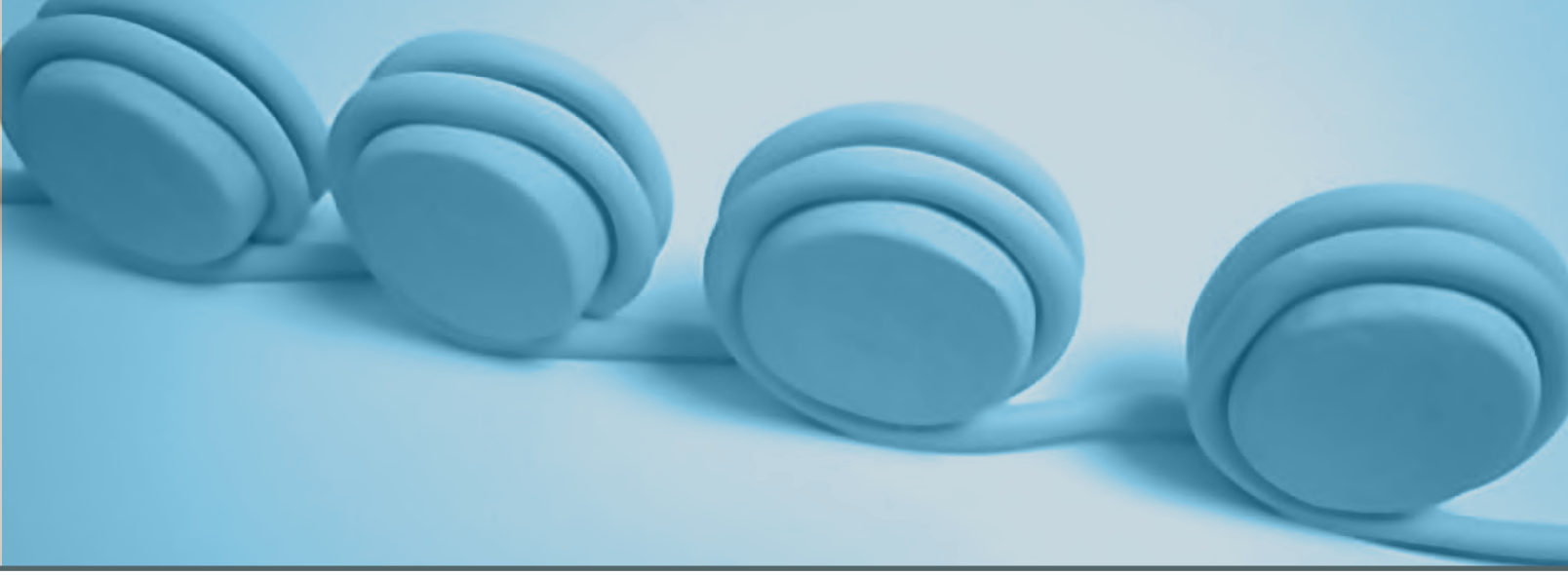


CHAPTER **FIVE** 5

DNA and Chromosomes

Life depends on the ability of cells to store, retrieve, and translate the genetic instructions required to make and maintain a living organism. This hereditary information is passed on from a cell to its daughter cells at cell division, and from generation to generation in multicellular organisms through the reproductive cells—eggs and sperm. These instructions are stored within every living cell in its *genes*—the information-containing elements that determine the characteristics of a species as a whole and of the individuals within it.

At the beginning of the twentieth century, when genetics emerged as a science, scientists became intrigued by the chemical nature of genes. The information in genes is copied and transmitted from cell to daughter cells millions of times during the life of a multicellular organism, and it survives the process essentially unchanged. What kind of molecule could be capable of such accurate and almost unlimited replication, and also be able to direct the development of an organism and the daily life of a cell? What kind of instructions does the genetic information contain? How are these instructions physically organized so that the enormous amount of information required for the development and maintenance of even the simplest organism can be contained within the tiny space of a cell?

The answers to some of these questions began to emerge in the 1940s, when it was discovered from studies in simple fungi that genetic information consists primarily of instructions for making proteins. Proteins perform most of the cell's functions: they serve as building blocks for cell structures; they form the enzymes that catalyze the cell's chemical reactions; they regulate the activity of genes; and they enable cells to

move and to communicate with one another. With hindsight, it is hard to imagine what other type of instructions the genetic information could have contained.

