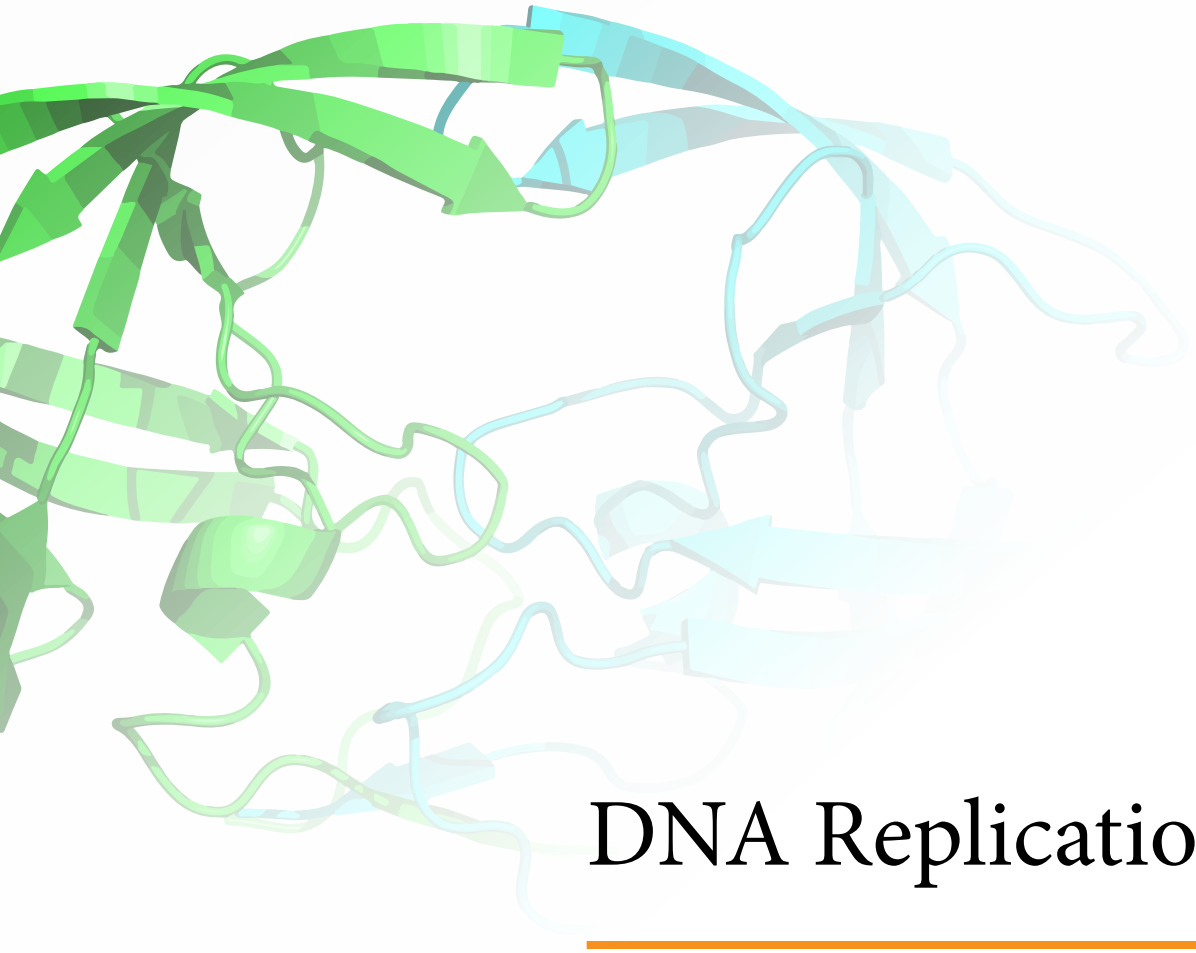

9

DNA Replication

Goal

To understand the chemistry of DNA synthesis and how DNA is replicated with high accuracy.

Objectives

After this chapter, you should be able to

- describe the experiment that proved that DNA replication is semi-conservative.
- diagram the reaction for phosphodiester bond formation.
- explain the energetics of DNA synthesis.
- explain why the 5'-to-3' rule creates a conundrum during replication.
- explain how DNA is replicated accurately.
- explain how and why damage to DNA is repaired.

Living cells must be able to duplicate their entire set of genetic instructions every time they divide. Likewise, multicellular organisms must be able to pass on complete copies of their genetic information to future generations. This requires that the instructions are stored in a form that is capable of being duplicated. As we have seen (and will return to in Chapter 11), genetic instructions are embedded in the order of nucleobases in the DNA. As we have also seen, the self-complementary nature of the double helix provides a simple (in principle) templating mechanism for replicating DNA into two identical copies. Only DNA (and in some instances RNA when, as in the case of the genomes of RNA viruses, it is used as a repository of genetic information) are self-complementary and hence capable of being replicated. The other three categories of macromolecules—carbohydrates, lipids, and proteins—are not self-complementary and do not serve as templates for their own production. Instead, the synthesis of these macromolecules is ultimately directed by information stored in DNA through the action of enzymes and other proteins. Here we focus on the chemical and enzymatic mechanisms by which DNA acts as a template for its own duplication and how this replication process is carried out accurately and rapidly.

DNA replication is semi-conservative

As we saw in Chapter 8, the self-complementarity of DNA suggested a mechanism for its replication. Specifically, Watson and Crick imagined that base pairing would make it possible for each polynucleotide strand of the double helix to serve as a template for the synthesis of a new strand.

That is, and as depicted in Figure 10 of Chapter 8, the parental double helix would separate into two strands, each of which would serve as a template for the synthesis of a new, complementary strand. In this scenario, known as **semi-conservative replication**, each daughter molecule is a hybrid helix in which one strand is parental (*conserved* from the parental helix) and the other newly synthesized. At the same time, an alternative mechanism known as **conservative replication** was a formal possibility. In this scenario the original double helix is left intact (both of its original strands are conserved) and an entirely new helix is somehow generated that consists of two newly synthesized strands. Certainly, semi-conservative replication was more appealing because it was hard to imagine a mechanism by which conservative replication could take place. But this didn't mean that DNA is not replicated by a conservative mechanism. How then was it determined which of the two models is correct?

In 1958, Matthew Meselson and Franklin Stahl reported an experiment that elegantly distinguished between semi-conservative and conservative replication (Figures 1 and 2). Thanks to its simplicity and decisiveness, the Meselson-Stahl experiment has been called "the most beautiful experiment in biology." The key to the Meselson-Stahl experiment was growing cells of *Escherichia coli* in medium containing a heavy isotope of nitrogen, ^{15}N, for many generations. *E. coli* needs nitrogen as a nutrient and incorporates the isotope from the growth medium into its proteins and nucleic acids.

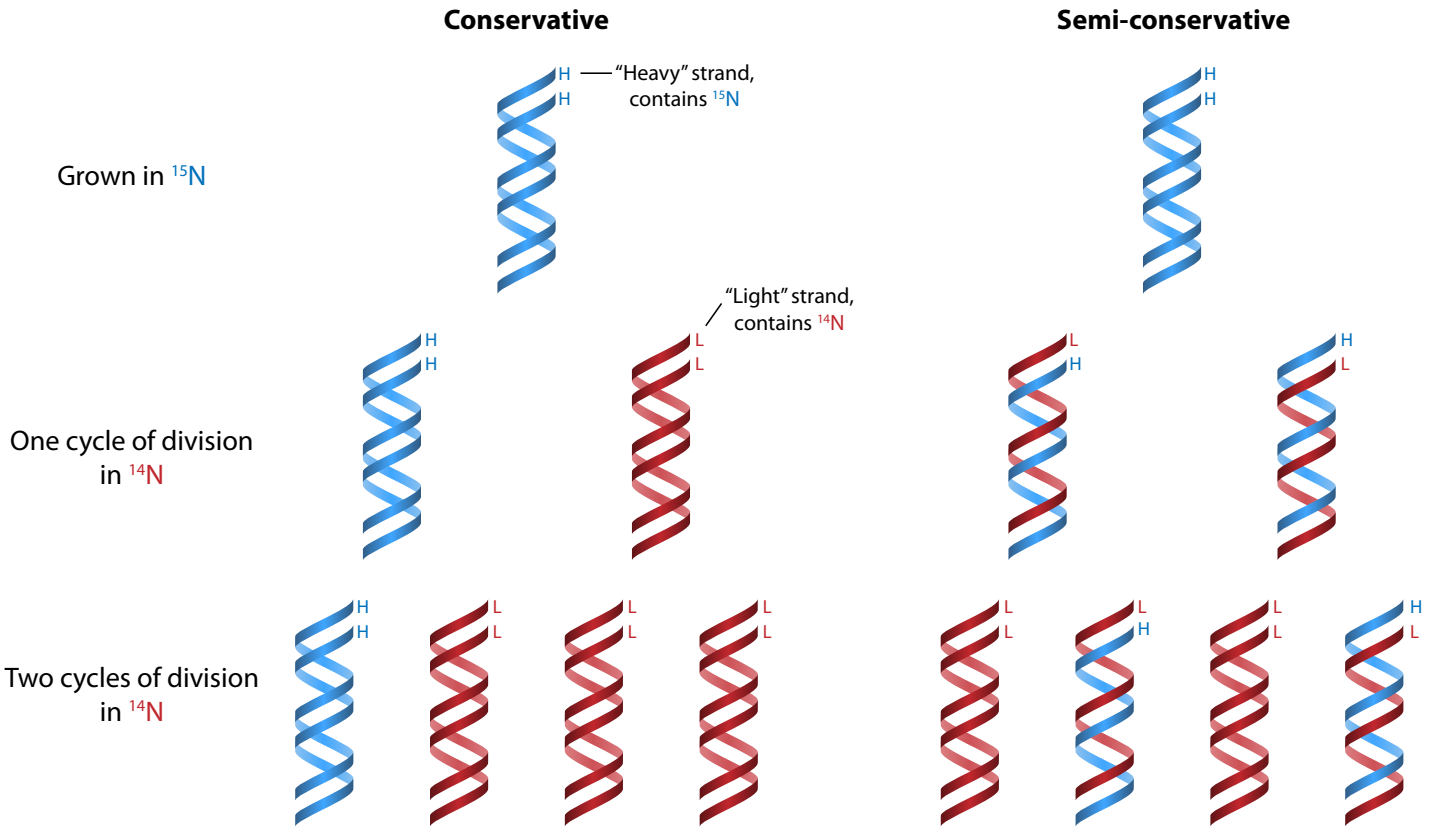

Figure 1 The Meselson-Stahl experiment showed that DNA replication is semi-conservative

Cells were grown in medium with the heavy isotope of nitrogen (^{15}N), which led to the synthesis of "heavy" DNA strands, shown in blue. These cells were then grown for either one or two division cycles in medium containing the common isotope of nitrogen (^{14}N). Newly synthesized "light" DNA strands, shown in red, are produced using the ^{14}N.

