreading activities of the enzyme. The Klenow fragment shown here is from the thermophilic bacterium *Bacillus stearothermophilus* (PDB ID 3BDP). The active site for addition of nucleotides is deep in the crevice at the far end of the bound DNA. The dark blue strand is the template.

another set of subunits, a clamp-loading complex, or  $\gamma$  complex, consisting of five subunits of four different types,  $\tau_2\gamma\delta\delta'$ . The core polymerases are linked through the  $\tau$  (tau) subunits. Two additional subunits,  $\chi$  (chi) and  $\psi$  (psi), are bound to the clamp-loading complex. The entire assembly of 13 protein subunits (nine different types) is called DNA polymerase III\* (Fig. 25-10a).

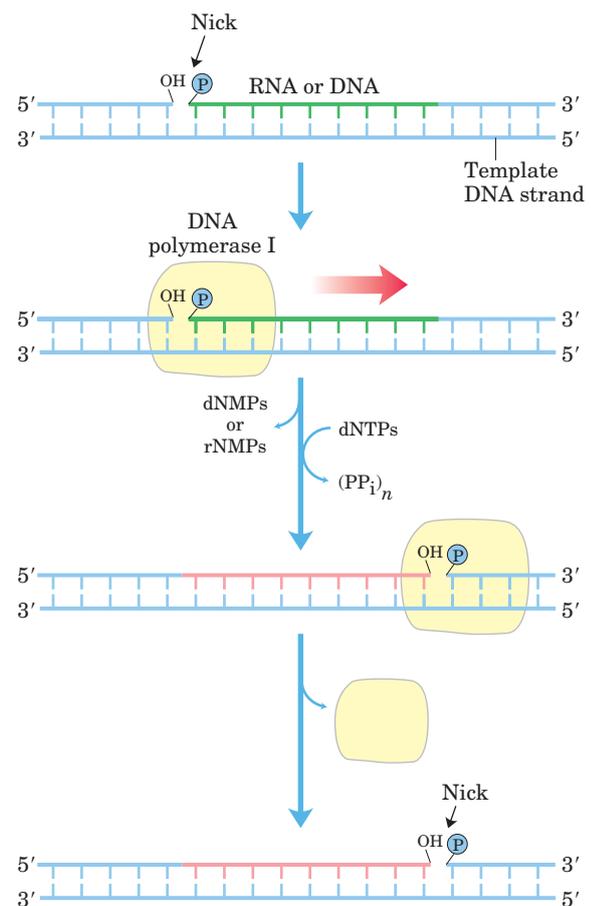
DNA polymerase III\* can polymerize DNA, but with a much lower processivity than one would expect for the organized replication of an entire chromosome. The necessary increase in processivity is provided by the addition of the  $\beta$  subunits, four of which complete the DNA polymerase III holoenzyme. The  $\beta$  subunits associate in pairs to form donut-shaped structures that encircle the DNA and act like clamps (Fig. 25-10b). Each dimer associates with a core subassembly of polymerase III\* (one dimeric clamp per core subassembly) and slides along the DNA as replication proceeds. The  $\beta$  sliding clamp prevents the dissociation of DNA polymerase III from DNA, dramatically increasing processivity—to greater than 500,000 (Table 25-1).

### DNA Replication Requires Many Enzymes and Protein Factors

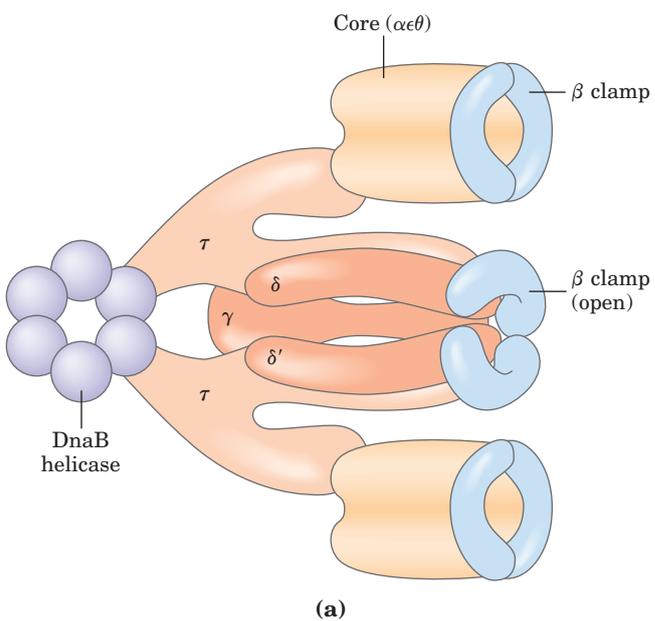
Replication in *E. coli* requires not just a single DNA polymerase but 20 or more different enzymes and proteins, each performing a specific task. The entire complex has been termed the **DNA replicase system** or **replisome**. The enzymatic complexity of replication reflects the constraints imposed by the structure of DNA

and by the requirements for accuracy. The main classes of replication enzymes are considered here in terms of the problems they overcome.

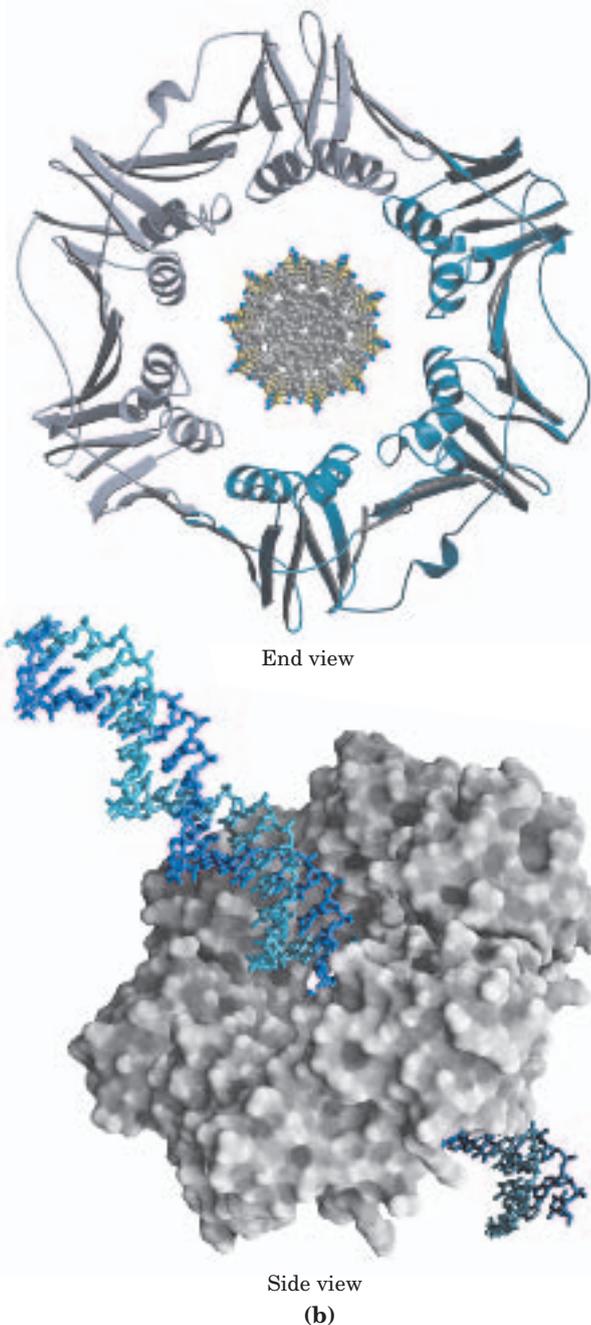
Access to the DNA strands that are to act as templates requires separation of the two parent strands. This is generally accomplished by **helicases**, enzymes that move along the DNA and separate the strands, using chemical energy from ATP. Strand separation creates topological stress in the helical DNA structure (see Fig. 24-12), which is relieved by the action of **topoisomerases**. The separated strands are stabilized by **DNA-binding proteins**. As noted earlier, before DNA polymerases can begin synthesizing DNA, primers must be present on the template—generally short segments



**FIGURE 25-9** Nick translation. In this process, an RNA or DNA strand paired to a DNA template is simultaneously degraded by the 5'→3' exonuclease activity of DNA polymerase I and replaced by the polymerase activity of the same enzyme. These activities have a role in both DNA repair and the removal of RNA primers during replication (both described later). The strand of nucleic acid to be removed (either DNA or RNA) is shown in green, the replacement strand in red. DNA synthesis begins at a nick (a broken phosphodiester bond, leaving a free 3' hydroxyl and a free 5' phosphate). Polymerase I extends the nontemplate DNA strand and moves the nick along the DNA—a process called nick translation. A nick remains where DNA polymerase I dissociates, and is later sealed by another enzyme.



**FIGURE 25-10 DNA polymerase III.** (a) Architecture of bacterial DNA polymerase III. Two core domains, composed of subunits  $\alpha$ ,  $\epsilon$ , and  $\theta$ , are linked by a five-subunit  $\gamma$  complex (also known as the clamp-loading complex) with the composition  $\tau_2\gamma\delta\delta'$ . The  $\gamma$  and  $\tau$  subunits are encoded by the same gene. The  $\gamma$  subunit is a shortened version of  $\tau$ ; the  $\tau$  subunit thus contains a domain identical to  $\gamma$ , along with an additional segment that interacts with the core polymerase. The other two subunits of DNA polymerase III\*,  $\chi$  and  $\psi$  (not shown), also bind to the  $\gamma$  complex. Two  $\beta$  clamps interact with the two-core subassembly, each clamp a dimer of the  $\beta$  subunit. The complex interacts with the DnaB helicase through the  $\tau$  subunit. (b) Two  $\beta$  subunits of *E. coli* polymerase III form a circular clamp that surrounds the DNA. The clamp slides along the DNA molecule, increasing the processivity of the polymerase III holoenzyme to greater than 500,000 by preventing its dissociation from the DNA. The end-on view shows the two  $\beta$  subunits as gray and light-blue ribbon structures surrounding a space-filling model of DNA. In the side view, surface contour models of the  $\beta$  subunits (gray) surround a stick representation of a DNA double helix (light and dark blue) (derived from PDB ID 2POL).



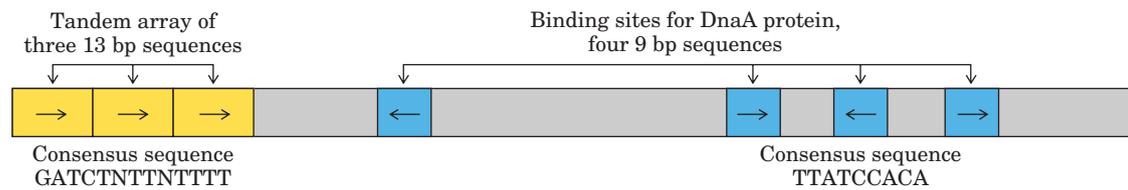
of RNA synthesized by enzymes known as **primases**. Ultimately, the RNA primers are removed and replaced by DNA; in *E. coli*, this is one of the many functions of DNA polymerase I. After an RNA primer is removed and the gap is filled in with DNA, a nick remains in the DNA backbone in the form of a broken phosphodiester bond. These nicks are sealed by **DNA ligases**. All these processes require coordination and regulation, an interplay best characterized in the *E. coli* system.

### Replication of the *E. coli* Chromosome Proceeds in Stages

The synthesis of a DNA molecule can be divided into three stages: initiation, elongation, and termination,

distinguished both by the reactions taking place and by the enzymes required. As you will find here and in the next two chapters, synthesis of the major information-containing biological polymers—DNAs, RNAs, and proteins—can be understood in terms of these same three stages, with the stages of each pathway having unique characteristics. The events described below reflect information derived primarily from in vitro experiments using purified *E. coli* proteins, although the principles are highly conserved in all replication systems.

**Initiation** The *E. coli* replication origin, *oriC*, consists of 245 bp; it bears DNA sequence elements that are highly conserved among bacterial replication origins. The general arrangement of the conserved sequences is



**FIGURE 25-11** Arrangement of sequences in the *E. coli* replication origin, *oriC*. Although the repeated sequences (shaded in color) are not identical, certain nucleotides are particularly common in each po-

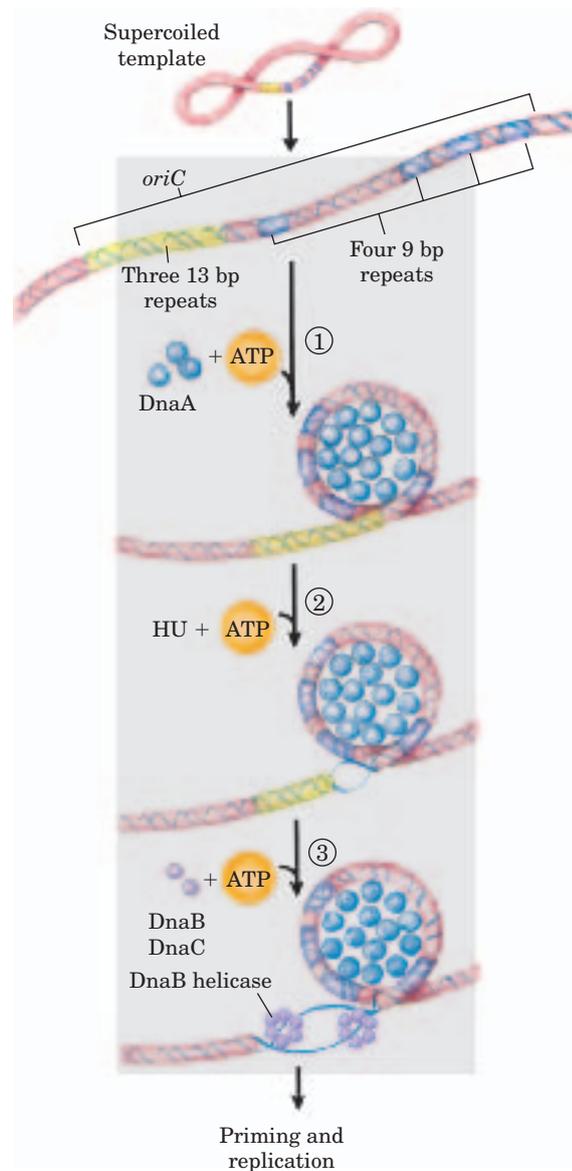
sition, forming a consensus sequence. In positions where there is no consensus, N represents any of the four nucleotides. The arrows indicate the orientations of the nucleotide sequences.

illustrated in Figure 25-11. The key sequences of interest here are two series of short repeats: three repeats of a 13 bp sequence and four repeats of a 9 bp sequence.

At least nine different enzymes or proteins (summarized in Table 25-3) participate in the initiation phase of replication. They open the DNA helix at the origin and establish a prepriming complex for subsequent reactions. The crucial component in the initiation process is the DnaA protein. A single complex of four to five DnaA protein molecules binds to the four 9 bp repeats in the origin (Fig. 25-12, step ①), then recognizes and successively denatures the DNA in the region of the three 13 bp repeats, which are rich in A=T pairs (step ②). This process requires ATP and the bacterial histonelike protein HU. The DnaC protein then loads the DnaB protein onto the unwound region. Two ring-shaped hexamers of DnaB, one loaded onto each DNA strand, act as helicases, unwinding the DNA bidirectionally and creating two potential replication forks. If the *E. coli* single-stranded DNA-binding protein (SSB) and DNA gyrase (DNA topoisomerase II) are now added in vitro, thousands of base pairs are rapidly unwound by the DnaB helicase, proceeding out from the origin. Many molecules of SSB bind cooperatively to single-stranded DNA, stabilizing the separated strands and preventing reannealing while gyrase relieves the topological stress produced by the DnaB helicase. When additional replication proteins are included in the in vitro system, the DNA unwinding mediated by DnaB is coupled to replication, as described below.

Initiation is the only phase of DNA replication that is known to be regulated, and it is regulated such that replication occurs only once in each cell cycle. The mechanism of regulation is not yet well understood, but genetic and biochemical studies have provided a few insights.

The timing of replication initiation is affected by DNA methylation and interactions with the bacterial plasma membrane. The *oriC* DNA is methylated by the Dam methylase (Table 25-3), which methylates the  $N^6$  position of adenine within the palindromic sequence (5')GATC. (Dam is not a biochemical expletive; it stands for DNA adenine methylation.) The *oriC* region of *E. coli* is highly enriched in GATC sequences—it has 11 of them in its 245 bp, whereas the average frequency of GATC in the *E. coli* chromosome as a whole is 1 in 256 bp.



**FIGURE 25-12** Model for initiation of replication at the *E. coli* origin, *oriC*. ① About 20 DnaA protein molecules, each with a bound ATP, bind at the four 9 bp repeats. The DNA is wrapped around this complex. ② The three A=T-rich 13 bp repeats are denatured sequentially. ③ Hexamers of the DnaB protein bind to each strand, with the aid of DnaC protein. The DnaB helicase activity further unwinds the DNA in preparation for priming and DNA synthesis.

**TABLE 25-3** Proteins Required to Initiate Replication at the *E. coli* Origin

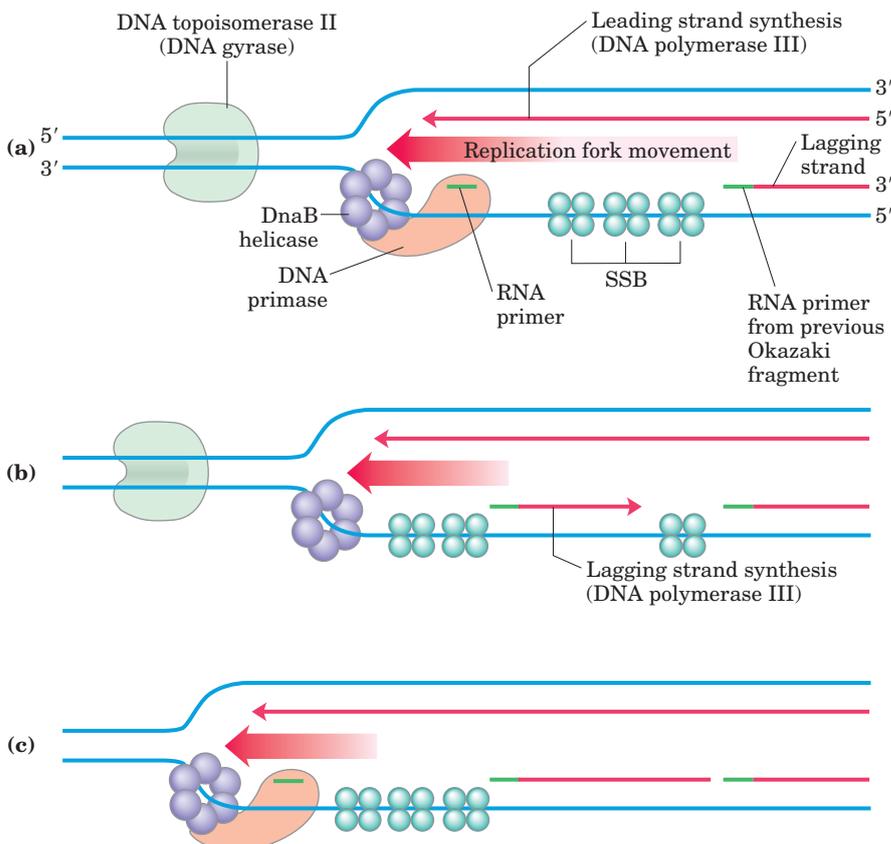
Protein	$M_r$	Number of subunits	Function
DnaA protein	52,000	1	Recognizes <i>ori</i> sequence; opens duplex at specific sites in origin
DnaB protein (helicase)	300,000	6*	Unwinds DNA
DnaC protein	29,000	1	Required for DnaB binding at origin
HU	19,000	2	Histonelike protein; DNA-binding protein; stimulates initiation
Primase (DnaG protein)	60,000	1	Synthesizes RNA primers
Single-stranded DNA-binding protein (SSB)	75,600	4*	Binds single-stranded DNA
RNA polymerase	454,000	5	Facilitates DnaA activity
DNA gyrase (DNA topoisomerase II)	400,000	4	Relieves torsional strain generated by DNA unwinding
Dam methylase	32,000	1	Methylates (5')GATC sequences at <i>oriC</i>

\*Subunits in these cases are identical.

Immediately after replication, the DNA is hemimethylated: the parent strands have methylated *oriC* sequences but the newly synthesized strands do not. The hemimethylated *oriC* sequences are now sequestered for a period by interaction with the plasma membrane (the mechanism is unknown). After a time, *oriC* is released from the plasma membrane, and it must be fully methylated by Dam methylase before it can again bind DnaA. Regulation of initiation also involves the slow hydrolysis of ATP by DnaA protein, which cycles the pro-

tein between active (with bound ATP) and inactive (with bound ADP) forms on a timescale of 20 to 40 minutes.

**Elongation** The elongation phase of replication includes two distinct but related operations: leading strand synthesis and lagging strand synthesis. Several enzymes at the replication fork are important to the synthesis of both strands. Parent DNA is first unwound by DNA helicases, and the resulting topological stress is relieved by topoisomerases. Each separated strand is then stabilized by



**FIGURE 25-13** Synthesis of Okazaki fragments. (a) At intervals, primase synthesizes an RNA primer for a new Okazaki fragment. Note that if we consider the two template strands as lying side by side, lagging strand synthesis formally proceeds in the opposite direction from fork movement. (b) Each primer is extended by DNA polymerase III. (c) DNA synthesis continues until the fragment extends as far as the primer of the previously added Okazaki fragment. A new primer is synthesized near the replication fork to begin the process again.

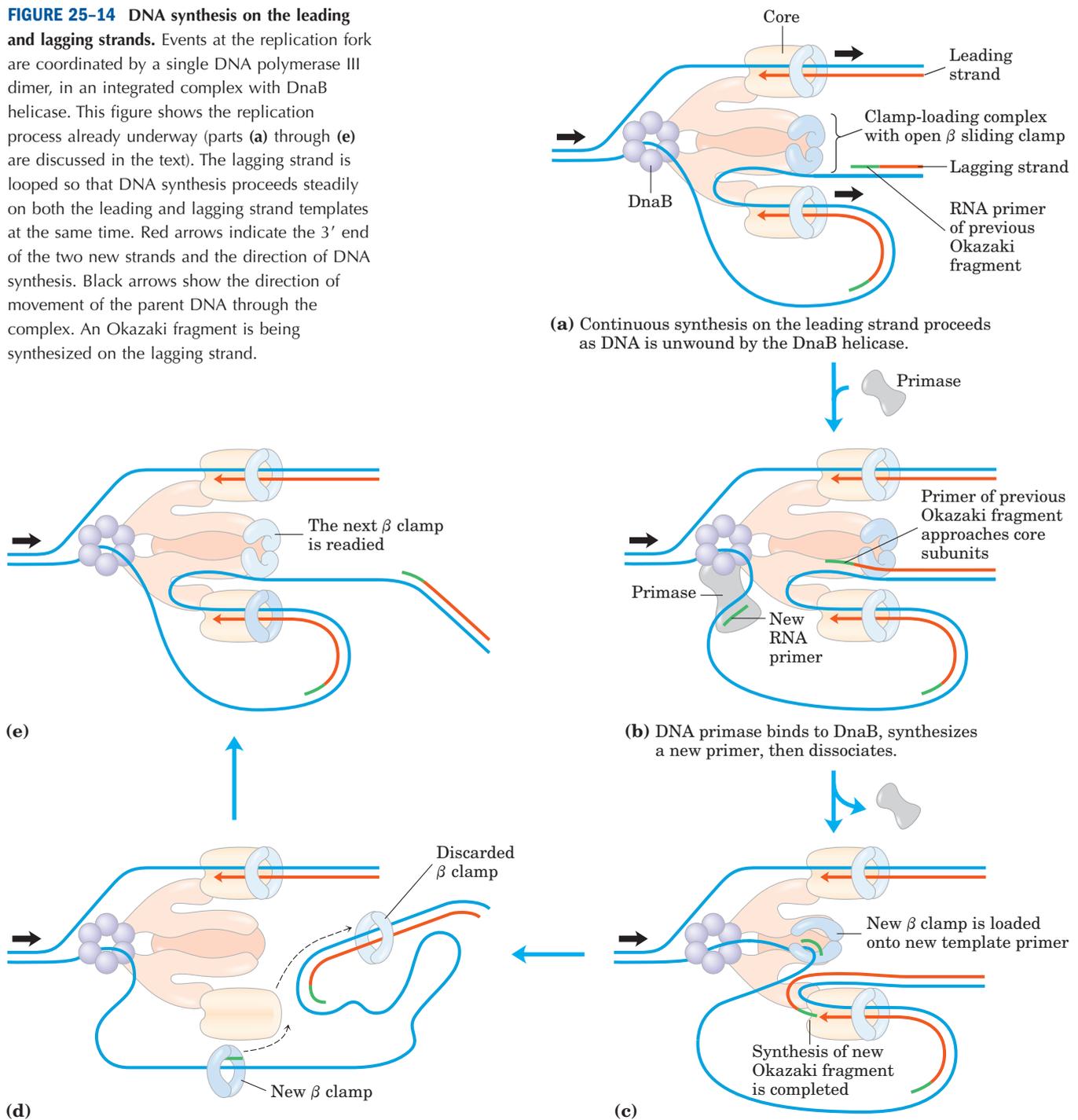
SSB. From this point, synthesis of leading and lagging strands is sharply different.

