CHAPTER **SIX** 6

DNA Replication and Repair

DNA REPLICATION

DNA REPAIR

For a cell to survive and proliferate in a chaotic environment, it must be able to accurately copy the vast quantity of genetic information carried in its DNA. This fundamental process, called **DNA replication**, must occur before a cell can divide to produce two genetically identical daughter cells. In addition to carrying out this painstaking task with stunning accuracy and efficiency, a cell must also continuously monitor and repair its genetic material, as DNA is subjected to unavoidable damage by chemicals and radiation in the environment and by reactive molecules that are generated inside the cell.

Yet despite the molecular safeguards that have evolved to protect a cell's DNA from copying errors and accidental damage, permanent changes—or **mutations**—sometimes do occur. Although most mutations do not affect the organism in any noticeable way, some have profound consequences. Occasionally, these changes can benefit the organism: for example, mutations can make bacteria resistant to antibiotics that are used to kill them. What is more, changes in DNA sequence can produce small variations that underlie the differences between individuals of the same species (**Figure 6–1**); such changes, when they accumulate over hundreds of millions of years, provide the variety in genetic material that makes one species distinct from another, as we discuss in Chapter 9.

Unfortunately, as mutations occur randomly, they are more likely to be detrimental than beneficial: they are responsible for thousands of human diseases, including cancer. The survival of a cell or organism, therefore, depends on keeping the changes in its DNA to a minimum. Without the systems that are continually inspecting and repairing damage to DNA, it is questionable whether life could exist at all. In this chapter, we describe the protein machines that replicate and repair the cell's DNA. These

Figure 6–1 Differences in DNA can produce the variations that underlie the differences between individuals of the same species—even within the same family. Over evolutionary time, these genetic changes give rise to the differences that distinguish one species from another.

machines catalyze some of the most rapid and elegant processes that take place within cells, and uncovering the strategies they employ to achieve these marvelous feats represents a triumph of scientific investigation.

DNA REPLICATION

At each cell division, a cell must copy its genome with extraordinary accuracy. In this section, we explore how the cell achieves this feat, while replicating its DNA at rates as high as 1000 nucleotides per second.

Base-Pairing Enables DNA Replication

In the preceding chapter, we saw that each strand of a DNA double helix contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand. Each strand can therefore serve as a **template**, or mold, for the synthesis of a new complementary strand. In other words, if we designate the two DNA strands as S and S′, strand S can serve as a template for making a new strand S′, while strand S′ can serve as a template for making a new strand S (**Figure 6–2**). Thus, the genetic information in DNA can be accurately copied by the beautifully simple process in which strand S separates from strand S′, and each separated strand then serves as a template for the production of a new complementary partner strand that is identical to its former partner.

The ability of each strand of a DNA molecule to act as a template for producing a complementary strand enables a cell to copy, or *replicate*, its genes before passing them on to its descendants. Although simple in principle, the process is awe-inspiring, as it can involve the copying of billions of nucleotide pairs with incredible speed and accuracy: a human cell undergoing division will copy the equivalent of 1000 books like this one in about 8 hours and, on average, get no more than a few letters wrong. This impressive feat is performed by a cluster of proteins that together form a *replication machine*.

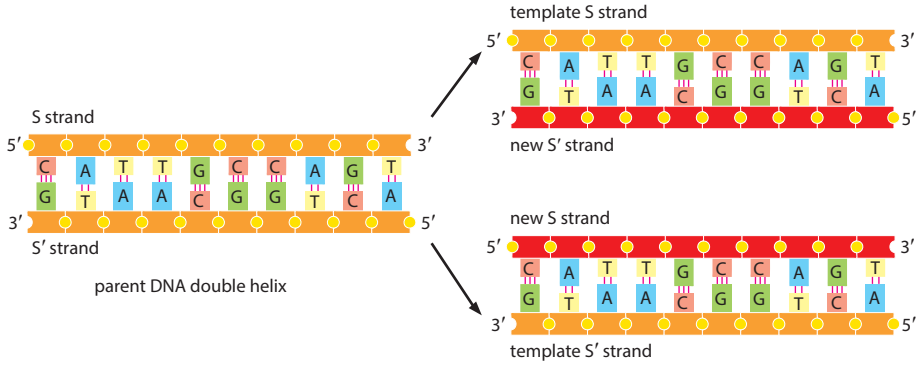


Figure 6–2 DNA acts as a template for its own replication. Because the nucleotide A will successfully pair only with T, and G with C, each strand of a DNA double helix—labeled here as the S strand and its complementary S′ strand—can serve as a template to specify the sequence of nucleotides in its complementary strand. In this way, both strands of a DNA double helix can be copied with precision.

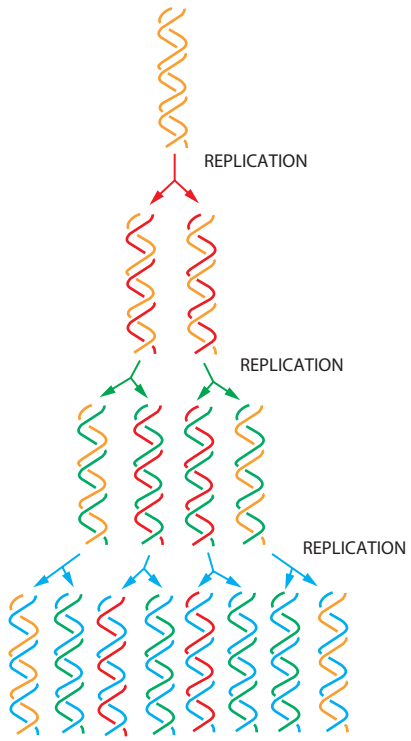


Figure 6–3 In each round of DNA replication, each of the two strands of DNA is used as a template for the formation of a new, complementary strand. DNA replication is "semiconservative" because each daughter DNA double helix is composed of one conserved (old) strand and one newly synthesized strand.

DNA replication produces two complete double helices from the original DNA molecule, with each new DNA helix being identical in nucleotide sequence (except for rare copying errors) to the original DNA double helix (see Figure 6–2). Because each parental strand serves as the template for one new strand, each of the daughter DNA double helices ends up with one of the original (old) strands plus one strand that is completely new; this style of replication is said to be *semiconservative* (**Figure 6–3**). We describe the inventive experiments that revealed this feature of DNA replication in **How We Know**, pp. 202–204.

DNA Synthesis Begins at Replication Origins

The DNA double helix is normally very stable: the two DNA strands are locked together firmly by the large numbers of hydrogen bonds between the bases on both strands (see Figure 5–2). As a result, only temperatures approaching those of boiling water provide enough thermal energy to separate the two strands. To be used as a template, however, the double helix must first be opened up and the two strands separated to expose the nucleotide bases. How does this separation occur at the temperatures found in living cells?

The process of DNA synthesis is begun by *initiator proteins* that bind to specific DNA sequences called **replication origins**. Here, the initiator proteins pry the two DNA strands apart, breaking the hydrogen bonds between the bases (**Figure 6–4**). Although the hydrogen bonds collectively make the DNA helix very stable, individually each hydrogen bond is weak (as discussed in Chapter 2). Separating a short length of DNA a few base pairs at a time therefore does not require a large energy input, and the initiator proteins can readily unzip short regions of the double helix at normal temperatures.

In simple cells such as bacteria or yeast, replication origins span approximately 100 nucleotide pairs. They are composed of DNA sequences that attract the initiator proteins and are especially easy to open. We saw in Chapter 5 that an A-T base pair is held together by fewer hydrogen bonds than is a G-C base pair. Therefore, DNA rich in A-T base pairs is easier to pull apart, and A-T-rich stretches of DNA are typically found at replication origins.

A bacterial genome, which is typically contained in a circular DNA molecule of several million nucleotide pairs, has a single replication origin. The human genome, which is very much larger, has approximately 10,000 such origins—an average of 220 origins per chromosome. Beginning DNA replication at many places at once greatly shortens the time a cell needs to copy its entire genome.

Once an initiator protein binds to DNA at a replication origin and locally opens up the double helix, it attracts a group of proteins that carry out DNA replication. These proteins form a replication machine, in which each protein carries out a specific function.

Two Replication Forks Form at Each Replication Origin

DNA molecules in the process of being replicated contain Y-shaped junctions called **replication forks**. Two replication forks are formed at each

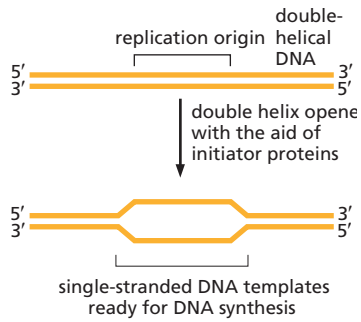


Figure 6–4 A DNA double helix is opened at replication origins. DNA sequences at replication origins are recognized by initiator proteins (not shown), which locally pull apart the two strands of the double helix. The exposed single strands can then serve as templates for copying the DNA.

HOW WE KNOW

THE NATURE OF REPLICATION

In 1953, James Watson and Francis Crick published their famous two-page paper describing a model for the structure of DNA. In this report, they proposed that complementary bases—adenine and thymine, guanine and cytosine—pair with one another along the center of the double helix, holding together the two strands of DNA (see Figure 5–2). At the very end of this succinct scientific blockbuster, they comment, almost as an aside, "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material."

Indeed, one month after the classic paper appeared in print in the journal *Nature*, Watson and Crick published a second article, suggesting how DNA might be replicated. In this paper, they proposed that the two strands of the double helix unwind, and that each strand serves as a template for the synthesis of a complementary daughter strand. In their model, dubbed semiconservative replication, each new DNA molecule consists of one strand derived from the original parent molecule and one newly synthesized strand (**Figure 6–5A**).

We now know that Watson and Crick's model for DNA replication was correct—but it was not universally accepted at first. Respected physicist-turned-geneticist Max Delbrück, for one, got hung up on what he termed "the untwiddling problem"; that is: How could the two strands of a double helix, twisted around each other so many times all along their great length, possibly be unwound without making a big tangled mess? Watson and Crick's conception of the DNA helix opening up like a zipper seemed, to Delbrück, physically unlikely and simply "too inelegant to be efficient."

Instead, Delbrück proposed that DNA replication proceeds through a series of breaks and reunions, in which the DNA backbone is broken and the strands are copied in short segments—perhaps only 10 nucleotides at a time—before being rejoined. In this model, which was later dubbed dispersive, the resulting copies would be patchwork collections of old and new DNA, each strand containing a mixture of both (**Figure 6–5B**). No unwinding was necessary.

Yet a third camp promoted the idea that DNA replication might be *conservative*: that the parent helix would somehow remain entirely intact after copying, and the daughter molecule would contain two entirely new DNA strands (**Figure 6–5C**). To determine which of these models was correct, an experiment was needed—one that would reveal the composition of the newly synthesized DNA strands. That's where Matt Meselson and Frank Stahl came in.

Heavy DNA

As a graduate student working with Linus Pauling, Meselson was toying with a method for telling the difference between old and new proteins. After chatting with Delbrück about Watson and Crick's replication

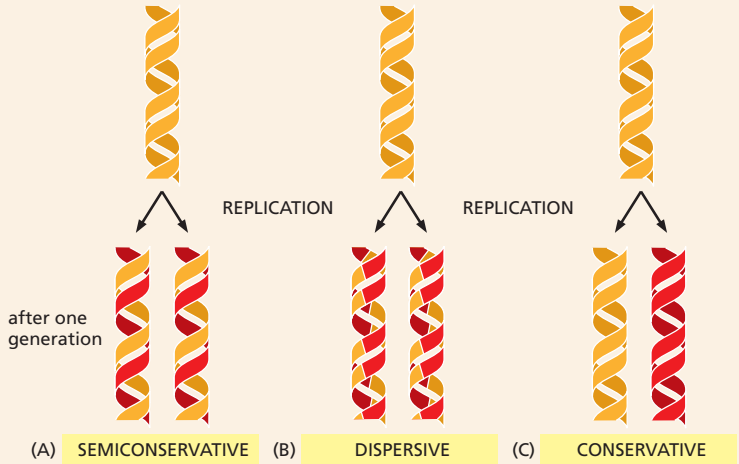


Figure 6–5 Three models for DNA replication make different predictions. (A) In the semiconservative model, each parent strand serves as a template for the synthesis of a new daughter strand. The first round of replication would produce two hybrid molecules, each containing one strand from the original parent and one newly synthesized strand. A subsequent round of replication would yield two hybrid molecules and two molecules that contain none of the original parent DNA (see Figure 6–3). (B) In the dispersive model, each generation of replicated DNA molecules will be a mosaic of DNA from the parent strands and the newly synthesized DNA. (C) In the conservative model, the parent molecule remains intact after being copied. In this case, the first round of replication would yield the original parent double helix and an entirely new double helix. For each model, parent DNA molecules are shown in *orange*; newly replicated DNA is *red*. Note that only a very small segment of DNA is shown for each model.

model, it occurred to Meselson that the approach he'd envisaged for exploring protein synthesis might also work for studying DNA. In the summer of 1954, Meselson met Stahl, who was then a graduate student in Rochester, NY, and they agreed to collaborate. It took a few years to get everything working, but the two eventually performed what has come to be known as "the most beautiful experiment in biology."

Their approach, in retrospect, was stunningly straightforward. They started by growing two batches of *Escherichia coli* bacteria, one in a medium containing a heavy isotope of nitrogen, ^{15}N, the other in a medium containing the normal, lighter ^{14}N. The nitrogen in the nutrient medium gets incorporated into the nucleotide bases and, from there, makes its way into the DNA of the organism. After growing bacterial cultures for many generations in either the ^{15}N- or ^{14}N-containing medium, the researchers had two flasks of bacteria, one with heavy DNA (containing *E. coli* that had incorporated the heavy isotope), the other with DNA that was light. Meselson and Stahl then broke open the bacterial cells and loaded the DNA into tubes containing a high concentration of the salt cesium chloride. When these tubes are centrifuged at high speed, the cesium chloride forms a density gradient, and the DNA molecules float or sink within the solution until they reach the point at which their density equals that of the salt solution that surrounds them (see Panel 4–3, pp. 164–165). Using this method, called equilibrium density centrifugation, Meselson and Stahl found that they could distinguish between heavy (^{15}N-containing) DNA and light (^{14}N-containing) DNA by observing the positions of the DNA within the cesium chloride gradient. Because the heavy DNA was denser than the light DNA, it collected at a position nearer to the bottom of the centrifuge tube (**Figure 6–6**).

And the winner is...

Once they had established this method for differentiating between light and heavy DNA, Meselson and Stahl set out to test the various hypotheses proposed for DNA replication. To do this, they took a flask of bacteria that had been grown in heavy nitrogen and transferred the bacteria into a medium containing the light isotope. At the start of the experiment, all the DNA would be heavy. But, as the bacteria divided, the newly synthesized DNA would be light. They could then monitor the accumulation of light DNA and see which model, if any, best fit their data. After one generation of growth, the researchers found that the parental, heavy DNA molecules—those made of two strands containing ^{15}N—had disappeared and were replaced by a new species of DNA that banded at a density halfway between those of ^{15}N-DNA and ^{14}N-DNA (**Figure 6–7**). These newly synthesized daughter helices, Meselson and Stahl reasoned, must be hybrids—containing both heavy and light isotopes.