The other crucial advance made in the 1940s was the recognition that deoxyribonucleic acid (DNA) was the likely carrier of this genetic information. But the mechanism whereby the hereditary information is copied for transmission from one generation of cells to the next, and how proteins are specified by the instructions in DNA, remained completely mysterious until 1953, when the structure of DNA was determined by James Watson and Francis Crick. The structure immediately revealed how DNA might be copied, or replicated, and it provided the first clues about how a molecule of DNA might encode the instructions for making proteins. Today, the fact that DNA is the genetic material is so fundamental to our understanding of life that it is difficult to appreciate what an enormous intellectual gap this discovery filled.

In this chapter, we begin by describing the structure of DNA. We see how, despite its chemical simplicity, the structure and chemical properties of DNA make it ideally suited for carrying genetic information. The genes of every cell on Earth are made of DNA, and insights into the relationship between DNA and genes have come from experiments in a wide variety of organisms. We then consider how genes and other important segments of DNA are arranged in the single, long DNA molecule that forms the core of each chromosome in the cell. Finally, we discuss how eukaryotic cells fold these long DNA molecules into compact chromosomes inside the nucleus. This packing has to be done in an orderly fashion so that the chromosomes can be duplicated and apportioned correctly between the two daughter cells at each cell division. It must also allow the DNA to be accessed by the proteins that replicate and repair DNA, and regulate the activity of its many genes.

This is the first of five chapters that deal with basic genetic mechanisms—the ways in which the cell maintains and makes use of the genetic information carried in its DNA. In Chapter 6, we discuss the mechanisms by which the cell accurately replicates and repairs its DNA. In Chapter 7, we consider gene expression—how genes are used to produce RNA and protein molecules. In Chapter 8, we describe how a cell controls gene expression to ensure that each of the many thousands of proteins encoded in its DNA is manufactured at the proper time and place. In Chapter 9, we discuss how present-day genes evolved from distant ancestors, and, in Chapter 10, we consider some of the experimental techniques used to study both DNA and its role in fundamental cell processes.

An enormous amount has been learned about these subjects in the past 60 years. Much less obvious, but equally important, is that our knowledge is very incomplete; thus a great deal still remains to be discovered about how DNA provides the instructions to build living things.

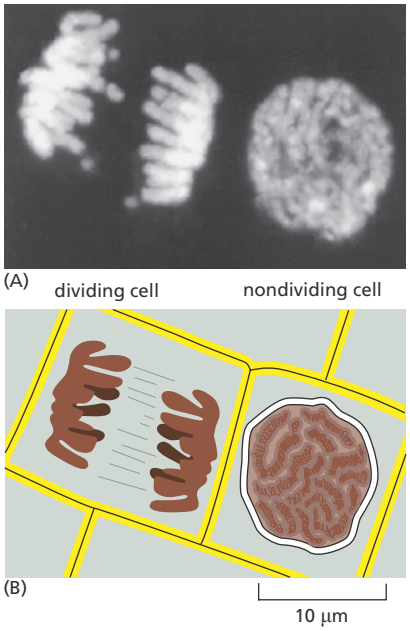


Figure 5–1 Chromosomes become visible as eukaryotic cells prepare to divide. (A) Two adjacent plant cells photographed in a fluorescence microscope. The DNA is labeled with a fluorescent dye (DAPI) that binds to it. The DNA is packaged into chromosomes, which become visible as distinct structures only when they condense in preparation for cell division, as shown on the *left*. The cell on the *right*, which is not dividing, contains the identical chromosomes, but they cannot be distinguished as individual entities because the DNA is in a much more extended conformation at this phase in the cell's life cycle. (B) Schematic diagram of the outlines of the two cells and their chromosomes. (A, courtesy of Peter Shaw.)

THE STRUCTURE OF DNA

Well before biologists understood the structure of DNA, they had recognized that inherited traits and the genes that determine them were associated with the chromosomes. Chromosomes (named from the Greek *chroma*, "color," because of their staining properties) were discovered in the nineteenth century as threadlike structures in the nucleus of eukaryotic cells that become visible as the cells begin to divide (**Figure 5–1**). As biochemical analysis became possible, researchers learned that chromosomes contain both DNA and protein. But which of these components encoded the organism's genetic information was not clear.

We now know that the DNA carries the hereditary information of the cell and that the protein components of chromosomes function largely to package and control the enormously long DNA molecules. But biologists in the 1940s had difficulty accepting DNA as the genetic material because of the apparent simplicity of its chemistry (see **How We Know**, pp. 174–176). DNA, after all, is simply a long polymer composed of only four types of nucleotide subunits, which are chemically very similar to one another.