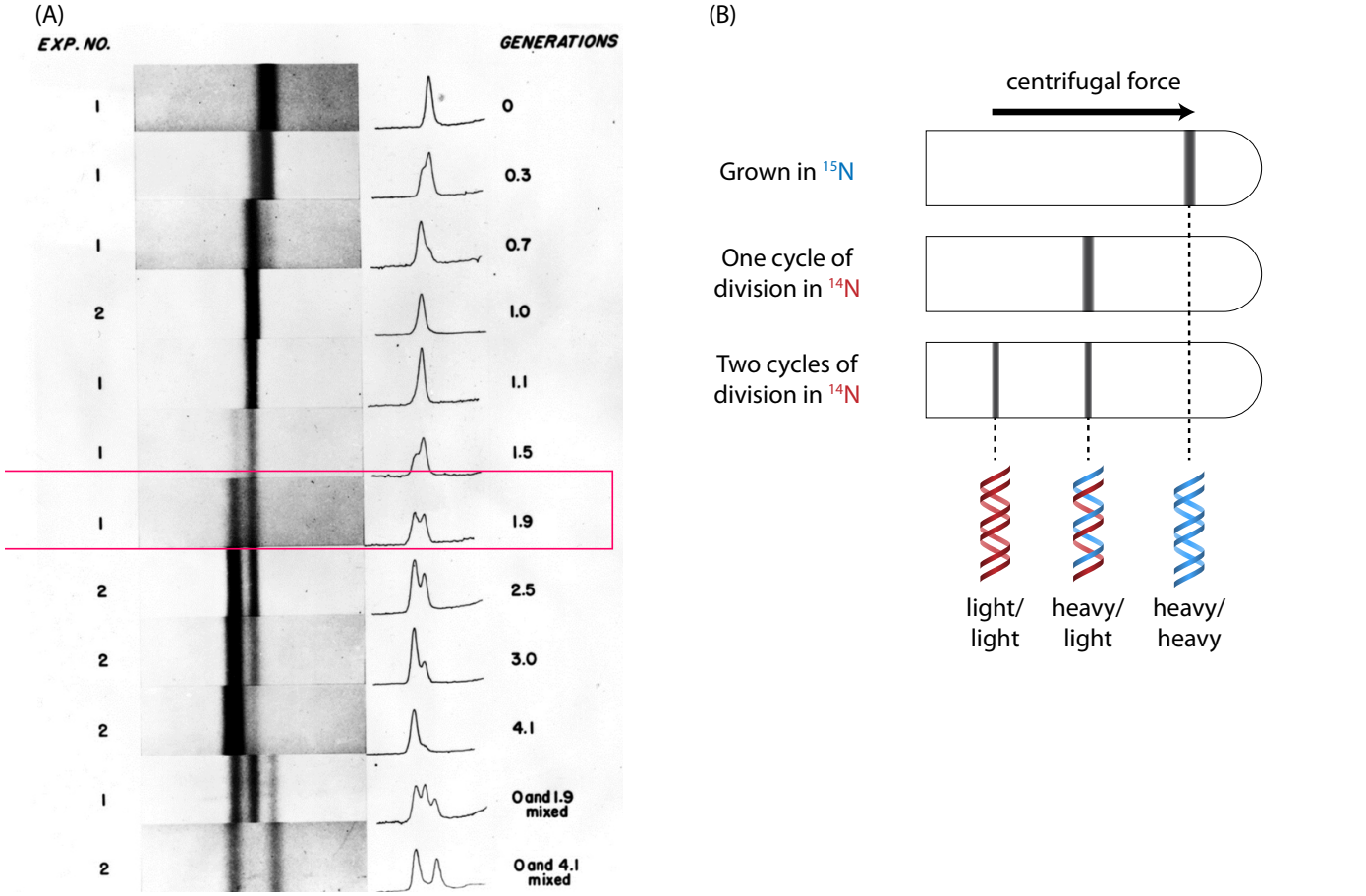

Figure 2 The Meselson-Stahl experiment analyzed the isotopic composition of DNA by measuring its buoyant density

Shown in (A) are the original experiments from the publication of Meselson and Stahl [*Proc. Natl. Acad. Sci.* 44: 671 (1958)] and in (B) a cartoon that depicts how centrifugation separates DNA molecules according to their bouyant density in a cesium gradient. The dark vertical bands in the left panel of (A) show the positions of the DNA molecules that were visualized by staining. The column of adjacent plots show quantitative distributions of the DNA. DNA was extracted from cells after the indicated number of generations (rounds of cell division) after transfer from growth in medium with ^{15}N to medium with ^{14}N (from generation 0 to generation 4.1). The red rectangle highlights the densities after approxiately two (generation 1.9) rounds of division.

Next, they transferred the bacteria for various periods of time to medium containing ^{14}N (the common form of nitrogen with an atomic mass of 14) such that any DNA synthesized after the transfer would not contain the heavy nitrogen isotope. Finally, they extracted DNA from the cells and subjected the DNA to high centrifugal forces in tubes containing cesium chloride (using an instrument called an ultracentrifuge, in which tubes are spun at high velocity). Cesium forms a density gradient under such conditions; at equilibrium, DNA molecules in the gradient localize to a position that matches their own buoyant density. Thus, parental DNA in which both strands contain the heavy isotope would have a high density (blue-blue in Figures 1 and 2) and localize near the bottom of the gradient. DNA in which both strands are newly synthesized and hence light (red-red) would localize near the top. Finally, hybrid molecules (blue-red) in which one strand was heavy and one light would localize at an intermediate position.

What did Meselson and Stahl observe? When *E. coli* cells were shifted to ^{14}N-containing medium for enough time for the DNA molecules to undergo

one round of replication, the DNA was found to exhibit an intermediate buoyant density consistent with that expected for hybrid DNA instead of the high density of the parental, heavy-heavy DNA (Figure 2). This fit with the expectation from the semi-conservative model that each strand was serving as a template for the other strand. Next, Meselson and Stahl went a step further: they extracted DNA from cells that had been grown for two cycles of division in the light, ^{14}N-medium. Half of the DNA from these cells exhibited the same buoyant density as that seen after a single round of division (halfway between the densities of ^{14}N- and ^{15}N-labeled DNA), and the other half exhibited a density corresponding to that of DNA that only contained ^{14}N (Figure 2). Thus after two rounds of replication, DNA was generated in which both strands were newly synthesized. *In toto*, these results met the expectation from a model in which parental DNA contributes one of its two strands to each of the double helices generated during each round of replication but were inconsistent with a model in which replication generates a double helix in which both strands are newly synthesized. Ergo, DNA replication is **semi-conservative**.

DNA is synthesized by the repetitive addition of nucleotides to the 3' end of the growing polynucleotide chain

DNA is synthesized by an iterative process in which nucleotides are added (one after another) to the 3' end of a growing polynucleotide chain. This growing strand, which we also refer to as the **daughter** strand, is base paired in anti-parallel orientation to a second strand, the **template** strand, which provides instructions via base pairing for the successive addition of nucleotides (Figure 3). Each incoming nucleotide must properly pair with the corresponding base in the template in order for a phosphodiester bond to form and for the process to proceed on to the addition of the next nucleotide. Because nucleotides are added at the 3' end of the growing strand, DNA synthesis proceeds in a 5'-to-3' direction.

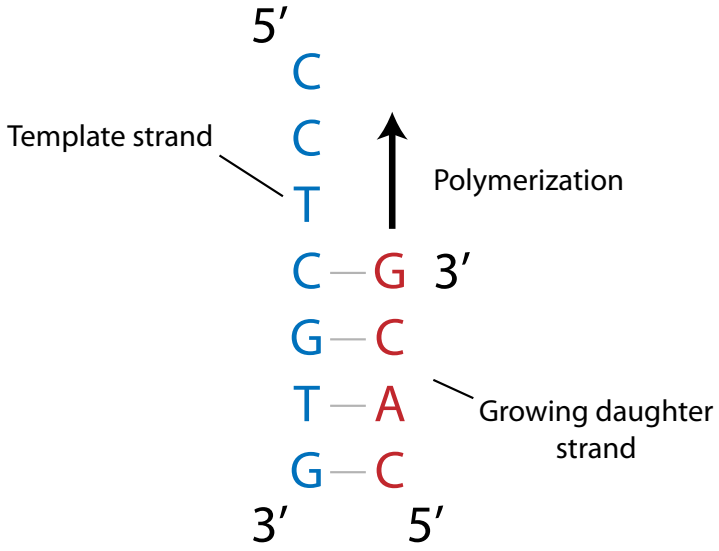

Figure 3 **DNA synthesis proceeds in a 5'-to-3' direction**

During DNA synthesis one strand of the helix serves as a template to specify incoming nucleotides that are added at the 3' end of the growing daughter strand.