Right away, this observation ruled out the conservative model of DNA replication, which predicted that the

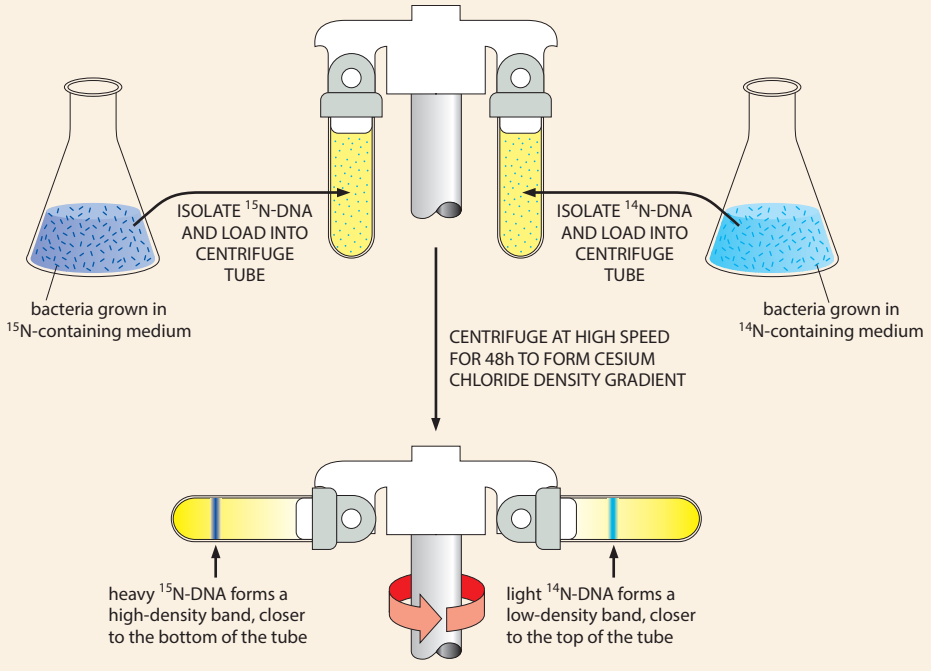


Figure 6–6 Centrifugation in a cesium chloride gradient allows the separation of heavy and light DNA. Bacteria are grown for several generations in a medium containing either ^{15}N (the heavy isotope) or ^{14}N (the light isotope) to label their DNA. The cells are then broken open, and the DNA is loaded into an ultracentrifuge tube containing a cesium chloride salt solution (*yellow*). These tubes are centrifuged at high speed for two days to allow the cesium chloride to form a gradient with low density at the top of the tube and high density at the bottom. As the gradient forms, the DNA will migrate to the region where its density matches that of the salt surrounding it. The heavy and light DNA molecules thus collect in different positions in the tube.

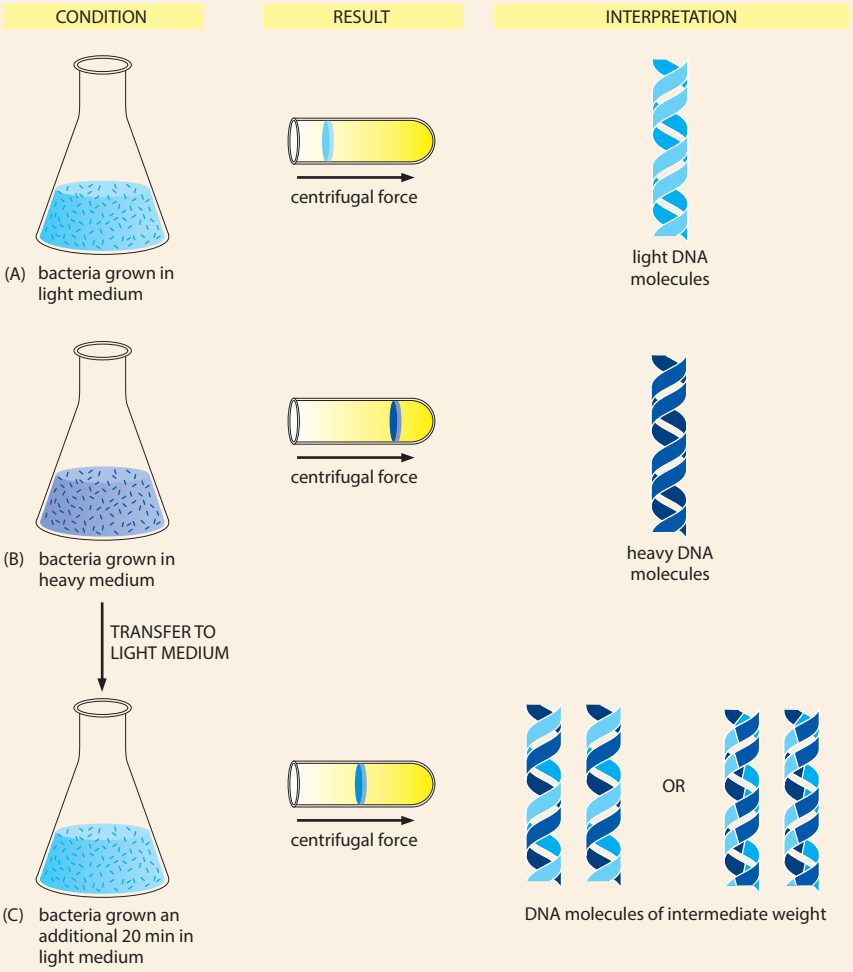


Figure 6–7 The first part of the Meselson–Stahl experiment ruled out the conservative model of DNA replication. (A) Bacteria grown in light medium (containing ^{14}N) yield DNA that forms a band near the top of the centrifuge tube, whereas bacteria grown in ^{15}N-containing heavy medium (B) produce DNA that reaches a position further down the tube. (C) When bacteria grown in a heavy medium are transferred to a light medium and allowed to divide for one hour (the time needed for one generation), they produce a band that is positioned about midway between the heavy and light DNA. These results rule out the conservative model of replication but do not distinguish between the semiconservative and dispersive models, both of which predict the formation of daughter DNA molecules with intermediate densities.

The fact that the results came out looking so clean—with discrete bands forming at the expected positions for newly replicated hybrid DNA molecules—was a happy accident of the experimental protocol. The researchers used a hypodermic syringe to load their DNA samples into the ultracentrifuge tubes (see Figure 6–6). In the process, they unwittingly sheared the large bacterial chromosome into smaller fragments. Had the chromosomes remained whole, the researchers might have isolated DNA molecules that were only partially replicated, because many cells would have been caught in the middle of copying their DNA. Molecules in such an intermediate stage of replication would not have separated into such beautifully discrete bands. But because the researchers were instead working with smaller pieces of DNA, the likelihood that any given fragment had been fully replicated—and contained a complete parent and daughter strand—was high, thus yielding clean, easy-to-interpret results.

parental DNA would remain entirely heavy, while the daughter DNA would be entirely light (see Figure 6–5C). The data supported the semiconservative model, which predicted the formation of hybrid molecules containing one strand of heavy DNA and one strand of light (see Figure 6–5A). The results, however, were also consistent with the dispersive model, in which hybrid DNA strands would contain a mixture of heavy and light DNA (see Figure 6–5B).

To distinguish between the remaining two models, Meselson and Stahl turned up the heat. When DNA is subjected to high temperature, the hydrogen bonds holding the two strands together break and the helix comes apart, leaving a collection of single-stranded DNAs. When the researchers heated the hybrid molecules before centrifuging, they discovered that one strand of the DNA was heavy, whereas the other was light. This observation ruled out the dispersive model; if this model were correct, the resulting strands, each containing a mottled assembly of heavy and light DNA, would have all banded together at an intermediate density.

According to historian Frederic Lawrence Holmes, the experiment was so elegant and the results so clean that Stahl—when being interviewed for a position at Yale University—was unable to fill the 50 minutes allotted for his talk. "I was finished in 25 minutes," said Stahl, "because that is all it takes to tell that experiment. It's so totally simple and contained." Stahl did not get the job at Yale, but the experiment convinced biologists that Watson and Crick had been correct. In fact, the results were accepted so widely and rapidly that the experiment was described in a textbook before Meselson and Stahl had even published the data.

replication origin (Figure 6–8). At each fork, a replication machine moves along the DNA, opening up the two strands of the double helix and using each strand as a template to make a new daughter strand. The two forks move away from the origin in opposite directions, unzipping the DNA double helix and copying the DNA as they go (Figure 6–9). DNA replication—in both bacterial and eukaryotic chromosomes—is therefore termed *bidirectional*. The forks move very rapidly: at about 1000 nucleotide pairs per second in bacteria and 100 nucleotide pairs per second in humans. The slower rate of fork movement in humans (indeed, in all eukaryotes) may be due to the difficulties in replicating DNA through the more complex chromatin structure of eukaryotic chromosomes (discussed in Chapter 5).

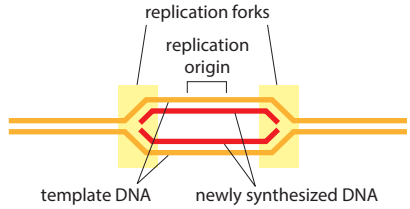


Figure 6–8 DNA synthesis occurs at Y-shaped junctions called replication forks. Two replication forks form at each replication origin and subsequently move away from each other as replication proceeds.

DNA Polymerase Synthesizes DNA Using a Parental Strand as a Template

The movement of a replication fork is driven by the action of the replication machine, at the heart of which is an enzyme called **DNA polymerase**. This enzyme catalyzes the addition of nucleotides to the 3′ end of a growing DNA strand, using one of the original, parental DNA strands as a template. Base-pairing between an incoming nucleotide and the template strand determines which of the four nucleotides (A, G, T, or C) will be selected. The final product is a new strand of DNA that is complementary in nucleotide sequence to the template (Figure 6–10).

The polymerization reaction involves the formation of a phosphodiester bond between the 3′ end of the growing DNA chain and the 5′-phosphate group of the incoming nucleotide, which enters the reaction as a *deoxyribonucleoside triphosphate*. The energy for polymerization is provided

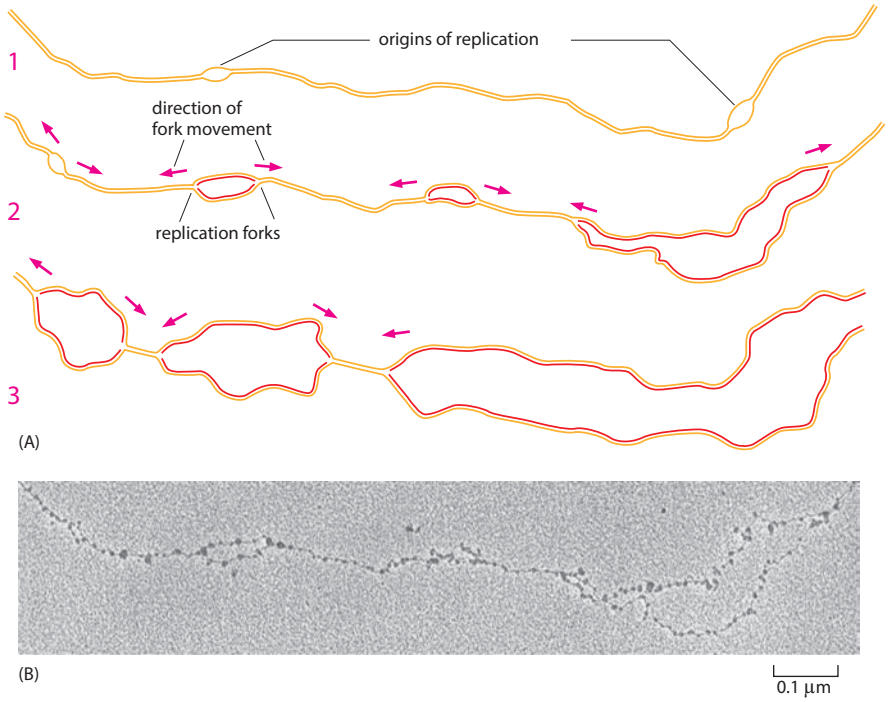


Figure 6–9 The two replication forks formed at a replication origin move away in opposite directions. (A) These drawings represent the same portion of a DNA molecule as it might appear at different times during replication. The *orange* lines represent the two parental DNA strands; the *red* lines represent the newly synthesized DNA strands. (B) An electron micrograph showing DNA replicating in an early fly embryo. The particles visible along the DNA are nucleosomes, structures made of DNA and the histone protein complexes around which the DNA is wrapped (discussed in Chapter 5). The chromosome in this micrograph is the same one that was redrawn in sketch (2) of (A). (B, courtesy of Victoria Foe.)

QUESTION 6–1

Look carefully at the micrograph and corresponding sketch (2) in Figure 6–9.

A. Using the scale bar, estimate the lengths of the DNA double helices between the replication forks. Numbering the replication forks sequentially from the left, how long will it take until forks 4 and 5, and forks 7 and 8, respectively, collide with each other? (Recall that the distance between the bases in DNA is 0.34 nm, and eukaryotic replication forks move at about 100 nucleotides per second.) For this question, disregard the nucleosomes seen in the micrograph and assume that the DNA is fully extended.

B. The fly genome is about 1.8×10^8 nucleotide pairs in size. What fraction of the genome is shown in the micrograph?

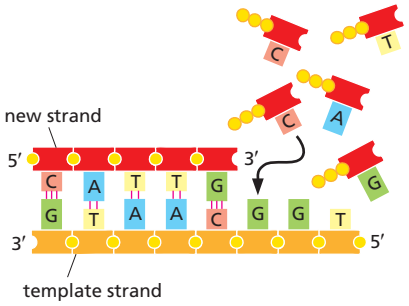


Figure 6–10 A new DNA strand is synthesized in the 5′-to-3′ direction. At each step, the appropriate incoming nucleoside triphosphate is selected by forming base pairs with the next nucleotide in the template strand: A with T, T with A, C with G, and G with C. Each is added to the 3′ end of the growing new strand, as indicated.

by the incoming deoxyribonucleoside triphosphate itself: hydrolysis of one of its high-energy phosphate bonds fuels the reaction that links the nucleotide monomer to the chain, releasing pyrophosphate (**Figure 6–11**). Pyrophosphate is further hydrolyzed to inorganic phosphate (P_i), which makes the polymerization reaction effectively irreversible (see Figure 3–42).

DNA polymerase does not dissociate from the DNA each time it adds a new nucleotide to the growing strand; rather, it stays associated with the DNA and moves along the template strand stepwise for many cycles of the polymerization reaction (**Movie 6.1**). We will see later that a special protein keeps the polymerase attached to DNA as it repeatedly adds new nucleotides to the growing strand.