Then, early in the 1950s, DNA was examined by X-ray diffraction analysis, a technique for determining the three-dimensional atomic structure of a molecule (see Figure 4–52). The early results indicated that DNA is composed of two strands wound into a helix. The observation that DNA is double-stranded was of crucial significance. It provided one of the major clues that led, in 1953, to a correct model for the structure of DNA. This structure immediately suggested how DNA could encode the instructions necessary for life, and how these instructions could be copied and passed along when cells divide. In this section, we examine the structure of DNA and explain in general terms how it is able to store hereditary information.

A DNA Molecule Consists of Two Complementary Chains of Nucleotides

A molecule of **deoxyribonucleic acid** (**DNA**) consists of two long polynucleotide chains. Each *chain*, or *strand*, is composed of four types of nucleotide subunits, and the two strands are held together by hydrogen bonds between the base portions of the nucleotides (**Figure 5–2**).

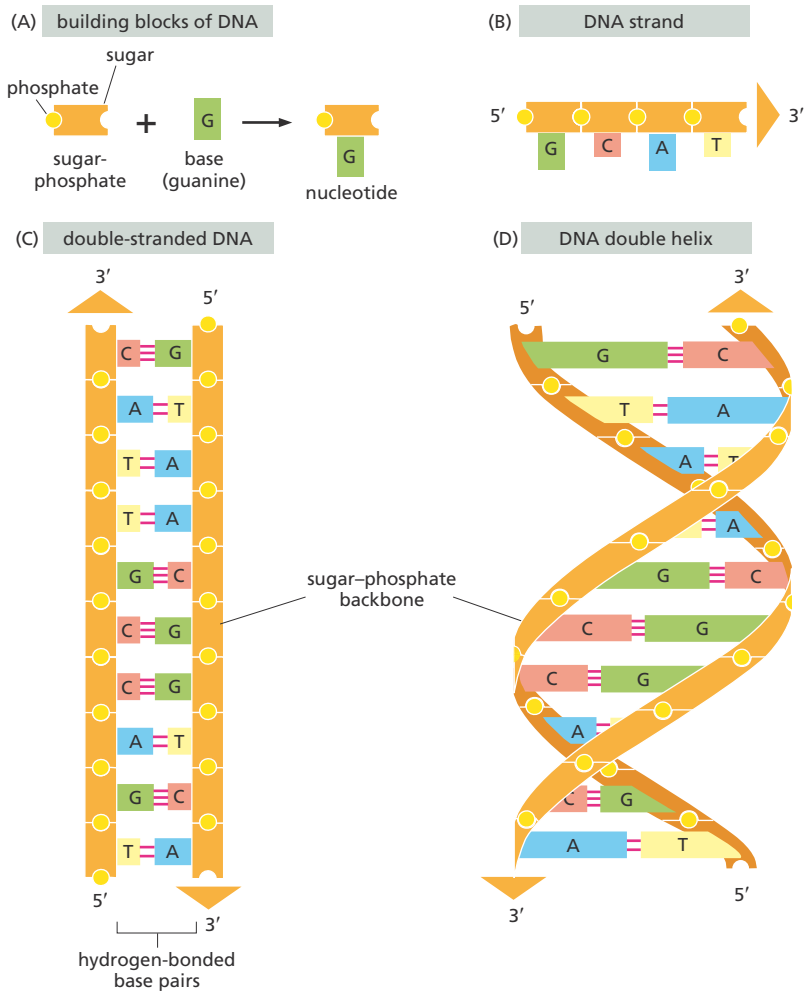


Figure 5–2 DNA is made of four nucleotide building blocks. (A) Each nucleotide is composed of a sugar–phosphate covalently linked to a base—guanine (G) in this figure. (B) The nucleotides are covalently linked together into polynucleotide chains, with a sugar–phosphate backbone from which the bases (A, C, G, and T) extend. (C) A DNA molecule is composed of two polynucleotide chains (DNA strands) held together by hydrogen bonds between the paired bases. The *arrows* on the DNA strands indicate the polarities of the two strands, which run antiparallel to each other in the DNA molecule. (D) Although the DNA is shown straightened out in (C), in reality, it is wound into a double helix, as shown here.

GENES ARE MADE OF DNA

By the 1920s, scientists generally agreed that genes reside on chromosomes, and they knew that chromosomes are composed of both DNA and proteins. But because DNA is so chemically simple, they naturally assumed that genes had to be made of proteins, which are much more chemically diverse than DNA molecules. Even when the experimental evidence suggested otherwise, this assumption proved hard to shake.