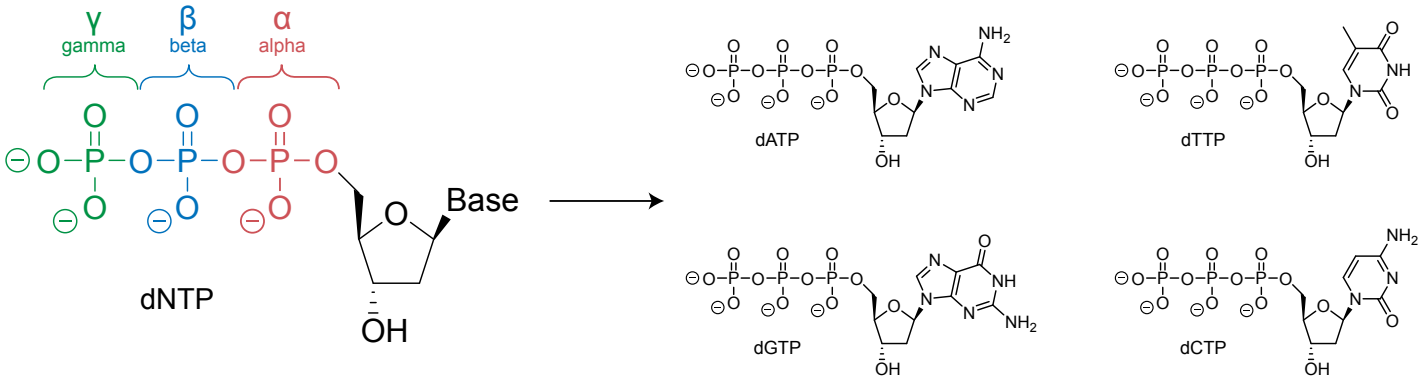

Figure 4 **2'-deoxynucleoside triphosphates are substrates for DNA replication**

The nucleotide substrates for DNA synthesis are 2'-deoxynucleoside triphosphates

The substrates for DNA synthesis are **2'-deoxynucleoside triphosphates** (Figure 4). As we learned in Chapter 8, the term nucleoside refers to a base and a sugar, in this case the sugar 2'-deoxyribose. The substrates for DNA synthesis are 2'-deoxynucleosides that additionally carry a chain of three phosphates at the 5' carbon of the nucleoside. Hence these substrates can be referred to as 2'-deoxynucleo*side triphosphates* or more simply as 2'-deoxynucleo*tides* (in that they consist of a base, a sugar, and phosphates) or even more simply as nucleotides, a shorthand that is frequently used. The three phosphate groups in 2'-deoxynucleoside triphosphates are designated using Greek letters as alpha (**α**), beta (**β**), and gamma (**γ**), with the phosphate that is directly attached to the sugar being the α-phosphate, the middle phosphate being the β-phosphate, and the phosphate most distal to the sugar being the γ-phosphate. As we will explain, the β- and γ-phosphates are released from the 2'-deoxynucleoside triphosphate substrates during the chemical reaction in which a 2'-deoxynucleotide is added to the 3' end of a growing chain; that is, only the α-phosphate is retained in the growing chain.

Henceforth, for simplicity we will use the abbreviations **dNMP**, **dNDP**, and **dNTP** for 2'-deoxynucleosides bearing one (mono), two (di) and three (tri) phosphates, respectively.

Polymerization of 2'-deoxynucleoside triphosphates involves the release of the β- and γ-phosphates as pyrophosphate

The addition of a dNTP (2'-deoxynucleoside triphosphate) to the 3' end of the growing strand during DNA synthesis involves the release of **pyrophosphate** (Figure 5). Pyrophosphate is a diphosphate ($P_2O_7^{4-}$, abbreviated **PPi**), and in DNA synthesis it is derived from the β- and γ-phosphates. Pyrophosphate is released from the substrate when a covalent bond is formed between the 3' oxygen of the growing polynucleotide and the α-phosphate of the incoming nucleotide. Pyrophosphate is a leaving group (Chapter 4) for the DNA synthesis reaction.

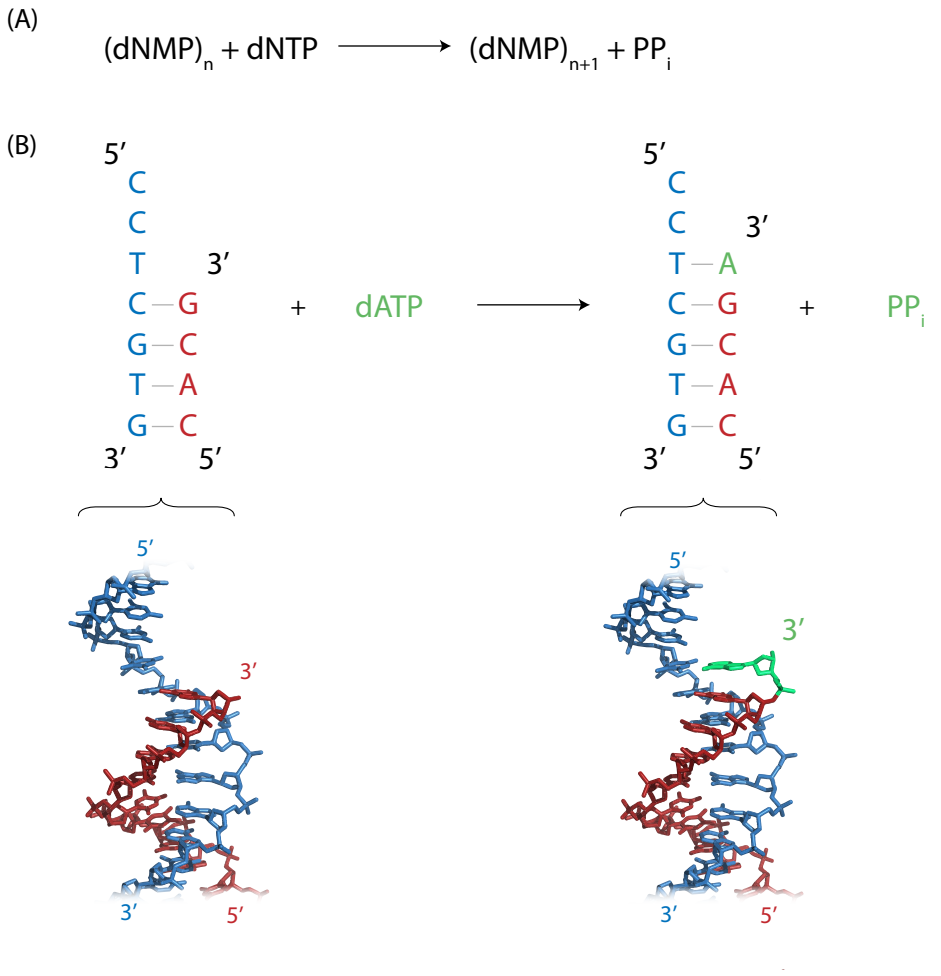

Figure 5 Each step of DNA synthesis consumes a dNTP molecule and releases a pyrophosphate molecule

(A) The chemical reaction for DNA synthesis in which the subscript "n" represents the length of the growing daughter strand $(dNMP)_n$. (B) A new nucleotide is added to the 3' end of a growing DNA strand (*red*) during each DNA synthesis reaction. The identity of the nucleotide is specified by complementarity with the template strand (*blue*). The incoming nucleotide is incorporated from a dNTP (*green*), and the addition of the nucleotide to the growing strand releases a molecule of pyrophosphate (PPi).

Phosphodiester bonds are formed by nucleophilic attack of the 3' oxygen on the α-phosphate of the incoming dNTP

The mechanism for the DNA polymerization reaction is shown in Figure 6. The terminal 3' oxygen of the growing strand acts as a nucleophile and forms a covalent bond with the phosphorus atom of the α-phosphate group of the incoming dNTP. Because phosphorus cannot form six bonds, it must lose one of its bonds, namely the bond to the oxygen that connects it to the β- and γ-phosphates, resulting as we have seen in the release of pyrophosphate.

Notice also in Figure 6, and as introduced above, that the incoming nucleotide, dATP (*green*), is specified by base pairing with a thymine (*blue*) at the corresponding position on the complementary template strand and that the 3' oxygen that acts as a nucleophile arises from a dGMP (*red*) that had been incorporated in the previous round of phosphodiester bond formation by pairing with a cytosine in the template strand.