The Replication Fork Is Asymmetrical

The 5′-to-3′ direction of the DNA polymerization reaction poses a problem at the replication fork. As illustrated in Figure 5–2, the sugar–phosphate backbone of each strand of a DNA double helix has a unique chemical direction, or polarity, determined by the way each sugar residue is linked to the next, and the two strands in the double helix are antiparallel; that is, they run in opposite directions. As a consequence, at each replication fork, one new DNA strand is being made on a template that runs in one direction (3′ to 5′), whereas the other new strand is being made on a template that runs in the opposite direction (5′ to 3′). The replication fork is therefore asymmetrical (**Figure 6–12**). Figure 6–9A, however, makes it look like both of the new DNA strands are growing in the same direction;

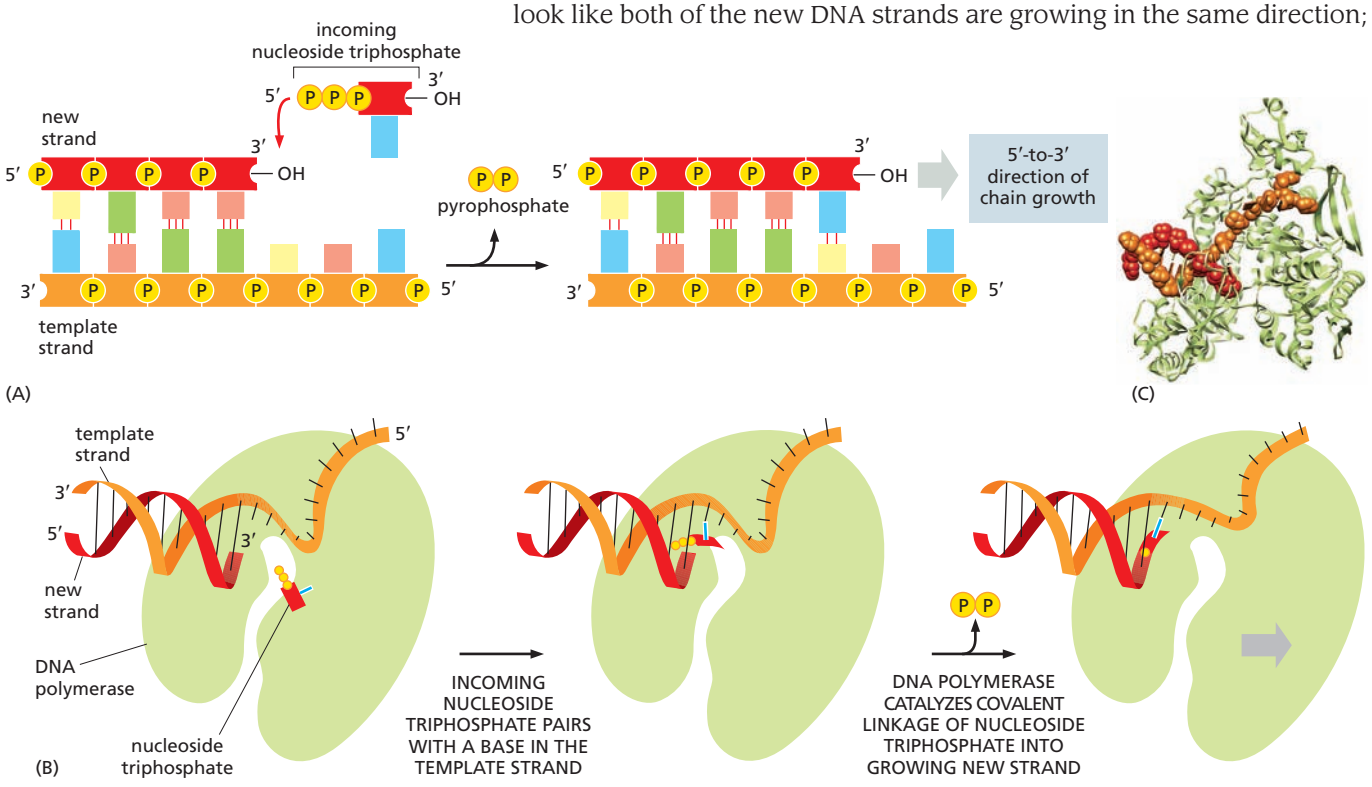


Figure 6–11 DNA polymerase adds a deoxyribonucleotide to the 3′ end of a growing DNA strand. (A) Nucleotides enter the reaction as deoxyribonucleoside triphosphates. An incoming nucleoside triphosphate forms a base pair with its partner in the template strand. It is then covalently attached to the free 3′ hydroxyl on the growing DNA strand. The new DNA strand is therefore synthesized in the 5′-to-3′ direction. The energy for the polymerization reaction comes from the hydrolysis of a high-energy phosphate bond in the incoming nucleoside triphosphate and the release of pyrophosphate, which is subsequently hydrolyzed to yield two molecules of inorganic phosphate (not shown). (B) The reaction is catalyzed by the enzyme DNA polymerase (*light green*). The polymerase guides the incoming nucleoside triphosphate to the template strand and positions it such that its 5′ triphosphate will be able to react with the 3′-hydroxyl group on the newly synthesized strand. The *gray* arrow indicates the direction of polymerase movement. (C) Structure of DNA polymerase, as determined by x-ray crystallography, also showing the replicating DNA. The template strand is the longer, *orange* strand, and the newly synthesized DNA strand is colored *red* (**Movie 6.1**).

that is, the direction in which the replication fork is moving. For that to be true, one strand would have to be synthesized in the 5′-to-3′ direction and the other in the 3′-to-5′ direction.

Does the cell have two types of DNA polymerase, one for each direction? The answer is no: all DNA polymerases add new subunits only to the 3′ end of a DNA strand (see Figure 6–11A). As a result, a new DNA chain can be synthesized only in a 5′-to-3′ direction. This can easily account for the synthesis of one of the two strands of DNA at the replication fork, but what happens on the other? This conundrum is solved by the use of a "backstitching" maneuver. The DNA strand that appears to grow in the incorrect 3′-to-5′ direction is actually made *discontinuously*, in successive, separate, small pieces—with the DNA polymerase moving backward with respect to the direction of replication-fork movement so that each new DNA fragment can be polymerized in the 5′-to-3′ direction.

The resulting small DNA pieces—called **Okazaki fragments** after the pair of biochemists who discovered them—are later joined together to form a continuous new strand. The DNA strand that is made discontinuously in this way is called the **lagging strand**, because the cumbersome backstitching mechanism imparts a slight delay to its synthesis; the other strand, which is synthesized continuously, is called the **leading strand** (**Figure 6–13**).

Although they differ in subtle details, the replication forks of all cells, prokaryotic and eukaryotic, have leading and lagging strands. This common feature arises from the fact that all DNA polymerases work only in the 5′-to-3′ direction—a restriction that allows DNA polymerase to "check its work," as we discuss next.

DNA Polymerase Is Self-correcting

DNA polymerase is so accurate that it makes only about one error in every 10^7 nucleotide pairs it copies. This error rate is much lower than can be explained simply by the accuracy of complementary base-pairing. Although A-T and C-G are by far the most stable base pairs, other, less stable base pairs—for example, G-T and C-A—can also be formed. Such incorrect base pairs are formed much less frequently than correct ones, but, if allowed to remain, they would result in an accumulation of

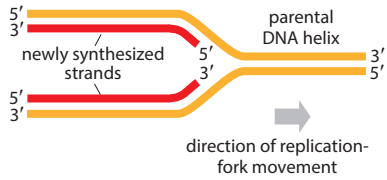


Figure 6–12 At a replication fork, the two newly synthesized DNA strands are of opposite polarities. This is because the two template strands are oriented in opposite directions.

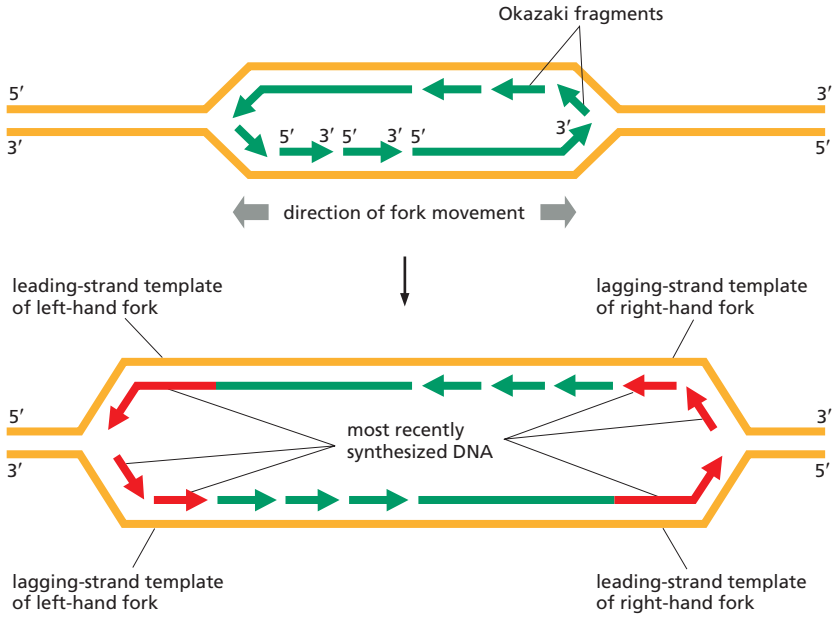


Figure 6–13 At each replication fork, the lagging DNA strand is synthesized in pieces. Because both of the new strands at a replication fork are synthesized in the 5′-to-3′ direction, the lagging strand of DNA must be made initially as a series of short DNA strands, which are later joined together. The upper diagram shows two replication forks moving in opposite directions; the lower diagram shows the same forks a short time later. To replicate the lagging strand, DNA polymerase uses a backstitching mechanism: it synthesizes short pieces of DNA (called Okazaki fragments) in the 5′-to-3′ direction and then moves back along the template strand (toward the fork) before synthesizing the next fragment.

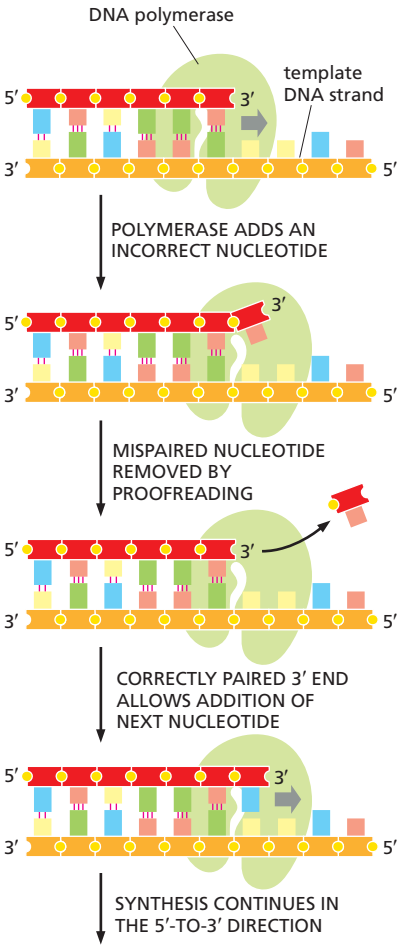


Figure 6–14 During DNA synthesis, DNA polymerase proofreads its own work. If an incorrect nucleotide is accidentally added to a growing strand, the DNA polymerase cleaves it from the strand and replaces it with the correct nucleotide before continuing.

mutations. This disaster is avoided because DNA polymerase has two special qualities that greatly increase the accuracy of DNA replication. First, the enzyme carefully monitors the base-pairing between each incoming nucleoside triphosphate and the template strand. Only when the match is correct does DNA polymerase undergo a small structural rearrangement that allows it to catalyze the nucleotide-addition reaction. Second, when DNA polymerase does make a rare mistake and adds the wrong nucleotide, it can correct the error through an activity called **proofreading**.

Proofreading takes place at the same time as DNA synthesis. Before the enzyme adds the next nucleotide to a growing DNA strand, it checks whether the previously added nucleotide is correctly base-paired to the template strand. If so, the polymerase adds the next nucleotide; if not, the polymerase clips off the mispaired nucleotide and tries again (**Figure 6–14**). Polymerization and proofreading are tightly coordinated, and the two reactions are carried out by different catalytic domains in the same polymerase molecule (**Figure 6–15**).

This proofreading mechanism is possible only for DNA polymerases that synthesize DNA exclusively in the 5′-to-3′ direction. If a DNA polymerase were able to synthesize in the 3′-to-5′ direction (circumventing the need for backstitching on the lagging strand), it would be unable to proofread. That's because if this "backward" polymerase were to remove an incorrectly paired nucleotide from the 5′ end, it would create a chemical dead end—a strand that could no longer be elongated (**Figure 6–16**). Thus, for a DNA polymerase to function as a self-correcting enzyme that removes its own polymerization errors as it moves along the DNA, it must proceed only in the 5′-to-3′ direction. The cumbersome backstitching mechanism on the lagging strand can be seen as a necessary consequence of maintaining this crucial proofreading activity.

Short Lengths of RNA Act as Primers for DNA Synthesis

We have seen that the accuracy of DNA replication depends on the requirement of the DNA polymerase for a correctly base-paired 3′ end before it can add more nucleotides to a growing DNA strand. How then can the polymerase begin a completely new DNA strand? To get the process started, a different enzyme is needed—one that can begin a new polynucleotide strand simply by joining two nucleotides together without the need for a base-paired end. This enzyme does not, however, synthesize DNA. It makes a short length of a closely related type of nucleic acid—**RNA** (**ribonucleic acid**)—using the DNA strand as a template. This short length of RNA, about 10 nucleotides long, is base-paired to the template strand and provides a base-paired 3′ end as a starting point for DNA polymerase (**Figure 6–17**). An RNA fragment thus serves as a *primer* for DNA synthesis, and the enzyme that synthesizes the RNA primer is known as **primase**.

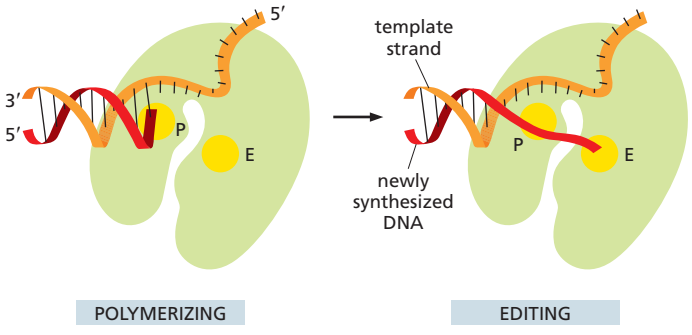


Figure 6–15 DNA polymerase contains separate sites for DNA synthesis and proofreading. The diagrams are based on the structure of an *E. coli* DNA polymerase molecule, as determined by x-ray crystallography. The DNA polymerase, which cradles the DNA molecule being replicated, is shown in the polymerizing mode (*left*) and in the proofreading, or editing, mode (*right*). The catalytic sites for the polymerization activity (P) and editing activity (E) are indicated. When the polymerase adds an incorrect nucleotide, the newly synthesized DNA strand (*red*) transiently unpairs from the template strand (*orange*), and its 3′ end moves into the editing site (E) to allow the incorrect nucleotide to be removed.