Messages from the dead

The case for DNA began to emerge in the late 1920s, when a British medical officer named Fred Griffith made an astonishing discovery. He was studying *Streptococcus pneumoniae* (pneumococcus), a bacterium that causes pneumonia. As antibiotics had not yet been discovered, infection with this organism was usually fatal. When

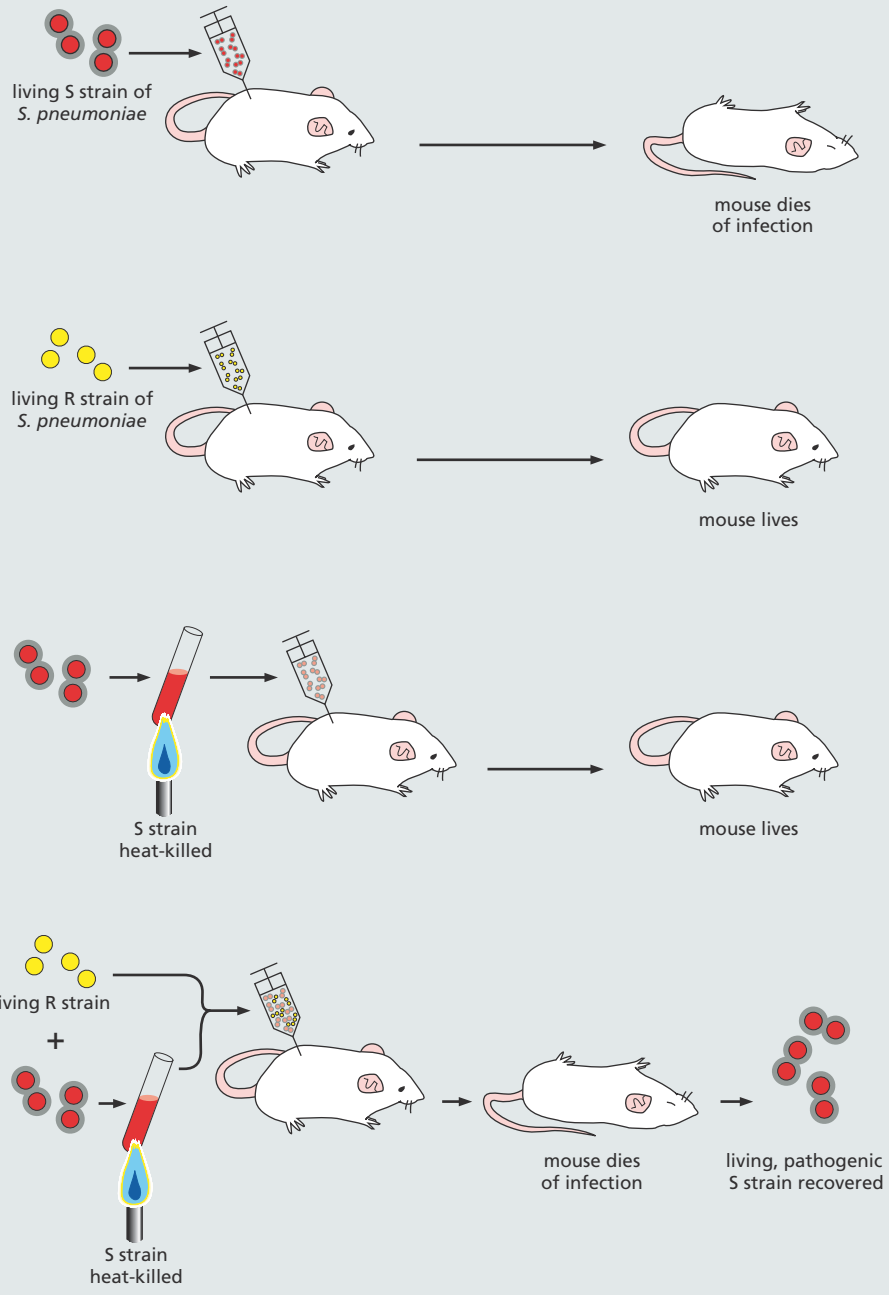


Figure 5–3 Griffith showed that heat-killed, infectious bacteria can transform harmless, living bacteria into pathogenic ones. The bacterium *Streptococcus pneumoniae* comes in two forms that differ from one another in their microscopic appearance and in their ability to cause disease. Cells of the pathogenic strain, which are lethal when injected into mice, are encased in a slimy, glistening polysaccharide capsule. When grown on a plate of nutrients in the laboratory, this disease-causing bacterium forms colonies that look dome-shaped and smooth; hence it is designated the S form. The harmless strain of the pneumococcus, on the other hand, lacks this protective coat; it forms colonies that appear flat and rough—hence, it is referred to as the R form. As illustrated, Griffith found that a substance present in the pathogenic S strain could permanently change, or transform, the nonlethal R strain into the deadly S strain.