DNA polymerization is coupled to pyrophosphate hydrolysis

DNA synthesis is catalyzed by an enzyme called DNA polymerase. It turns out that the reaction catalyzed by DNA polymerase—the formation of a phosphodiester bond and the release of pyrophosphate—is only modestly favorable (Figure 7, Reaction 1). Yet, the cell needs to replicate DNA

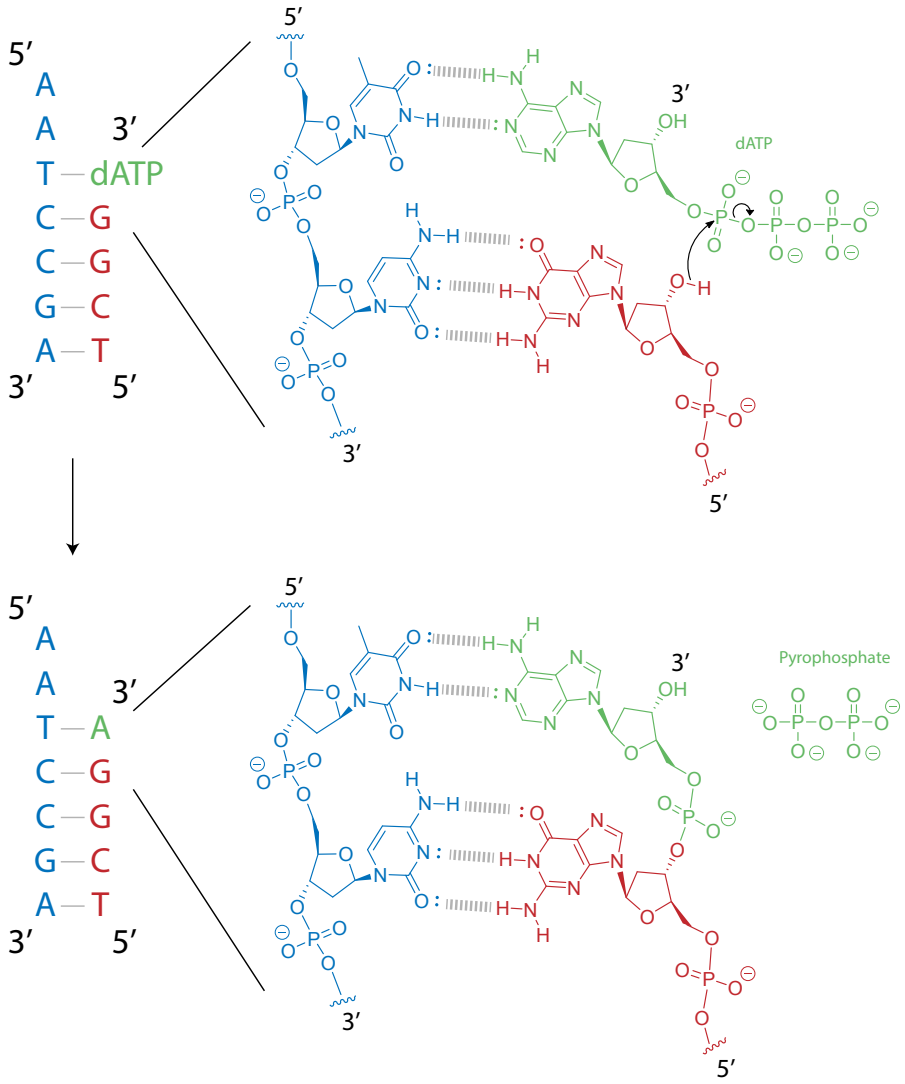

Figure 6 The 3' oxygen acts as a nucleophile during DNA polymerization

The arrow-pushing mechanism for DNA polymerization is shown. The nucleophile is the 3' oxygen at the 3' end of the growing DNA strand. Electrons from this oxygen are used to form a new bond with the phosphorus atom of the α-phosphate group of the incoming dNTP. The β- and γ-phosphate groups leave together, forming a molecule of pyrophosphate.

efficiently and completely. How then does the cell succeed in replicating its genetic material if the fundamental step in DNA synthesis is not energetically highly favorable? The answer is that phosphodiester bond formation is coupled to a second, subsequent reaction in which the newly released pyrophosphate is hydrolyzed to give rise to two molecules of inorganic phosphate (PO_4^{3-} or Pi) by the enzyme pyrophosphatase (Figure 7, Reaction 2). The hydrolysis of pyrophosphate is also moderately favorable. However, in sum, the thermodynamics of these two successive reactions is highly favorable (Figure 7, combined reaction). That is, the sum of the $\Delta G°_{rxn}$ for

Reaction 1	$(dNMP)_n + dNTP \longrightarrow (dNMP)_{n+1} + PP_i$	$\Delta G°_{rxn} = -3.5$ kcal/mol
Reaction 2	$PP_i + H_2O \longrightarrow 2\ P_i$	$\Delta G°_{rxn} = -4.0$ kcal/mol
Combined reaction	$(dNMP)_n + dNTP + H_2O \longrightarrow (dNMP)_{n+1} + 2\ P_i$	$\Delta G°_{rxn} = -7.5$ kcal/mol

Figure 7 The favorability of DNA polymerization is enhanced by the hydrolysis of pyrophosphate

Reactions 1 and 2 equals −7.5 kcal/mol. Using the equation that we learned in Chapter 3 relating free energy to the equilibrium constant ($\Delta G°_{rxn} = -RT \ln K_{eq}$), we can calculate that the K_{eq} for the combined reaction is about 300,000, meaning that the products of the reaction are much more stable than the reactants.

A simple mnemonic for quickly carrying out such calculations is that a $\Delta G°_{rxn}$ of ~−2.7 kcal/mol is roughly equivalent to a K_{eq} of ~10^2. Ergo, −7.5 kcal/mol would correspond to an equilibrium constant of $(10^2)^{(7.5/2.7)}$ or ~3 x 10^5. Because many reactions in living systems have a $\Delta G°_{rxn}$ of only a few kilocalories/mole, this shorthand is a simple and convenient way to estimate equilibrium constants.

Finally, we note that this coupling of phosphodiester bond formation to pyrophosphate hydrolysis is a striking example of Le Châtelier's principle (analogous to the case of photosynthesis we considered in Box 1 of Chapter 3) operating in a living system. By depleting the cell of pyrophosphate, pyrophosphatase perturbs the equilibrium for Reaction 1, draining pyrophosphate and driving the reaction further in the direction of phosphodiester bond formation.

DNA synthesis takes place at moving replication forks

DNA replication requires that the two strands of the helix are pulled apart and that both are copied into daughter strands. This strand separation creates a moving **replication fork** downstream of which the DNA remains wound in a double helix and upstream of which the two strands are separated and are each serving as a template for the synthesis of daughter strands. It is a *moving* replication fork (as we will see, it moves very fast!) in that downstream DNA is continuously being unwound and upstream DNA is continuously being replicated into hybrid helices of old and new DNA strands.

But this creates a conundrum! If DNA only grows in a 5'-to-3' direction and if the two strands of the double helix have an anti-parallel orientation, then how can DNA be copied continuously on both template strands without violating one of the two fundamental rules of nucleic acid chemistry?

DNA synthesis at the replication fork is continuous and discontinuous

The answer to the conundrum is that DNA is copied continuously on one template strand and discontinuously on the other (Figure 8). One template strand is oriented such that the 5'-to-3' direction in which the daughter strand synthesis is taking place is aligned with the direction of fork movement; that is, it points toward the base of the fork where the helix is being unwound and where fresh, single-stranded template DNA is being generated. This is known as **leading strand synthesis**, and it take places in a continuous manner.

DNA copied from the other template strand, on the other hand, is discontinuous. It occurs in short bursts that point away from the fork. This is known as **lagging strand synthesis**. Lagging strand synthesis takes place in a 5'-to-3' direction. So the rule that polynucleotide synthesis proceeds

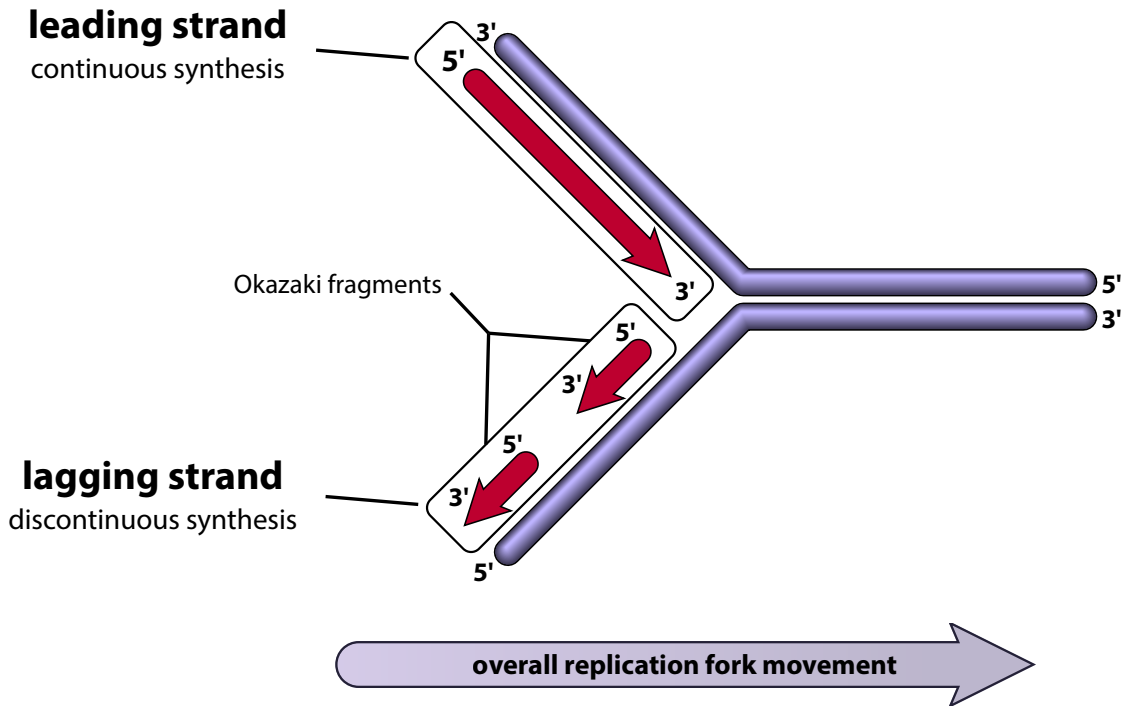

Figure 8 **The leading strand is synthesized continuously, but the lagging strand is not**

5' to 3' is not violated. Instead, short stretches of DNA are synthesized and then joined to each other by an enzyme known as a **DNA ligase** to create long, continuous DNA strands. A ligase is an enzyme that joins two polynucleotide chains together by creating a phosphodiester bond between the 5' end of one chain and the 3' end of the other.