Leading strand synthesis, the more straightforward of the two, begins with the synthesis by primase (DnaG protein) of a short (10 to 60 nucleotide) RNA primer at the replication origin. Deoxyribonucleotides are added to this primer by DNA polymerase III. Leading strand synthesis then proceeds continuously, keeping pace with the unwinding of DNA at the replication fork.

Lagging strand synthesis, as we have noted, is accomplished in short Okazaki fragments. First, an RNA

primer is synthesized by primase and, as in leading strand synthesis, DNA polymerase III binds to the RNA primer and adds deoxyribonucleotides (Fig. 25-13). On this level, the synthesis of each Okazaki fragment seems straightforward, but the reality is quite complex. The complexity lies in the *coordination* of leading and lagging strand synthesis: both strands are produced by a *single* asymmetric DNA polymerase III dimer, which is accomplished by looping the DNA of the lagging strand as shown in Figure 25-14, bringing together the two points of polymerization.

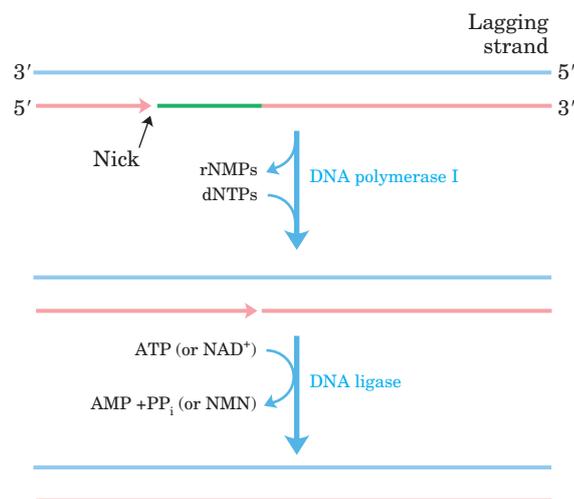
**FIGURE 25-14 DNA synthesis on the leading and lagging strands.** Events at the replication fork are coordinated by a single DNA polymerase III dimer, in an integrated complex with DnaB helicase. This figure shows the replication process already underway (parts (a) through (e) are discussed in the text). The lagging strand is looped so that DNA synthesis proceeds steadily on both the leading and lagging strand templates at the same time. Red arrows indicate the 3' end of the two new strands and the direction of DNA synthesis. Black arrows show the direction of movement of the parent DNA through the complex. An Okazaki fragment is synthesized on the lagging strand.



The synthesis of Okazaki fragments on the lagging strand entails some elegant enzymatic choreography. The DnaB helicase and DnaG primase constitute a functional unit within the replication complex, the **primosome**. DNA polymerase III uses one set of its core subunits (the core polymerase) to synthesize the leading strand continuously, while the other set of core subunits cycles from one Okazaki fragment to the next on the looped lagging strand. The DnaB helicase unwinds the DNA at the replication fork (Fig. 25-14a) as it travels along the lagging strand template in the 5'→3' direction. DNA primase occasionally associates with DnaB helicase and synthesizes a short RNA primer (Fig. 25-14b). A new  $\beta$  sliding clamp is then positioned at the primer by the clamp-loading complex of DNA polymerase III (Fig. 25-14c). When synthesis of an Okazaki fragment has been completed, replication halts, and the core subunits of DNA polymerase III dissociate from their  $\beta$  sliding clamp (and from the completed Okazaki fragment) and associate with the new clamp (Fig. 25-14d, e). This initiates synthesis of a new Okazaki fragment. As noted earlier, the entire complex responsible for coordinated DNA synthesis at a replication fork is a **replisome**. The proteins acting at the replication fork are summarized in Table 25-4.

The replisome promotes rapid DNA synthesis, adding ~1,000 nucleotides/s to each strand (leading and lagging). Once an Okazaki fragment has been completed, its RNA primer is removed and replaced with DNA by DNA polymerase I, and the remaining nick is sealed by DNA ligase (Fig. 25-15).

DNA ligase catalyzes the formation of a phosphodiester bond between a 3' hydroxyl at the end of one DNA strand and a 5' phosphate at the end of another strand. The phosphate must be activated by adenylylation. DNA ligases isolated from viruses and eukaryotes use ATP for this purpose. DNA ligases from bacteria are unusual in that they generally use  $\text{NAD}^+$ —a cofactor that normally functions in hydride transfer reactions



**FIGURE 25-15** Final steps in the synthesis of lagging strand segments. RNA primers in the lagging strand are removed by the 5'→3' exonuclease activity of DNA polymerase I and replaced with DNA by the same enzyme. The remaining nick is sealed by DNA ligase. The role of ATP or  $\text{NAD}^+$  is shown in Figure 25-16.

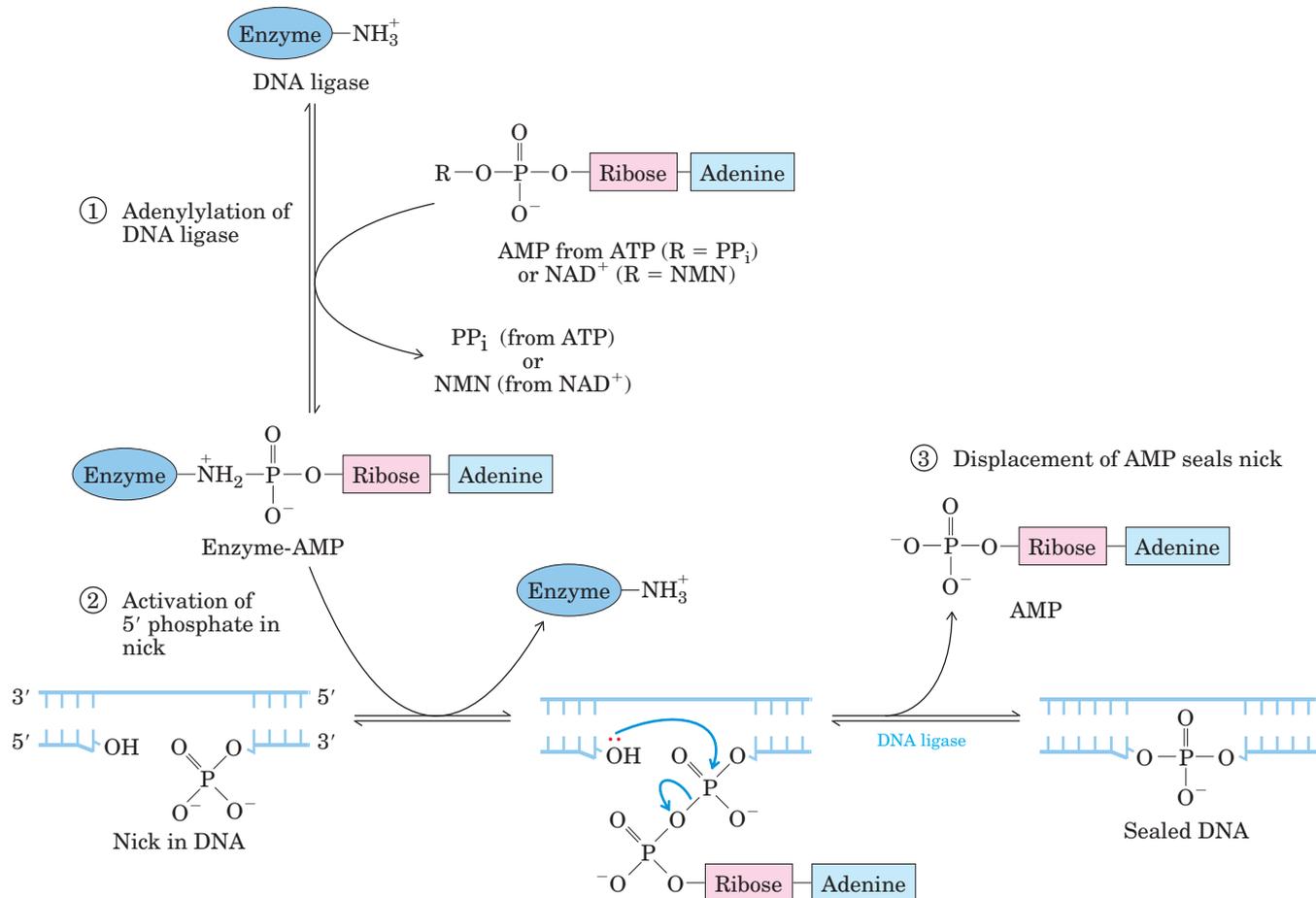
(see Fig. 13-15)—as the source of the AMP activating group (Fig. 25-16). DNA ligase is another enzyme of DNA metabolism that has become an important reagent in recombinant DNA experiments (see Fig. 9-1).

**Termination** Eventually, the two replication forks of the circular *E. coli* chromosome meet at a terminus region containing multiple copies of a 20 bp sequence called Ter (for *terminus*) (Fig. 25-17a). The Ter sequences are arranged on the chromosome to create a sort of trap that a replication fork can enter but cannot leave. The Ter sequences function as binding sites for a protein called Tus (*terminus utilization substance*). The Tus-Ter complex can arrest a replication fork from only one direction. Only one Tus-Ter complex functions per replication cycle—the complex first encountered by either

**TABLE 25-4** Proteins at the *E. coli* Replication Fork

Protein	$M_r$	Number of subunits	Function
SSB	75,600	4	Binding to single-stranded DNA
DnaB protein (helicase)	300,000	6	DNA unwinding; primosome constituent
Primase (DnaG protein)	60,000	1	RNA primer synthesis; primosome constituent
DNA polymerase III	791,500	17	New strand elongation
DNA polymerase I	103,000	1	Filling of gaps; excision of primers
DNA ligase	74,000	1	Ligation
DNA gyrase (DNA topoisomerase II)	400,000	4	Supercoiling

Modified from Kornberg, A. (1982) *Supplement to DNA Replication*, Table S11-2, W. H. Freeman and Company, New York.



**FIGURE 25-16 Mechanism of the DNA ligase reaction.** In each of the three steps, one phosphodiester bond is formed at the expense of another. Steps ① and ② lead to activation of the 5' phosphate in the nick. An AMP group is transferred first to a Lys residue on the enzyme and then to the 5' phosphate in the nick. In step ③, the 3'-hydroxyl group attacks this phosphate and displaces AMP, producing a

phosphodiester bond to seal the nick. In the *E. coli* DNA ligase reaction, AMP is derived from NAD<sup>+</sup>. The DNA ligases isolated from a number of viral and eukaryotic sources use ATP rather than NAD<sup>+</sup>, and they release pyrophosphate rather than nicotinamide mononucleotide (NMN) in step ①.

replication fork. Given that opposing replication forks generally halt when they collide, Ter sequences do not seem essential, but they may prevent overreplication by one replication fork in the event that the other is delayed or halted by an encounter with DNA damage or some other obstacle.

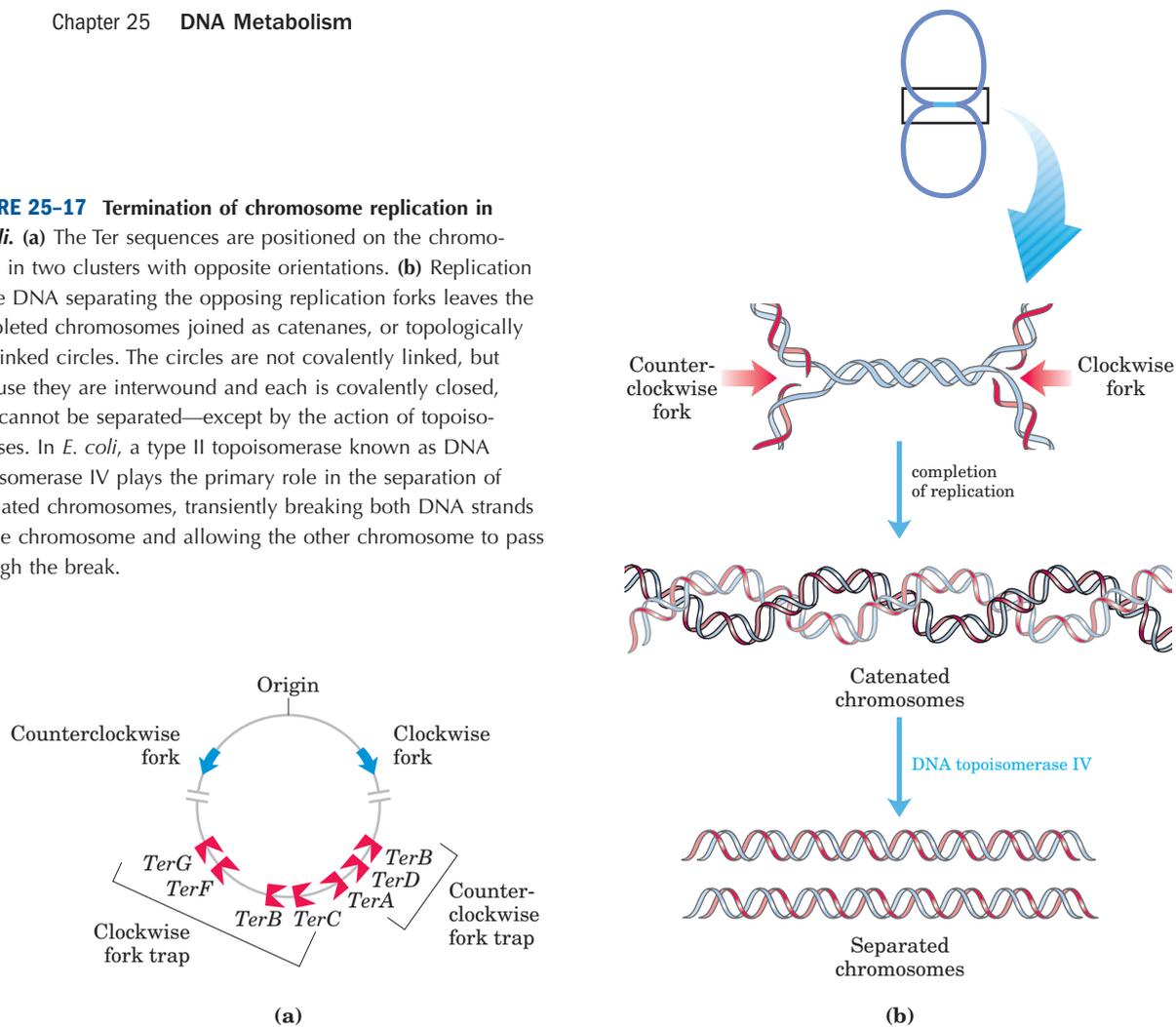
So, when either replication fork encounters a functional Tus-Ter complex, it halts; the other fork halts when it meets the first (arrested) fork. The final few hundred base pairs of DNA between these large protein complexes are then replicated (by an as yet unknown mechanism), completing two topologically interlinked (catenated) circular chromosomes (Fig. 25-17b). DNA circles linked in this way are known as **catenanes**. Separation of the catenated circles in *E. coli* requires topoisomerase IV (a type II topoisomerase). The separated chromosomes then segregate into daughter cells at cell

division. The terminal phase of replication of other circular chromosomes, including many of the DNA viruses that infect eukaryotic cells, is similar.

### Bacterial Replication Is Organized in Membrane-Bound Replication Factories

The replication of a circular bacterial chromosome is highly organized. Once bidirectional replication is initiated at the origin, the two replisomes do not travel away from each other along the DNA. Instead, the replisomes are linked together and tethered to one point on the bacterial inner membrane, and the DNA substrate is fed through this “replication factory” (Fig. 25-18a). The tethering point is at the center of the elongated bacterial cell. After initiation, each of the two newly synthesized replication origins is partitioned into one half of

**FIGURE 25-17 Termination of chromosome replication in *E. coli*.** (a) The *Ter* sequences are positioned on the chromosome in two clusters with opposite orientations. (b) Replication of the DNA separating the opposing replication forks leaves the completed chromosomes joined as catenanes, or topologically interlinked circles. The circles are not covalently linked, but because they are interwound and each is covalently closed, they cannot be separated—except by the action of topoisomerases. In *E. coli*, a type II topoisomerase known as DNA topoisomerase IV plays the primary role in the separation of catenated chromosomes, transiently breaking both DNA strands of one chromosome and allowing the other chromosome to pass through the break.



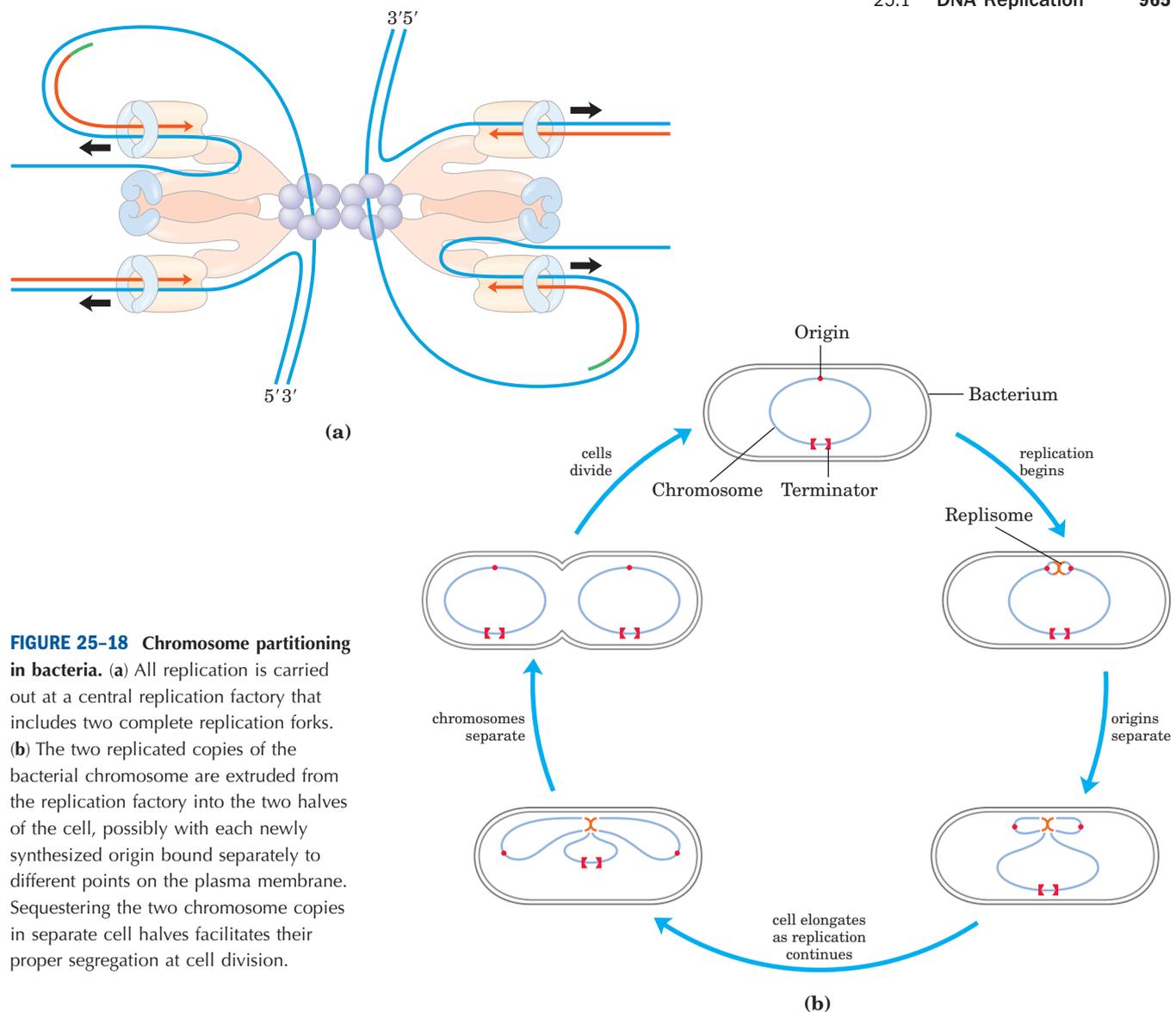
the cell, and continuing replication extrudes each new chromosome into that half (Fig. 25-18b). The elaborate spatial organization of the newly replicated chromosomes is orchestrated and maintained by many proteins, including bacterial homologs of the SMC proteins and topoisomerases (Chapter 24). Once replication is terminated, the cell divides, and the chromosomes sequestered in the two halves of the original cell are accurately partitioned into the daughter cells. When replication commences in the daughter cells, the origin of replication is sequestered in new replication factories formed at a point on the membrane at the center of the cell, and the entire process is repeated.

### Replication in Eukaryotic Cells Is More Complex

The DNA molecules in eukaryotic cells are considerably larger than those in bacteria and are organized into complex nucleoprotein structures (chromatin; p. 938). The essential features of DNA replication are the same in eukaryotes and prokaryotes, and many of the protein complexes are functionally and structurally conserved. However, some interesting variations on the general principles discussed above promise new insights into the regulation of replication and its link with the cell cycle.

Origins of replication, called *autonomously replicating sequences* (ARS) or **replicators**, have been identified and best studied in yeast. Yeast replicators span ~150 bp and contain several essential conserved sequences. About 400 replicators are distributed among the 16 chromosomes in a haploid yeast genome. Initiation of replication in all eukaryotes requires a multi-subunit protein, the *origin recognition complex* (ORC), which binds to several sequences within the replicator. ORC interacts with and is regulated by a number of other proteins involved in control of the eukaryotic cell cycle. Two other proteins, CDC6 (discovered in a screen for genes affecting the *cell division cycle*) and CDT1 (*Cdc10-dependent transcript 1*), bind to ORC and mediate the loading of a heterohexameric of *minichromosome maintenance proteins* (MCM2 to MCM7). The MCM complex is a ring-shaped replicative helicase, analogous to the bacterial DnaB helicase. The CDC6 and CDT1 proteins have a role comparable to that of the bacterial DnaC protein, loading the MCM helicase onto the DNA near the replication origin.

The rate of replication fork movement in eukaryotes (~50 nucleotides/s) is only one-twentieth that observed in *E. coli*. At this rate, replication of an average human chromosome proceeding from a single origin



**FIGURE 25-18** Chromosome partitioning in bacteria. (a) All replication is carried out at a central replication factory that includes two complete replication forks. (b) The two replicated copies of the bacterial chromosome are extruded from the replication factory into the two halves of the cell, possibly with each newly synthesized origin bound separately to different points on the plasma membrane. Sequestering the two chromosome copies in separate cell halves facilitates their proper segregation at cell division.

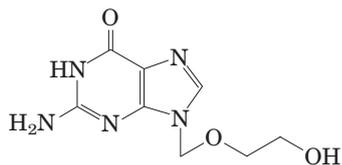
would take more than 500 hours. Replication of human chromosomes in fact proceeds bidirectionally from many origins, spaced 30,000 to 300,000 bp apart. Eukaryotic chromosomes are almost always much larger than bacterial chromosomes, so multiple origins are probably a universal feature in eukaryotic cells.