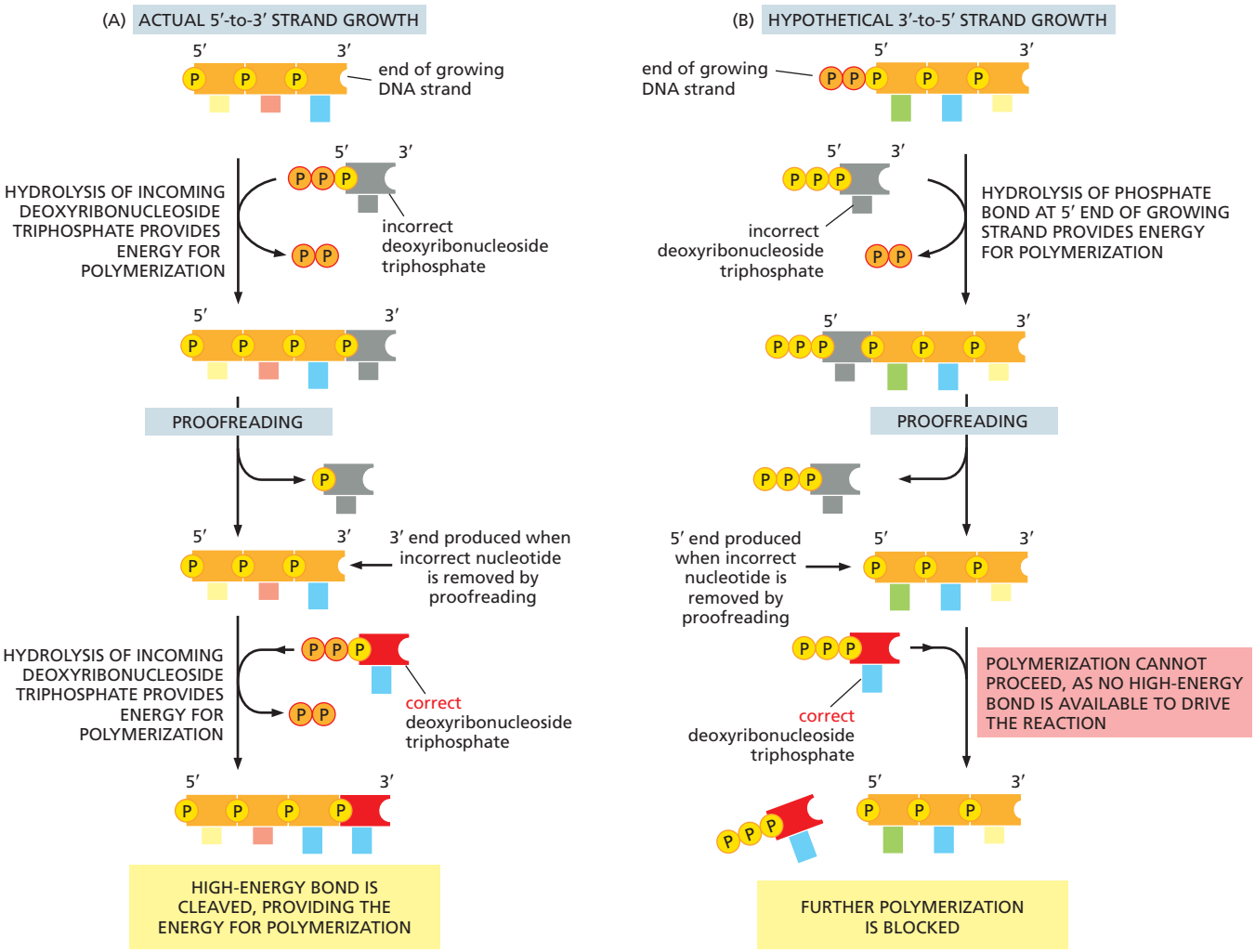


Figure 6–16 For proofreading to take place, DNA polymerization must proceed in the 5′-to-3′ direction. (A) Polymerization in the normal 5′-to-3′ direction allows the DNA strand to continue to be elongated after an incorrectly added nucleotide (*gray*) has been removed by proofreading (see Figure 6–14). (B) If DNA synthesis instead proceeded in the backward 3′-to-5′ direction, the energy for polymerization would come from the hydrolysis of the phosphate groups at the 5′ end of the growing chain (*orange*), rather than the 5′ end of the incoming nucleoside triphosphate. Removal of an incorrect nucleotide would block the addition of the correct nucleotide (*red*), as there are no high-energy phosphodiester bonds remaining at the 5′ end of the growing strand.

Primase is an example of an *RNA polymerase*, an enzyme that synthesizes RNA using DNA as a template. A strand of RNA is very similar chemically to a single strand of DNA except that it is made of ribonucleotide subunits, in which the sugar is ribose, not deoxyribose; RNA also differs from DNA in that it contains the base uracil (U) instead of thymine (T) (see Panel 2–7, pp. 78–79). However, because U can form a base pair with A, the RNA primer is synthesized on the DNA strand by complementary base-pairing in exactly the same way as is DNA.

For the leading strand, an RNA primer is needed only to start replication at a replication origin; at that point, the DNA polymerase simply takes over, extending this primer with DNA synthesized in the 5′-to-3′ direction. But on the lagging strand, where DNA synthesis is discontinuous, new primers are continuously needed to keep polymerization going (see Figure 6–13). The movement of the replication fork continually exposes unpaired bases on the lagging-strand template, and new RNA primers must be laid down at intervals along the newly exposed, single-stranded

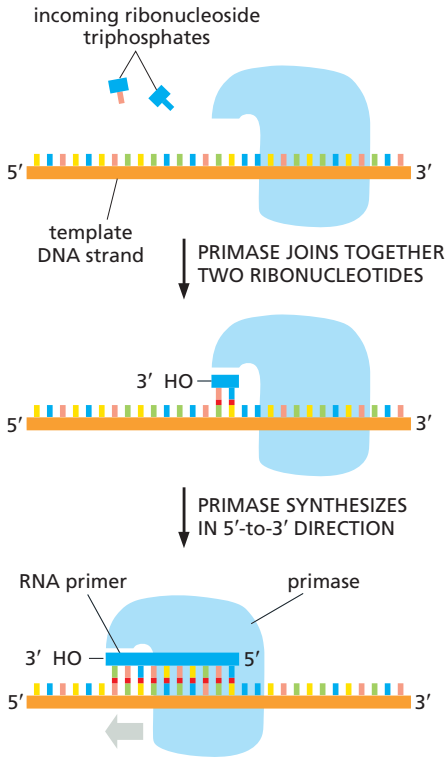


Figure 6–17 RNA primers are synthesized by an RNA polymerase called primase, which uses a DNA strand as a template. Like DNA polymerase, primase synthesizes in the 5′-to-3′ direction. Unlike DNA polymerase, however, primase can start a new polynucleotide chain by joining together two nucleoside triphosphates without the need for a base-paired 3′ end as a starting point. Primase uses ribonucleoside triphosphate rather than deoxyribonucleoside triphosphate.

stretch. DNA polymerase then adds a deoxyribonucleotide to the 3′ end of each new primer to produce another Okazaki fragment, and it will continue to elongate this fragment until it runs into the previously synthesized RNA primer (**Figure 6–18**).

To produce a continuous new DNA strand from the many separate pieces of nucleic acid made on the lagging strand, three additional enzymes are needed. These act quickly to remove the RNA primer, replace it with DNA, and join the remaining DNA fragments together. A nuclease degrades the RNA primer, a DNA polymerase called a *repair polymerase* replaces the RNA primers with DNA (using the end of the adjacent Okazaki fragment as its primer), and the enzyme **DNA ligase** joins the 5′-phosphate end of one DNA fragment to the adjacent 3′-hydroxyl end of the next (**Figure 6–19**). Because it was discovered first, the repair polymerase involved in this process is often called DNA polymerase I; the polymerase that carries out the bulk of DNA replication at the forks is known as DNA polymerase III.

Unlike DNA polymerases I and III, primase does not proofread its work. As a result, primers frequently contain mistakes. But because primers are made of RNA instead of DNA, they stand out as "suspect copy" to be automatically removed and replaced by DNA. The repair polymerase that makes this DNA, like the replicative polymerase, proofreads as it synthesizes. In this way, the cell's replication machinery is able to begin new DNA strands and, at the same time, ensure that all of the DNA is copied faithfully.

Proteins at a Replication Fork Cooperate to Form a Replication Machine

DNA replication requires the cooperation of a large number of proteins that act in concert to synthesize new DNA. These proteins form part of a remarkably complex replication machine. The first problem faced by the replication machine is accessing the nucleotides that lie ahead of the replication fork and are thus buried within the double helix. For DNA replication to occur, the double helix must be continuously pried apart so that the incoming nucleoside triphosphates can form base pairs with

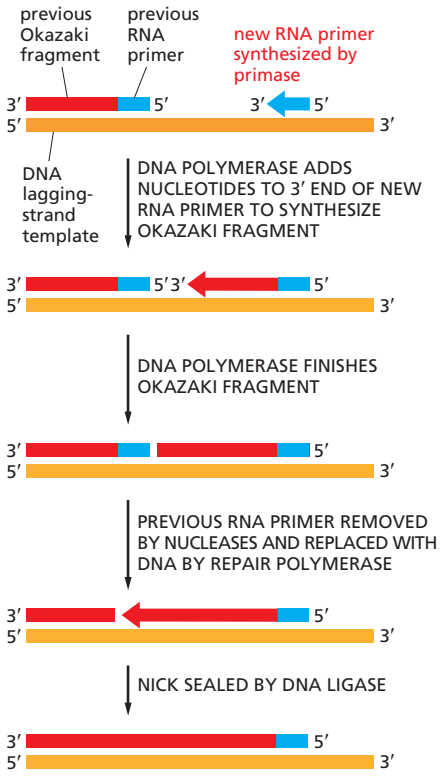


Figure 6–18 Multiple enzymes are required to synthesize the lagging DNA strand. In eukaryotes, RNA primers are made at intervals of about 200 nucleotides on the lagging strand, and each RNA primer is approximately 10 nucleotides long. These primers are extended by a replicative DNA polymerase to produce Okazaki fragments. The primers are subsequently removed by nucleases that recognize the RNA strand in an RNA–DNA hybrid helix and degrade it; this leaves gaps that are filled in by a repair DNA polymerase that can proofread as it fills in the gaps. The completed DNA fragments are finally joined together by an enzyme called DNA ligase, which catalyzes the formation of a phosphodiester bond between the 3′-hydroxyl end of one fragment and the 5′-phosphate end of the next, thus linking up the sugar–phosphate backbones. This nick-sealing reaction requires an input of energy in the form of ATP (see Figure 6–19).

Figure 6–19 DNA ligase joins together Okazaki fragments on the lagging strand during DNA synthesis. The ligase enzyme uses a molecule of ATP to activate the 5′ phosphate of one fragment (step 1) before forming a new bond with the 3′ hydroxyl of the other fragment (step 2).

each template strand. Two types of replication proteins—*DNA helicases* and *single-strand DNA-binding proteins*—cooperate to carry out this task. A helicase sits at the very front of the replication machine, where it uses the energy of ATP hydrolysis to propel itself forward, prying apart the double helix as it speeds along the DNA (**Figure 6–20** and **Movie 6.2**). Single-strand DNA-binding proteins then latch onto the single-stranded DNA exposed by the helicase, preventing the strands from re-forming base pairs and keeping them in an elongated form so that they can serve as efficient templates.

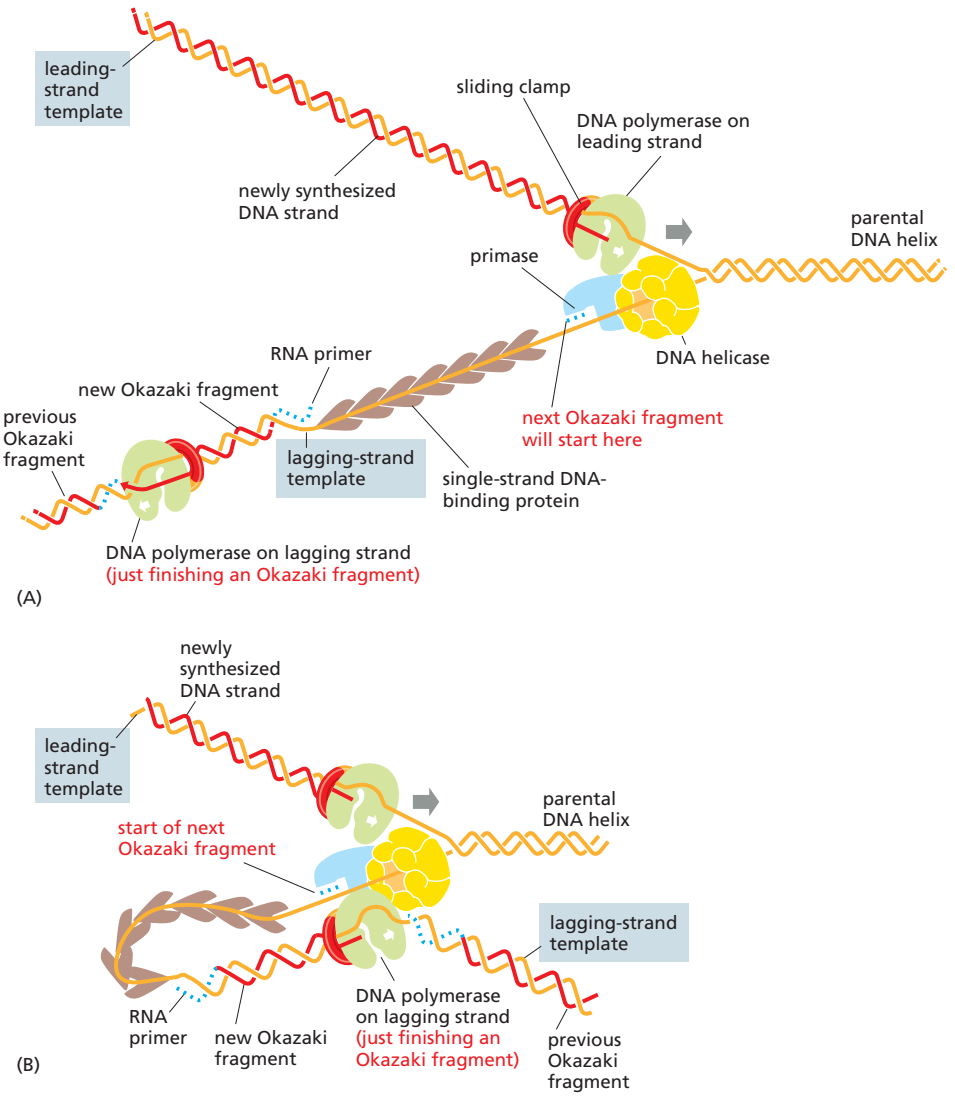


Figure 6–20 DNA synthesis is carried out by a group of proteins that act together as a replication machine. (A) DNA polymerases are held on the leading- and lagging-strand templates by circular protein clamps that allow the polymerases to slide. On the lagging-strand template, the clamp detaches each time the polymerase completes an Okazaki fragment. A clamp loader (not shown) is required to attach a sliding clamp each time a new Okazaki fragment is synthesized. At the head of the fork, a DNA helicase unwinds the strands of the parental DNA double helix. Single-strand DNA-binding proteins keep the DNA strands apart to provide access for the primase and polymerase. For simplicity, this diagram shows the proteins working independently; in the cell, they are held together in a large replication machine, as shown in (B).

(B) This diagram shows a current view of how the replication proteins are arranged when a replication fork is moving. To generate this structure, the lagging strand shown in (A) has been folded to bring its DNA polymerase in contact with the leading-strand DNA polymerase. This folding process also brings the 3′ end of each completed Okazaki fragment close to the start site for the next Okazaki fragment. Because the lagging-strand DNA polymerase is bound to the rest of the replication proteins, the same polymerase can be reused to synthesize successive Okazaki fragments; in this diagram, the lagging-strand DNA polymerase is about to let go of its completed Okazaki fragment and move to the next RNA primer being synthesized by the nearby primase. To watch the replication complex in action, see **Movie 6.3** and **Movie 6.4**.