The short stretches of DNA generated during lagging strand synthesis are commonly known as **Okazaki fragments** after their discoverer Reiji Okazaki. In bacteria, Okazaki fragments are 1,000-2,000 nucleotides in length, whereas in the cells of higher organisms they are typically only 100-200 nucleotides long.

To sum up, as the replication fork proceeds in the downstream direction, additional single-stranded templates are exposed for leading strand synthesis and lagging strand synthesis. Both take place in a 5'-to-3' direction. But in the case of lagging strand synthesis, this synthesis is discontinuous, with each newly synthesized segment of DNA being joined to the previously synthesized segments by ligase to create an intact daughter strand.

DNA polymerase catalyzes template-directed DNA synthesis

As we learned in Chapter 4, thermodynamically favorable reactions in the cell often require a protein catalyst in order for the reaction to proceed rapidly. This is because even highly favorable reactions often must overcome a kinetic barrier before the reactants can proceed on to products in a time frame compatible with life. In other words, many biological reactions have a high activation energy or $\Delta G^{\ddagger}$.

DNA polymerase is the enzyme that accelerates the rate of phosphodiester bond formation between the 3' end of a polynucleotide chain and a dNTP

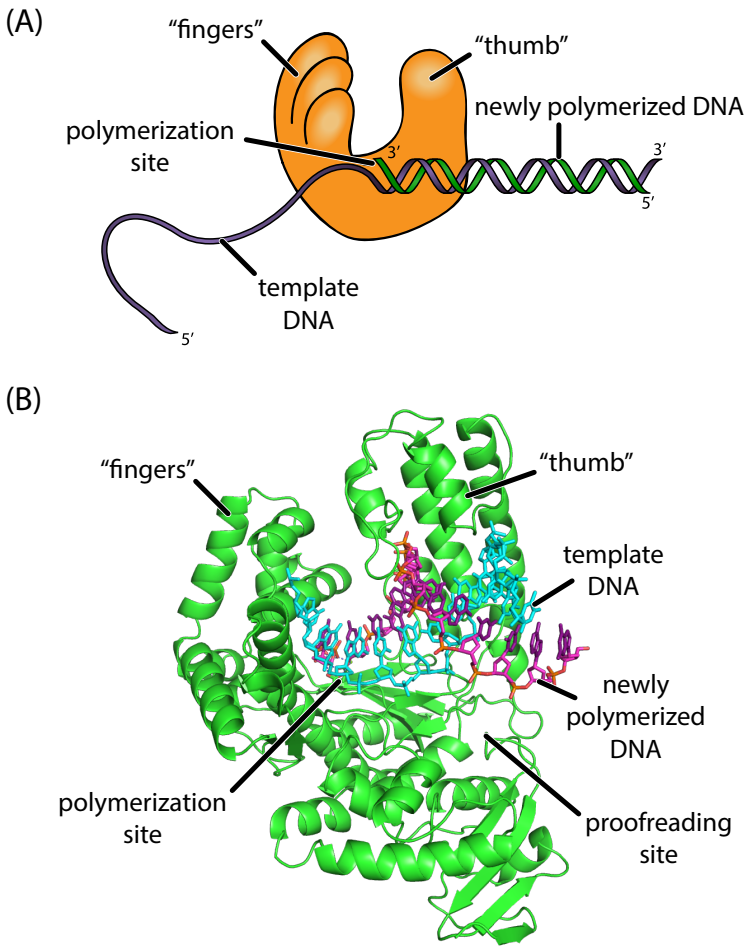

Figure 9 DNA polymerase catalyzes DNA polymerization

(A) Cartoon diagram of DNA polymerase, showing the fingers, thumb, and polymerization site. (B) Crystal structure of DNA polymerase bound to a piece of DNA. The fingers, thumb, and polymerization site are indicated. The location of the proofreading site or editing pocket, which we will discuss shortly, is also shown.

molecule. The substrate for DNA polymerase is a polynucleotide strand annealed to a longer template strand. The polynucleotide provides the 3' hydroxyl at which the dNTP is attached. The template strand, which extends past the 3' hydroxyl, specifies the specific nucleotide that will be appended to the growing strand through base pairing (as we have seen).

Figure 9B shows the atomic structure of DNA polymerase as determined by X-ray crystallography. Roughly speaking, it resembles a hand in which the polymerization site or catalytic center is in the palm and is surrounded by thumb-like and finger-like features of the enzyme (Figure 9A). The 3' end of the growing DNA strand is located near this catalytic center during replication. The catalytic center is formed from a set of two to three aspartate residues that bind two magnesium ions and hold them in specific positions (Figure 10). These magnesium ions accelerate the polymerization reaction by making the α-phosphate more electrophilic and the 3' hydroxyl group more nucleophilic. The magnesium ions bind the triphosphate portion of the dNTP and thereby remove some of the negative charge that otherwise shields the phosphorus atom of the α-phosphate from incoming nucleophiles. In addition, one of the magnesium ions forms an ion-dipole interaction with the 3' hydroxyl group at the 3' end of the growing strand. By donating some of its electron density to form this interaction, the oxygen atom of this 3' hydroxyl group becomes more positive. To counter this withdrawal of electrons, oxygen, as the more electronegative partner, attracts the electrons it shares with hydrogen even more strongly than

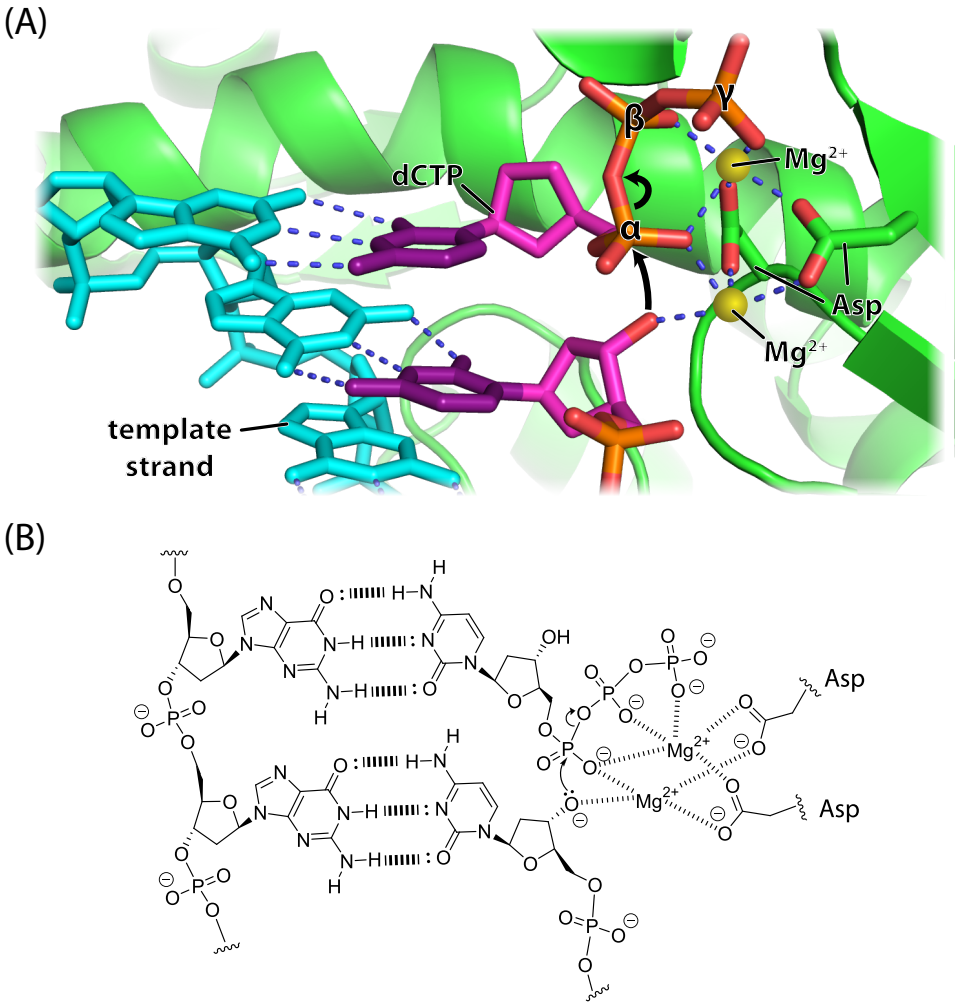

Figure 10 The active site of DNA polymerase accelerates the rate of the polymerization reaction

(A) Shown is the X-ray crystal structure of DNA polymerase showing the polymerization active site. The template strand is shown in cyan, the growing strand is shown in magenta, and DNA polymerase is shown in green. (B) Shown is a two-dimensional representation of the structure shown in (A).

usual, ultimately breaking the O-H bond and creating an oxygen anion. The additional negative charge carried by oxygen makes it a more effective nucleophile than it would have been had it been part of a hydroxyl group.

Proper alignment of the substrate dNTP with the oxygen anion requires that the incoming nucleotide pair with the corresponding base on the template strand. Conversely, an incoming nucleotide with a non-complementary base would be unable to align properly in the active site and hence fail to react with the oxygen anion. In other words, pairing with the template strand is intimately involved in the chemical reaction that promotes phosphodiester bond formation. Therefore, the ability of DNA polymerase to ensure that the correct, complementary nucleotide is incorporated at the 3' hydroxyl of the growing strand is embedded in the architecture of the active site of the enzyme.

In sum, DNA polymerase lowers the activation energy for phosphodiester bond formation (reduces $\Delta G^{\ddagger}$) by aligning the substrates for the reaction in its catalytic center and by promoting nucleophilic attack of the 3' oxygen on the α-phosphate of the incoming dNTP.