Like bacteria, eukaryotes have several types of DNA polymerases. Some have been linked to particular functions, such as the replication of mitochondrial DNA. The replication of nuclear chromosomes involves DNA polymerase  $\alpha$ , in association with DNA polymerase  $\delta$ . **DNA polymerase  $\alpha$**  is typically a multisubunit enzyme with similar structure and properties in all eukaryotic cells. One subunit has a primase activity, and the largest subunit ( $M_r$  ~180,000) contains the polymerization activity. However, this polymerase has no proofreading  $3' \rightarrow 5'$  exonuclease activity, making it unsuitable for high-fidelity DNA replication. DNA polymerase  $\alpha$  is believed to function only in the synthesis of short primers (containing either RNA or DNA) for Okazaki fragments on the lagging strand. These primers

are then extended by the multisubunit **DNA polymerase  $\delta$** . This enzyme is associated with and stimulated by a protein called *proliferating cell nuclear antigen* (PCNA;  $M_r$  29,000), found in large amounts in the nuclei of proliferating cells. The three-dimensional structure of PCNA is remarkably similar to that of the  $\beta$  subunit of *E. coli* DNA polymerase III (Fig. 25-10b), although primary sequence homology is not evident. PCNA has a function analogous to that of the  $\beta$  subunit, forming a circular clamp that greatly enhances the processivity of the polymerase. DNA polymerase  $\delta$  has a  $3' \rightarrow 5'$  proofreading exonuclease activity and appears to carry out both leading and lagging strand synthesis in a complex comparable to the dimeric bacterial DNA polymerase III.

Yet another polymerase, **DNA polymerase  $\epsilon$** , replaces DNA polymerase  $\delta$  in some situations, such as in DNA repair. DNA polymerase  $\epsilon$  may also function at the replication fork, perhaps playing a role analogous to that of the bacterial DNA polymerase I, removing the primers of Okazaki fragments on the lagging strand.

Many DNA viruses encode their own DNA polymerases, and some of these have become targets for pharmaceuticals. For example, the DNA polymerase of the herpes simplex virus is inhibited by acyclovir, a compound developed by Gertrude Elion (p. 876). Acyclovir consists of guanine attached to an incomplete ribose ring. It is phosphorylated by a virally encoded thymidine kinase; acyclovir binds to this viral enzyme with an affinity 200-fold greater than its binding to the cellular thymidine kinase. This ensures that phosphorylation occurs mainly in virus-infected cells. Cellular kinases convert the resulting acyclo-GMP to acyclo-GTP, which is both an inhibitor and a substrate of DNA polymerases, and which competitively inhibits the herpes DNA polymerase more strongly than cellular DNA polymerases. Because it lacks a 3' hydroxyl, acyclo-GTP also acts as a chain terminator when incorporated into DNA. Thus viral replication is inhibited at several steps.



Two other protein complexes also function in eukaryotic DNA replication. RPA (*r*eplication *p*rotein *A*) is a eukaryotic single-stranded DNA-binding protein, equivalent in function to the *E. coli* SSB protein. RFC (*r*eplication *f*actor *C*) is a clamp loader for PCNA and facilitates the assembly of active replication complexes. The subunits of the RFC complex have significant sequence similarity to the subunits of the bacterial clamp-loading ( $\gamma$ ) complex.

The termination of replication on linear eukaryotic chromosomes involves the synthesis of special structures called **telomeres** at the ends of each chromosome, as discussed in the next chapter.

### SUMMARY 25.1 DNA Replication

- Replication of DNA occurs with very high fidelity and at a designated time in the cell cycle. Replication is semiconservative, each strand acting as template for a new daughter strand. It is carried out in three identifiable phases: initiation, elongation, and termination. The reaction starts at the origin and usually proceeds bidirectionally.
- DNA is synthesized in the 5'→3' direction by DNA polymerases. At the replication fork, the leading strand is synthesized continuously in the same direction as replication fork movement; the lagging strand is synthesized discontinuously as Okazaki fragments, which are subsequently ligated.

- The fidelity of DNA replication is maintained by (1) base selection by the polymerase, (2) a 3'→5' proofreading exonuclease activity that is part of most DNA polymerases, and (3) specific repair systems for mismatches left behind after replication.
- Most cells have several DNA polymerases. In *E. coli*, DNA polymerase III is the primary replication enzyme. DNA polymerase I is responsible for special functions during replication, recombination, and repair.
- Replication of the *E. coli* chromosome involves many enzymes and protein factors organized in replication factories, in which template DNA is spooled through two replisomes tethered to the bacterial plasma membrane.
- Replication is similar in eukaryotic cells, but eukaryotic chromosomes have many replication origins.

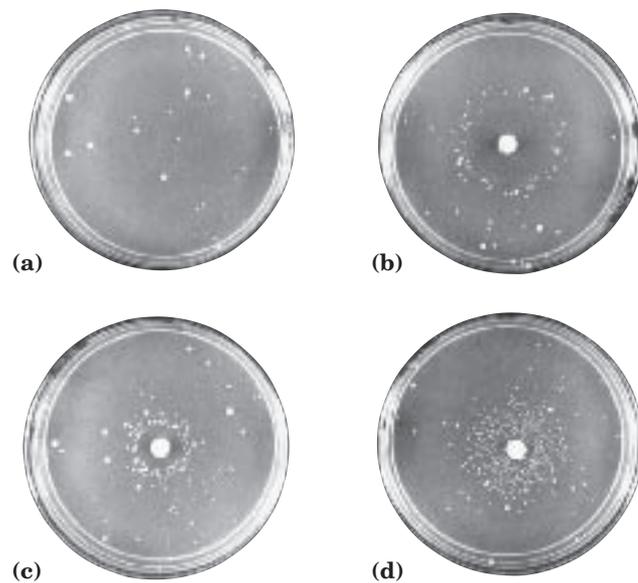
## 25.2 DNA Repair

A cell generally has only one or two sets of genomic DNA. Damaged proteins and RNA molecules can be quickly replaced by using information encoded in the DNA, but DNA molecules themselves are irreplaceable. Maintaining the integrity of the information in DNA is a cellular imperative, supported by an elaborate set of DNA repair systems. DNA can become damaged by a variety of processes, some spontaneous, others catalyzed by environmental agents (Chapter 8). Replication itself can very occasionally damage the information content in DNA when errors introduce mismatched base pairs (such as G paired with T).

The chemistry of DNA damage is diverse and complex. The cellular response to this damage includes a wide range of enzymatic systems that catalyze some of the most interesting chemical transformations in DNA metabolism. We first examine the effects of alterations in DNA sequence and then consider specific repair systems.

### Mutations Are Linked to Cancer

The best way to illustrate the importance of DNA repair is to consider the effects of *unrepaired* DNA damage (a lesion). The most serious outcome is a change in the base sequence of the DNA, which, if replicated and transmitted to future cell generations, becomes permanent. A permanent change in the nucleotide sequence of DNA is called a **mutation**. Mutations can involve the replacement of one base pair with another (substitution mutation) or the addition or deletion of one or more base pairs (insertion or deletion mutations). If the mutation affects nonessential DNA or if it has a negligible



**FIGURE 25-19** Ames test for carcinogens, based on their mutagenicity. A strain of *Salmonella typhimurium* having a mutation that inactivates an enzyme of the histidine biosynthetic pathway is plated on a histidine-free medium. Few cells grow. (a) The few small colonies of *S. typhimurium* that do grow on a histidine-free medium carry spontaneous back-mutations that permit the histidine biosynthetic pathway to operate. Three identical nutrient plates (b), (c), and (d) have been inoculated with an equal number of cells. Each plate then receives a disk of filter paper containing progressively lower concentrations of a mutagen. The mutagen greatly increases the rate of back-mutation and hence the number of colonies. The clear areas around the filter paper indicate where the concentration of mutagen is so high that it is lethal to the cells. As the mutagen diffuses away from the filter paper, it is diluted to sublethal concentrations that promote back-mutation. Mutagens can be compared on the basis of their effect on mutation rate. Because many compounds undergo a variety of chemical transformations after entering a cell, compounds are sometimes tested for mutagenicity after first incubating them with a liver extract. Some substances have been found to be mutagenic only after this treatment.

effect on the function of a gene, it is known as a **silent mutation**. Rarely, a mutation confers some biological advantage. Most nonsilent mutations, however, are deleterious.

In mammals there is a strong correlation between the accumulation of mutations and cancer. A simple test developed by Bruce Ames measures the potential of a given chemical compound to promote certain easily detected mutations in a specialized bacterial strain (Fig. 25–19). Few of the chemicals that we encounter in daily life score as mutagens in this test. However, of the compounds known to be carcinogenic from extensive animal trials, more than 90% are also found to be mutagenic in the Ames test. Because of this strong correlation between mutagenesis and carcinogenesis, the Ames test for bacterial mutagens is widely used as a rapid and inexpensive screen for potential human carcinogens.

The genome of a typical mammalian cell accumulates many thousands of lesions during a 24-hour period. However, as a result of DNA repair, fewer than 1 in 1,000 becomes a mutation. DNA is a relatively stable molecule, but in the absence of repair systems, the cumulative effect of many infrequent but damaging reactions would make life impossible.

### All Cells Have Multiple DNA Repair Systems

The number and diversity of repair systems reflect both the importance of DNA repair to cell survival and the diverse sources of DNA damage (Table 25–5). Some common types of lesions, such as pyrimidine dimers (see Fig. 8–34), can be repaired by several distinct systems. Many DNA repair processes also appear to be extraordinarily inefficient energetically—an exception to

**TABLE 25-5** Types of DNA Repair Systems in *E. coli*

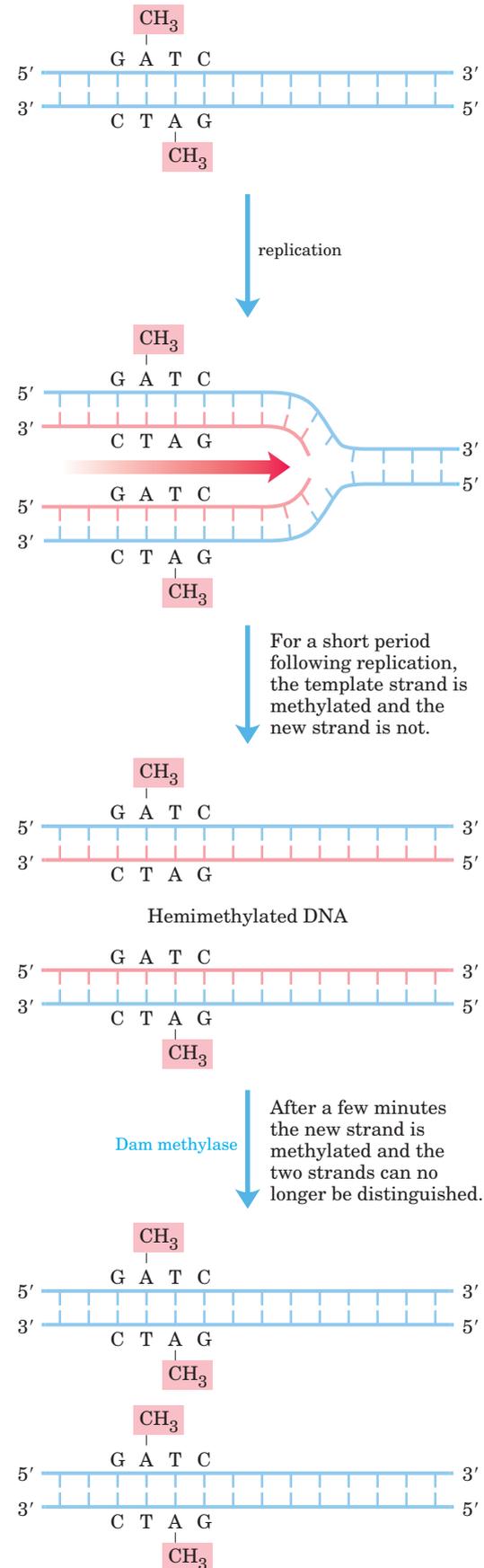
Enzymes/proteins	Type of damage
<b>Mismatch repair</b>	
Dam methylase	Mismatches
MutH, MutL, MutS proteins	
DNA helicase II	
SSB	
DNA polymerase III	
Exonuclease I	
Exonuclease VII	
RecJ nuclease	Abnormal bases (uracil, hypoxanthine, xanthine); alkylated bases; in some other organisms, pyrimidine dimers
Exonuclease X	
DNA ligase	
DNA glycosylases	
<b>Base-excision repair</b>	
AP endonucleases	DNA lesions that cause large structural changes (e.g., pyrimidine dimers)
DNA polymerase I	
DNA ligase	
<b>Nucleotide-excision repair</b>	
ABC excinuclease	DNA lesions that cause large structural changes (e.g., pyrimidine dimers)
DNA polymerase I	
DNA ligase	
<b>Direct repair</b>	
DNA photolyases	Pyrimidine dimers
<i>O</i> <sup>6</sup> -Methylguanine-DNA methyltransferase	
AlkB protein	1-Methylguanine, 3-methylcytosine

the pattern observed in the metabolic pathways, where every ATP is generally accounted for and used optimally. When the integrity of the genetic information is at stake, the amount of chemical energy invested in a repair process seems almost irrelevant.

DNA repair is possible largely because the DNA molecule consists of two complementary strands. DNA damage in one strand can be removed and accurately replaced by using the undamaged complementary strand as a template. We consider here the principal types of repair systems, beginning with those that repair the rare nucleotide mismatches that are left behind by replication.

**Mismatch Repair** Correction of the rare mismatches left after replication in *E. coli* improves the overall fidelity of replication by an additional factor of  $10^2$  to  $10^3$ . The mismatches are nearly always corrected to reflect the information in the old (template) strand, so the repair system must somehow discriminate between the template and the newly synthesized strand. The cell accomplishes this by tagging the template DNA with methyl groups to distinguish it from newly synthesized strands. The mismatch repair system of *E. coli* includes at least 12 protein components (Table 25-5) that function either in strand discrimination or in the repair process itself.

The strand discrimination mechanism has not been worked out for most bacteria or eukaryotes, but is well understood for *E. coli* and some closely related bacteria. In these prokaryotes, strand discrimination is based on the action of Dam methylase (Table 25-3), which, as you will recall, methylates DNA at the  $N^6$  position of all adenines within (5')GATC sequences. Immediately after passage of the replication fork, there is a short period (a few seconds or minutes) during which the template strand is methylated but the newly synthesized strand is not (Fig. 25-20). The transient unmethylated state of GATC sequences in the newly synthesized strand permits the new strand to be distinguished from the template strand. Replication mismatches in the vicinity of a hemimethylated GATC sequence are then repaired according to the information in the methylated parent (template) strand. Tests in vitro show that if both strands are methylated at a GATC sequence, few mismatches are repaired; if neither strand is methylated, repair occurs but does not favor either strand. The cell's



**FIGURE 25-20 Methylation and mismatch repair.** Methylation of DNA strands can serve to distinguish parent (template) strands from newly synthesized strands in *E. coli* DNA, a function that is critical to mismatch repair (see Fig. 25-21). The methylation occurs at the  $N^6$  of adenines in (5')GATC sequences. This sequence is a palindrome (see Fig. 8-20), present in opposite orientations on the two strands.

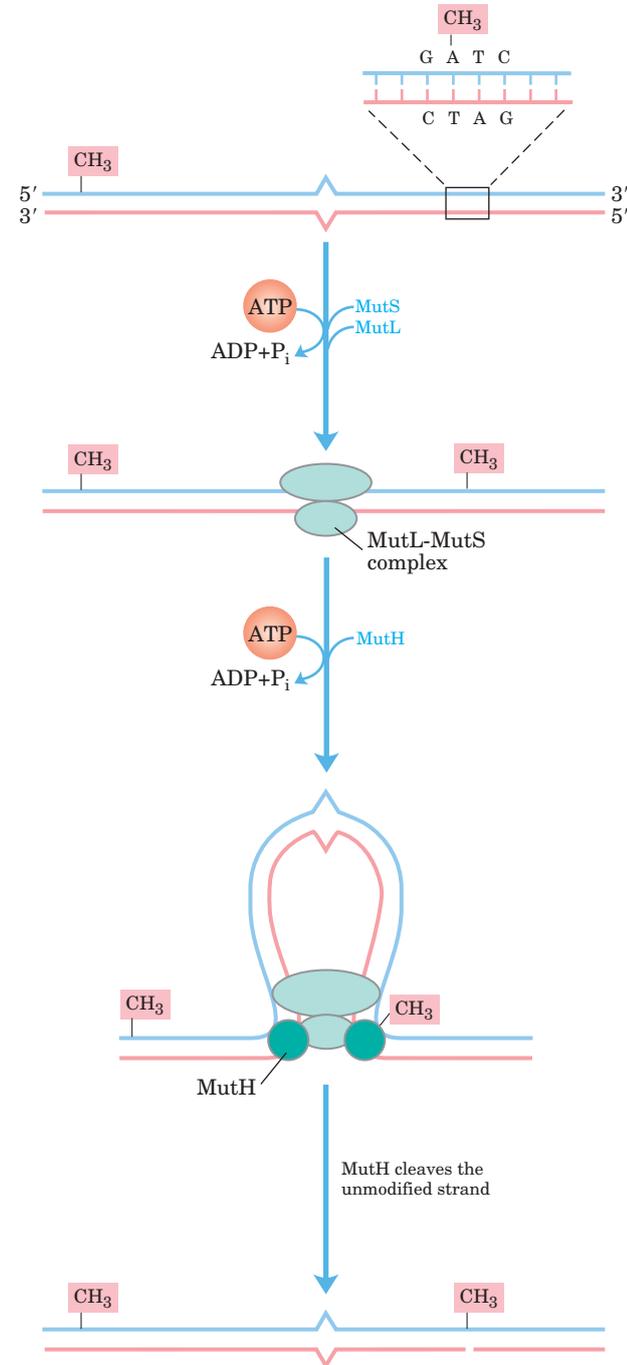
methyl-directed mismatch repair system efficiently repairs mismatches up to 1,000 bp from a hemimethylated GATC sequence. For many bacterial species, the mechanism of strand discrimination during mismatch repair has not been determined.

How is the mismatch correction process directed by relatively distant GATC sequences? A mechanism is illustrated in Figure 25–21. MutL protein forms a complex with MutS protein, and the complex binds to all mismatched base pairs (except C–C). MutH protein binds to MutL and to GATC sequences encountered by the MutL–MutS complex. DNA on both sides of the mismatch is threaded through the MutL–MutS complex, creating a DNA loop; simultaneous movement of both legs of the loop through the complex is equivalent to the complex moving in both directions at once along the DNA. MutH has a site-specific endonuclease activity that is inactive until the complex encounters a hemimethylated GATC sequence. At this site, MutH catalyzes cleavage of the unmethylated strand on the 5' side of the G in GATC, which marks the strand for repair. Further steps in the pathway depend on where the mismatch is located relative to this cleavage site (Fig. 25–22).

When the mismatch is on the 5' side of the cleavage site, the unmethylated strand is unwound and degraded in the 3'→5' direction from the cleavage site through the mismatch, and this segment is replaced with new DNA. This process requires the combined action of DNA helicase II, SSB, exonuclease I or exonuclease X (both of which degrade strands of DNA in the 3'→5' direction), DNA polymerase III, and DNA ligase. The pathway for repair of mismatches on the 3' side of the cleavage site is similar, except that the exonuclease is either exonuclease VII (which degrades single-stranded DNA in the 5'→3' or 3'→5' direction) or RecJ nuclease (which degrades single-stranded DNA in the 5'→3' direction).

Mismatch repair is a particularly expensive process for *E. coli* in terms of energy expended. The mismatch may be 1,000 bp or more from the GATC sequence. The degradation and replacement of a strand segment of this length require an enormous investment in activated deoxynucleotide precursors to repair a *single* mismatched base. This again underscores the importance to the cell of genomic integrity.

All eukaryotic cells have several proteins structurally and functionally analogous to the bacterial MutS and MutL (but not MutH) proteins. Alterations in human genes encoding proteins of this type produce some of the most common inherited cancer-susceptibility syndromes (Box 25–1), further demonstrating the value to the organism of DNA repair systems. The main MutS homologs in most eukaryotes, from yeast to humans, are MSH2 (*MutS* homolog 2), MSH3, and MSH6. Heterodimers of MSH2 and MSH6 generally bind to single



**FIGURE 25-21** A model for the early steps of methyl-directed mismatch repair. The proteins involved in this process in *E. coli* have been purified (see Table 25–5). Recognition of the sequence (5')GATC and of the mismatch are specialized functions of the MutH and MutS proteins, respectively. The MutL protein forms a complex with MutS at the mismatch. DNA is threaded through this complex such that the complex moves simultaneously in both directions along the DNA until it encounters a MutH protein bound at a hemimethylated GATC sequence. MutH cleaves the unmethylated strand on the 5' side of the G in this sequence. A complex consisting of DNA helicase II and one of several exonucleases then degrades the unmethylated DNA strand from that point toward the mismatch (see Fig. 25–22).


**BOX 25-1 BIOCHEMISTRY IN MEDICINE**
**DNA Repair and Cancer**

Human cancer develops when certain genes that regulate normal cell division (oncogenes and tumor suppressor genes; Chapter 12) fail to function, are activated at the wrong time, or are altered. As a consequence, cells may grow out of control and form a tumor. The genes controlling cell division can be damaged by spontaneous mutation or overridden by the invasion of a tumor virus (Chapter 26). Not surprisingly, alterations in DNA-repair genes that result in an increase in the rate of mutation can greatly increase an individual's susceptibility to cancer. Defects in the genes encoding the proteins involved in nucleotide-excision repair, mismatch repair, recombinational repair, and error-prone translesion synthesis have all been linked to human cancers. Clearly, DNA repair can be a matter of life and death.

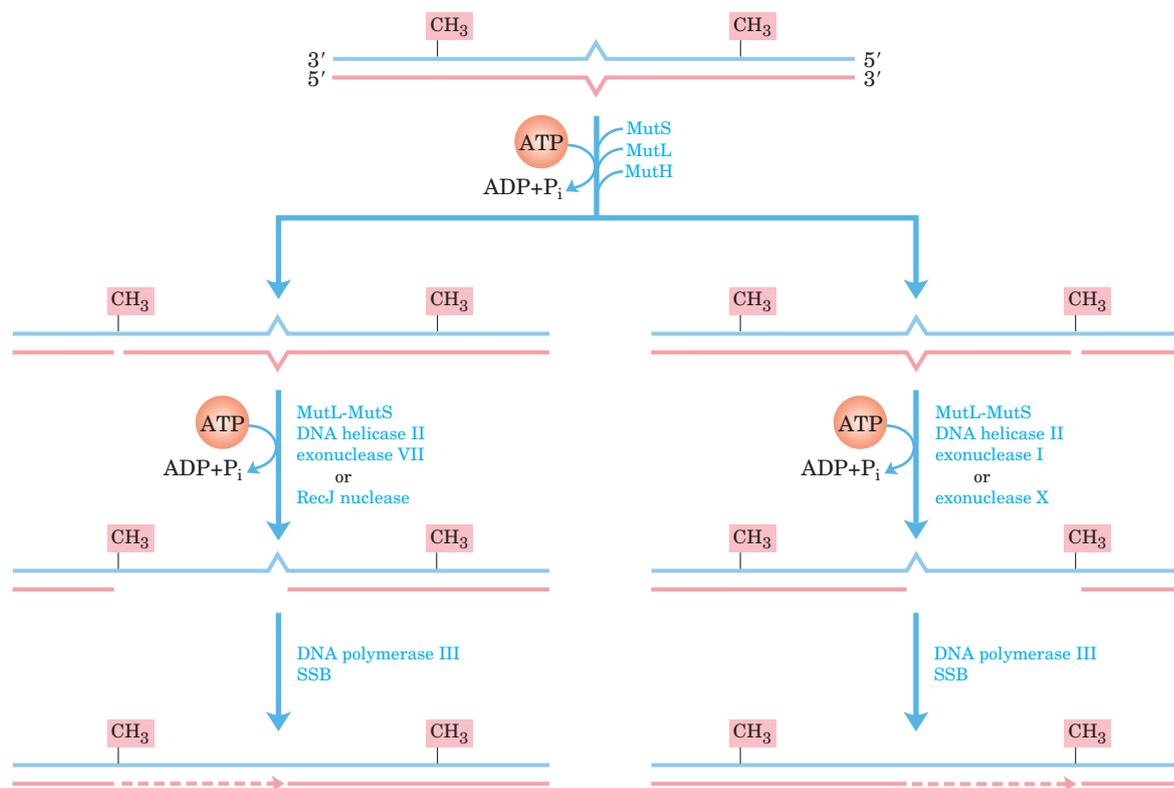
Nucleotide-excision repair requires a larger number of proteins in humans than in bacteria, although the overall pathways are very similar. Genetic defects that inactivate nucleotide-excision repair have been associated with several genetic diseases, the best-studied of which is xeroderma pigmentosum, or XP. Because nucleotide-excision repair is the sole repair pathway for pyrimidine dimers in humans, people with XP are extremely light sensitive and readily develop sunlight-induced skin cancers. Most people with XP also have neurological abnormalities, presumably because of their inability to repair certain lesions caused by the high rate of oxidative metabolism in neurons. Defects in the genes encoding any of at least seven different protein components of the nucleotide-excision repair system can result in XP, giving rise to seven different genetic groups denoted XPA to XPG. Several of these proteins (notably XPB, XPD, and XPG) also play roles in transcription-coupled base-excision repair of oxidative lesions, described in Chapter 26.