grown in the laboratory, pneumococci come in two forms: a pathogenic form that causes a lethal infection when injected into animals, and a harmless form that is easily conquered by the animal's immune system and does not produce an infection.

In the course of his investigations, Griffith injected various preparations of these bacteria into mice. He showed that pathogenic pneumococci that had been killed by heating were no longer able to cause infection. The surprise came when Griffith injected both heat-killed pathogenic bacteria and live harmless bacteria into the same mouse. This combination proved lethal: not only did the animals die of pneumonia, but Griffith found that their blood was teeming with live bacteria of the pathogenic form (**Figure 5–3**). The heat-killed pneumococci had somehow converted the harmless bacteria into the lethal form. What's more, Griffith found that the change was permanent: he could grow these "transformed" bacteria in culture, and they remained pathogenic. But what was this mysterious material that turned harmless bacteria into killers? And how was this change passed on to progeny bacteria?

Transformation

Griffith's remarkable finding set the stage for the experiments that would provide the first strong evidence that genes are made of DNA. The American bacteriologist Oswald Avery, following up on Griffith's work, discovered that the harmless pneumococcus could be transformed into a pathogenic strain in a culture tube by exposing it to an extract prepared from the pathogenic strain. It would take another 15 years, however, for Avery and his colleagues Colin MacLeod and Maclyn McCarty to successfully purify the "transforming principle" from this soluble extract and to demonstrate that the active ingredient was DNA. Because the transforming principle caused a heritable change in the bacteria that received it, DNA must be the very stuff of which genes are made.

The 15-year delay was in part a reflection of the academic climate—and the widespread supposition that the genetic material was likely to be made of protein. Because of the potential ramifications of their work, the researchers wanted to be absolutely certain that the transforming principle was DNA before they announced their findings. As Avery noted in a letter to his brother, also a bacteriologist, "It's lots of fun to blow bubbles, but it's wiser to prick them yourself before someone else tries to." So the researchers subjected the transforming material to a battery of chemical tests (**Figure 5–4**). They found that it exhibited all the chemical properties

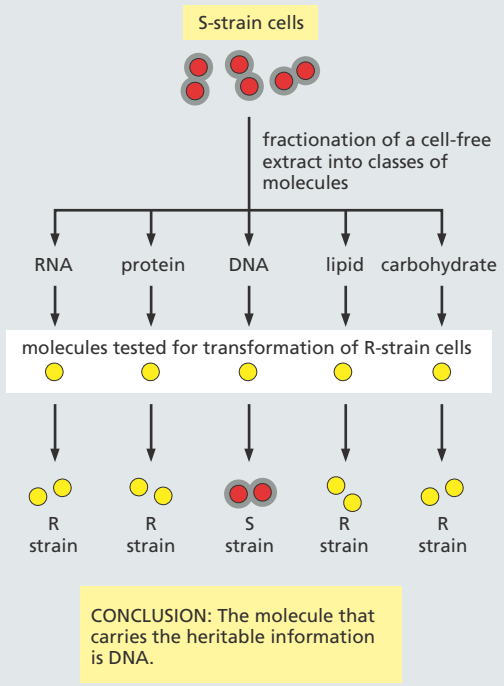


Figure 5–4 Avery, MacLeod, and McCarty demonstrated that DNA is the genetic material. The researchers prepared an extract from the disease-causing S strain of pneumococci and showed that the "transforming principle" that would permanently change the harmless R-strain pneumococci into the pathogenic S strain is DNA. This was the first evidence that DNA could serve as the genetic material.

characteristic of DNA; furthermore, they showed that enzymes that destroy proteins and RNA did not affect the ability of the extract to transform bacteria, while enzymes that destroy DNA inactivated it. And like Griffith before them, the investigators found that their purified preparation changed the bacteria permanently: DNA from the pathogenic species was taken up by the harmless species, and this change was faithfully passed on to subsequent generations of bacteria.