DNA polymerase is processive

Not only does DNA polymerase catalyze phosphodiester bond formation, but it also does so repeatedly. Indeed, the same DNA polymerase molecule

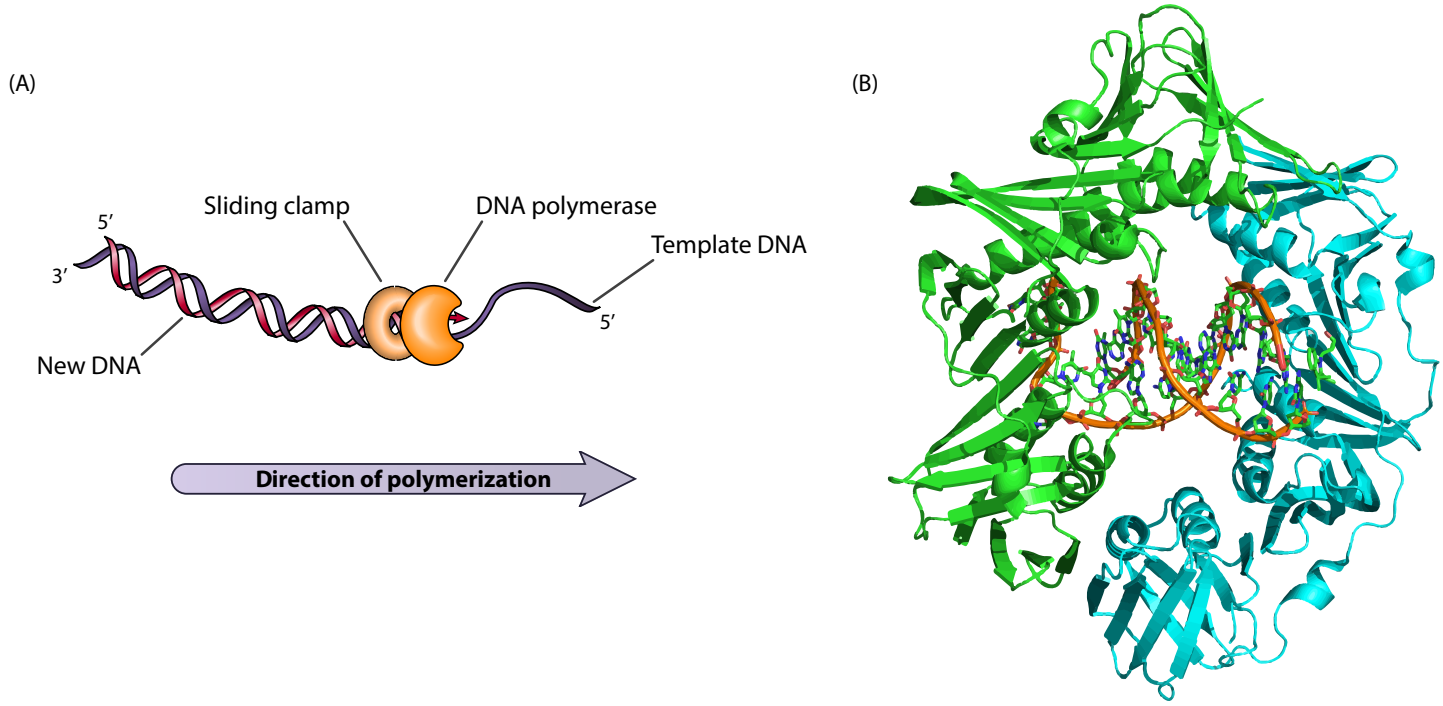

Figure 11 The sliding clamp allows DNA polymerase to be processive

(A) Cartoon representation of the ring-shaped sliding clamp surrounding the DNA helix and anchoring DNA polymerase to the DNA. (B) X-ray structure of a sliding clamp with a cartoon representation of DNA passing through the hole in the center. The clamp is composed of two proteins, which are shown in green and cyan.

successively attaches one nucleotide after another to the growing daughter strand much faster than it dissociates from the DNA. This property of remaining attached to the DNA through many rounds of nucleotide addition is referred to as **processivity**. Processivity maximizes the speed of DNA synthesis. If the polymerase frequently fell off the DNA and had to re-bind in order to resume nucleotide incorporation, the rate of DNA synthesis would be much slower.

What keeps the enzyme anchored to the DNA? The answer is geometrically elegant and simple: DNA polymerase is tethered to the DNA by a **sliding clamp** that fully encircles the DNA helix (Figure 11A). The sliding clamp is a complex of proteins that forms a complete circle around the DNA. In effect, it resembles a doughnut, with the DNA passing through the hole in the doughnut. The sliding clamp slides on the DNA along with the DNA polymerase to which it is attached and travels with it as DNA synthesis proceeds. Thus, the sliding clamp tethers the polymerase to the DNA by means of a topologically closed structure. The beautiful X-ray structure of the clamp in Figure 11B shows how the hole in the doughnut accommodates the DNA.

DNA polymerase and the sliding clamp are part of a larger complex of proteins that constitutes a DNA replication machine

DNA polymerase and the sliding clamp act in concert with a larger cohort of proteins that drive movement of the replication fork and coordinate DNA synthesis on the leading and lagging strands. For example, an enzyme

known as **DNA helicase** unwinds double-stranded DNA downstream of the replication fork to separate the two strands that will serve as templates for daughter strand synthesis. Another protein is involved in priming the synthesis of the Okazaki fragments on the lagging strand. DNA polymerase can only extend an existing polynucleotide chain. Bursts of DNA synthesis on the lagging strand require a special enzyme known as **primase** that produces a short stretch of nucleotides that can be extended by the DNA polymerase. A **clamp loader** then assembles a sliding clamp on the DNA behind the DNA polymerase that will produce the Okazaki fragment. And, as we have seen, a DNA ligase will merge the Okazaki fragments into an uninterrupted polynucleotide strand.

The entire set of proteins that mediate DNA replication can be thought of as a molecular machine with multiple working parts that acts in a coordinated fashion to bring about the replication of the genetic material. And it does so remarkably rapidly. The replication machine synthesizes DNA at a rate of 800 nucleotides per second. Consider that the chromosome of *E. coli* consists of almost five million base pairs. Because DNA replication in *E. coli* takes place simultaneously from two replication forks, the overall rate of DNA synthesis is 1,600 nucleotides per second. Thus, *E. coli* is capable of duplicating its entire chromosome in as little as 40 minutes, and as we will see below, it does so with considerable accuracy.

The replication machine is only one of many molecular machines that carry out the workings of the cell. We will encounter others below and in later chapters. The very basis for life can be thought of as an ensemble of molecular machines that carry out the chemical transactions of living systems.

DNA is replicated with high fidelity

An extraordinary feature of DNA synthesis is its accuracy; it achieves an error rate of only about one mistake for every 10^{10} nucleotides incorporated. How is this accuracy achieved? Part of the answer is that DNA polymerase incorporates correct nucleotides more effectively than incorrect nucleotides. This selectivity is dependent on DNA polymerase's ability to recognize correctly matched base pairs in its active site. DNA polymerase does not recognize specific base pairs; instead, it recognizes the geometry (i.e., shape and size) of the base pair (Figure 12). As we have seen, A:T and G:C base pairs are similar in structure and are approximately the same width. When a correctly matched base pair is present in the enzyme active site, the α-phosphate of the incoming nucleotide is optimally positioned for the nucleophilic attack of the 3' hydroxyl at the end of the growing strand. In contrast, a mismatched pair of nucleotides has a different shape, and as a consequence, the α-phosphate is not positioned properly to facilitate the nucleophilic attack of the 3' hydroxyl. Because the reactants are not optimally positioned, the reaction proceeds slowly for mismatched nucleotides; this results in a lag time that gives the mismatched nucleotide time to dissociate from the polymerase.

Nonetheless, DNA synthesis is not error-free, allowing about one

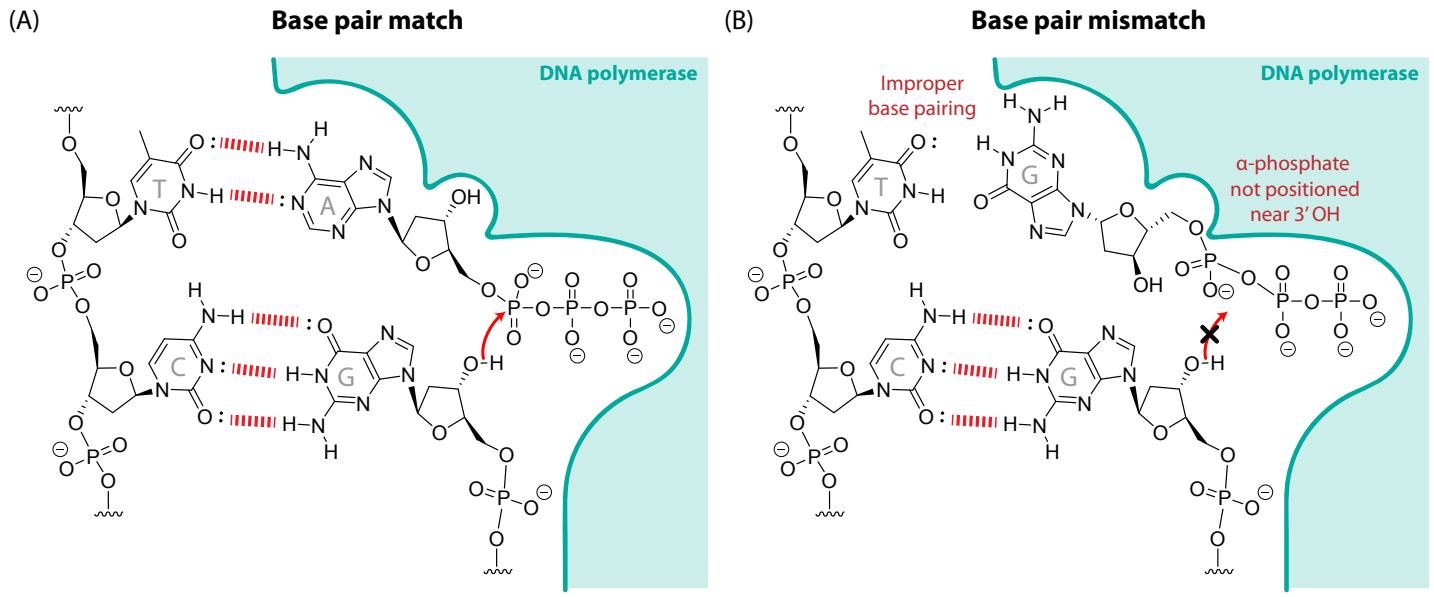

Figure 12 **DNA polymerase catalyzes the addition of correct base pairs more effectively than it catalyzes the addition of incorrect base pairs**

mistake for every 10^5 phosphodiester bonds formed. Therefore, the ability of DNA polymerase to recognize correctly paired nucleotides in its active site is insufficient to explain the extraordinary fidelity of DNA synthesis.