Most microorganisms have redundant pathways for the repair of cyclobutane pyrimidine dimers—making use of DNA photolyase and sometimes base-excision repair as alternatives to nucleotide-excision repair—but humans and other placental mammals do not. This lack of a back-up to nucleotide-excision repair for the removal of pyrimidine dimers has led to speculation that early mammalian evolution involved small, furry, nocturnal animals with little need to repair UV damage. However, mammals do have a path-

way for the translesion bypass of cyclobutane pyrimidine dimers, which involves DNA polymerase  $\eta$ . This enzyme preferentially inserts two A residues opposite a T–T pyrimidine dimer, minimizing mutations. People with a genetic condition in which DNA polymerase  $\eta$  function is missing exhibit an XP-like illness known as XP-variant or XP-V. Clinical manifestations of XP-V are similar to those of the classic XP diseases, although mutation levels are higher when cells are exposed to UV light. Apparently, the nucleotide-excision repair system works in concert with DNA polymerase  $\eta$  in normal human cells, repairing and/or bypassing pyrimidine dimers as needed to keep cell growth and DNA replication going. Exposure to UV light introduces a heavy load of pyrimidine dimers, requiring that some be bypassed by translesion synthesis to keep replication on track. When either system is missing, it is partly compensated for by the other. A loss of polymerase  $\eta$  activity leads to stalled replication forks and bypass of UV lesions by different, and more mutagenic, translesion synthesis (TLS) polymerases. As when other DNA repair systems are absent, the resulting increase in mutations often leads to cancer.

One of the most common inherited cancer-susceptibility syndromes is hereditary nonpolyposis colon cancer, or HNPCC. This syndrome has been traced to defects in mismatch repair. Human and other eukaryotic cells have several proteins analogous to the bacterial MutL and MutS proteins (see Fig. 25–21). Defects in at least five different mismatch repair genes can give rise to HNPCC. The most prevalent are defects in the *hMLH1* (human MutL homolog 1) and *hMSH2* (human MutS homolog 2) genes. In individuals with HNPCC, cancer generally develops at an early age, with colon cancers being most common.

Most human breast cancer occurs in women with no known predisposition. However, about 10% of cases are associated with inherited defects in two genes, *BRCA1* and *BRCA2*. *BRCA1* and *BRCA2* are large proteins (human *BRCA1* and *BRCA2* are 1834 and 3418 amino acid residues long, respectively). They both interact with a wide range of other proteins involved in transcription, chromosome maintenance, DNA repair, and control of the cell cycle. However, the precise molecular function of *BRCA1* and *BRCA2* in these various cellular processes is not yet clear. Women with defects in either the *BRCA1* or *BRCA2* gene have a greater than 80% chance of developing breast cancer.



**FIGURE 25-22** Completing methyl-directed mismatch repair. The combined action of DNA helicase II, SSB, and one of four different exonucleases removes a segment of the new strand between the MutH cleavage site and a point just beyond the mismatch. The exonuclease

that is used depends on the location of the cleavage site relative to the mismatch. The resulting gap is filled in by DNA polymerase III, and the nick is sealed by DNA ligase (not shown).

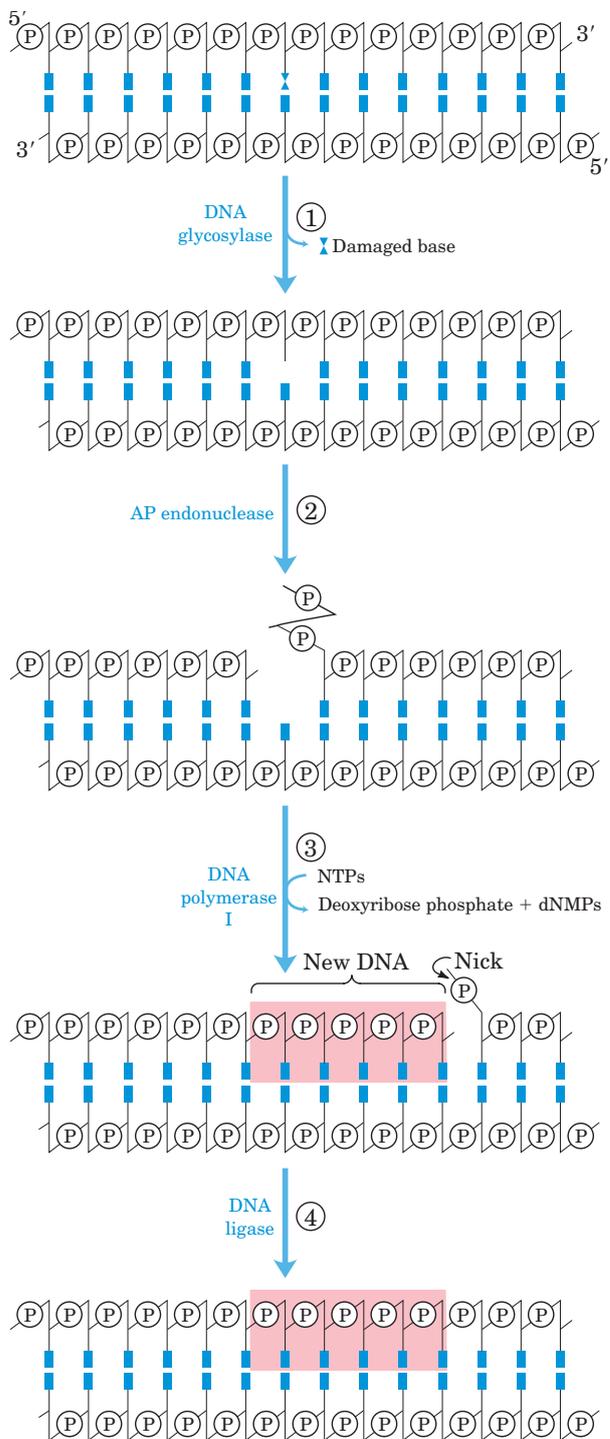
base-pair mismatches, and bind less well to slightly longer mismatched loops. In many organisms the longer mismatches (2 to 6 bp) may be bound instead by a heterodimer of MSH2 and MSH3, or are bound by both types of heterodimers in tandem. Homologs of MutL, predominantly a heterodimer of MLH1 and PMS1 (*post-meiotic segregation*), bind to and stabilize the MSH complexes. Many details of the subsequent events in eukaryotic mismatch repair remain to be worked out. In particular, we do not know the mechanism by which newly synthesized DNA strands are identified, although research has revealed that this strand identification does not involve GATC sequences.

**Base-Excision Repair** Every cell has a class of enzymes called **DNA glycosylases** that recognize particularly common DNA lesions (such as the products of cytosine and adenine deamination; see Fig. 8-33a) and remove the affected base by cleaving the *N*-glycosyl bond. This cleavage creates an apurinic or apyrimidinic site in the DNA, commonly referred to as an **AP site** or **abasic**

**site**. Each DNA glycosylase is generally specific for one type of lesion.

Uracil DNA glycosylases, for example, found in most cells, specifically remove from DNA the uracil that results from spontaneous deamination of cytosine. Mutant cells that lack this enzyme have a high rate of G≡C to A≡T mutations. This glycosylase does not remove uracil residues from RNA or thymine residues from DNA. The capacity to distinguish thymine from uracil, the product of cytosine deamination—necessary for the selective repair of the latter—may be one reason why DNA evolved to contain thymine instead of uracil (p. 293).

Bacteria generally have just one type of uracil DNA glycosylase, whereas humans have at least four types, with different specificities—an indicator of the importance of uracil removal from DNA. The most abundant human uracil glycosylase, UNG, is associated with the human replisome, where it eliminates the occasional U residue inserted in place of a T during replication. The deamination of C residues is 100-fold faster in single-stranded DNA than in double-stranded DNA, and



**FIGURE 25-23** DNA repair by the base-excision repair pathway.

① A DNA glycosylase recognizes a damaged base and cleaves between the base and deoxyribose in the backbone. ② An AP endonuclease cleaves the phosphodiester backbone near the AP site. ③ DNA polymerase I initiates repair synthesis from the free 3' hydroxyl at the nick, removing (with its 5'→3' exonuclease activity) a portion of the damaged strand and replacing it with undamaged DNA. ④ The nick remaining after DNA polymerase I has dissociated is sealed by DNA ligase.

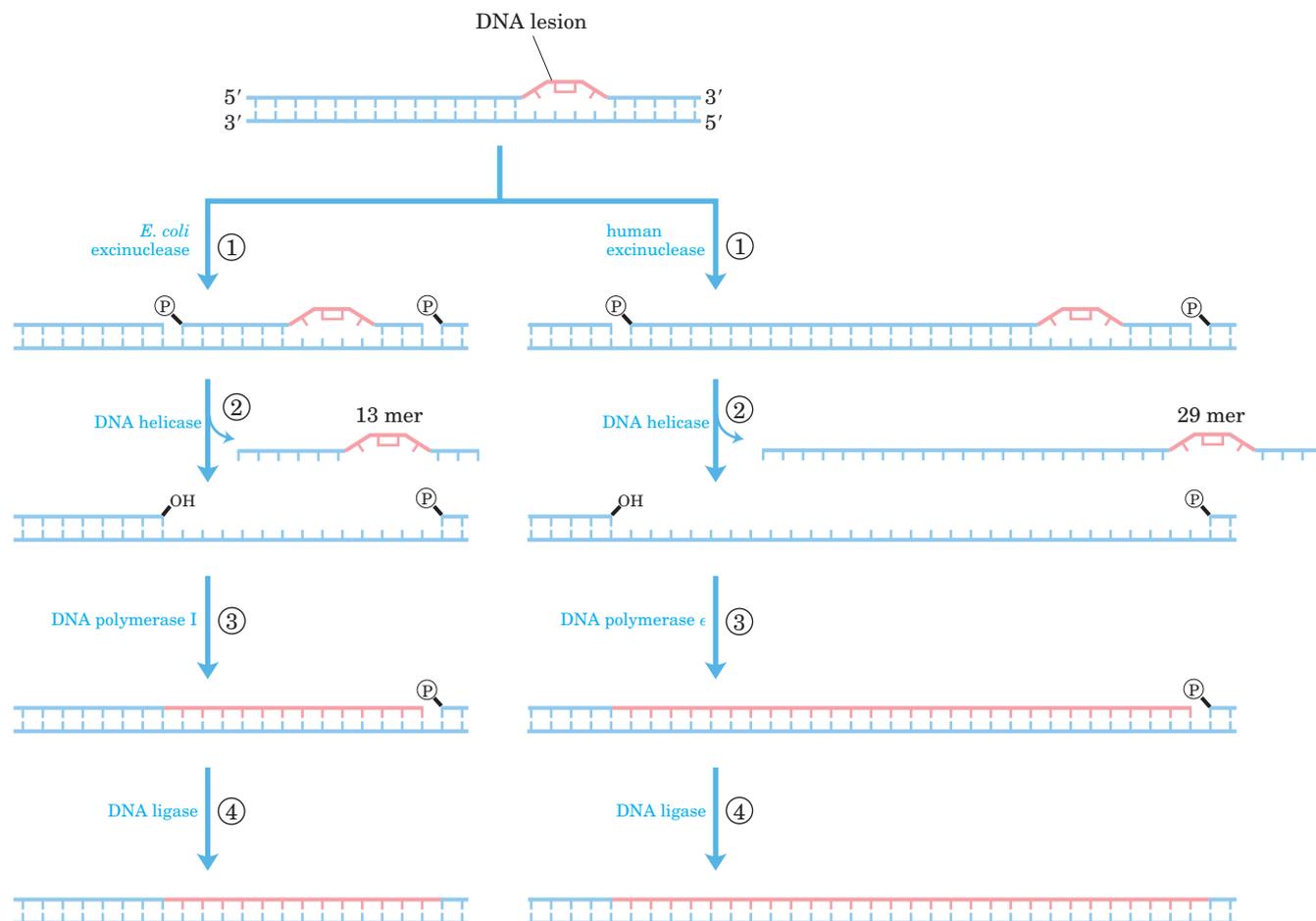
humans have the enzyme hSMUG1, which removes any U residues that occur in single-stranded DNA during replication or transcription. Two other human DNA glycosylases, TDG and MBD4, remove either U or T residues paired with G, generated by deamination of cytosine or 5-methylcytosine, respectively.

Other DNA glycosylases recognize and remove a variety of damaged bases, including formamidopyrimidine and 8-hydroxyguanine (both arising from purine oxidation), hypoxanthine (arising from adenine deamination), and alkylated bases such as 3-methyladenine and 7-methylguanine. Glycosylases that recognize other lesions, including pyrimidine dimers, have also been identified in some classes of organisms. Remember that AP sites also arise from the slow, spontaneous hydrolysis of the *N*-glycosyl bonds in DNA (see Fig. 8-33b).

Once an AP site has formed, another group of enzymes must repair it. The repair is *not* made by simply inserting a new base and re-forming the *N*-glycosyl bond. Instead, the deoxyribose 5'-phosphate left behind is removed and replaced with a new nucleotide. This process begins with **AP endonucleases**, enzymes that cut the DNA strand containing the AP site. The position of the incision relative to the AP site (5' or 3' to the site) varies with the type of AP endonuclease. A segment of DNA including the AP site is then removed, DNA polymerase I replaces the DNA, and DNA ligase seals the remaining nick (Fig. 25-23). In eukaryotes, nucleotide replacement is carried out by specialized polymerases, as described below.

**Nucleotide-Excision Repair** DNA lesions that cause large distortions in the helical structure of DNA generally are repaired by the nucleotide-excision system, a repair pathway critical to the survival of all free-living organisms. In nucleotide-excision repair (Fig. 25-24), a multisubunit enzyme hydrolyzes two phosphodiester bonds, one on either side of the distortion caused by the lesion. In *E. coli* and other prokaryotes, the enzyme system hydrolyzes the fifth phosphodiester bond on the 3' side and the eighth phosphodiester bond on the 5' side to generate a fragment of 12 to 13 nucleotides (depending on whether the lesion involves one or two bases). In humans and other eukaryotes, the enzyme system hydrolyzes the sixth phosphodiester bond on the 3' side and the twenty-second phosphodiester bond on the 5' side, producing a fragment of 27 to 29 nucleotides. Following the dual incision, the excised oligonucleotides are released from the duplex and the resulting gap is filled—by DNA polymerase I in *E. coli* and DNA polymerase  $\epsilon$  in humans. DNA ligase seals the nick.

In *E. coli*, the key enzymatic complex is the ABC excinuclease, which has three subunits, UvrA ( $M_r$  104,000), UvrB ( $M_r$  78,000), and UvrC ( $M_r$  68,000). The

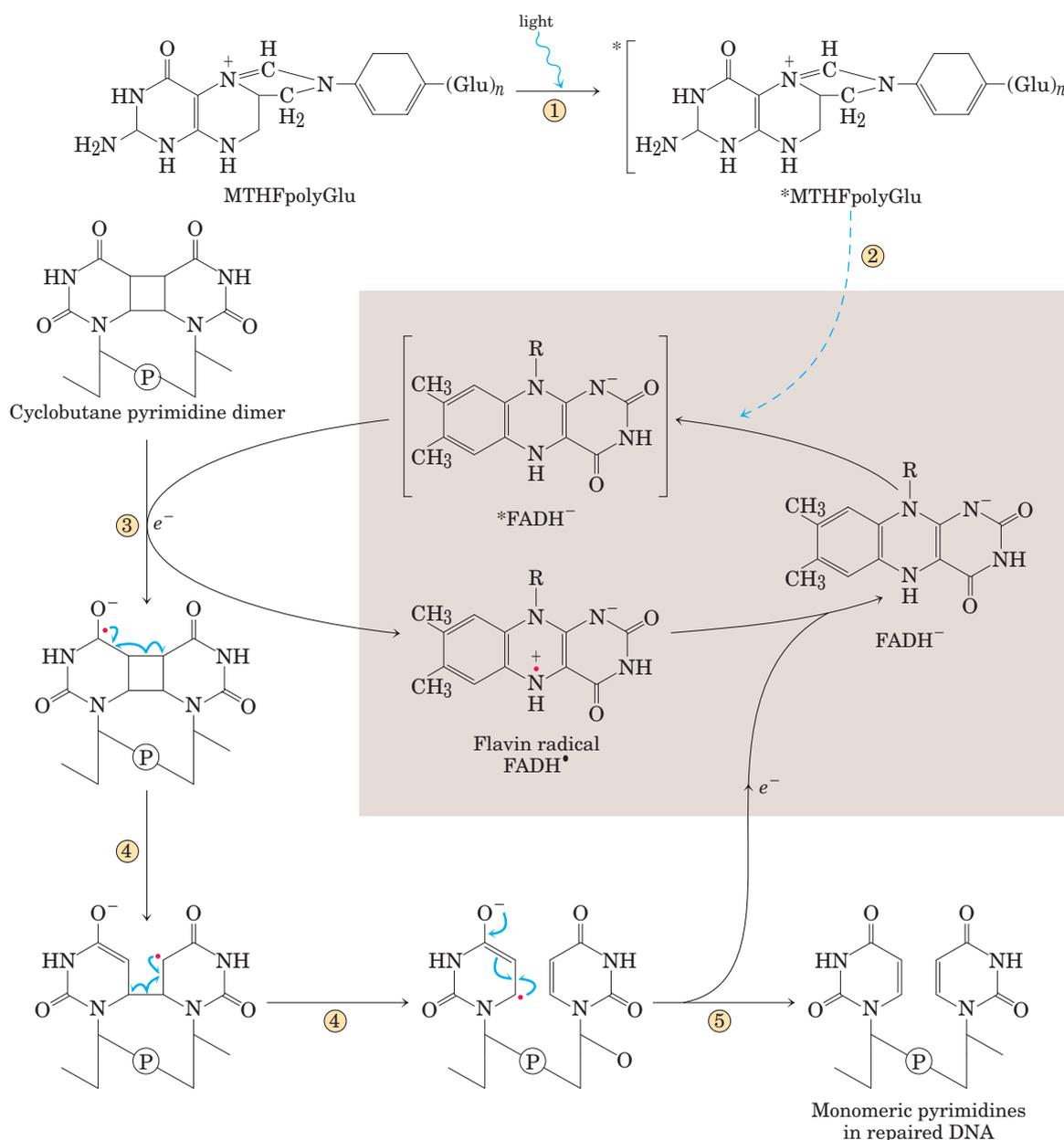


**FIGURE 25-24** Nucleotide-excision repair in *E. coli* and humans. The general pathway of nucleotide-excision repair is similar in all organisms. ① An excinuclease binds to DNA at the site of a bulky lesion and cleaves the damaged DNA strand on either side of the lesion. ② The DNA segment—of 13 nucleotides (13 mer) or 29 nucleotides (29 mer)—is removed with the aid of a helicase. ③ The gap is filled in by DNA polymerase, and ④ the remaining nick is sealed with DNA ligase.

term “excinuclease” is used to describe the unique capacity of this enzyme complex to catalyze two specific endonucleolytic cleavages, distinguishing this activity from that of standard endonucleases. A complex of the UvrA and UvrB proteins ( $A_2B$ ) scans the DNA and binds to the site of a lesion. The UvrA dimer then dissociates, leaving a tight UvrB-DNA complex. UvrC protein then binds to UvrB, and UvrB makes an incision at the fifth phosphodiester bond on the 3' side of the lesion. This is followed by a UvrC-mediated incision at the eighth phosphodiester bond on the 5' side. The resulting 12 to 13 nucleotide fragment is removed by UvrD helicase. The short gap thus created is filled in by DNA polymerase I and DNA ligase. This pathway is a primary repair route for many types of lesions, including cyclo-

butane pyrimidine dimers, 6-4 photoproducts (see Fig. 8-34), and several other types of base adducts including benzo[*a*]pyrene-guanine, which is formed in DNA by exposure to cigarette smoke. The nucleolytic activity of the ABC excinuclease is novel in the sense that two cuts are made in the DNA (Fig. 25-24).

The mechanism of eukaryotic excinucleases is quite similar to that of the bacterial enzyme, although 16 polypeptides with no similarity to the *E. coli* excinuclease subunits are required for the dual incision. As described in Chapter 26, some of the nucleotide-excision repair and base-excision repair in eukaryotes is closely tied to transcription. Genetic deficiencies in nucleotide-excision repair in humans give rise to a variety of serious diseases (Box 25-1).

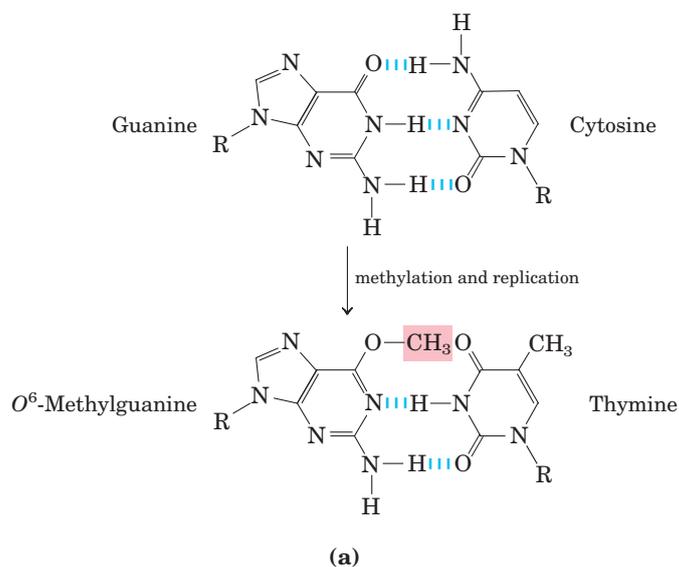


**MECHANISM FIGURE 25-25** Repair of pyrimidine dimers with photolyase. Energy derived from absorbed light is used to reverse the photoreaction that caused the lesion. The two chromophores in *E. coli* photolyase ( $M_r$  54,000),  $N^5,N^{10}$ -methenyltetrahydrofolylpolyglutamate (MTHFpolyGlu) and FADH<sup>-</sup>, perform complementary functions. On binding of photolyase to a pyrimidine dimer, repair proceeds as follows. ① A blue-light photon (300 to 500 nm wavelength) is ab-

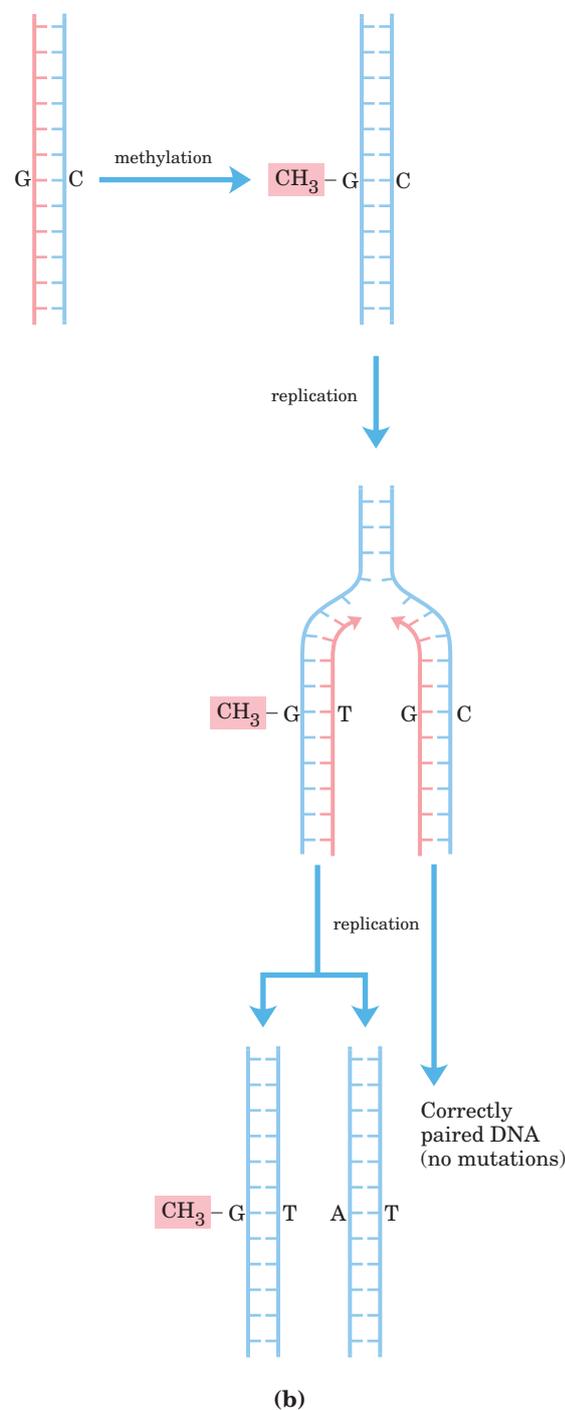
sorbed by the MTHFpolyGlu, which functions as a photoantenna. ② The excitation energy passes to FADH<sup>-</sup> in the active site of the enzyme. ③ The excited flavin (\*FADH<sup>-</sup>) donates an electron to the pyrimidine dimer (shown here in a simplified representation) to generate an unstable dimer radical. ④ Electronic rearrangement restores the monomeric pyrimidines, and ⑤ the electron is transferred back to the flavin radical to regenerate FADH<sup>-</sup>.