QUESTION 6–2

Discuss the following statement: "Primase is a sloppy enzyme that makes many mistakes. Eventually, the RNA primers it makes are removed and replaced with DNA synthesized by a polymerase with higher fidelity. This is wasteful. It would be more energy-efficient if a DNA polymerase were used to make an accurate primer in the first place."

This localized unwinding of the DNA double helix itself presents a problem. As the helicase moves forward, prying open the double helix, the DNA ahead of the fork gets wound more tightly. This excess twisting in front of the replication fork creates tension in the DNA that—if allowed to build—makes unwinding the double helix increasingly difficult and ultimately impedes the forward movement of the replication machinery (**Figure 6–21A**). Enzymes called *DNA topoisomerases* relieve this tension. A DNA topoisomerase produces a transient, single-strand nick in the DNA backbone, which temporarily releases the built-up tension; the enzyme then reseals the nick before falling off the DNA (**Figure 6–21B**).

Back at the replication fork, an additional protein, called a *sliding clamp*, keeps DNA polymerase firmly attached to the template while it is synthesizing new strands of DNA. Left on their own, most DNA polymerase molecules will synthesize only a short string of nucleotides before falling off the DNA template strand. The sliding clamp forms a ring around the newly formed DNA double helix and, by tightly gripping the polymerase, allows the enzyme to move along the template strand without falling off as it synthesizes new DNA (see Figure 6–20A and **Movie 6.5**).

Assembly of the clamp around DNA requires the activity of another replication protein, the *clamp loader*, which hydrolyzes ATP each time it locks a sliding clamp around a newly formed DNA double helix. This loading needs to occur only once per replication cycle on the leading strand; on the lagging strand, however, the clamp is removed and then reattached each time a new Okazaki fragment is made. In bacteria, this happens approximately once per second.

Most of the proteins involved in DNA replication are held together in a large multienzyme complex that moves as a unit along the parental DNA double helix, enabling DNA to be synthesized on both strands in a coordinated manner. This complex can be likened to a miniature sewing machine composed of protein parts and powered by nucleoside triphosphate hydrolysis (Figure 6–20B). The proteins involved in DNA replication are listed in **Table 6–1**.

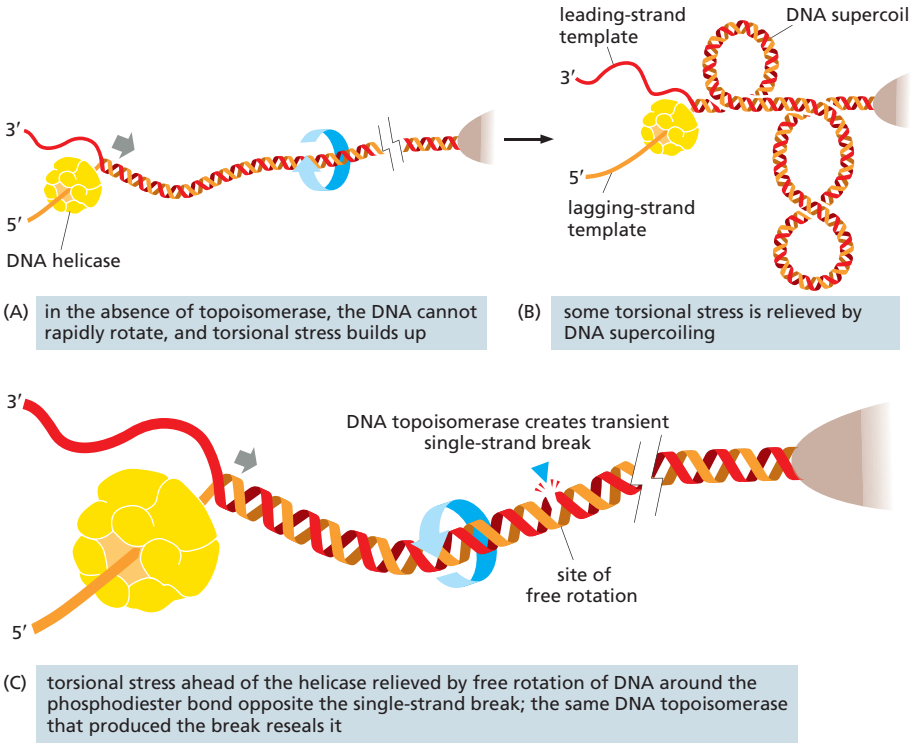


Figure 6–21 DNA topoisomerases relieve the tension that builds up in front of a replication fork. (A) As a DNA helicase moves forward, unwinding the DNA double helix, it generates a section of overwound DNA ahead of it. Tension builds up because the rest of the chromosome (shown in *brown*) is too large to rotate fast enough to relieve the buildup of torsional stress. The broken bars represent approximately 20 turns of DNA. (B) Some of this torsional stress is relieved by additional coiling of the DNA double helix to form supercoils. (C) DNA topoisomerases relieve this stress by generating temporary nicks in the DNA, which allow rapid rotation around the single strands opposite the nicks.

TABLE 6–1 PROTEINS INVOLVED IN DNA REPLICATION

Protein	Activity
DNA polymerase	catalyzes the addition of nucleotides to the 3′ end of a growing strand of DNA using a parental DNA strand as a template
DNA helicase	uses the energy of ATP hydrolysis to unwind the DNA double helix ahead of the replication fork
Single-strand DNA-binding protein	binds to single-stranded DNA exposed by DNA helicase, preventing base pairs from re-forming before the lagging strand can be replicated
DNA topoisomerase	produces transient nicks in the DNA backbone to relieve the tension built up by the unwinding of DNA ahead of the DNA helicase
Sliding clamp	keeps DNA polymerase attached to the template, allowing the enzyme to move along without falling off as it synthesizes new DNA
Clamp loader	uses the energy of ATP hydrolysis to lock the sliding clamp onto DNA
Primase	synthesizes RNA primers along the lagging-strand template
DNA ligase	uses the energy of ATP hydrolysis to join Okazaki fragments made on the lagging-strand template

Telomerase Replicates the Ends of Eukaryotic Chromosomes

Having discussed how DNA replication begins at origins and continues as the replication forks proceed, we now turn to the special problem of replicating the very ends of chromosomes. As we discussed previously, because DNA replication proceeds only in the 5′-to-3′ direction, the lagging strand of the replication fork must be synthesized in the form of discontinuous DNA fragments, each of which is initiated from an RNA primer laid down by a primase (see Figure 6–18). A serious problem arises, however, as the replication fork approaches the end of a chromosome: although the leading strand can be replicated all the way to the chromosome tip, the lagging strand cannot. When the final RNA primer on the lagging strand is removed, there is no enzyme that can replace it with DNA (**Figure 6–22**). Without a strategy to deal with this problem, the lagging strand would become shorter with each round of DNA replication and, after repeated cell divisions, the chromosomes themselves would shrink—eventually losing valuable genetic information.

Bacteria avoid this "end-replication" problem by having circular DNA molecules as chromosomes. Eukaryotes get around it by adding long, repetitive nucleotide sequences to the ends of every chromosome. These sequences, which are incorporated into structures called **telomeres**, attract an enzyme called **telomerase** to the chromosome ends. Telomerase carries its own RNA template, which it uses to add multiple copies of the same repetitive DNA sequence to the lagging-strand template. In many dividing cells, telomeres are continuously replenished, and the resulting extended templates can then be copied by conventional DNA replication, ensuring that no peripheral chromosomal sequences are lost (**Figure 6–23**).

In addition to allowing replication of chromosome ends, telomeres form structures that mark the true ends of a chromosome. These structures allow the cell to distinguish unambiguously between the natural ends of

QUESTION 6–3

A gene encoding one of the proteins involved in DNA replication has been inactivated by a mutation in a cell. In the absence of this protein, the cell attempts to replicate its DNA. What would happen during the DNA replication process if each of the following proteins were missing?

A. DNA polymerase
B. DNA ligase
C. Sliding clamp
D. Nuclease that removes RNA primers
E. DNA helicase
F. Primase

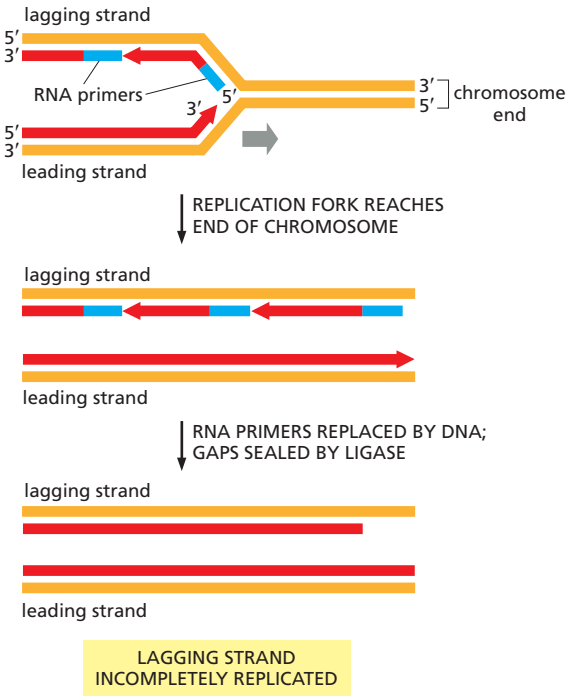


Figure 6–22 Without a special mechanism to replicate the ends of linear chromosomes, DNA would be lost during each round of cell division. DNA synthesis begins at origins of replication and continues until the replication machinery reaches the ends of the chromosome. The leading strand is synthesized in its entirety. But the ends of the lagging strand can't be completed, because once the final RNA primer has been removed, there is no mechanism for replacing it with DNA. Complete replication of the lagging strand requires a special mechanism to keep the chromosome ends from shrinking with each cell division.

chromosomes and the double-strand DNA breaks that sometimes occur accidentally in the middle of chromosomes. These breaks are dangerous and must be immediately repaired, as we will see shortly.

Telomere Length Varies by Cell Type and with Age

In addition to attracting telomerase, the repetitive DNA sequences found within telomeres attract other telomere-binding proteins that not only physically protect chromosome ends, but help maintain telomere length.

Cells that divide at a rapid rate throughout the life of the organism—those that line the gut or generate blood cells in the bone marrow, for example—keep their telomerase fully active. Many other cell types, however, gradually turn down their telomerase activity. After many rounds

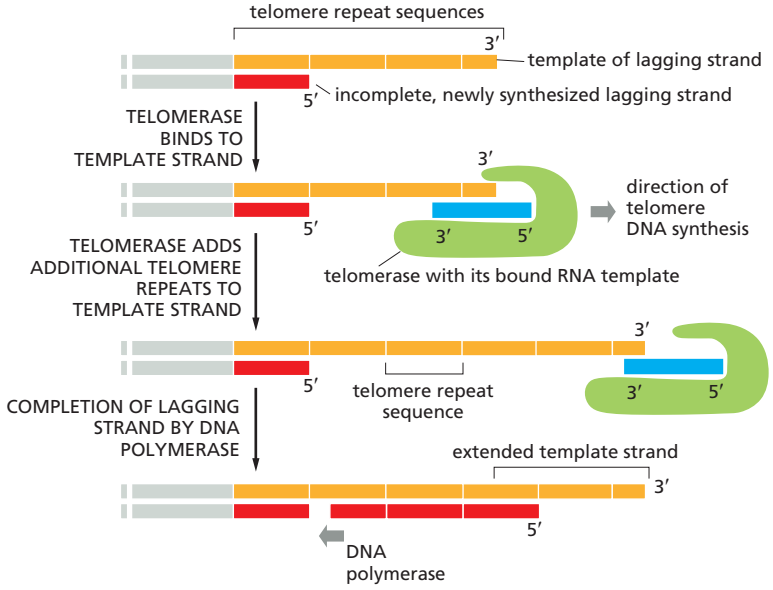


Figure 6–23 Telomeres and telomerase prevent linear eukaryotic chromosomes from shortening with each cell division. To complete the replication of the lagging strand at the ends of a chromosome, the template strand (*orange*) is first extended beyond the DNA that is to be copied. To achieve this, the enzyme telomerase adds to the telomere repeat sequences at the 3′ end of the template strand, which then allows the newly synthesized lagging strand (*red*) to be lengthened by DNA polymerase, as shown. The telomerase enzyme itself carries a short piece of RNA (*blue*) with a sequence that is complementary to the DNA repeat sequence; this RNA acts as the template for telomere DNA synthesis. After the lagging-strand replication is complete, a short stretch of single-stranded DNA remains at the ends of the chromosome; however, the newly synthesized lagging strand, at this point, contains all the information present in the original DNA. To see telomerase in action, view Movie 6.6.

of cell division, the telomeres in these descendent cells will shrink, until they essentially disappear. At this point, these cells will cease dividing. In theory, such a mechanism could provide a safeguard against the uncontrolled proliferation of cells—including abnormal cells that have accumulated mutations that could promote the development of cancer.

DNA REPAIR

The diversity of living organisms and their success in colonizing almost every part of the Earth's surface depend on genetic changes accumulated gradually over billions of years. A small subset of these changes will be beneficial, allowing the affected organisms to adapt to changing conditions and to thrive in new habitats. However, most of these changes will be of little consequence or even deleterious.

In the short term, and from the perspective of an individual organism, such genetic alterations—called mutations—are kept to a minimum: to survive and reproduce, individuals must be genetically stable. This stability is achieved not only through the extremely accurate mechanism for replicating DNA that we have just discussed, but also through the work of a variety of protein machines that continually scan the genome for DNA damage and fix it when it occurs. Although some changes arise from rare mistakes in the replication process, the majority of DNA damage is an unintended consequence of the vast number of chemical reactions that occur inside cells.

Most DNA damage is only temporary, because it is immediately corrected by processes collectively called **DNA repair**. The importance of these DNA repair processes is evident from the consequences of their malfunction. Humans with the genetic disease *xeroderma pigmentosum*, for example, cannot mend the damage done by ultraviolet (UV) radiation because they have inherited a defective gene for one of the proteins involved in this repair process. Such individuals develop severe skin lesions, including skin cancer, because of the DNA damage that accumulates in cells exposed to sunlight and the consequent mutations that arise in these cells.

In this section, we describe a few of the specialized mechanisms cells use to repair DNA damage. We then consider examples of what happens when these mechanisms fail—and we discuss how the evolutionary history of DNA replication and repair is reflected in our genome.

DNA Damage Occurs Continually in Cells

Just like any other molecule in the cell, DNA is continually undergoing thermal collisions with other molecules, often resulting in major chemical changes in the DNA. For example, in the time it takes to read this sentence, a total of about a trillion (10^{12}) purine bases (A and G) will be lost from DNA in the cells of your body by a spontaneous reaction called *depurination* (**Figure 6–24A**). Depurination does not break the DNA phosphodiester backbone but instead removes a purine base from a nucleotide, giving rise to lesions that resemble missing teeth (see Figure 6–26B). Another common reaction is the spontaneous loss of an amino group (*deamination*) from a cytosine in DNA to produce the base uracil (**Figure 6–24B**).

The ultraviolet radiation in sunlight is also damaging to DNA; it promotes covalent linkage between two adjacent pyrimidine bases, forming, for example, the *thymine dimer* shown in **Figure 6–25**. It is the failure to repair thymine dimers that spells trouble for individuals with the disease xeroderma pigmentosum.