DNA polymerase removes misincorporated nucleotides by proofreading

DNA polymerase is able to achieve further accuracy by proofreading its own work. That is, if an incorrect nucleotide is incorporated, DNA polymerase is often capable of catching the error and removing the misincorporated nucleotide. Proofreading takes advantage of the fact that it is difficult for DNA polymerase to add a new nucleotide onto an improperly incorporated nucleotide at the 3' end of the growing chain. In other words, the nucleotide at the 3' end must be correctly paired with the template strand in order to be properly aligned for polymerization to proceed rapidly to the next nucleotide. If not correctly paired, the 3' nucleotide in the stalled polymerase is able to slip out of the catalytic center and into an editing pocket on the enzyme (Figure 9). The editing pocket contains a catalytic activity that removes the nucleotide at the 3' end of the daughter strand (a 3' nuclease). The resulting 3' end of the growing chain with the misincorporated nucleotide removed is able to slip back into the catalytic center, where a fresh round of base pairing and phosphodiester bond formation can take place with a new incoming nucleotide.

Proofreading improves the accuracy of DNA synthesis by a factor of about 100. But what happens to errors that escape proofreading and are incorporated into the daughter strand of the double helix? Cells have a further back-up system that catches and repairs such errors known as

mismatch repair. Mismatch repair improves the accuracy of DNA synthesis by another two or three orders of magnitude. The proteins of the mismatch repair system scan the double helix for distortions (recall that only properly paired bases are accommodated within a diameter of 20 Å). The repair system then removes a stretch of DNA containing the mismatch from the daughter strand and replaces it with newly synthesized DNA using the parental strand as a template.

Mismatch repair is in a race against time. If the mismatch is not caught and repaired before the DNA is fully replicated and a new round of replication has commenced, then the misincorporated nucleotide will now serve as a template for a new daughter strand. At this point, the DNA will no longer contain a mismatch, and the misincorporated nucleotide will have been permanently incorporated into the DNA as a mutation.

DNA damage can occur independently of replication

DNA is susceptible to damage even after it is replicated. Chemicals and radiation in the environment can damage DNA. If left unrepaired, environmentally induced DNA damage results in mutations. Cells have multiple molecular machines for detecting and repairing such damage. For example, ultraviolet light from the sun damages pyrimidines. Individuals with a condition known as **xeroderma pigmentosum** are defective in the machine for repairing this damage. As a consequence, such individuals are hypersensitive to sunlight and readily contract skin cancer.

Some mutations arise without environmental influences. Remember that DNA is bathed in water at a concentration of about 55 M. Water, which we usually consider to be innocuous, can cause hydrolytic damage to DNA bases. One example of such damage is the deamination of the base cytosine (Figure 13). In this reaction, the exocyclic amino group (i.e., the amino group that is not part of the ring) is replaced with a carbonyl group, thereby converting cytosine to uracil, a base usually only found in RNA. Like cytosine, which pairs with guanine, uracil is a pyrimidine but unlike cytosine, uracil (like thymine) pairs with adenine. Thus, if left unrepaired, the resulting uracil will pair with adenine during replication, and after a further round of replication, that adenine will, in turn, pair with thymine, resulting in the permanent replacement of the original G:C base pair with an A:T base pair.

Cells have a repair system that detects uracil in DNA and replaces it with cytosine. This repair mechanism is only possible because uracil is not normally found in DNA. Indeed, the susceptibility of cytosine to deamination provides an appealing explanation for why DNA has thymine as a base rather than uracil. If uracil were normally found in DNA, then the cell could not distinguish uracil that arose aberrantly from cytosine by deamination from all the other uracils in the DNA. RNA, on the other hand, does not generally serve as a genetic information carrier, and so conversions of cytosine to uracil do not normally have permanent genetic (mutational) consequences.

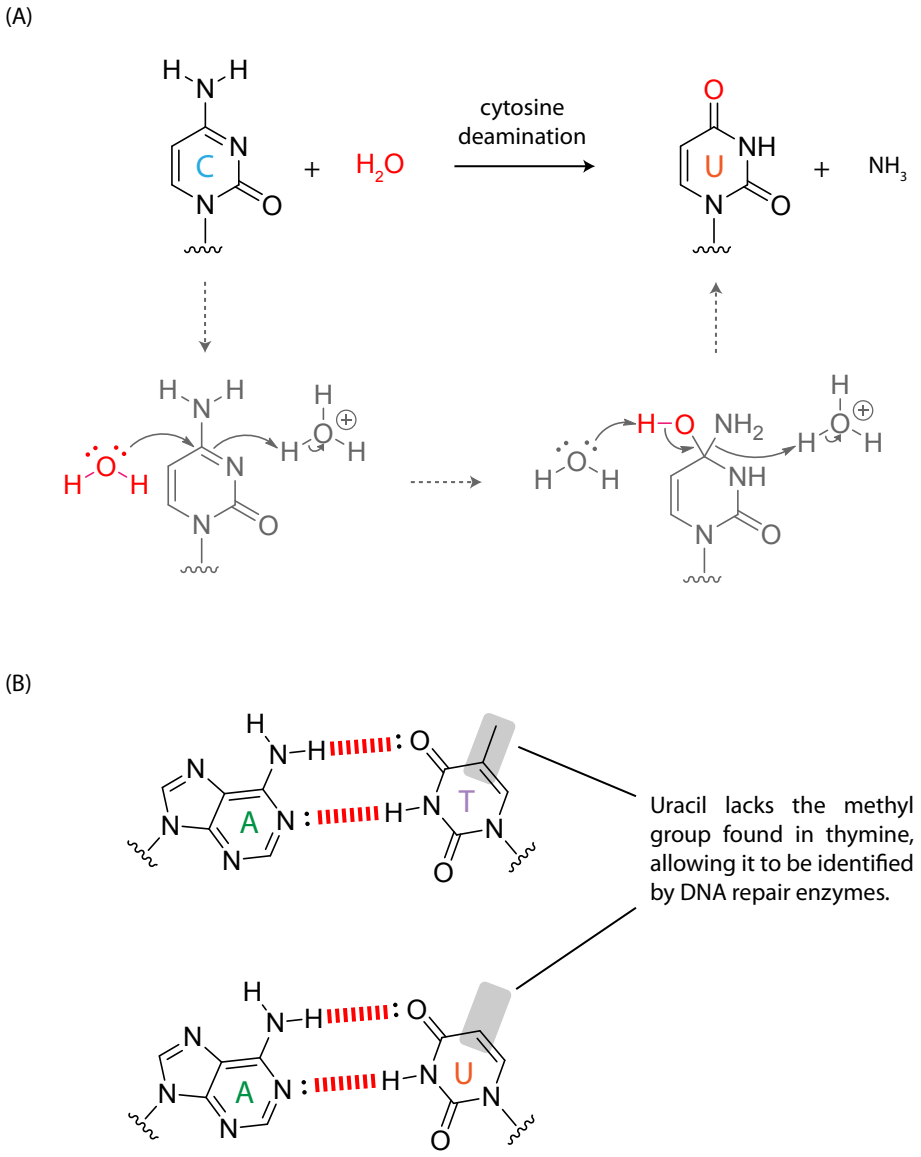

Figure 13 Cytosine deamination converts cytosine into uracil

(A) As shown in the arrow-pushing mechanism, water acts as a nucleophile, forming a covalent bond to carbon 4 ("C4") of cytosine. In this same step, the double bond between C4 and nitrogen 3 ("N3") breaks, and N3 becomes protonated. In the second step, a second water molecule acts as a base, abstracting a hydrogen atom from the hydroxyl group attached to C4. The electrons from the broken H-O bond form a new O-C4 double bond, and the C4-NH_2 bond is broken as a consequence. The -NH_2 leaving group becomes protonated to form ammonia (NH_3). Cytosine is converted to uracil by this reaction. (B) The absence of the methyl group at pyrimidine position 5 allows repair enzymes to distinguish uracil from thymine in DNA.

Box 1 The Polymerase Chain Reaction is a technique for amplifying specific DNA sequences exponentially

An extremely powerful technique that is based on the concepts we have been considering in this chapter is the Polymerase Chain Reaction or PCR. PCR makes it possible to amplify a desired stretch of DNA from a much larger DNA molecule. The DNA is amplified in an exponential fashion, meaning that each cycle of PCR, in principle, doubles the number of DNA molecules with a specific sequence present in a test tube. As a result, minute quantities of DNA, as little as a single molecule, can be amplified into billions of molecules in just a few hours using relatively simple equipment. PCR was invented by Kary Mullis, who earned the 1993 Nobel Prize in Chemistry for his achievement.