**Direct Repair** Several types of damage are repaired without removing a base or nucleotide. The best-characterized example is direct photoreactivation of cyclobutane pyrimidine dimers, a reaction promoted by **DNA photolyases**. Pyrimidine dimers result from an ultraviolet light-induced reaction, and photolyases use energy derived from absorbed light to reverse the dam-

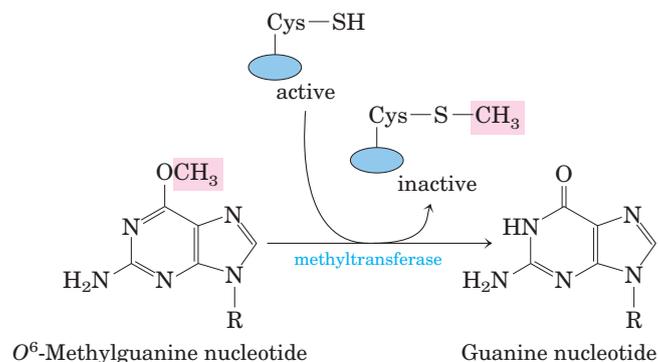
age (Fig. 25-25). Photolyases generally contain two cofactors that serve as light-absorbing agents, or chromophores. One of the chromophores is always FADH<sup>-</sup>. In *E. coli* and yeast, the other chromophore is a folate. The reaction mechanism entails the generation of free radicals. DNA photolyases are not present in the cells of placental mammals (which include humans).



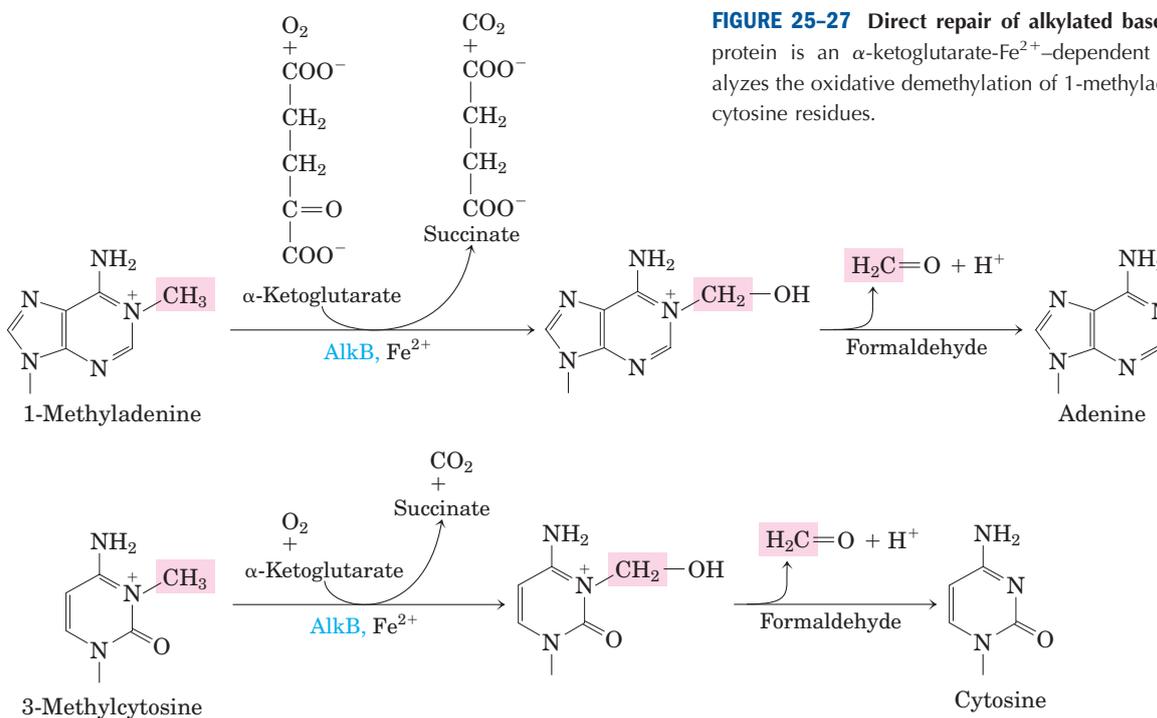
**FIGURE 25-26** Example of how DNA damage results in mutations. (a) The methylation product *O*<sup>6</sup>-methylguanine pairs with thymine rather than cytosine. (b) If not repaired, this leads to a G≡C to A=T mutation after replication.



Additional examples can be seen in the repair of nucleotides with alkylation damage. The modified nucleotide *O*<sup>6</sup>-methylguanine forms in the presence of alkylating agents and is a common and highly mutagenic lesion (p. 295). It tends to pair with thymine rather than cytosine during replication, and therefore causes G≡C to A=T mutations (Fig. 25-26). Direct repair of *O*<sup>6</sup>-methylguanine is carried out by *O*<sup>6</sup>-methylguanine-DNA methyltransferase, a protein that catalyzes transfer of the methyl group of *O*<sup>6</sup>-methylguanine to one of its own Cys residues. This methyltransferase is not strictly an enzyme, because a single methyl transfer event permanently methylates the protein, making it inactive in this pathway. The consumption of an entire protein molecule to correct a single damaged base is another vivid illustration of the priority given to maintaining the integrity of cellular DNA.



A very different but equally direct mechanism is used to repair 1-methyladenine and 3-methylcytosine. The amino groups of A and C residues are sometimes methylated when the DNA is single-stranded, and the methylation directly affects proper base pairing. In *E. coli*, oxidative demethylation of these alkylated nucleotides is mediated by the AlkB protein, a member of the  $\alpha$ -ketoglutarate-Fe<sup>2+</sup>-dependent dioxygenase superfamily (Fig. 25-27). (See Box 4-3 for a description of another member of this enzyme family.)



**FIGURE 25-27** Direct repair of alkylated bases by AlkB. The AlkB protein is an  $\alpha$ -ketoglutarate-Fe<sup>2+</sup>-dependent dioxygenase. It catalyzes the oxidative demethylation of 1-methyladenine and 3-methylcytosine residues.

### The Interaction of Replication Forks with DNA Damage Can Lead to Error-Prone Translesion DNA Synthesis

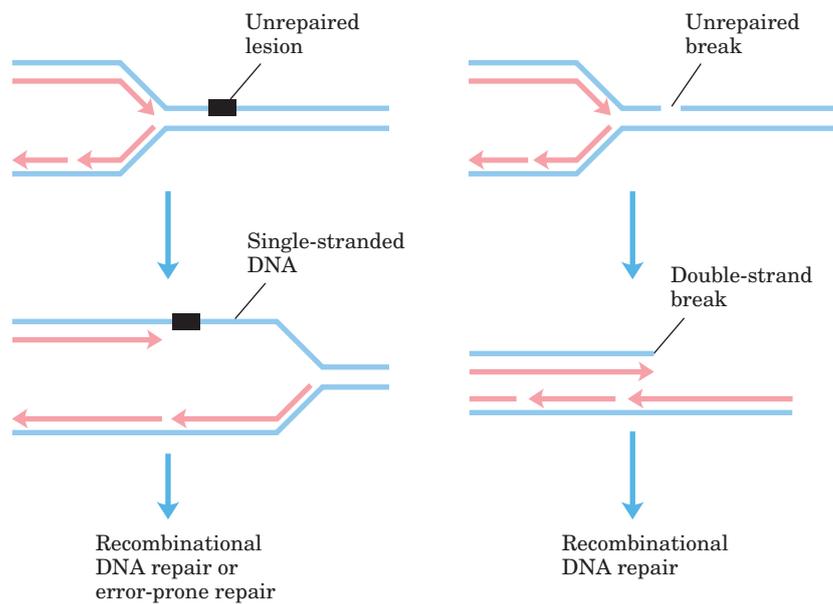
The repair pathways considered to this point generally work only for lesions in double-stranded DNA, the undamaged strand providing the correct genetic information to restore the damaged strand to its original state. However, in certain types of lesions, such as double-strand breaks, double-strand cross-links, or lesions in a single-stranded DNA, the complementary strand is itself damaged or is absent. Double-strand breaks and lesions in single-stranded DNA most often arise when a replication fork encounters an unrepaired DNA lesion (Fig. 25-28). Such lesions and DNA cross-links can also result from ionizing radiation and oxidative reactions.

At a stalled bacterial replication fork, there are two avenues for repair. In the absence of a second strand, the information required for accurate repair must come from a separate, homologous chromosome. The repair system thus involves homologous genetic recombination. This **recombinational DNA repair** is considered in detail in Section 25.3. Under some conditions, a second repair pathway, **error-prone translesion DNA synthesis** (often abbreviated TLS), becomes available. When this pathway is active, DNA repair becomes significantly less accurate and a high mutation rate can result. In bacteria, error-prone translesion DNA synthesis is part of a cellular stress response to extensive DNA damage known, appropriately enough, as the **SOS response**. Some SOS proteins, such as the UvrA and UvrB proteins already described (Table 25-6), are normally

present in the cell but are induced to higher levels as part of the SOS response. Additional SOS proteins participate in the pathway for error-prone repair; these include the UmuC and UmuD proteins (“Umu” from *unmutable*; lack of the *umu* gene function eliminates error-prone repair). The UmuD protein is cleaved in an SOS-regulated process to a shorter form called UmuD’, which forms a complex with UmuC to create a specialized DNA polymerase (DNA polymerase V) that can replicate past many of the DNA lesions that would normally block replication. Proper base pairing is often impossible at the site of such a lesion, so this translesion replication is error-prone.

Given the emphasis on the importance of genomic integrity throughout this chapter, the existence of a system that increases the rate of mutation may seem incongruous. However, we can think of this system as a desperation strategy. The *umuC* and *umuD* genes are fully induced only late in the SOS response, and they are not activated for translesion synthesis initiated by UmuD cleavage unless the levels of DNA damage are particularly high and all replication forks are blocked. The mutations resulting from DNA polymerase V-mediated replication kill some cells and create deleterious mutations in others, but this is the biological price an organism pays to overcome an otherwise insurmountable barrier to replication, as it permits at least a few mutant cells to survive.

In addition to DNA polymerase V, translesion replication requires the RecA protein, SSB, and some subunits derived from DNA polymerase III. Yet another DNA polymerase, DNA polymerase IV, is also induced during



**FIGURE 25-28 DNA damage and its effect on DNA replication.** If the replication fork encounters an unrepaired lesion or strand break, replication generally halts and the fork may collapse. A lesion is left behind in an unrepliated, single-stranded segment of the DNA; a strand break becomes a double-strand break. In each case, the damage to one strand cannot be repaired by mechanisms described earlier in this chapter, because the complementary strand required to direct accurate repair is damaged or absent. There are two possible avenues for repair: recombinational DNA repair (described in Fig. 25-37) or, when lesions are unusually numerous, error-prone repair. The latter mechanism involves a novel DNA polymerase (DNA polymerase V, encoded by the *umuC* and *umuD* genes) that can replicate, albeit inaccurately, over many types of lesions. The repair mechanism is referred to as error-prone because mutations often result.

the SOS response. Replication by DNA polymerase IV, a product of the *dinB* gene, is also highly error-prone. The bacterial DNA polymerases IV and V are part of a family of TLS polymerases found in all organisms. These enzymes lack a proofreading exonuclease activity, and the fidelity of replicative base selection can be reduced

by a factor of  $10^2$ , lowering overall replication fidelity to one error in ~1,000 nucleotides.

Mammals have many low-fidelity DNA polymerases of the TLS polymerase family. However, the presence of these enzymes does not necessarily translate into an unacceptable mutational burden, because most of the

**TABLE 25-6 Genes Induced as Part of the SOS Response in *E. coli***

Gene name	Protein encoded and/or role in DNA repair
<b>Genes of known function</b>	
<i>polB</i> ( <i>dinA</i> )	Encodes polymerization subunit of DNA polymerase II, required for replication restart in recombinational DNA repair
<i>uvrA</i> } <i>uvrB</i> }	Encode ABC excinuclease subunits UvrA and UvrB
<i>umuC</i> } <i>umuD</i> }	Encode DNA polymerase V
<i>sulA</i>	Encodes protein that inhibits cell division, possibly to allow time for DNA repair
<i>recA</i>	Encodes RecA protein, required for error-prone repair and recombinational repair
<i>dinB</i>	Encodes DNA polymerase IV
<b>Genes involved in DNA metabolism, but role in DNA repair unknown</b>	
<i>ssb</i>	Encodes single-stranded DNA-binding protein (SSB)
<i>uvrD</i>	Encodes DNA helicase II (DNA-unwinding protein)
<i>himA</i>	Encodes subunit of integration host factor (IHF), involved in site-specific recombination, replication, transposition, regulation of gene expression
<i>recN</i>	Required for recombinational repair
<b>Genes of unknown function</b>	
<i>dinD</i>	
<i>dinF</i>	

Note: Some of these genes and their functions are further discussed in Chapter 28.

enzymes also have specialized functions in DNA repair. DNA polymerase  $\eta$  (eta), for example, is a TLS polymerase found in all eukaryotes. It promotes translesion synthesis primarily across cyclobutane T–T dimers. Few mutations result in this case, because the enzyme preferentially inserts two A residues across from the linked T residues. Several other low-fidelity polymerases, including DNA polymerases  $\beta$ ,  $\iota$  (iota), and  $\lambda$ , have specialized roles in eukaryotic base-excision repair. Each of these enzymes has a 5'-deoxyribose phosphate lyase activity in addition to its polymerase activity. After base removal by a glycosylase and backbone cleavage by an AP endonuclease, these enzymes remove the abasic site (a 5'-deoxyribose phosphate) and fill in the very short gap with their polymerase activity. The frequency of mutations due to DNA polymerase  $\eta$  activity is minimized by the very short lengths (often one nucleotide) of DNA synthesized.

What emerges from research into cellular DNA repair systems is a picture of a DNA metabolism that maintains genomic integrity with multiple and often redundant systems. In the human genome, more than 130 genes encode proteins dedicated to the repair of DNA. In many cases, the loss of function of one of these proteins results in genomic instability and an increased occurrence of oncogenesis (Box 25–1). These repair systems are often integrated with the DNA replication systems and are complemented by the recombination systems that we turn to next.

### SUMMARY 25.2 DNA Repair

- Cells have many systems for DNA repair. Mismatch repair in *E. coli* is directed by transient nonmethylation of (5')GATC sequences on the newly synthesized strand.
- Base-excision repair systems recognize and repair damage caused by environmental agents (such as radiation and alkylating agents) and spontaneous reactions of nucleotides. Some repair systems recognize and excise only damaged or incorrect bases, leaving an AP (abasic) site in the DNA. This is repaired by excision and replacement of the DNA segment containing the AP site.
- Nucleotide-excision repair systems recognize and remove a variety of bulky lesions and pyrimidine dimers. They excise a segment of the DNA strand including the lesion, leaving a gap that is filled in by DNA polymerase and ligase activities.
- Some DNA damage is repaired by direct reversal of the reaction causing the damage:

pyrimidine dimers are directly converted to monomeric pyrimidines by a photolyase, and the methyl group of  $O^6$ -methylguanine is removed by a methyltransferase.

- In bacteria, error-prone translesion DNA synthesis, involving TLS DNA polymerases, occurs in response to very heavy DNA damage. In eukaryotes, similar polymerases have specialized roles in DNA repair that minimize the introduction of mutations.

## 25.3 DNA Recombination

The rearrangement of genetic information within and among DNA molecules encompasses a variety of processes, collectively placed under the heading of genetic recombination. The practical applications of DNA rearrangements in altering the genomes of increasing numbers of organisms are now being explored (Chapter 9).

Genetic recombination events fall into at least three general classes. **Homologous genetic recombination** (also called general recombination) involves genetic exchanges between any two DNA molecules (or segments of the same molecule) that share an extended region of nearly identical sequence. The actual sequence of bases is irrelevant, as long as it is similar in the two DNAs. In **site-specific recombination**, the exchanges occur only at a *particular* DNA sequence. **DNA transposition** is distinct from both other classes in that it usually involves a short segment of DNA with the remarkable capacity to move from one location in a chromosome to another. These “jumping genes” were first observed in maize in the 1940s by Barbara McClintock. There is in addition a wide range of unusual genetic rearrangements for which no mechanism or purpose has yet been proposed. Here we focus on the three general classes.



Barbara McClintock,  
1902–1992

The functions of genetic recombination systems are as varied as their mechanisms. They include roles in specialized DNA repair systems, specialized activities in DNA replication, regulation of expression of certain genes, facilitation of proper chromosome segregation during eukaryotic cell division, maintenance of genetic diversity, and implementation of programmed genetic rearrangements during embryonic development. In most cases, genetic recombination is closely integrated with other processes in DNA metabolism, and this becomes a theme of our discussion.

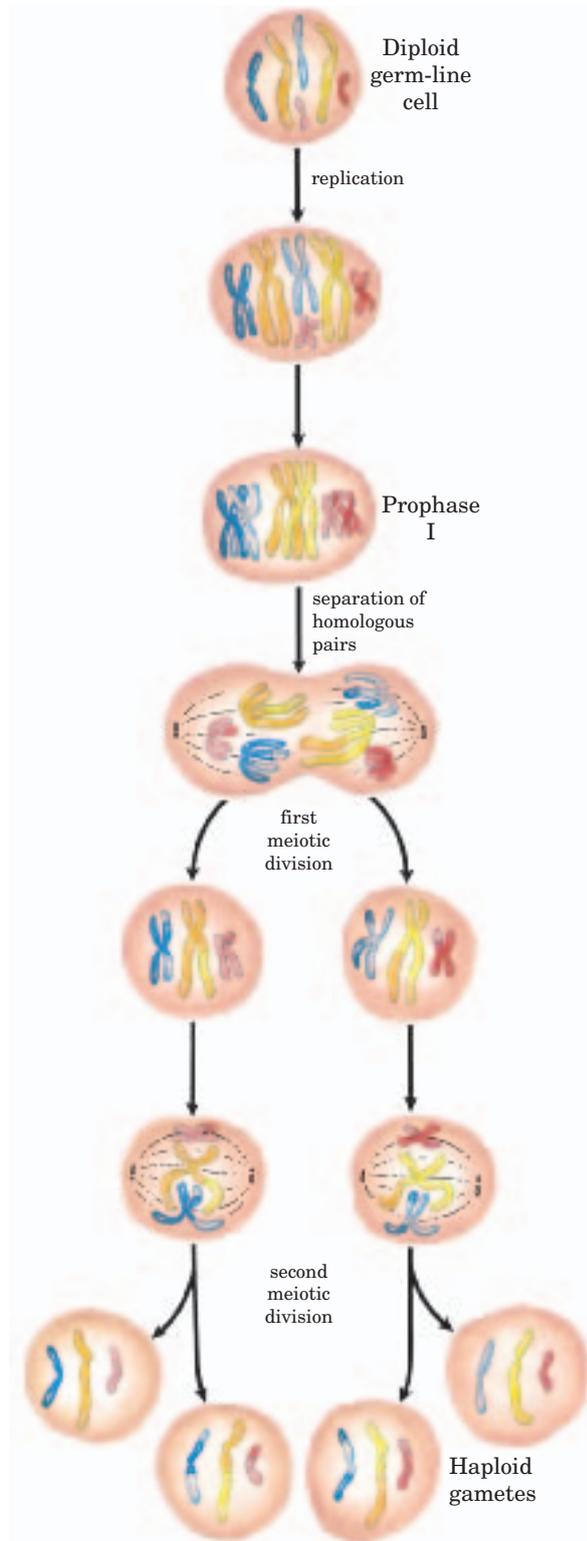
### Homologous Genetic Recombination Has Several Functions

In bacteria, homologous genetic recombination is primarily a DNA repair process and in this context (as noted in Section 25.2) is referred to as **recombinational DNA repair**. It is usually directed at the reconstruction of replication forks stalled at the site of DNA damage. Homologous genetic recombination can also occur during conjugation (mating), when chromosomal DNA is transferred from a donor to a recipient bacterial cell. Recombination during conjugation, although rare in wild bacterial populations, contributes to genetic diversity.

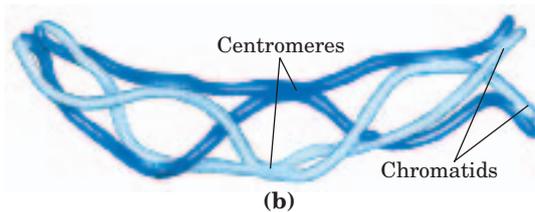
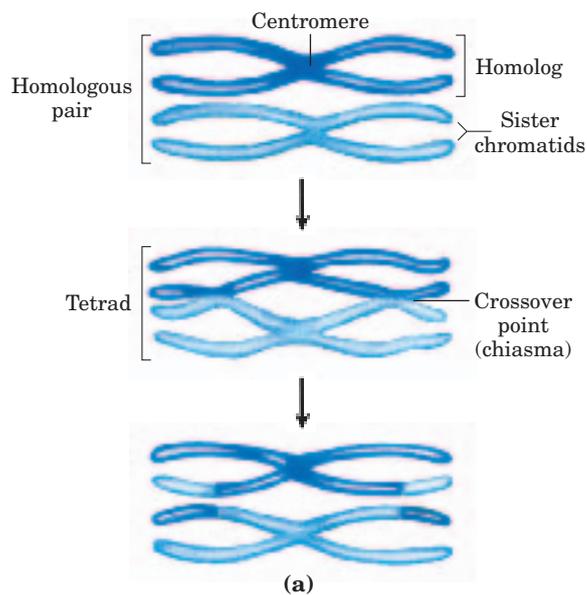
In eukaryotes, homologous genetic recombination can have several roles in replication and cell division, including the repair of stalled replication forks. Recombination occurs with the highest frequency during **meiosis**, the process by which diploid germ-line cells with two sets of chromosomes divide to produce haploid gametes—sperm cells or ova in higher eukaryotes—each gamete having only one member of each chromosome pair (Fig. 25–29). Meiosis begins with replication of the DNA in the germ-line cell so that each DNA molecule is present in four copies. The cell then goes through two rounds of cell division without an intervening round of DNA replication. This reduces the DNA content to the haploid level in each gamete.