QUESTION 6–4

Discuss the following statement: "The DNA repair enzymes that fix deamination and depurination damage must preferentially recognize such damage on newly synthesized DNA strands."

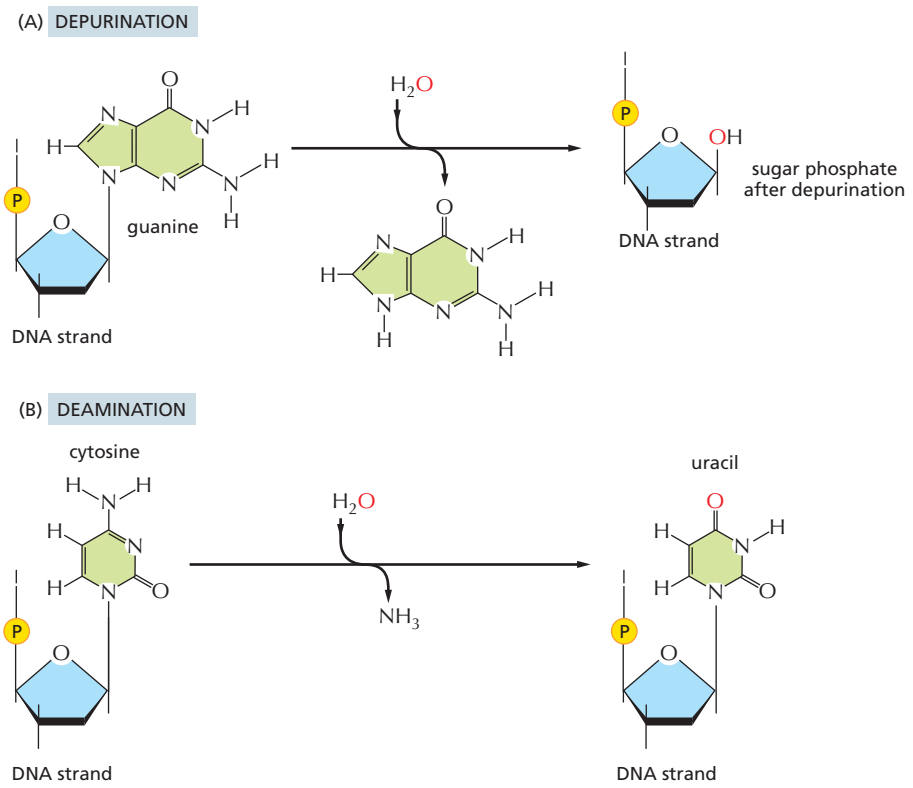


Figure 6–24 Depurination and deamination are the most frequent chemical reactions known to create serious DNA damage in cells. (A) Depurination can remove guanine (or adenine) from DNA. (B) The major type of deamination reaction converts cytosine to uracil, which, as we have seen, is not normally found in DNA. However, deamination can occur on other bases as well. Both depurination and deamination take place on double-helical DNA, and neither break the phosphodiester backbone.

These are only a few of many chemical changes that can occur in our DNA. Others are caused by reactive chemicals produced as a normal part of cell metabolism. If left unrepaired, DNA damage leads either to the substitution of one nucleotide pair for another as a result of incorrect base-pairing during replication (**Figure 6–26A**) or to deletion of one or more nucleotide pairs in the daughter DNA strand after DNA replication (**Figure 6–26B**). Some types of DNA damage (thymine dimers, for example) can stall the DNA replication machinery at the site of the damage.

In addition to this chemical damage, DNA can also be altered by replication itself. The replication machinery that copies the DNA can—albeit rarely—incorporate an incorrect nucleotide that it fails to correct via proofreading (see Figure 6–14).

For each of these forms of DNA damage, cells possess a mechanism for repair, as we discuss next.

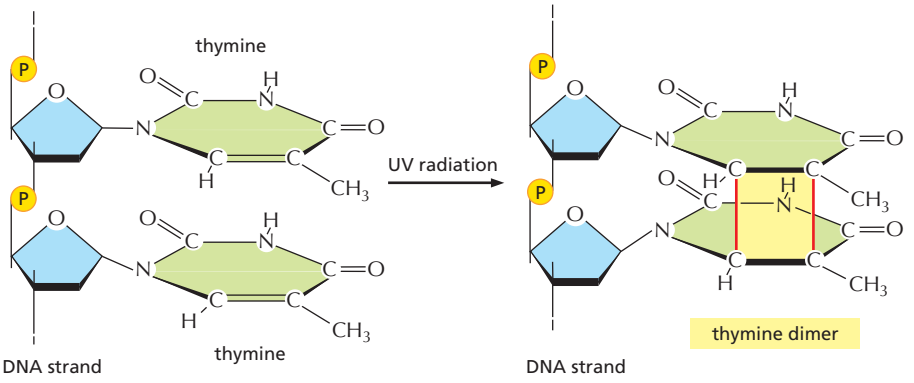


Figure 6–25 The ultraviolet radiation in sunlight can cause the formation of thymine dimers. Two adjacent thymine bases have become covalently attached to each other to form a thymine dimer. Skin cells that are exposed to sunlight are especially susceptible to this type of DNA damage.

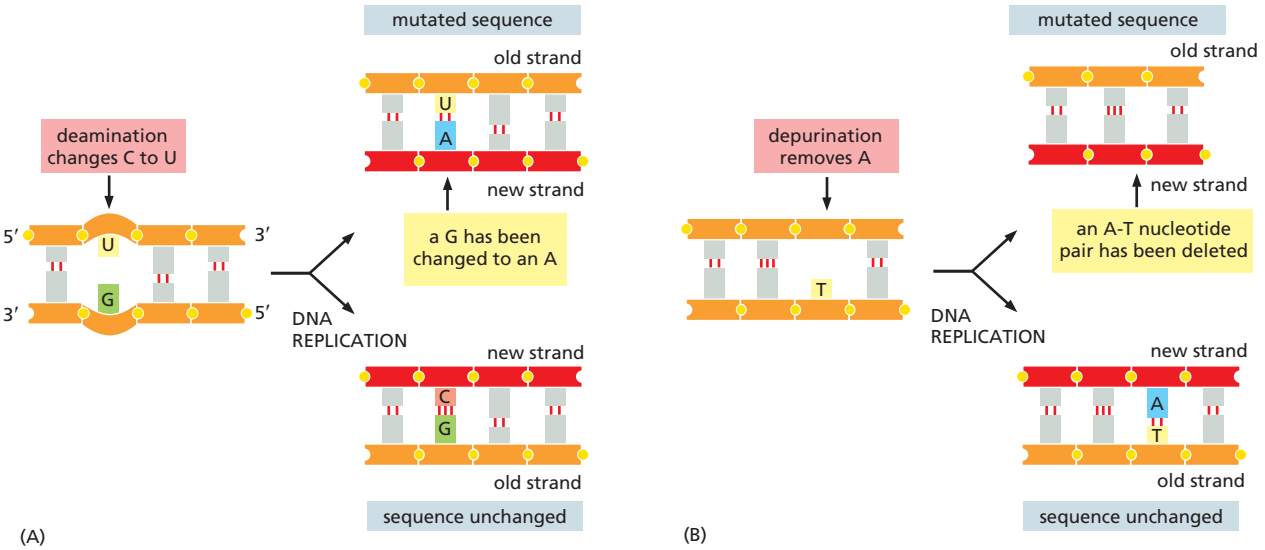


Figure 6–26 Chemical modifications of nucleotides, if left unrepaired, produce mutations. (A) Deamination of cytosine, if uncorrected, results in the substitution of one base for another when the DNA is replicated. As shown in Figure 6–24B, deamination of cytosine produces uracil. Uracil differs from cytosine in its base-pairing properties and preferentially base-pairs with adenine. The DNA replication machinery therefore inserts an adenine when it encounters a uracil on the template strand. (B) Depurination, if uncorrected, can lead to the loss of a nucleotide pair. When the replication machinery encounters a missing purine on the template strand, it can skip to the next complete nucleotide, as shown, thus producing a daughter DNA molecule that is missing one nucleotide pair. In other cases, the replication machinery places an incorrect nucleotide across from the missing base, again resulting in a mutation (not shown).

Cells Possess a Variety of Mechanisms for Repairing DNA

The thousands of random chemical changes that occur every day in the DNA of a human cell—through thermal collisions or exposure to reactive metabolic by-products, DNA-damaging chemicals, or radiation—are repaired by a variety of mechanisms, each catalyzed by a different set of enzymes. Nearly all these repair mechanisms depend on the double-helical structure of DNA, which provides two copies of the genetic information—one in each strand of the double helix. Thus, if the sequence in one strand is accidentally damaged, information is not lost irretrievably, because a backup version of the altered strand remains in the complementary sequence of nucleotides in the other, undamaged strand. Most DNA damage creates structures that are never encountered in an undamaged DNA strand; thus the good strand is easily distinguished from the bad.

The basic pathway for repairing damage to DNA, illustrated schematically in **Figure 6–27**, involves three basic steps:

1. The damaged DNA is recognized and removed by one of a variety of mechanisms. These involve nucleases, which cleave the covalent bonds that join the damaged nucleotides to the rest of the DNA strand, leaving a small gap on one strand of the DNA double helix.
2. A *repair DNA polymerase* binds to the 3′-hydroxyl end of the cut DNA strand. The enzyme then fills in the gap by making a complementary copy of the information present in the undamaged strand. Although they differ from the DNA polymerase that replicates DNA, repair DNA polymerases synthesize DNA strands in the same way. For example, they elongate chains in the 5′-to-3′ direction and have the same type of proofreading activity to ensure that the template strand is copied accurately. In many cells, the repair polymerase is the same enzyme that fills in the gaps left after the RNA primers are removed during the normal DNA replication process (see Figure 6–18).

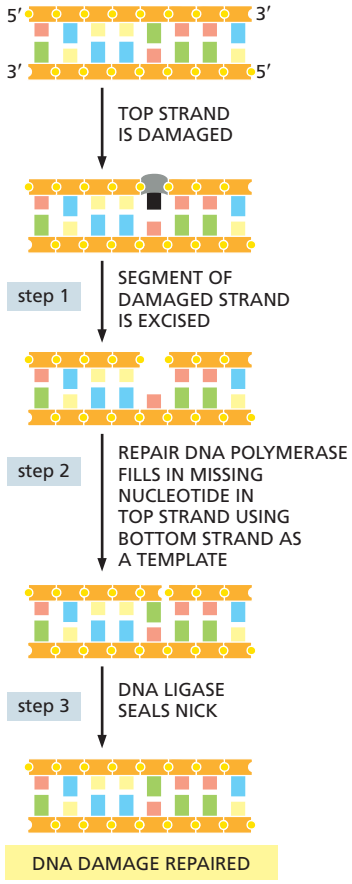


Figure 6–27 The basic mechanism of DNA repair involves three steps. In step 1 (excision), the damage is cut out by one of a series of nucleases, each specialized for a certain type of DNA damage. In step 2 (resynthesis), the original DNA sequence is restored by a repair DNA polymerase, which fills in the gap created by the excision events. In step 3 (ligation), DNA ligase seals the nick left in the sugar–phosphate backbone of the repaired strand. Nick sealing, which requires energy from ATP hydrolysis, remakes the broken phosphodiester bond between the adjacent nucleotides (see Figure 6–19).

3. When the repair DNA polymerase has filled in the gap, a break remains in the sugar–phosphate backbone of the repaired strand. This nick in the helix is sealed by DNA ligase, the same enzyme that joins the Okazaki fragments during replication of the lagging DNA strand (see Figure 6–19).

A DNA Mismatch Repair System Removes Replication Errors That Escape Proofreading

Although the high fidelity and proofreading abilities of the cell's replication machinery generally prevent replication errors from occurring, rare mistakes do happen. Fortunately, the cell has a backup system—called **mismatch repair**—that is dedicated to correcting these errors. The replication machine makes approximately one mistake per 10^7 nucleotides synthesized; DNA mismatch repair corrects 99% of these replication errors, increasing the overall accuracy to one mistake in 10^9 nucleotides synthesized. This level of accuracy is much, much higher than that generally encountered in our day-to-day lives (**Table 6–2**).

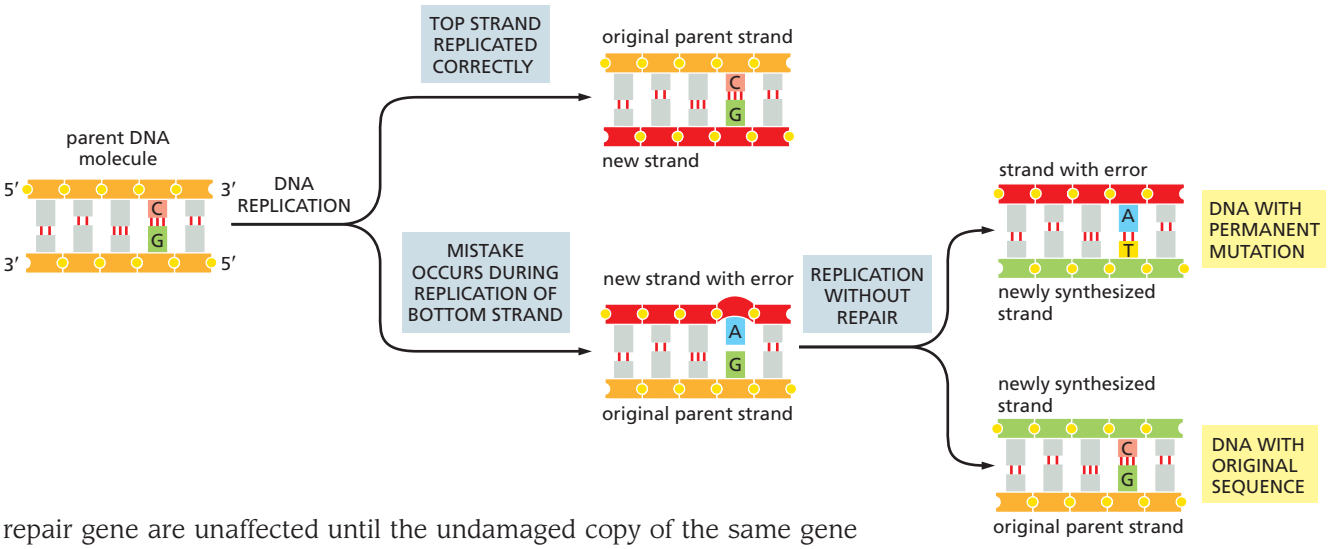
Whenever the replication machinery makes a copying mistake, it leaves behind a mispaired nucleotide (commonly called a *mismatch*). If left uncorrected, the mismatch will result in a permanent mutation in the next round of DNA replication (**Figure 6–28**). In most cases, however, a complex of mismatch repair proteins will detect the DNA mismatch, remove a portion of the DNA strand containing the error, and then resynthesize the missing DNA. This repair mechanism restores the correct sequence (**Figure 6–29**).

To be effective, the mismatch repair system must be able to recognize which of the DNA strands contains the error. Removing a segment from the strand that contains the correct sequence would only compound the mistake. The way the mismatch system solves this problem is by recognizing and removing only the newly made DNA. In bacteria, newly synthesized DNA lacks a type of chemical modification (a methyl group added to certain adenines) that is present on the preexisting parent DNA. Newly synthesized DNA is unmethylated for a short time, during which the new and template strands can be easily distinguished. Other cells use different strategies for distinguishing their parent DNA from a newly replicated strand.