A PCR reaction contains one or more DNA template molecules, many copies of each of two short single-stranded DNAs that will prime the reaction, each of the four dNTPs, and DNA polymerase. The sequences of the two DNA primers are chosen to anneal to each of the template strands at locations that flank the desired region to be amplified, with the 3' ends of the primers pointing at each other.

Figure 14 illustrates the PCR process for the amplification of a segment of DNA highlighted in gray. Single-stranded DNAs (*purple and light blue*) that are complementary to the ends of this region are used to prime

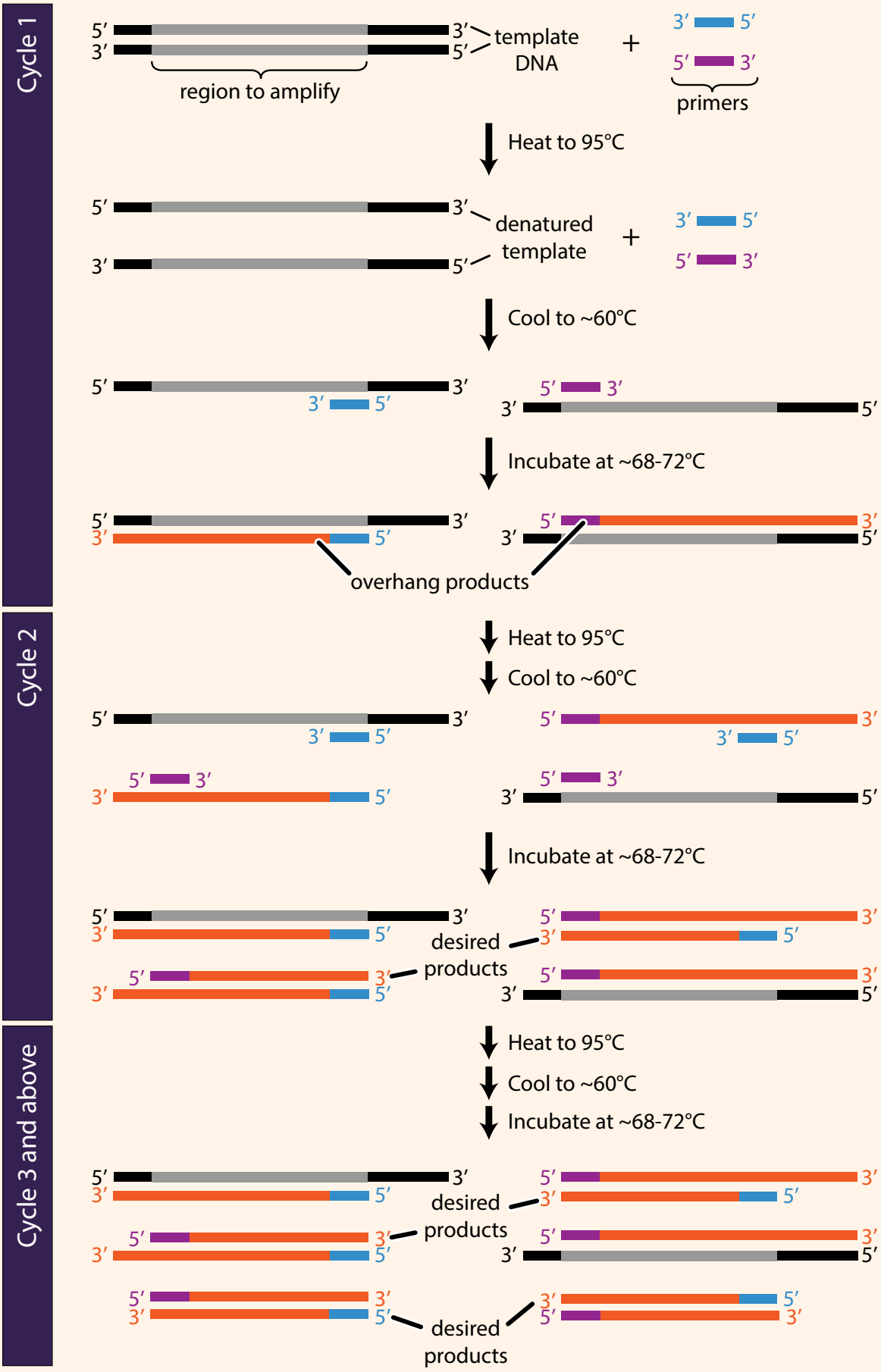

Figure 14 **PCR uses specially designed primers to amplify a desired DNA sequence**

the reaction. At the start of the first PCR cycle, the test tube containing all of these ingredients is heated to 95°C, which is close to the temperature at which water boils. At such a high temperature, the two strands of the DNA template denature into single strands. The test tube is then cooled, typically to about 60°C (the temperature of uncomfortably hot, but not scalding, water). At this intermediate temperature, the DNA primers anneal to the dissociated template strands, forming two template-primer complexes. So far no DNA synthesis has taken place. After the template-primer complexes form, the temperature is raised to 68°C-72°C. DNA polymerase begins to extend each primer along its template, polymerizing a strand complementary to the template in the 5'-to-3' direction and consuming dNTPs in the test tube. When both of the polymerization reactions are complete, the end result is that the desired region of the template molecule has been copied, along with additional DNA past the region of interest. We will call these pieces of extended DNA overhang products; they are longer than the segment of DNA we seek to amplify.

Next, the test tube is again heated to 95°C so that the double-stranded products from the first round of polymerization denature into single strands. Now these amplified single strands can serve as the templates for the next cycle of PCR. The test tube is once again cooled to about 60°C at which temperature the primers anneal to the templates and DNA polymerization takes place once again. Notice what happens: when an overhang product is used as a template for the next round of PCR, the primer anneals at the point where the overhang begins. When polymerization stops at the end of the DNA, the result is only the desired product, with no overhang. These correctly-sized products, as well as the overhang products, can be used as templates again in the next cycle.

In each cycle, the undesired overhang products continue to be produced from the original template DNA, but these increase only linearly with the number of cycles. Because every product can be used as a template to make the desired product, the desired product is amplified exponentially, doubling in number during each cycle. The net result of this process is that each cycle (template denaturation, primer annealing, and DNA polymerization) of PCR can double the number of desired double-stranded DNA molecules in the tube. After 26 PCR cycles, one single-stranded template molecule can be theoretically copied more than 30 million (2^{26}) times. At this point, the desired product is vastly more abundant than the unwanted overhang products.

Summary

The Meselson-Stahl experiment showed that DNA is replicated by a semi-conservative mechanism in which each of the two strands of the double helix separately serve as templates for the synthesis of new daughter strands. Using the heavy isotope of nitrogen ^{15}N to density-label DNA, the Meselson-Stahl experiment showed that after one round of replication in the presence of ^{14}N all the DNA had a hybrid density, as predicted for semi-conservative replication.

DNA synthesis takes place in a 5'-to-3' direction using as substrates the four 2'-deoxynucleoside triphosphates (dNTPs). The 3' hydroxyl group of the growing chain acts as a nucleophile that attacks the α-phosphate of the incoming dNTP, resulting in the formation of a phosphodiester bond and the release of pyrophosphate as a leaving group. The complementary template strand to which the growing daughter strand is annealed specifies the identity of each nucleotide to be added by base pairing between the incoming nucleotide and the corresponding base on the template strand.

The enzyme for DNA synthesis is DNA polymerase. The reaction that it catalyzes, phosphodiester bond formation, is thermodynamically favorable

but only moderately so. Phosphodiester bond formation is, however, coupled to a second moderately favorable process, the hydrolysis of the pyrophosphate leaving group. Together, the two coupled reactions are quite favorable. The coupling of phosphodiester bond formation to pyrophosphate hydrolysis is an example of Le Châtelier's principle.

Replication takes place at replication forks, which are moving sites on DNA where the two strands are being separated so that each can serve as a template for daughter strand synthesis. Replication forks are asymmetric in that synthesis takes place continuously in the direction in which the fork is moving (leading strand synthesis) and discontinuously on the other strand (lagging strand synthesis). Lagging strand synthesis takes place in short bursts that point away from the movement of the fork. The resulting short segments of DNA are called Okazaki fragments and are stitched together by DNA ligase to generate an uninterrupted daughter strand.

DNA polymerase synthesizes DNA accurately by requiring that each incoming nucleotide pair with the corresponding base on the template. Incoming nucleotides that do not properly pair are disfavored for phosphodiester bond formation. Nucleotides that are nonetheless misincorporated are removed by a proofreading mechanism in which a 3' nuclease in the editing pocket of DNA polymerase releases the misincorporated nucleotide from the 3' end of the growing chain, allowing the polymerase to incorporate a fresh nucleotide. Misincorporated nucleotides that escape proofreading are, in turn, removed by a mismatch repair system.

Chemicals, radiation, and hydrolytic damage also contribute to introducing mutations in DNA. Dedicated enzyme systems scan the DNA for damage, such as damage to pyrimidines caused by ultraviolet light, and repair the lesions. Hydrolytic damage resulting from deamination of cytosine is repaired by a system that detects and removes uracil, the product of the deamination reaction. Cytosine deamination offers an explanation for why DNA contains thymine instead of uracil, which is normally only found in RNA; if DNA contained uracil, then the cell could not distinguish uracil arising from deamination from uracil normally present in DNA.

The PCR reaction is a powerful technique for amplifying specific segments of DNA exponentially from much larger DNA molecules by the use of short DNA primers.