After the DNA is replicated during prophase of the first meiotic division, the resulting sister chromatids remain associated at their centromeres. At this stage, each set of four homologous chromosomes exists as two pairs of chromatids. Genetic information is now exchanged between the closely associated homologous chromatids

by homologous genetic recombination, a process involving the breakage and rejoining of DNA (Fig. 25–30). This exchange, also referred to as crossing over, can be observed with the light microscope. Crossing over links the two pairs of sister chromatids together at points called chiasmata (singular, chiasma).



**FIGURE 25–29 Meiosis in eukaryotic germ-line cells.** The chromosomes of a hypothetical diploid germ-line cell (six chromosomes; three homologous pairs) replicate and are held together at their centromeres. Each replicated double-stranded DNA molecule is called a chromatid (sister chromatid). In prophase I, just before the first meiotic division, the three homologous sets of chromatids align to form tetrads, held together by covalent links at homologous junctions (chiasmata). Crossovers occur within the chiasmata (see Fig. 25–30). These transient associations between homologs ensure that the two tethered chromosomes segregate properly in the next step, when they migrate toward opposite poles of the dividing cell in the first meiotic division. The products of this division are two daughter cells, each with three pairs of chromatids. The pairs now line up across the equator of the cell in preparation for separation of the chromatids (now called chromosomes). The second meiotic division produces four haploid daughter cells that can serve as gametes. Each has three chromosomes, half the number of the diploid germ-line cell. The chromosomes have recombined and recombined.



**FIGURE 25-30 Crossing over.** (a) Crossing over often produces an exchange of genetic material. (b) The homologous chromosomes of a grasshopper are shown during prophase I of meiosis. Many points of

joining (chiasmata) are evident between the two homologous pairs of chromatids. These chiasmata are the physical manifestation of prior homologous recombination (crossing over) events.

Crossing over effectively links together all four homologous chromatids, a linkage that is essential to the proper segregation of chromosomes in the subsequent meiotic cell divisions. Crossing over is not an entirely random process, and “hot spots” have been identified on many eukaryotic chromosomes. However, the assumption that crossing over can occur with equal probability at almost any point along the length of two homologous chromosomes remains a reasonable approximation in many cases, and it is this assumption that permits the genetic mapping of genes. The frequency of homologous recombination in any region separating two points on a chromosome is roughly proportional to the distance between the points, and this allows determination of the relative positions of and distances between different genes.

Homologous recombination thus serves at least three identifiable functions: (1) it contributes to the repair of several types of DNA damage; (2) it provides, in eukaryotic cells, a transient physical link between chromatids that promotes the orderly segregation of chromosomes at the first meiotic cell division; and (3) it enhances genetic diversity in a population.

### Recombination during Meiosis Is Initiated with Double-Strand Breaks

A likely pathway for homologous recombination during meiosis is outlined in Figure 25-31a. The model has four key features. First, homologous chromosomes are aligned. Second, a double-strand break in a DNA mole-

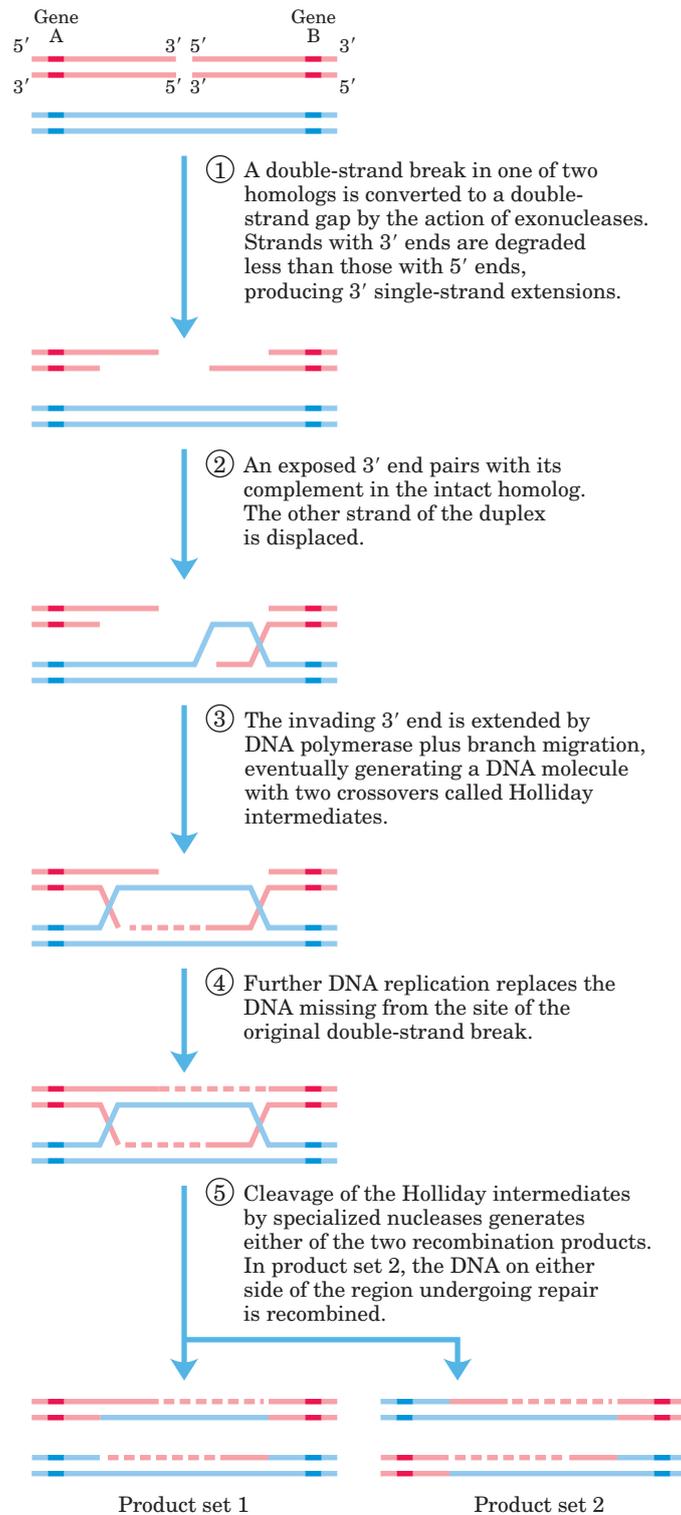
cule is enlarged by an exonuclease, leaving a single-strand extension with a free 3′-hydroxyl group at the broken end (step ①). Third, the exposed 3′ ends invade the intact duplex DNA, and this is followed by **branch migration** (Fig. 25-32) and/or replication to create a pair of crossover structures, called Holliday junctions (Fig. 25-31a, steps ② to ④). Fourth, cleavage of the two crossovers creates two complete recombinant products (step ⑤).

In this **double-strand break repair model** for recombination, the 3′ ends are used to initiate the genetic exchange. Once paired with the complementary strand on the intact homolog, a region of hybrid DNA is created that contains complementary strands from two different parent DNAs (the product of step ② in Fig. 25-31a). Each of the 3′ ends can then act as a primer for DNA replication. The structures thus formed, **Holliday intermediates** (Fig. 25-31b), are a feature of homologous genetic recombination pathways in all organisms.

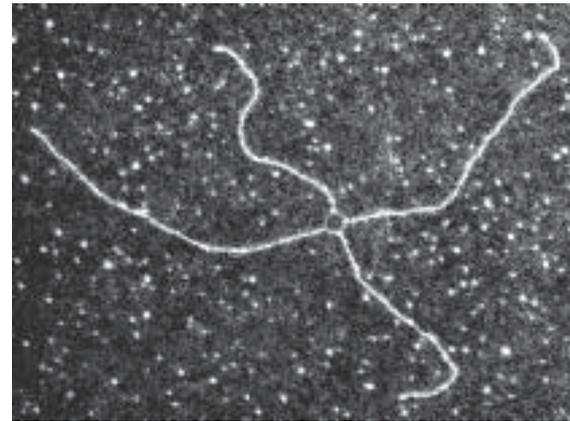
Homologous recombination can vary in many details from one species to another, but most of the steps outlined above are generally present in some form. There are two ways to cleave, or “resolve,” the Holliday intermediate so that the two recombinant products carry genes in the same linear order as in the substrates—the original, unrecombined chromosomes (step ⑤ of Fig. 25-31a). If cleaved one way, the DNA flanking the region containing the hybrid DNA is not recombined; if cleaved the other way, the flanking DNA is recombined. Both outcomes are observed *in vivo* in eukaryotes and prokaryotes.

The homologous recombination illustrated in Figure 25-31 is a very elaborate process with subtle molecular consequences for the generation of genetic diversity. To understand how this process contributes to diversity, we should keep in mind that the two homologous chromo-

somes that undergo recombination are not necessarily *identical*. The linear array of genes may be the same, but the base sequences in some of the genes may differ slightly (in different alleles). In a human, for example, one chromosome may contain the allele for hemoglobin A



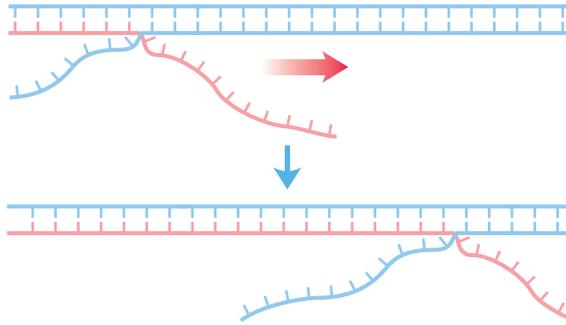
(a)



(b)

**FIGURE 25-31** Recombination during meiosis.

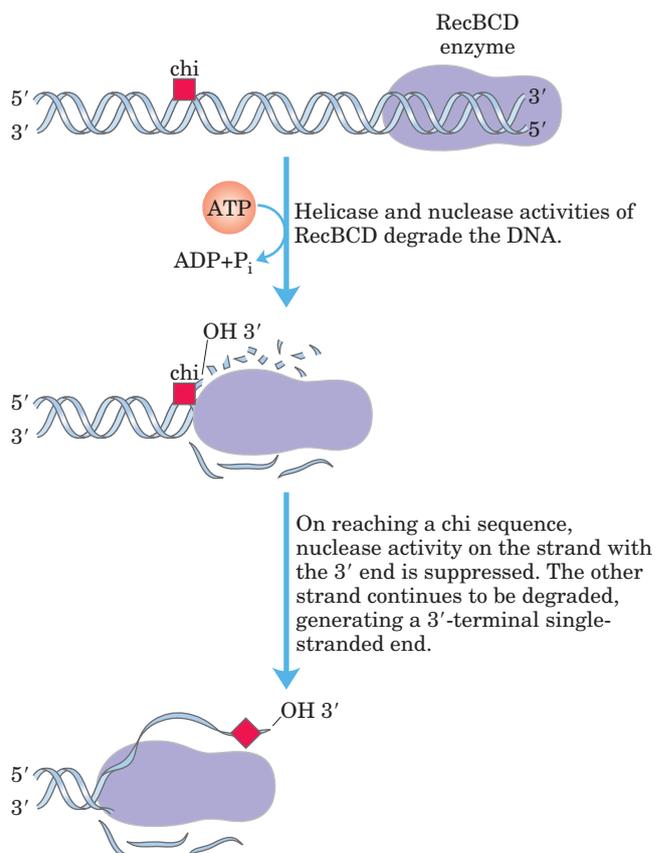
(a) Model of double-strand break repair for homologous genetic recombination. The two homologous chromosomes involved in this recombination event have similar sequences. Each of the two genes shown has different alleles on the two chromosomes. The DNA strands and alleles are colored differently so that their fate is evident. The steps are described in the text. (b) A Holliday intermediate formed between two bacterial plasmids *in vivo*, as seen with the electron microscope. The intermediates are named for Robin Holliday, who first proposed their existence in 1964.



(normal hemoglobin) while the other contains the allele for hemoglobin S (the sickle-cell mutation). The difference may consist of no more than one base pair among millions. Homologous recombination does not change the linear array of genes, but it can determine which alleles become linked together on a single chromosome.

### Recombination Requires a Host of Enzymes and Other Proteins

Enzymes that promote various steps of homologous recombination have been isolated from both prokaryotes and eukaryotes. In *E. coli*, the *recB*, *recC*, and *recD* genes encode the RecBCD enzyme, which has both helicase and nuclease activities. The RecA protein pro-



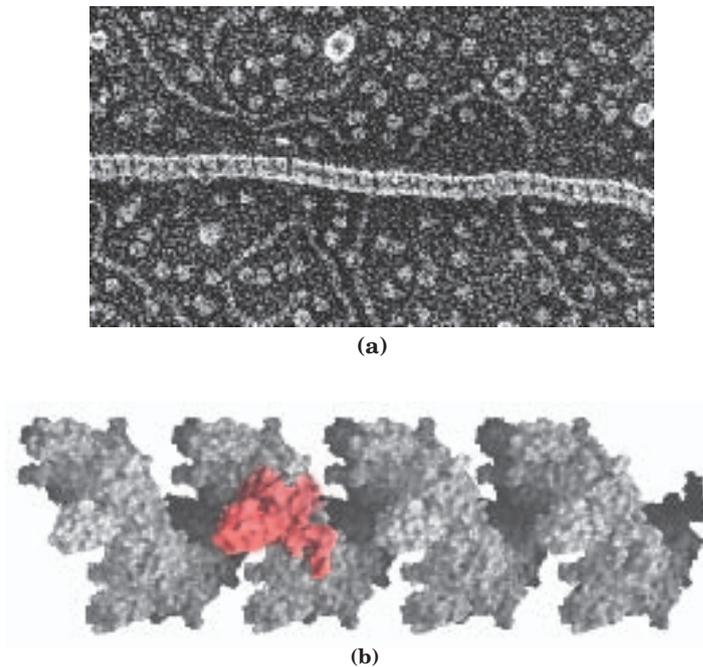
**FIGURE 25-32 Branch migration.** When a template strand pairs with two different complementary strands, a branch is formed at the point where the three complementary strands meet. The branch “migrates” when base pairing to one of the two complementary strands is broken and replaced with base pairing to the other complementary strand. In the absence of an enzyme to direct it, this process can move the branch spontaneously in either direction. Spontaneous branch migration is blocked wherever one of the otherwise complementary strands has a sequence nonidentical to the other strand.

motes all the central steps in the homologous recombination process: the pairing of two DNAs, formation of Holliday intermediates, and branch migration (as described below). The RuvA and RuvB proteins (repair of UV damage) form a complex that binds to Holliday intermediates, displaces RecA protein, and promotes branch migration at higher rates than does RecA. Nucleases that specifically cleave Holliday intermediates, often called resolvases, have been isolated from bacteria and yeast. The RuvC protein is one of at least two such nucleases in *E. coli*.

The RecBCD enzyme binds to linear DNA at a free (broken) end and moves inward along the double helix, unwinding and degrading the DNA in a reaction coupled to ATP hydrolysis (Fig. 25-33). The activity of the enzyme is altered when it interacts with a sequence referred to as chi, (5')GCTGGTGG. From that point, degradation of the strand with a 3' terminus is greatly reduced, but degradation of the 5'-terminal strand is increased. This process creates a single-stranded DNA with a 3' end, which is used during subsequent steps in recombination (Fig. 25-31). The 1,009 chi sequences scattered throughout the *E. coli* genome enhance the frequency of recombination about five- to tenfold within 1,000 bp of the chi site. The enhancement declines as the distance from the site increases. Sequences that enhance recombination frequency have also been identified in several other organisms.

RecA is unusual among the proteins of DNA metabolism in that its active form is an ordered, helical filament of up to several thousand RecA monomers that assemble cooperatively on DNA (Fig. 25-34). This fila-

**FIGURE 25-33 Helicase and nuclease activities of the RecBCD enzyme.** Entering at a double-stranded end, RecBCD unwinds and degrades the DNA until it encounters a chi sequence. The interaction with chi alters the activity of RecBCD so that it generates a single-stranded DNA with a 3' end, suitable for subsequent steps in recombination. Movement of the enzyme requires ATP hydrolysis. This enzyme is believed to help initiate homologous genetic recombination in *E. coli*. It is also involved in the repair of double-strand breaks at collapsed replication forks.

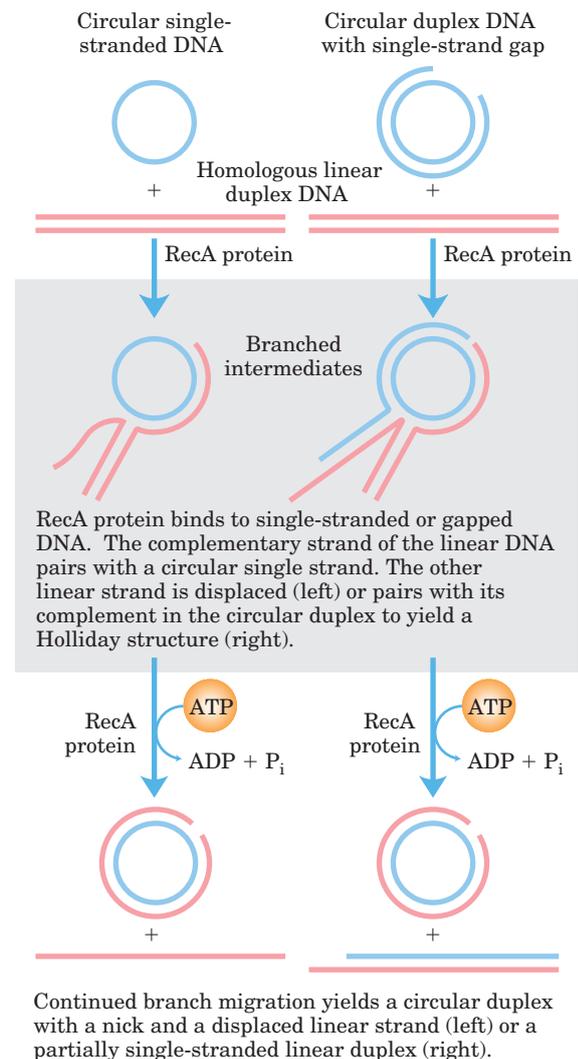


**FIGURE 25-34 RecA.** (a) Nucleoprotein filament of RecA protein on single-stranded DNA, as seen with the electron microscope. The striations indicate the right-handed helical structure of the filament. (b) Surface contour model of a 24-subunit RecA filament. The filament has six subunits per turn. One subunit is colored red to provide perspective (derived from PDB ID 2REB).

ment normally forms on single-stranded DNA, such as that produced by the RecBCD enzyme. The filament will also form on a duplex DNA with a single-strand gap; in this case, the first RecA monomers bind to the single-stranded DNA in the gap, after which the assembled filament rapidly envelops the neighboring duplex. The RecF, RecO, and RecR proteins regulate the assembly and disassembly of RecA filaments.

A useful model to illustrate the recombination activities of the RecA filament is the *in vitro* DNA strand exchange reaction (Fig. 25-35). A single strand of DNA is first bound by RecA to establish the nucleoprotein filament. The RecA filament then takes up a homologous duplex DNA and aligns it with the bound single strand. Strands are then exchanged between the two DNAs to create hybrid DNA. The exchange occurs at a rate of 6 bp/s and progresses in the 5'→3' direction relative to the single-stranded DNA within the RecA filament. This reaction can involve either three or four strands (Fig. 25-35); in the latter case, a Holliday intermediate forms during the process.

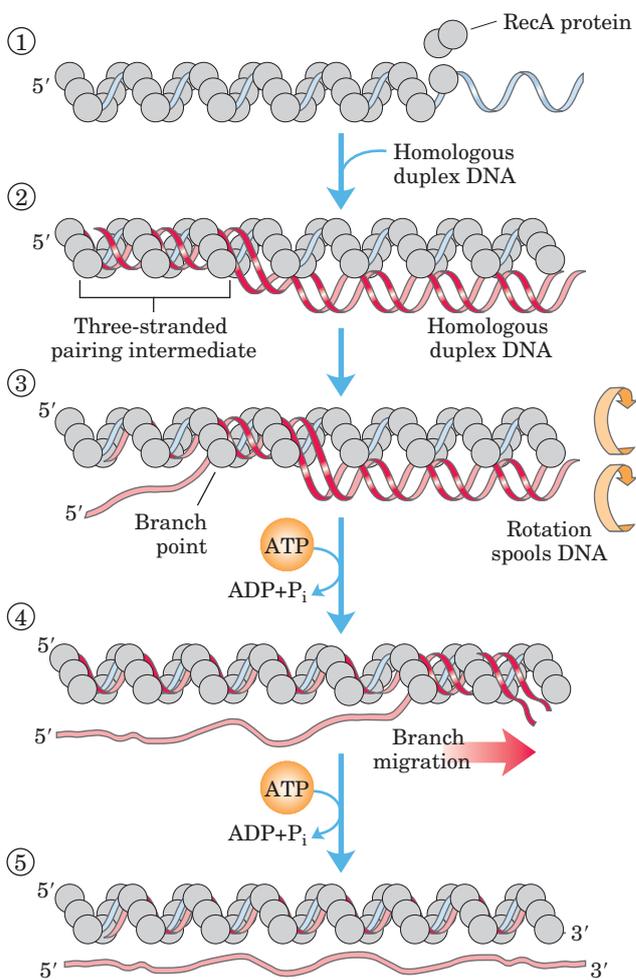
As the duplex DNA is incorporated within the RecA filament and aligned with the bound single-stranded DNA over regions of hundreds of base pairs, one strand of the duplex switches pairing partners (Fig. 25-36,



**FIGURE 25-35 DNA strand-exchange reactions promoted by RecA protein *in vitro*.** Strand exchange involves the separation of one strand of a duplex DNA from its complement and transfer of the strand to an alternative complementary strand to form a new duplex (heteroduplex) DNA. The transfer forms a branched intermediate. Formation of the final product depends on branch migration, which is facilitated by RecA. The reaction can involve three strands (left) or a reciprocal exchange between two homologous duplexes—four strands in all (right). When four strands are involved, the branched intermediate that results is a Holliday intermediate. RecA protein promotes the branch-migration phases of these reactions, using energy derived from ATP hydrolysis.

step ②). Because DNA is a helical structure, continued strand exchange requires an ordered rotation of the two aligned DNAs. This brings about a spooling action (steps ③ and ④) that shifts the branch point along the helix. ATP is hydrolyzed by RecA protein during this reaction.

Once a Holliday intermediate has formed, a host of enzymes—topoisomerases, the RuvAB branch migration protein, a resolvase, other nucleases, DNA polymerase



**FIGURE 25-36** Model for DNA strand exchange mediated by RecA protein. A three-strand reaction is shown. The balls representing RecA protein are undersized relative to the thickness of DNA to clarify the fate of the DNA strands. ① RecA protein forms a filament on the single-stranded DNA. ② A homologous duplex incorporates into this complex. ③ As spooling shifts the three-stranded region from left to right, one of the strands in the duplex is transferred to the single strand originally bound in the filament. The other strand of the duplex is displaced, and a new duplex forms within the filament. As rotation continues (④ and ⑤), the displaced strand separates entirely. In this model, hydrolysis of ATP by RecA protein rotates the two DNA molecules relative to each other and thus directs the strand exchange from left to right as shown.

I or III, and DNA ligase—are required to complete recombination. The RuvC protein ( $M_r$  20,000) of *E. coli* cleaves Holliday intermediates to generate full-length, unbranched chromosome products.