In humans, mismatch repair plays an important role in preventing cancer. An inherited predisposition to certain cancers (especially some types of colon cancer) is caused by mutations in genes that encode mismatch repair proteins. Human cells have two copies of these genes (one from each parent), and individuals who inherit one damaged mismatch

TABLE 6–2 ERROR RATES

A professional typist typing at 120 words per minute	1 mistake per 250 characters
Airline luggage system	1 bag lost, damaged, or delayed per 400 passengers
Driving a car in the United States	1 death per 10^4 people per year
DNA replication (without proofreading)	1 mistake per 10^5 nucleotides copied
DNA replication (with proofreading; without mismatch repair)	1 mistake per 10^7 nucleotides copied
DNA replication (with mismatch repair)	1 mistake per 10^9 nucleotides copied



repair gene are unaffected until the undamaged copy of the same gene is randomly mutated in a somatic cell. This mutant cell—and all of its progeny—are then deficient in mismatch repair; they therefore accumulate mutations more rapidly than do normal cells. Because cancers arise from cells that have accumulated multiple mutations, a cell deficient in mismatch repair has a greatly enhanced chance of becoming cancerous. Thus, inheriting a single damaged mismatch repair gene strongly predisposes an individual to cancer.

Figure 6–28 Errors made during DNA replication must be corrected to avoid mutations. If uncorrected, a mismatch will lead to a permanent mutation in one of the two DNA molecules produced during the next round of DNA replication.

Double-Strand DNA Breaks Require a Different Strategy for Repair

The repair mechanisms we have discussed thus far rely on the genetic redundancy built into every DNA double helix. If nucleotides on one strand are damaged, they can be repaired using the information present in the complementary strand. This feature makes the DNA double helix especially well-suited for stably carrying genetic information from one generation to the next.

But what happens when both strands of the double helix are damaged at the same time? Mishaps at the replication fork, radiation, and various chemical assaults can all fracture DNA, creating a *double-strand break*. Such lesions are particularly dangerous, because they can lead to the fragmentation of chromosomes and the subsequent loss of genes.

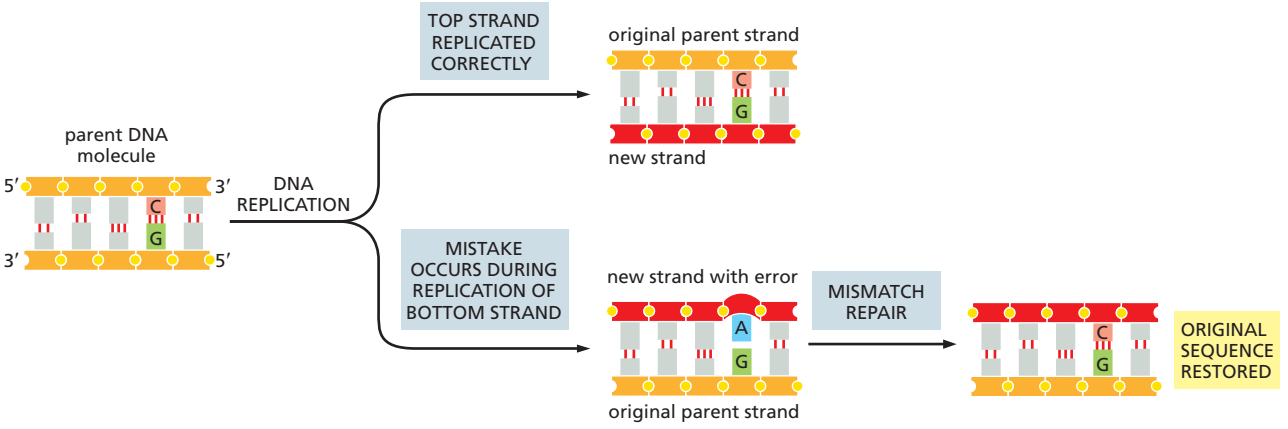


Figure 6–29 Mismatch repair eliminates replication errors and restores the original DNA sequence. When mistakes occur during DNA replication, the repair machinery must replace the incorrect nucleotide on the newly synthesized strand, using the original parent strand as its template. This mechanism eliminates the error, and allows the original sequence to be copied during subsequent rounds of replication.

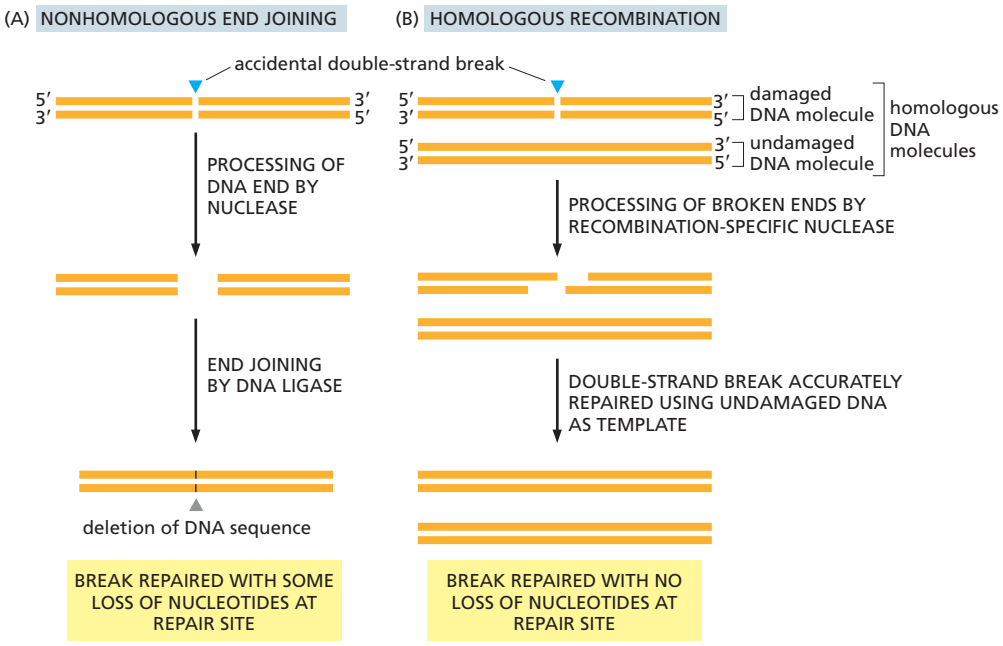


Figure 6–30 Cells can repair double-strand breaks in one of two ways. (A) In nonhomologous end joining, the break is first "cleaned" by a nuclease that chews back the broken ends to produce flush ends. The flush ends are then stitched together by a DNA ligase. Some nucleotides are usually lost in the repair process, as indicated by the *black* lines in the repaired DNA. (B) If a double-strand break occurs in one of two duplicated DNA double helices after DNA replication has occurred, but before the chromosome copies have been separated, the undamaged double helix can be readily used as a template to repair the damaged double helix through homologous recombination. Although more complicated than nonhomologous end joining, this process accurately restores the original DNA sequence at the site of the break. The detailed mechanism is presented in Figure 6–31.

This type of damage is especially difficult to repair. Every chromosome contains unique information; if a chromosome experiences a double-strand break, and the broken pieces become separated, the cell has no spare copy it can use to reconstruct the information that is now missing.

To handle this potentially disastrous type of DNA damage, cells have evolved two basic strategies. The first involves hurriedly sticking the broken ends back together, before the DNA fragments drift apart and get lost. This repair mechanism, called **nonhomologous end joining**, occurs in many cell types and is carried out by a specialized group of enzymes that "clean" the broken ends and rejoin them by DNA ligation. This "quick and dirty" mechanism rapidly seals the break, but it comes with a price: in "cleaning" the break to make it ready for ligation, nucleotides are often lost at the site of repair (**Figure 6–30A** and **Movie 6.7**). If this imperfect repair disrupts the activity of a gene, the cell could suffer serious consequences. Thus, nonhomologous end joining can be a risky strategy for fixing broken chromosomes. Fortunately, cells have an alternative, error-free strategy for repairing double-strand breaks, called *homologous recombination* (**Figure 6–30B**), as we discuss next.

Homologous Recombination Can Flawlessly Repair DNA Double-Strand Breaks

The challenge in repairing a double-strand break, as mentioned previously, is finding an intact template to guide the repair. However, if a double-strand break occurs in a double helix shortly after that stretch of DNA has been replicated, the undamaged copy can serve as a template to guide the repair of both broken strands of DNA. The information on the undamaged strands of the intact double helix can be used to repair the complementary strands in the broken DNA. Because the two DNA molecules are homologous—they have identical or nearly identical nucleotide sequences outside the broken region—this mechanism is known as **homologous recombination**. It results in a flawless repair of the double-strand break, with no loss of genetic information (see Figure 6–30B).

Homologous recombination most often occurs shortly after a cell's DNA has been replicated before cell division, when the duplicated helices are still physically close to each other (**Figure 6–31A**). To initiate

the repair, a recombination-specific nuclease chews back the 5′ ends of the two broken strands at the break (**Figure 6–31B**). Then, with the help of specialized enzymes (called recA in bacteria and Rad52 in eukaryotes), one of the broken 3′ ends "invades" the unbroken homologous DNA duplex and searches for a complementary sequence through base-pairing (**Figure 6–31C**). Once an extensive, accurate match is made, the invading strand is elongated by a repair DNA polymerase, using the complementary undamaged strand as a template (**Figure 6–31D**). After the repair polymerase has passed the point where the break occurred, the newly elongated strand rejoins its original partner, forming base pairs that hold the two strands of the broken double helix together (**Figure 6–31E**). Repair is then completed by additional DNA synthesis at the 3′ ends of both strands of the broken double helix (**Figure 6–31F**), followed by DNA ligation (**Figure 6–31G**). The net result is two intact DNA helices, for which the genetic information from one was used as a template to repair the other.

Homologous recombination can also be used to repair many other types of DNA damage, making it perhaps the most handy DNA repair mechanism available to the cell: all that is needed is an intact homologous

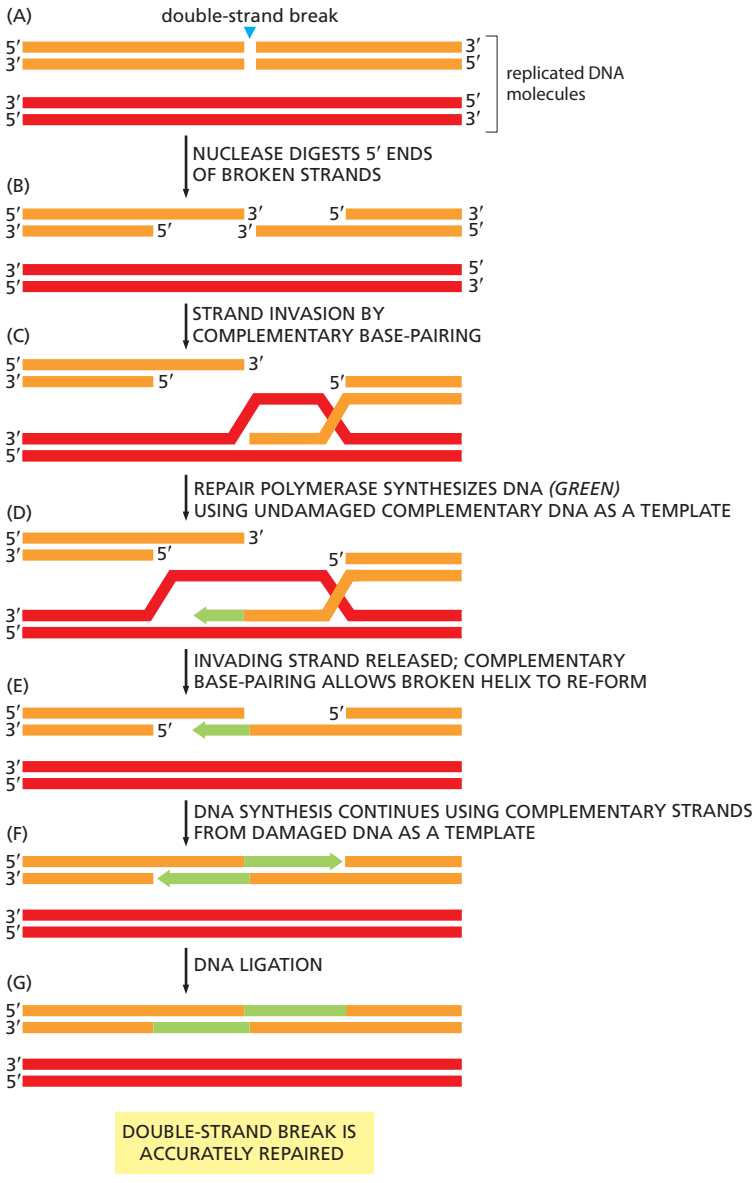


Figure 6–31 Homologous recombination flawlessly repairs DNA double-strand breaks. This is the preferred method for repairing double-strand breaks that arise shortly after the DNA has been replicated but before the cell has divided. See text for details. (Adapted from M. McVey et al., *Proc. Natl. Acad. Sci. U.S.A.* 101: 15694–15699, 2004.)

chromosome to use as a partner—a situation that occurs transiently each time a chromosome is duplicated. The "all-purpose" nature of homologous recombinational repair probably explains why this mechanism, and the proteins that carry it out, have been conserved in virtually all cells on Earth.

Homologous recombination is versatile, and it also has a crucial role in the exchange of genetic information that occurs during the formation of the gametes—sperm and eggs. This exchange, during the specialized form of cell division called *meiosis*, enhances the generation of genetic diversity within a species during sexual reproduction. We will discuss it when we talk about sex in Chapter 19.

Failure to Repair DNA Damage Can Have Severe Consequences for a Cell or Organism

On occasion, the cell's DNA replication and repair processes fail and allow a mutation to arise. This permanent change in the DNA sequence can have profound consequences. If the change occurs in a particular position in the DNA sequence, it could alter the amino acid sequence of a protein in a way that reduces or eliminates that protein's ability to function. For example, mutation of a single nucleotide in the human hemoglobin gene can cause the disease *sickle-cell anemia*. The hemoglobin protein is used to transport oxygen in the blood (see Figure 4–24). Mutations in the hemoglobin gene can produce a protein that is less soluble than normal hemoglobin and forms fibrous intracellular precipitates, which produce the characteristic sickle shape of affected red blood cells (**Figure 6–32**). Because these cells are more fragile and frequently tear as they travel through the bloodstream, patients with this potentially life-threatening disease have fewer red blood cells than usual—that is, they are anemic. Moreover, the abnormal red blood cells that remain can aggregate and block small vessels, causing pain and organ failure. We know about sickle-cell hemoglobin because individuals with the mutation survive; the mutation even provides a benefit—an increased resistance to malaria, as we discuss in Chapter 19.

The example of sickle-cell anemia, which is an inherited disease, illustrates the consequences of mutations arising in the reproductive *germ-line cells*. A mutation in a germ-line cell will be passed on to all the cells in the body of the multicellular organism that develop from it, including the gametes responsible for the production of the next generation.