### All Aspects of DNA Metabolism Come Together to Repair Stalled Replication Forks

Like all cells, bacteria sustain high levels of DNA damage even under normal growth conditions. Most DNA

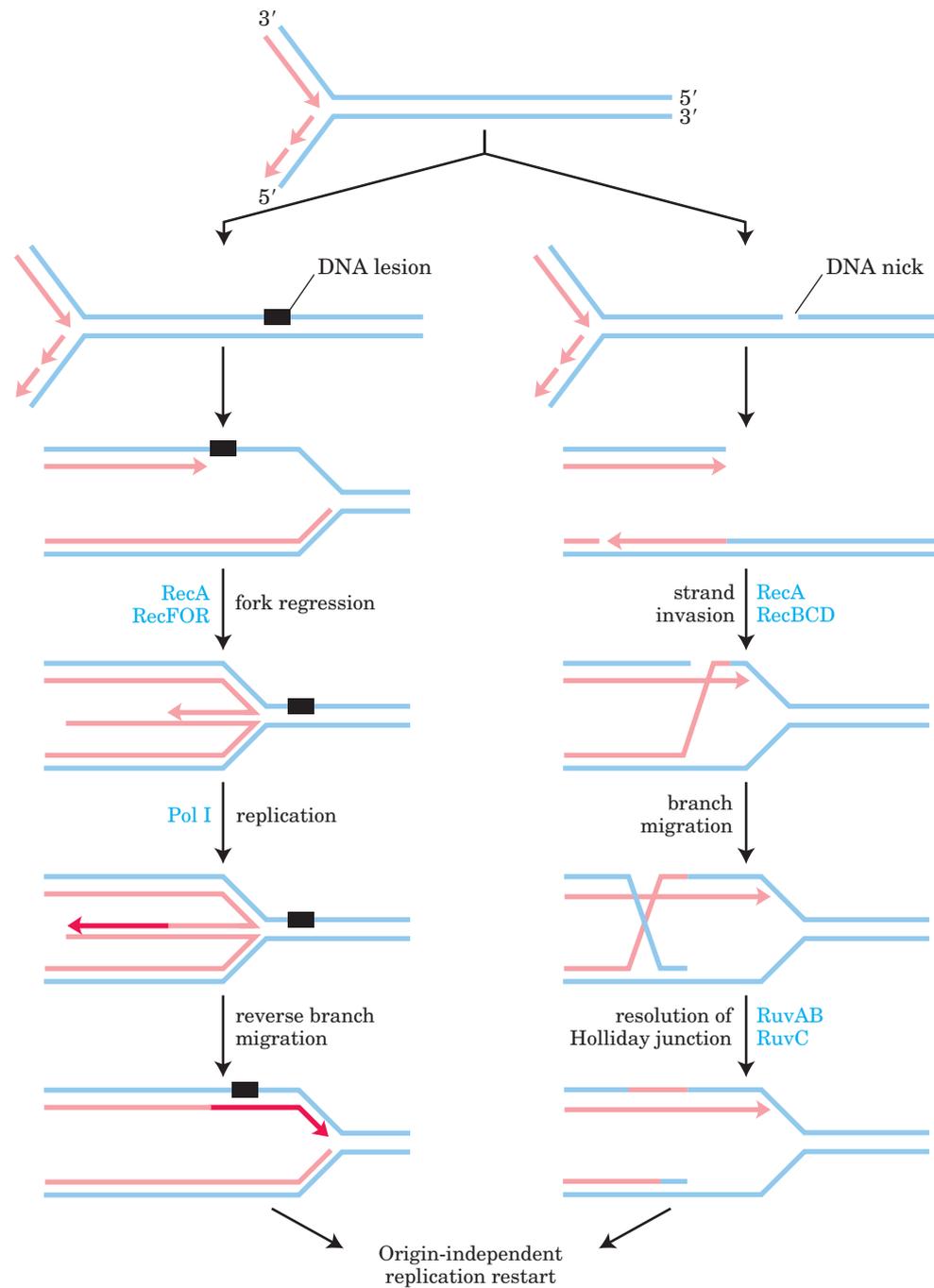
lesions are repaired rapidly by base-excision repair, nucleotide-excision repair, and the other pathways described earlier. Nevertheless, almost every bacterial replication fork encounters an unrepaired DNA lesion or break at some point in its journey from the replication origin to the terminus (Fig. 25–28). DNA polymerase III cannot proceed past many types of DNA lesions, and these encounters tend to leave the lesion in a single-strand gap. An encounter with a DNA strand break creates a double-strand break. Both situations require recombinational DNA repair (Fig. 25–37). Under normal growth conditions, stalled replication forks are reactivated by an elaborate repair pathway encompassing recombinational DNA repair, the restart of replication, and the repair of any lesions left behind. All aspects of DNA metabolism come together in this process.

After a replication fork has been halted, it can be restored by at least two major paths, both of which require the RecA protein. The repair pathway for lesion-containing DNA gaps also requires the RecF, RecO, and RecR proteins. Repair of double-strand breaks requires the RecBCD enzyme (Fig. 25–37). Additional recombination steps are followed by a process called **origin-independent restart of replication**, in which the replication fork reassembles with the aid of a complex of seven proteins (PriA, B, and C, and DnaB, C, G, and T). This complex, originally discovered as a component required for the replication of  $\phi$ X174 DNA in vitro, is now termed the **replication restart primosome**. Restart of the replication fork also requires DNA polymerase II, in a role not yet defined; this polymerase II activity gives way to DNA polymerase III for the extensive replication generally required to complete the chromosome.

The repair of stalled replication forks entails a coordinated transition from replication to recombination and back to replication. The recombination steps function to fill the DNA gap or rejoin the broken DNA branch to recreate the branched DNA structure at the replication fork. Lesions left behind in what is now duplex DNA are repaired by pathways such as base-excision or nucleotide-excision repair. Thus a wide range of enzymes encompassing every aspect of DNA metabolism ultimately take part in the repair of a stalled replication fork. This type of repair process is clearly a primary function of the homologous recombination system of every cell, and defects in recombinational DNA repair play an important role in human disease (Box 25–1).

### Site-Specific Recombination Results in Precise DNA Rearrangements

Homologous genetic recombination can involve any two homologous sequences. The second general type of recombination, site-specific recombination, is a very different type of process: recombination is limited to spe-

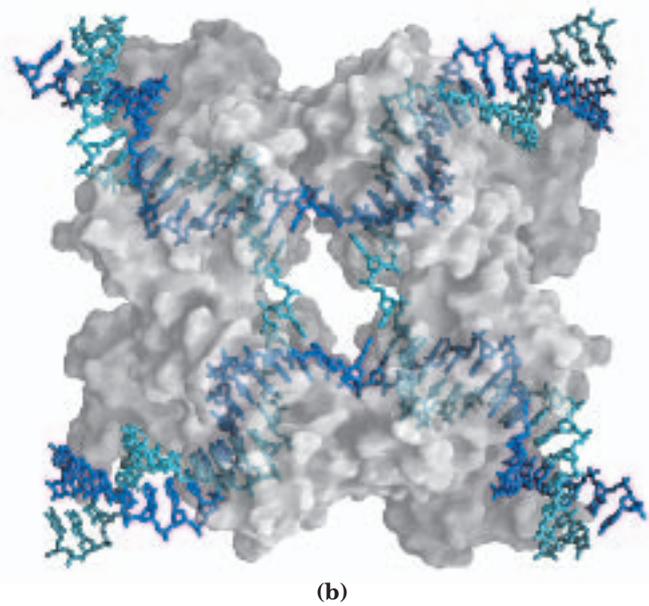
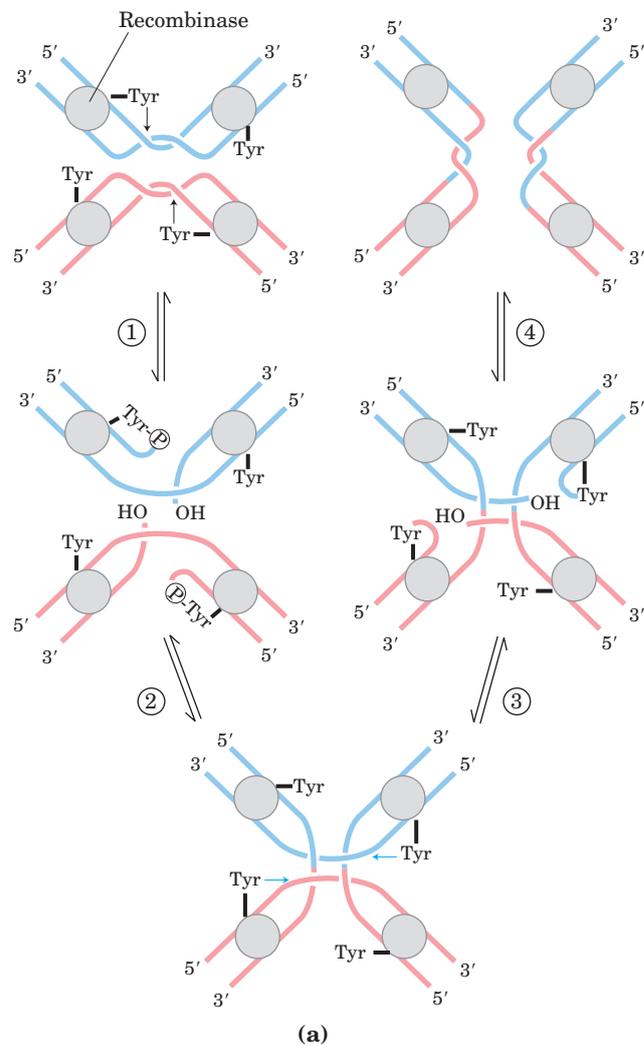


**FIGURE 25-37 Models for recombinational DNA repair of stalled replication forks.** The replication fork collapses on encountering a DNA lesion (left) or strand break (right). Recombination enzymes promote the DNA strand transfers needed to repair the branched DNA structure at the replication fork. A lesion in a single-strand gap is repaired in a reaction requiring the RecF, RecO, and RecR proteins. Double-strand breaks are repaired in a pathway requiring the RecBCD enzyme. Both pathways require RecA. Recombination intermediates

are processed by additional enzymes (e.g., RuvA, RuvB, and RuvC, which process Holliday intermediates). Lesions in double-stranded DNA are repaired by nucleotide-excision repair or other pathways. The replication fork re-forms with the aid of enzymes catalyzing origin-independent replication restart, and chromosomal replication is completed. The overall process requires an elaborate coordination of all aspects of bacterial DNA metabolism.

sific sequences. Recombination reactions of this type occur in virtually every cell, filling specialized roles that vary greatly from one species to another. Examples include regulation of the expression of certain genes and promotion of programmed DNA rearrangements in em-

bryonic development or in the replication cycles of some viral and plasmid DNAs. Each site-specific recombination system consists of an enzyme called a recombinase and a short (20 to 200 bp), unique DNA sequence where the recombinase acts (the recombination site). One or



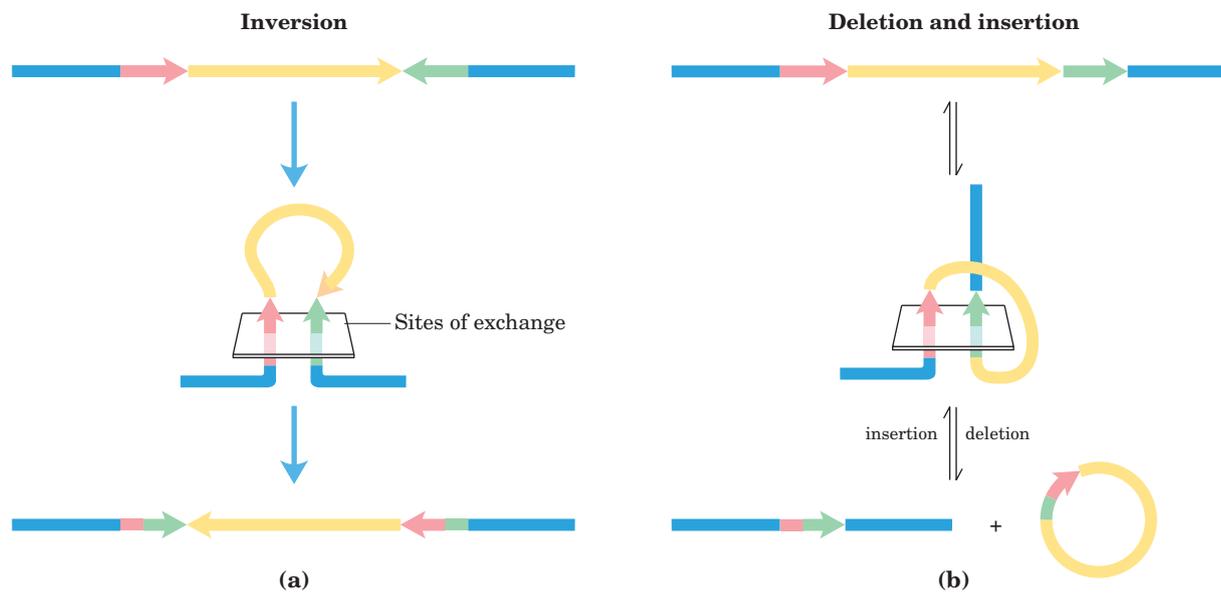
**FIGURE 25-38** A site-specific recombination reaction. **(a)** The reaction shown here is for a common class of site-specific recombinases called integrase-class recombinases (named after bacteriophage  $\lambda$  integrase, the first recombinase characterized). The reaction is carried out within a tetramer of identical subunits. Recombinase subunits bind to a specific sequence, often called simply the recombination site. ① One strand in each DNA is cleaved at particular points within the sequence. The nucleophile is the OH group of an active-site Tyr residue, and the product is a covalent phosphotyrosine link between protein and DNA. ② The cleaved strands join to new partners, producing a Holliday intermediate. Steps ③ and ④ complete the reaction by a process similar to the first two steps. The original sequence of the recombination site is regenerated after recombining the DNA flanking the site. These steps occur within a complex of multiple recombinase subunits that sometimes includes other proteins not shown here. **(b)** A surface contour model of a four-subunit integrase-class recombinase called the Cre recombinase, bound to a Holliday intermediate (shown with light blue and dark blue helix strands). The protein has been rendered transparent so that the bound DNA is visible (derived from PDB ID 3CRX).

more auxiliary proteins may regulate the timing or outcome of the reaction.

In vitro studies of many site-specific recombination systems have elucidated some general principles, including the fundamental reaction pathway (Fig. 25-38a). A separate recombinase recognizes and binds to each of two recombination sites on two different DNA molecules or within the same DNA. One DNA strand in each site is cleaved at a specific point within the site, and the recombinase becomes covalently linked to the DNA at the cleavage site through a phosphotyrosine (or phosphoserine) bond (step ①). The transient protein-DNA linkage preserves the phosphodiester bond that is lost in cleaving the DNA, so high-energy cofactors such as ATP are unnecessary in subsequent steps. The cleaved DNA strands are rejoined to new partners to form a Holliday intermediate, with new phosphodiester bonds created at the expense of the protein-DNA linkage (step

②). To complete the reaction, the process must be repeated at a second point within each of the two recombination sites (steps ③ and ④). In some systems, both strands of each recombination site are cut concurrently and rejoined to new partners without the Holliday intermediate. The exchange is always reciprocal and precise, regenerating the recombination sites when the reaction is complete. We can view a recombinase as a site-specific endonuclease and ligase in one package.

The sequences of the recombination sites recognized by site-specific recombinases are partially asymmetric (nonpalindromic), and the two recombining sites align in the same orientation during the recombinase reaction. The outcome depends on the location and orientation of the recombination sites (Fig. 25-39). If the two sites are on the same DNA molecule, the reaction either inverts or deletes the intervening DNA, determined by whether the recombination sites have the opposite or the same



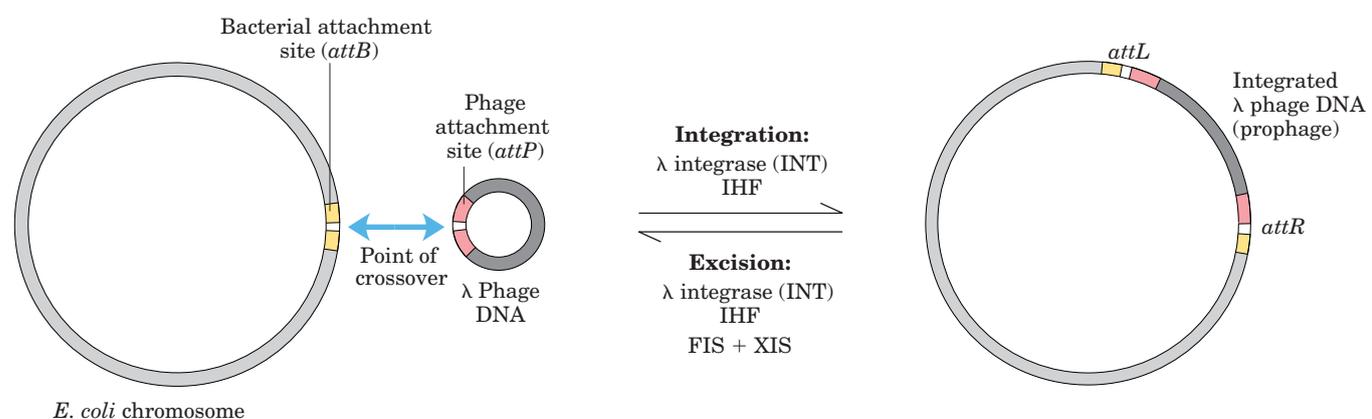
**FIGURE 25-39** Effects of site-specific recombination. The outcome of site-specific recombination depends on the location and orientation of the recombination sites (red and green) in a double-stranded DNA molecule. Orientation here (shown by arrowheads) refers to the order of nucleotides in the recombination site, not the 5'→3' direction.

(a) Recombination sites with opposite orientation in the same DNA molecule. The result is an inversion. (b) Recombination sites with the same orientation, either on one DNA molecule, producing a deletion, or on two DNA molecules, producing an insertion.

orientation, respectively. If the sites are on different DNAs, the recombination is intermolecular; if one or both DNAs are circular, the result is an insertion. Some recombinase systems are highly specific for one of these reaction types and act only on sites with particular orientations.

The first site-specific recombination system studied in vitro was that encoded by bacteriophage  $\lambda$ . When  $\lambda$  phage DNA enters an *E. coli* cell, a complex series of regulatory events commits the DNA to one of two fates.

The  $\lambda$  DNA either replicates and produces more bacteriophages (destroying the host cell) or integrates into the host chromosome, replicating passively along with the chromosome for many cell generations. Integration is accomplished by a phage-encoded recombinase ( $\lambda$  integrase) that acts at recombination sites on the phage and bacterial DNAs—at attachment sites *attP* and *attB*, respectively (Fig. 25-40). The role of site-specific recombination in regulating gene expression is considered in Chapter 28.

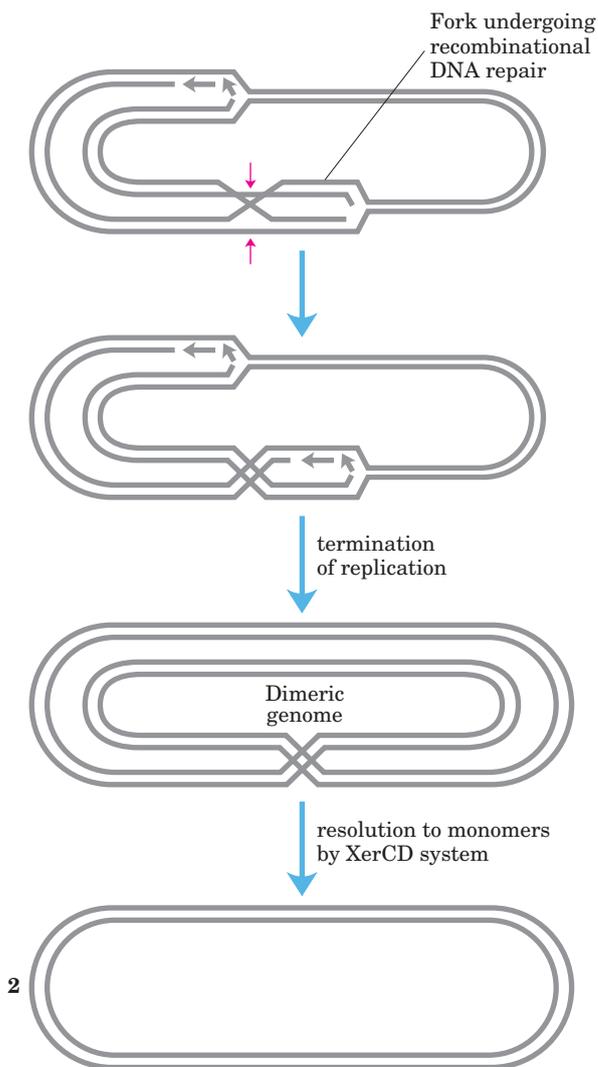


**FIGURE 25-40** Integration and excision of bacteriophage  $\lambda$  DNA at the chromosomal target site. The attachment site on the  $\lambda$  phage DNA (*attP*) shares only 15 bp of complete homology with the bacterial site (*attB*) in the region of the crossover. The reaction generates two new attachment sites (*attR* and *attL*) flanking the integrated phage DNA.

The recombinase is the  $\lambda$  integrase (or INT protein). Integration and excision use different attachment sites and different auxiliary proteins. Excision uses the proteins XIS, encoded by the bacteriophage, and FIS, encoded by the bacterium. Both reactions require the protein IHF (*integration host factor*), encoded by the bacterium.

### Complete Chromosome Replication Can Require Site-Specific Recombination

Recombinational DNA repair of a circular bacterial chromosome, while essential, sometimes generates deleterious byproducts. The resolution of a Holliday junction at a replication fork by a nuclease such as RuvC, followed by completion of replication, can give rise to one of two products: the usual two monomeric chromosomes or a contiguous dimeric chromosome (Fig. 25–41). In the latter case, the covalently linked chromosomes cannot be segregated to daughter cells at cell division and the dividing cells become “stuck.” A specialized site-specific recombination system in *E. coli*, the XerCD system, converts the dimeric chromosomes to monomeric chromosomes so that cell division can proceed. The reaction is a site-specific deletion reaction (Fig. 25–39b). This is another example of the close coordination between DNA recombination processes and other aspects of DNA metabolism.



### Transposable Genetic Elements Move from One Location to Another

We now consider the third general type of recombination system: recombination that allows the movement of transposable elements, or **transposons**. These segments of DNA, found in virtually all cells, move, or “jump,” from one place on a chromosome (the donor site) to another on the same or a different chromosome (the target site). DNA sequence homology is not usually required for this movement, called **transposition**; the new location is determined more or less randomly. Insertion of a transposon in an essential gene could kill the cell, so transposition is tightly regulated and usually very infrequent. Transposons are perhaps the simplest of molecular parasites, adapted to replicate passively within the chromosomes of host cells. In some cases they carry genes that are useful to the host cell, and thus exist in a kind of symbiosis with the host.

Bacteria have two classes of transposons. **Insertion sequences** (simple transposons) contain only the sequences required for transposition and the genes for proteins (transposases) that promote the process. **Complex transposons** contain one or more genes in addition to those needed for transposition. These extra genes might, for example, confer resistance to antibiotics and thus enhance the survival chances of the host cell. The spread of antibiotic-resistance elements among disease-causing bacterial populations that is rendering some antibiotics ineffectual (pp. 925–926) is mediated in part by transposition.

Bacterial transposons vary in structure, but most have short repeated sequences at each end that serve as binding sites for the transposase. When transposition occurs, a short sequence at the target site (5 to 10 bp) is duplicated to form an additional short repeated sequence that flanks each end of the inserted transposon (Fig. 25–42). These duplicated segments result from the cutting mechanism used to insert a transposon into the DNA at a new location.