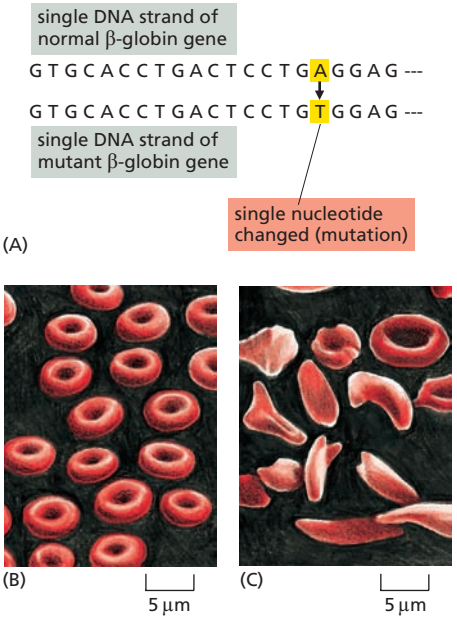


Figure 6–32 A single nucleotide change causes the disease sickle-cell anemia. (A) β-globin is one of the two types of protein subunits that form hemoglobin (see Figure 4–24). A single mutation in the β-globin gene produces a β-globin subunit that differs from normal β-globin by a change from glutamic acid to valine at the sixth amino acid position. (Only a portion of the gene is shown here; the β-globin subunit contains a total of 146 amino acids. The complete sequence of the β-globin gene is shown in Figure 5–11.) Humans carry two copies of each gene (one inherited from each parent); a sickle-cell mutation in one of the two β-globin genes generally causes no harm to the individual, as it is compensated for by the normal gene. However, an individual who inherits two copies of the mutant β-globin gene will have sickle-cell anemia. (B and C) Normal red blood cells are shown in (B), and those from an individual suffering from sickle-cell anemia in (C). Although sickle-cell anemia can be a life-threatening disease, the responsible mutation can also be beneficial. People with the disease, or those who carry one normal gene and one sickle-cell gene, are more resistant to malaria than unaffected individuals, because the parasite that causes malaria grows poorly in red blood cells that contain the sickle-cell form of hemoglobin.

The many other cells in a multicellular organism (its *somatic cells*) must also be protected against mutation—in this case, against mutations that arise during the life of the individual. Nucleotide changes that occur in somatic cells can give rise to variant cells, some of which grow and divide in an uncontrolled fashion at the expense of the other cells in the organism. In the extreme case, an unchecked cell proliferation known as **cancer** results. Cancers are responsible for about 30% of the deaths that occur in Europe and North America, and they are caused primarily by a gradual accumulation of random mutations in a somatic cell and its descendants (**Figure 6–33**). Increasing the mutation frequency even two- or threefold could cause a disastrous increase in the incidence of cancer by accelerating the rate at which such somatic cell variants arise.

Thus, the high fidelity with which DNA sequences are replicated and maintained is important both for germ-line cells, which transmit the genes to the next generation, and for somatic cells, which normally function as carefully regulated members of the complex community of cells in a multicellular organism. We should therefore not be surprised to find that all cells possess a very sophisticated set of mechanisms to reduce the number of mutations that occur in their DNA, devoting hundreds of genes to these repair processes.

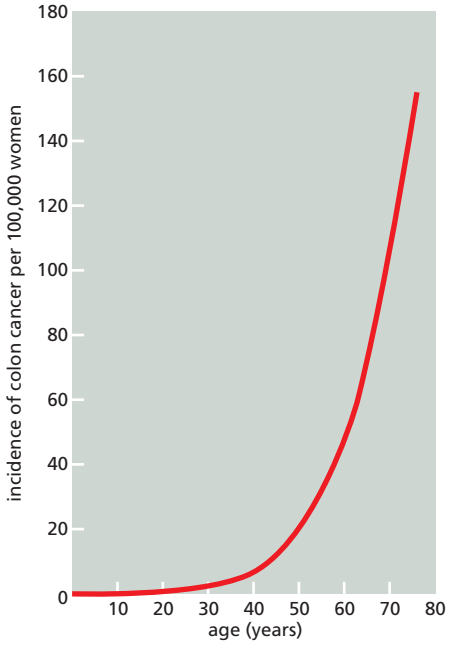


Figure 6–33 Cancer incidence increases dramatically with age. The number of newly diagnosed cases of colon cancer in women in England and Wales in a single year is plotted as a function of age at diagnosis. Colon cancer, like most human cancers, is caused by the accumulation of multiple mutations. Because cells are continually experiencing accidental changes to their DNA—which accumulate and are passed on to progeny cells when the mutated cells divide—the chance that a cell will become cancerous increases greatly with age. (Data from C. Muir et al., Cancer Incidence in Five Continents, Vol. V. Lyon: International Agency for Research on Cancer, 1987.)

A Record of the Fidelity of DNA Replication and Repair Is Preserved in Genome Sequences

Although the majority of mutations do neither harm nor good to an organism, those that have severely harmful consequences are usually eliminated through natural selection; individuals carrying the altered DNA may die or experience decreased fertility, in which case these changes will be gradually lost from the population. By contrast, favorable changes will tend to persist and spread.

But even where no selection operates—at the many sites in the DNA where a change of nucleotide has no effect on the fitness of the organism—the genetic message has been faithfully preserved over tens of millions of years. Thus humans and chimpanzees, after about 5 million years of divergent evolution, still have DNA sequences that are at least 98% identical. Even humans and whales, after 10 or 20 times this amount of time, have chromosomes that are unmistakably similar in their DNA sequence (**Figure 6–34**). Thus our genome—and those of our relatives—contains a message from the distant past. Thanks to the faithfulness of DNA replication and repair, 100 million years of evolution have scarcely changed its essential content.

whale GTGTGGTCTCGTGATCAAAGGCGAAAGGTGGCTCTAGAGAATCCC
human GTGTGGTCTCGCGATCAGAGGCGCAAGATGGCTCTAGAGAATCCC

Figure 6–34 The sex-determination genes from humans and whales are noticeably similar. Despite the many millions of years that have passed since humans and whales diverged from a common ancestor, the nucleotide sequences of many of their genes remain closely related. The DNA sequences of a part of the gene that determines maleness in both humans and whales are lined up, one above the other; the positions where the two sequences are identical are shaded in *gray*.

ESSENTIAL CONCEPTS

- Before a cell divides, it must accurately replicate the vast quantity of genetic information carried in its DNA.
- Because the two strands of a DNA double helix are complementary, each strand can act as a template for the synthesis of the other. Thus DNA replication produces two identical, double-helical DNA molecules, enabling genetic information to be copied and passed on from a cell to its daughter cells and from a parent to its offspring.
- During replication, the two strands of a DNA double helix are pulled apart at a replication origin to form two Y-shaped replication forks. DNA polymerases at each fork produce a new, complementary DNA strand on each parental strand.
- DNA polymerase replicates a DNA template with remarkable fidelity, making only about one error in every 10^7 nucleotides copied. This accuracy is made possible, in part, by a proofreading process in which the enzyme corrects its own mistakes as it moves along the DNA.
- Because DNA polymerase synthesizes new DNA in the 5′-to-3′ direction, only the leading strand at the replication fork can be synthesized in a continuous fashion. On the lagging strand, DNA is synthesized in a discontinuous backstitching process, producing short fragments of DNA that are later joined together by DNA ligase.
- DNA polymerase is incapable of starting a new DNA strand from scratch. Instead, DNA synthesis is primed by an RNA polymerase called primase, which makes short lengths of RNA primers that are then elongated by DNA polymerase. These primers are subsequently removed and replaced with DNA.
- DNA replication requires the cooperation of many proteins that form a multienzyme replication machine that pries open the double helix and copies the information contained in both DNA strands.
- In eukaryotes, a special enzyme called telomerase replicates the DNA at the ends of the chromosomes, particularly in rapidly dividing cells.
- The rare copying mistakes that escape proofreading are dealt with by mismatch repair proteins, which increase the accuracy of DNA replication to one mistake per 10^9 nucleotides copied.
- Damage to one of the two DNA strands, caused by unavoidable chemical reactions, is repaired by a variety of DNA repair enzymes that recognize damaged DNA and excise a short stretch of the damaged strand. The missing DNA is then resynthesized by a repair DNA polymerase, using the undamaged strand as a template.
- If both DNA strands are broken, the double-strand break can be rapidly repaired by nonhomologous end joining. Nucleotides are often lost in the process, altering the DNA sequence at the repair site.
- Homologous recombination can flawlessly repair double-strand breaks (and many other types of DNA damage) using an undamaged homologous double helix as a template.
- Highly accurate DNA replication and DNA repair processes play a key role in protecting us from the uncontrolled growth of somatic cells known as cancer.

KEY TERMS

cancer
DNA ligase
DNA polymerase
DNA repair
DNA replication
homologous recombination
lagging strand
leading strand
mismatch repair
mutation
nonhomologous end joining
Okazaki fragment
primase
proofreading
replication fork
replication origin
RNA (ribonucleic acid)
telomerase
telomere
template

QUESTIONS

QUESTION 6–5

DNA mismatch repair enzymes preferentially repair bases on the newly synthesized DNA strand, using the old DNA strand as a template. If mismatches were simply repaired without regard for which strand served as template, would this reduce replication errors as effectively? Explain your answer.

QUESTION 6–6

Suppose a mutation affects an enzyme that is required to repair the damage to DNA caused by the loss of purine bases. The loss of a purine occurs about 5000 times in the DNA of each of your cells per day. As the average difference in DNA sequence between humans and chimpanzees is about 1%, how long will it take you to turn into an ape? Or would this transformation be unlikely to occur?

QUESTION 6–7

Which of the following statements are correct? Explain your answers.

A. A bacterial replication fork is asymmetrical because it contains two DNA polymerase molecules that are structurally distinct.

B. Okazaki fragments are removed by a nuclease that degrades RNA.

C. The error rate of DNA replication is reduced both by proofreading by DNA polymerase and by DNA mismatch repair.

D. In the absence of DNA repair, genes become less stable.

E. None of the aberrant bases formed by deamination occur naturally in DNA.

F. Cancer can result from the accumulation of mutations in somatic cells.

QUESTION 6–8

The speed of DNA replication at a replication fork is about 100 nucleotides per second in human cells. What is the minimum number of origins of replication that a human cell must have if it is to replicate its DNA once every 24 hours?

Recall that a human cell contains two copies of the human genome—one inherited from the mother, the other from the father—each consisting of 3×10^9 nucleotide pairs.

QUESTION 6–9

Look carefully at Figure 6–11 and at the structures of the compounds shown in Figure Q6–9.

A. What would you expect if ddCTP were added to a DNA replication reaction in large excess over the concentration of the available dCTP, the normal deoxycytidine triphosphate?

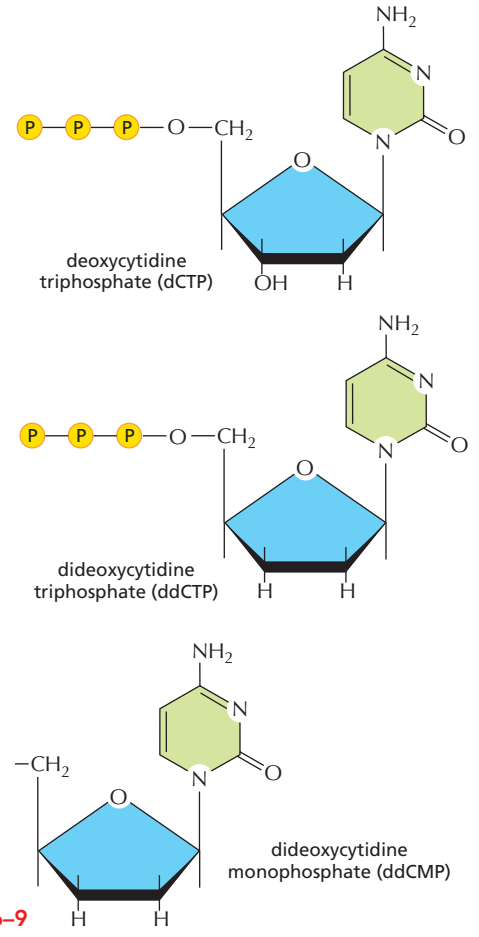


Figure Q6–9

B. What would happen if it were added at 10% of the concentration of the available dCTP?

C. What effects would you expect if ddCMP were added under the same conditions?

QUESTION 6–10

Figure Q6–10 shows a snapshot of a replication fork in which the RNA primer has just been added to the lagging strand. Using this diagram as a guide, sketch the path of the DNA as the next Okazaki fragment is synthesized. Indicate the sliding clamp and the single-strand DNA-binding protein as appropriate.

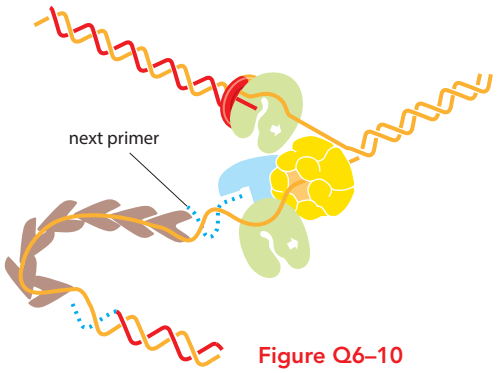


Figure Q6–10

QUESTION 6–11

Approximately how many high-energy bonds does DNA polymerase use to replicate a bacterial chromosome (ignoring helicase and other enzymes associated with the replication fork)? Compared with its own dry weight of 10^{-12} g, how much glucose does a single bacterium need to provide enough energy to copy its DNA once? The number of nucleotide pairs in the bacterial chromosome is 3×10^6. Oxidation of one glucose molecule yields about 30 high-energy phosphate bonds. The molecular weight of glucose is 180 g/mole. (Recall from Figure 2–3 that a mole consists of 6×10^{23} molecules.)

QUESTION 6–12

What, if anything, is wrong with the following statement: "DNA stability in both reproductive cells and somatic cells is essential for the survival of a species." Explain your answer.

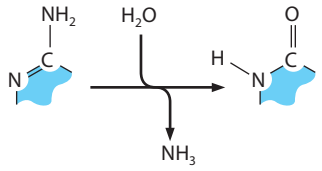


Figure Q6–13

QUESTION 6–13

A common type of chemical damage to DNA is produced by a spontaneous reaction termed *deamination*, in which a nucleotide base loses an amino group (NH_2). The amino group is replaced with a keto group (C=O) by the general reaction shown in Figure Q6–13. Write the structures of the bases A, G, C, T, and U and predict the products that will be produced by deamination. By looking at the products of this reaction—and remembering that, in the cell, these will need to be recognized and repaired—can you propose an explanation for why DNA does not contain uracil?

QUESTION 6–14

A. Explain why telomeres and telomerase are needed for replication of eukaryotic chromosomes but not for replication of circular bacterial chromosomes. Draw a diagram to illustrate your explanation.

B. Would you still need telomeres and telomerase to complete eukaryotic chromosome replication if primase always laid down the RNA primer at the very 3′ end of the template for the lagging strand?

QUESTION 6–15

Describe the consequences that would arise if a eukaryotic chromosome:

A. contained only one origin of replication:

   (i) at the exact center of the chromosome.

   (ii) at one end of the chromosome.

B. lacked telomeres.

C. lacked a centromere.

Assume that the chromosome is 150 million nucleotide pairs in length, a typical size for an animal chromosome, and that DNA replication in animal cells proceeds at about 100 nucleotides per second.