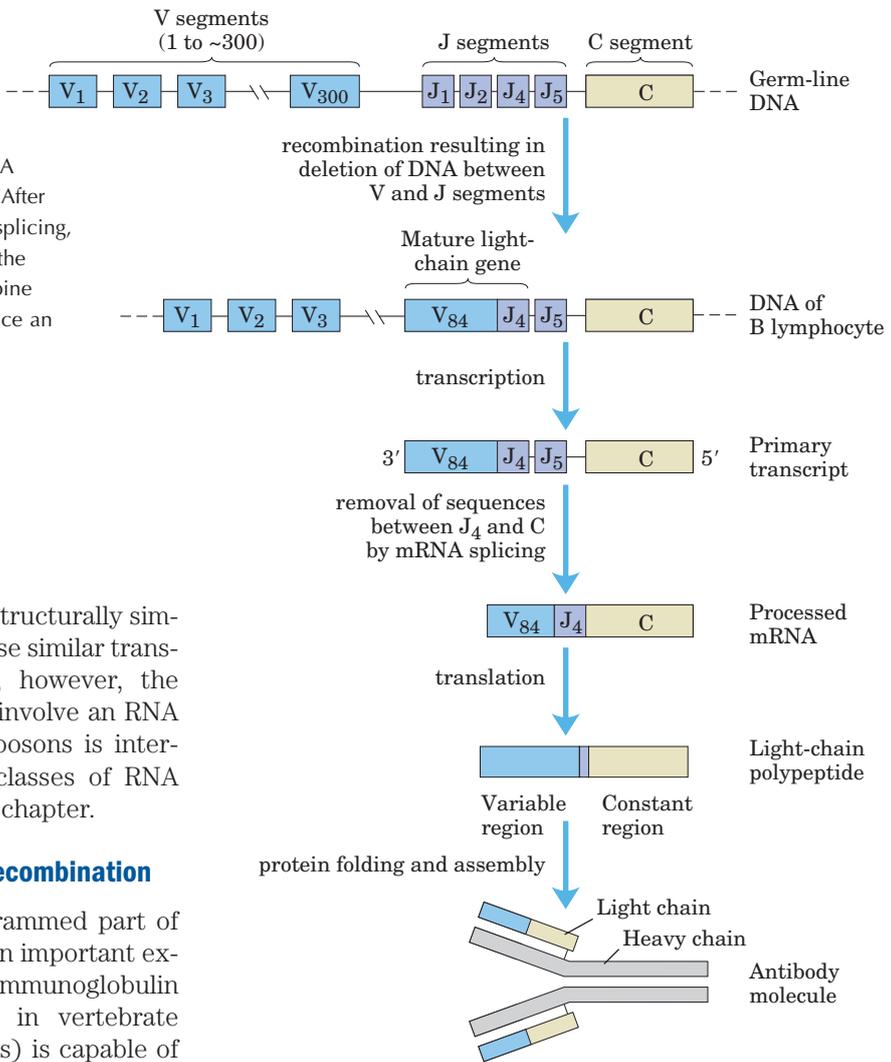
There are two general pathways for transposition in bacteria. In direct or simple transposition (Fig. 25–43, left), cuts on each side of the transposon excise it, and the transposon moves to a new location. This leaves a double-strand break in the donor DNA that must be

**FIGURE 25–41** DNA deletion to undo a deleterious effect of recombinational DNA repair. The resolution of a Holliday intermediate during recombinational DNA repair (if cut at the points indicated by red arrows) can generate a contiguous dimeric chromosome. A specialized site-specific recombinase in *E. coli*, XerCD, converts the dimer to monomers, allowing chromosome segregation and cell division to proceed.



**FIGURE 25–44** Recombination of the V and J gene segments of the human IgG kappa light chain.

This process is designed to generate antibody diversity. At the top is shown the arrangement of IgG-coding sequences in a bone marrow stem cell. Recombination deletes the DNA between a particular V segment and a J segment. After transcription, the transcript is processed by RNA splicing, as described in Chapter 26; translation produces the light-chain polypeptide. The light chain can combine with any of 5,000 possible heavy chains to produce an antibody molecule.



Eukaryotes also have transposons, structurally similar to bacterial transposons, and some use similar transposition mechanisms. In other cases, however, the mechanism of transposition appears to involve an RNA intermediate. Evolution of these transposons is intertwined with the evolution of certain classes of RNA viruses. Both are described in the next chapter.

### Immunoglobulin Genes Assemble by Recombination

Some DNA rearrangements are a programmed part of development in eukaryotic organisms. An important example is the generation of complete immunoglobulin genes from separate gene segments in vertebrate genomes. A human (like other mammals) is capable of producing *millions* of different immunoglobulins (antibodies) with distinct binding specificities, even though the human genome contains only ~35,000 genes. Recombination allows an organism to produce an extraordinary diversity of antibodies from a limited DNA-coding capacity. Studies of the recombination mechanism reveal a close relationship to DNA transposition and suggest that this system for generating antibody diversity may have evolved from an ancient cellular invasion of transposons.

We can use the human genes that encode proteins of the immunoglobulin G (IgG) class to illustrate how antibody diversity is generated. Immunoglobulins consist of two heavy and two light polypeptide chains (see Fig. 5–23). Each chain has two regions, a variable region, with a sequence that differs greatly from one immunoglobulin to another, and a region that is virtually constant within a class of immunoglobulins. There are also two distinct families of light chains, kappa and lambda, which differ somewhat in the sequences of their constant regions. For all three types of polypeptide chain (heavy chain, and kappa and lambda light chains), diversity in the variable regions is generated by a simi-

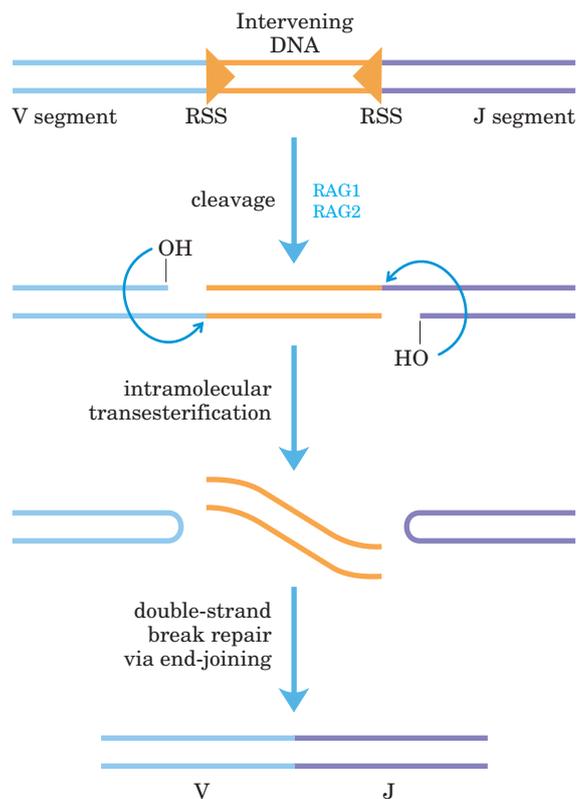
lar mechanism. The genes for these polypeptides are divided into segments, and the genome contains clusters with multiple versions of each segment. The joining of one version of each of the segments creates a complete gene.

Figure 25–44 depicts the organization of the DNA encoding the kappa light chains of human IgG and shows how a mature kappa light chain is generated. In undifferentiated cells, the coding information for this polypeptide chain is separated into three segments. The V (variable) segment encodes the first 95 amino acid residues of the variable region, the J (joining) segment encodes the remaining 12 residues of the variable region, and the C segment encodes the constant region. The genome contains ~300 different V segments, 4 different J segments, and 1 C segment.

As a stem cell in the bone marrow differentiates to form a mature B lymphocyte, one V segment and one J segment are brought together by a specialized recombination system (Fig. 25–44). During this programmed DNA deletion, the intervening DNA is discarded. There are about  $300 \times 4 = 1,200$  possible V–J combinations.

The recombination process is not as precise as the site-specific recombination described earlier, so additional variation occurs in the sequence at the V–J junction. This increases the overall variation by a factor of at least 2.5, thus the cells can generate about  $2.5 \times 1,200 = 3,000$  different V–J combinations. The final joining of the V–J combination to the C region is accomplished by an RNA-splicing reaction after transcription, a process described in Chapter 26.

The recombination mechanism for joining the V and J segments is illustrated in Figure 25–45. Just beyond each V segment and just before each J segment lie recombination signal sequences (RSS). These are bound by proteins called RAG1 and RAG2 (*recombination activating gene*). The RAG proteins catalyze the formation of a double-strand break between the signal sequences and the V (or J) segments to be joined. The V and J segments are then joined with the aid of a second complex of proteins.



**FIGURE 25–45** Mechanism of immunoglobulin gene rearrangement.

The RAG1 and RAG2 proteins bind to the recombination signal sequences (RSS) and cleave one DNA strand between the RSS and the V (or J) segments to be joined. The liberated 3' hydroxyl then acts as a nucleophile, attacking a phosphodiester bond in the other strand to create a double-strand break. The resulting hairpin bends on the V and J segments are cleaved, and the ends are covalently linked by a complex of proteins specialized for end-joining repair of double-strand breaks. The steps in the generation of the double-strand break catalyzed by RAG1 and RAG2 are chemically related to steps in transposition reactions.

The genes for the heavy chains and the lambda light chains form by similar processes. Heavy chains have more gene segments than light chains, with more than 5,000 possible combinations. Because any heavy chain can combine with any light chain to generate an immunoglobulin, each human has at least  $3,000 \times 5,000 = 1.5 \times 10^7$  possible IgGs. And additional diversity is generated by high mutation rates (of unknown mechanism) in the V sequences during B-lymphocyte differentiation. Each mature B lymphocyte produces only one type of antibody, but the range of antibodies produced by different cells is clearly enormous.

Did the immune system evolve in part from ancient transposons? The mechanism for generation of the double-strand breaks by RAG1 and RAG2 does mirror several reaction steps in transposition (Fig. 25–45). In addition, the deleted DNA, with its terminal RSS, has a sequence structure found in most transposons. In the test tube, RAG1 and RAG2 can associate with this deleted DNA and insert it, transposonlike, into other DNA molecules (probably a rare reaction in B lymphocytes). Although we cannot know for certain, the properties of the immunoglobulin gene rearrangement system suggest an intriguing origin in which the distinction between host and parasite has become blurred by evolution.

### SUMMARY 25.3 DNA Recombination

- DNA sequences are rearranged in recombination reactions, usually in processes tightly coordinated with DNA replication or repair.
- Homologous genetic recombination can take place between any two DNA molecules that share sequence homology. In meiosis (in eukaryotes), this type of recombination helps to ensure accurate chromosomal segregation and create genetic diversity. In both bacteria and eukaryotes it serves in the repair of stalled replication forks. A Holliday intermediate forms during homologous recombination.
- Site-specific recombination occurs only at specific target sequences, and this process can also involve a Holliday intermediate. Recombinases cleave the DNA at specific points and ligate the strands to new partners. This type of recombination is found in virtually all cells, and its many functions include DNA integration and regulation of gene expression.
- In virtually all cells, transposons use recombination to move within or between chromosomes. In vertebrates, a programmed recombination reaction related to transposition joins immunoglobulin gene segments to form immunoglobulin genes during B-lymphocyte differentiation.

## Key Terms

*Terms in bold are defined in the glossary.*

<b>template</b> 950	<b>processivity</b> 954	<b>mutation</b> 966	<b>site-specific recombination</b> 978
semiconservative	<b>proofreading</b> 955	base-excision repair	DNA transposition 978
replication 950	DNA polymerase III	971	<b>meiosis</b> 979
<b>replication fork</b> 951	955	DNA glycosylases 971	<b>branch migration</b> 980
<b>origin</b> 952	<b>replisome</b> 957	AP site 971	double-strand break repair
Okazaki fragments 952	<b>helicases</b> 957	AP endonucleases 972	model 980
<b>leading strand</b> 952	<b>topoisomerases</b> 957	DNA photolyases 974	<b>Holliday intermediate</b>
<b>lagging strand</b> 952	<b>primases</b> 958	<b>recombinational DNA</b>	980
<b>nucleases</b> 952	<b>DNA ligase</b> 958	<b>repair</b> 976	<b>transposons</b> 988
<b>exonuclease</b> 952	<b>primosome</b> 962	error-prone translesion	<b>transposition</b> 988
<b>endonuclease</b> 952	<b>catenane</b> 963	DNA synthesis 976	<b>insertion sequence</b>
DNA polymerase I 952	DNA polymerase $\alpha$ 965	<b>SOS response</b> 976	988
<b>primer</b> 954	DNA polymerase $\delta$ 965	<b>homologous genetic re-</b>	<b>cointegrate</b> 989
<b>primer terminus</b> 954	DNA polymerase $\epsilon$ 965	<b>combination</b> 978	

## Further Reading

### General

**Friedberg, E.C., Walker, G.C., & Siede, W.** (1995) *DNA Repair and Mutagenesis*, American Society for Microbiology, Washington, DC.

A thorough treatment of DNA metabolism and a good place to start exploring this field.

**Kornberg, A. & Baker, T.A.** (1991) *DNA Replication*, 2nd edn, W. H. Freeman and Company, New York.

Excellent primary source for all aspects of DNA metabolism.

### DNA Replication

**Benkovic, S.J., Valentine, A.M., & Salinas, F.** (2001) Replisome-mediated DNA replication. *Annu. Rev. Biochem.* **70**, 181–208.

This review describes the similar strategies and enzymes of DNA replication in different classes of organisms.

**Boye, E., Lobner-Olesen, A., & Skarstad, K.** (2000) Limiting DNA replication to once and only once. *EMBO Rep.* **1**, 479–483.

Good summary of the mechanisms by which replication initiation is regulated.

**Davey, M.J. & O'Donnell, M.** (2000) Mechanisms of DNA replication. *Curr. Opin. Chem. Biol.* **4**, 581–586.

**Ellison, V. & Stillman, B.** (2001) Opening of the clamp: an intimate view of an ATP-driven biological machine. *Cell* **106**, 655–660.

**Frick, D.N. & Richardson, C.C.** (2001) DNA primases. *Annu. Rev. Biochem.* **70**, 39–80.

**Hübscher, U., Maga, G., & Spadari, S.** (2002) Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* **71**, 133–163.

Good summary of the properties and roles of the more than one dozen known eukaryotic DNA polymerases.

**Jeruzalmski, D., O'Donnell, M., & Kuriyan, J.** (2002) Clamp loaders and sliding clamps. *Curr. Opin. Struct. Biol.* **12**, 217–224.

Summary of some of the elegant work elucidating how clamp loaders function.

**Kamada, K., Horiuchi, T., Ohsumi, K., Shimamoto, N., & Morikawa, K.** (1996) Structure of a replication-terminator protein complexed with DNA. *Nature* **383**, 598–603.

The report revealing the structure of the Tus-Ter complex.

**Katayama, T.** (2001) Feedback controls restrain the initiation of *Escherichia coli* chromosomal replication. *Mol. Microbiol.* **41**, 9–17.

**Kool, E.T.** (2002) Active site tightness and substrate fit in DNA replication. *Annu. Rev. Biochem.* **71**, 191–219.

Excellent summary of the molecular basis of replication fidelity by a DNA polymerase—base-pair geometry as well as hydrogen bonding.

**Lemon, K.P. & Grossman, A.D.** (2001) The extrusion-capture model for chromosome partitioning in bacteria. *Genes Dev.* **15**, 2031–2041.

Report describing the replication factory model for bacterial DNA replication.

**Nishitani, H. & Lygerou, Z.** (2002) Control of DNA replication licensing in a cell cycle. *Genes Cells* **7**, 523–534.

A good summary of recent advances in the understanding of how eukaryotic DNA replication is initiated.

**Toyn, J.H., Toone, M.W., Morgan, B.A., & Johnston, L.H.** (1995) The activation of DNA replication in yeast. *Trends Biochem. Sci.* **20**, 70–73.

### DNA Repair

**Begley, T.J. & Samson, L.D.** (2003) AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. *Trends Biochem. Sci.* **28**, 2–5.

**Friedberg, E.C., Fischhaber, P.L., & Kisker, C.** (2001) Error-prone DNA polymerases: novel structures and the benefits of infidelity. *Cell* **107**, 9–12.

**Goodman, M.F.** (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* **71**, 17–50.

Review of a class of DNA polymerases that continues to grow.

**Kolodner, R.D.** (1995) Mismatch repair: mechanisms and relationship to cancer susceptibility. *Trends Biochem. Sci.* **20**, 397–401.

**Lindahl, T. & Wood, R.D.** (1999) Quality control by DNA repair. *Science* **286**, 1897–1905.

**Marnett, L.J. & Plataras, J.P.** (2001) Endogenous DNA damage and mutation. *Trends Genet.* **17**, 214–221.

**McCullough, A.K., Dodson, M.L., & Lloyd, R.S.** (1999) Initiation of base excision repair: glycosylase mechanisms and structures. *Annu. Rev. Biochem.* **68**, 255–286.

**Modrich, P. & Lahue, R.** (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* **65**, 101–133.

**Sancar, A.** (1996) DNA excision repair. *Annu. Rev. Biochem.* **65**, 43–81.

**Sutton, M.D., Smith, B.T., Godoy, V.G., & Walker, G.C.** (2000) The SOS response: recent insights into umuDC-dependent mutagenesis and DNA damage tolerance. *Annu. Rev. Genet.* **34**, 479–497.

**Wood, R.D., Mitchell, M., Sgouros, J., & Lindahl T.** (2001) Human DNA repair genes. *Science* **291**, 1284–1289.

Description of what an early look at the human genome reveals about DNA repair.

### DNA Recombination

**Cox, M.M.** (2001) Historical overview: searching for replication help in all of the rec places. *Proc. Natl. Acad. Sci. USA* **98**, 8173–8180.

A review of how recombination was shown to be a replication fork repair process.

**Craig, N.L.** (1995) Unity in transposition reactions. *Science* **270**, 253–254.

**Eggleston, A.K. & West, S.C.** (1996) Exchanging partners: recombination in *E. coli*. *Trends Genet.* **12**, 20–26.

**Gellert, M.** (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.* **71**, 101–132.

**Hallet, B. & Sherratt, D.J.** (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.* **21**, 157–178.

**Kogoma, T.** (1996) Recombination by replication. *Cell* **85**, 625–627.

**Lieber, M.** (1996) Immunoglobulin diversity: rearranging by cutting and repairing. *Curr. Biol.* **6**, 134–136.

**Lusetti, S.L. & Cox, M.M.** (2002) The bacterial RecA protein and the recombinational DNA repair of stalled replication forks. *Annu. Rev. Biochem.* **71**, 71–100.

**Marians, K.J.** (2000) PriA-directed replication fork restart in *Escherichia coli*. *Trends Biochem. Sci.* **25**, 185–189.

**Pâques, F. & Haber, J.E.** (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **63**, 349–404.

**Van Duyne, G.D.** (2001) A structural view of Cre-loxP site-specific recombination. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 87–104.

## Problems

### 1. Conclusions from the Meselson-Stahl Experiment

The Meselson-Stahl experiment (see Fig. 25–2) proved that DNA undergoes semiconservative replication in *E. coli*. In the “dispersive” model of DNA replication, the parent DNA strands are cleaved into pieces of random size, then joined with pieces of newly replicated DNA to yield daughter duplexes. In the Meselson-Stahl experiment, each strand would contain random segments of heavy and light DNA. Explain how the results of Meselson and Stahl’s experiment ruled out such a model.

### 2. Heavy Isotope Analysis of DNA Replication

A culture of *E. coli* growing in a medium containing  $^{15}\text{NH}_4\text{Cl}$  is switched to a medium containing  $^{14}\text{NH}_4\text{Cl}$  for three generations (an eightfold increase in population). What is the molar ratio of hybrid DNA ( $^{15}\text{N}$ – $^{14}\text{N}$ ) to light DNA ( $^{14}\text{N}$ – $^{14}\text{N}$ ) at this point?

### 3. Replication of the *E. coli* Chromosome

The *E. coli* chromosome contains 4,639,221 bp.

(a) How many turns of the double helix must be unwound during replication of the *E. coli* chromosome?

(b) From the data in this chapter, how long would it take to replicate the *E. coli* chromosome at 37 °C if two replica-

tion forks proceeded from the origin? Assume replication occurs at a rate of 1,000 bp/s. Under some conditions *E. coli* cells can divide every 20 min. How might this be possible?

(c) In the replication of the *E. coli* chromosome, about how many Okazaki fragments would be formed? What factors guarantee that the numerous Okazaki fragments are assembled in the correct order in the new DNA?

### 4. Base Composition of DNAs Made from Single-Stranded Templates

Predict the base composition of the total DNA synthesized by DNA polymerase on templates provided by an equimolar mixture of the two complementary strands of bacteriophage  $\phi\text{X174}$  DNA (a circular DNA molecule). The base composition of one strand is A, 24.7%; G, 24.1%; C, 18.5%; and T, 32.7%. What assumption is necessary to answer this problem?

### 5. DNA Replication

Kornberg and his colleagues incubated soluble extracts of *E. coli* with a mixture of dATP, dTTP, dGTP, and dCTP, all labeled with  $^{32}\text{P}$  in the  $\alpha$ -phosphate group. After a time, the incubation mixture was treated with trichloroacetic acid, which precipitates the DNA but not the nucleotide precursors. The precipitate was collected, and the extent of precursor incorporation into DNA was determined

from the amount of radioactivity present in the precipitate.

(a) If any one of the four nucleotide precursors were omitted from the incubation mixture, would radioactivity be found in the precipitate? Explain.

(b) Would  $^{32}\text{P}$  be incorporated into the DNA if only dTTP were labeled? Explain.

(c) Would radioactivity be found in the precipitate if  $^{32}\text{P}$  labeled the  $\beta$  or  $\gamma$  phosphate rather than the  $\alpha$  phosphate of the deoxyribonucleotides? Explain.

**6. Leading and Lagging Strands** Prepare a table that lists the names and compares the functions of the precursors, enzymes, and other proteins needed to make the leading versus lagging strands during DNA replication in *E. coli*.

**7. Function of DNA Ligase** Some *E. coli* mutants contain defective DNA ligase. When these mutants are exposed to  $^3\text{H}$ -labeled thymine and the DNA produced is sedimented on an alkaline sucrose density gradient, two radioactive bands appear. One corresponds to a high molecular weight fraction, the other to a low molecular weight fraction. Explain.

**8. Fidelity of Replication of DNA** What factors promote the fidelity of replication during the synthesis of the leading strand of DNA? Would you expect the lagging strand to be made with the same fidelity? Give reasons for your answers.

**9. Importance of DNA Topoisomerases in DNA Replication** DNA unwinding, such as that occurring in replication, affects the superhelical density of DNA. In the absence of topoisomerases, the DNA would become overwound ahead of a replication fork as the DNA is unwound behind it. A bacterial replication fork will stall when the superhelical density ( $\sigma$ ) of the DNA ahead of the fork reaches +0.14 (see Chapter 24).

Bidirectional replication is initiated at the origin of a 6,000 bp plasmid in vitro, in the absence of topoisomerases. The plasmid initially has a  $\sigma$  of  $-0.06$ . How many base pairs will be unwound and replicated by each replication fork before the forks stall? Assume that each fork travels at the same rate and that each includes all components necessary for elongation except topoisomerase.

**10. The Ames Test** In a nutrient medium that lacks histidine, a thin layer of agar containing  $\sim 10^9$  *Salmonella typhimurium* histidine auxotrophs (mutant cells that require histidine to survive) produces  $\sim 13$  colonies over a two-day incubation period at  $37^\circ\text{C}$  (see Fig. 25–19). How do these colonies arise in the absence of histidine? The experiment is repeated in the presence of  $0.4\ \mu\text{g}$  of 2-aminoanthracene. The number of colonies produced over two days exceeds 10,000. What does this indicate about 2-aminoanthracene? What can you surmise about its carcinogenicity?

**11. DNA Repair Mechanisms** Vertebrate and plant cells often methylate cytosine in DNA to form 5-methylcytosine (see Fig. 8–5a). In these same cells, a specialized repair system recognizes G–T mismatches and repairs them to G=C base pairs. How might this repair system be advantageous to the cell? (Explain in terms of the presence of 5-methylcytosine in the DNA.)



**12. DNA Repair in People with Xeroderma Pigmentosum**

The condition known as xeroderma pigmentosum (XP) arises from mutations in at least seven different human genes. The deficiencies are generally in genes encoding enzymes involved in some part of the pathway for human nucleotide-excision repair. The various types of XP are labeled A through G (XPA, XPB, etc.), with a few additional variants lumped under the label XPV.

Cultures of cells from healthy individuals and from patients with XPG are irradiated with ultraviolet light. The DNA is isolated and denatured, and the resulting single-stranded DNA is characterized by analytical ultracentrifugation.

(a) Samples from the normal fibroblasts show a significant reduction in the average molecular weight of the single-stranded DNA after irradiation, but samples from the XPG fibroblasts show no such reduction. Why might this be?

(b) If you assume that a nucleotide-excision repair system is operative, which step might be defective in the fibroblasts from the patients with XPG? Explain.

**13. Holliday Intermediates** How does the formation of Holliday intermediates in homologous genetic recombination differ from their formation in site-specific recombination?