This landmark study offered rigorous proof that purified DNA can act as genetic material. But the resulting paper, published in 1944, drew remarkably little attention. Despite the meticulous care with which these experiments were performed, geneticists were not immediately convinced that DNA is the hereditary material. Many argued that the transformation might have been caused by some trace protein contaminant in the preparations. Or that the extract might contain a mutagen that alters the genetic material of the harmless bacteria—converting them to the pathogenic form—rather than containing the genetic material itself.

Virus cocktails

The debate was not settled definitively until 1952, when Alfred Hershey and Martha Chase fired up their laboratory blender and demonstrated, once and for all, that genes are made of DNA. The researchers were studying T2—a virus that infects and eventually destroys the bacterium *E. coli*. These bacteria-killing viruses behave like little molecular syringes: they inject their genetic material into the bacterial host cell, while the empty virus heads remain attached outside (**Figure 5–5A**). Once inside the bacterial cell, the viral genes direct the formation of new virus particles. In less than an hour, the infected cells explode, spewing thousands of new viruses into the medium. These then infect neighboring bacteria, and the process begins again.

The beauty of T2 is that these viruses contain only two kinds of molecules: DNA and protein. So the genetic material had to be one or the other. But which? The experiment was fairly straightforward. Because the viral DNA enters the bacterial cell, while the rest of the virus particle remains outside, the researchers decided to radioactively label the protein in one batch of virus and the DNA in another. Then, all they had to do was follow the radioactivity to see whether viral DNA or viral protein wound up inside the bacteria. To do this, Hershey and Chase incubated their radiolabeled viruses with *E. coli*; after allowing a few minutes for infection to take place, they poured the mix into a Waring blender and hit "puree." The blender's spinning blades sheared the empty virus heads from the surfaces of the bacterial cells. The researchers then centrifuged the sample to separate the heavier, infected bacteria, which formed a pellet at the bottom of the centrifuge tube, from the empty viral coats, which remained in suspension (**Figure 5–5B**).

As you have probably guessed, Hershey and Chase found that the radioactive DNA entered the bacterial cells, while the radioactive proteins remained outside with the empty virus heads. They found that the radioactive DNA was also incorporated into the next generation of virus particles.

This experiment demonstrated conclusively that viral DNA enters bacterial host cells, whereas viral protein does not. Thus, the genetic material in this virus had to be made of DNA. Together with the studies done by Avery, MacLeod, and McCarty, this evidence clinched the case for DNA as the agent of heredity.

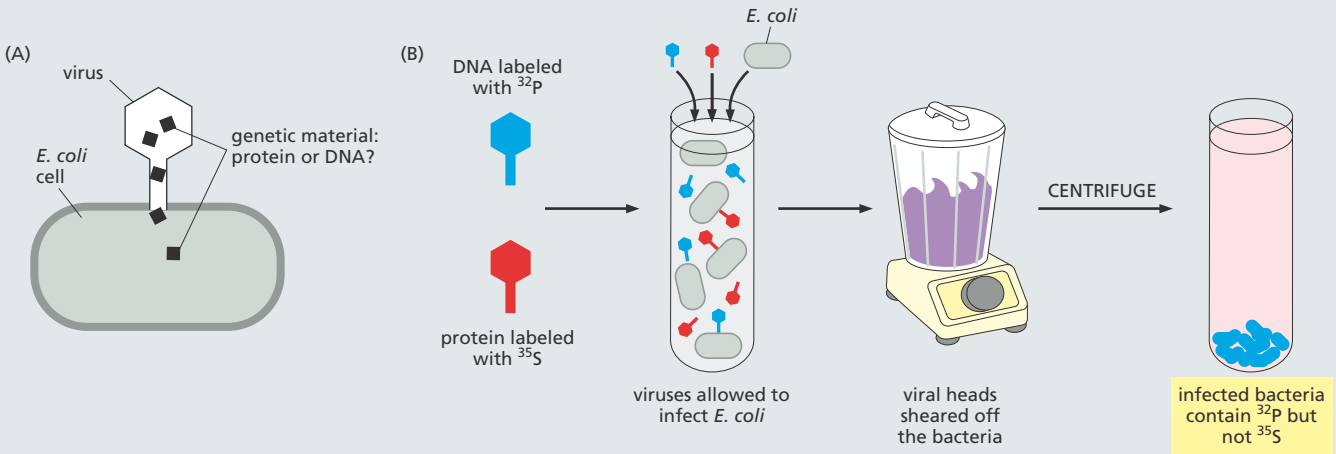


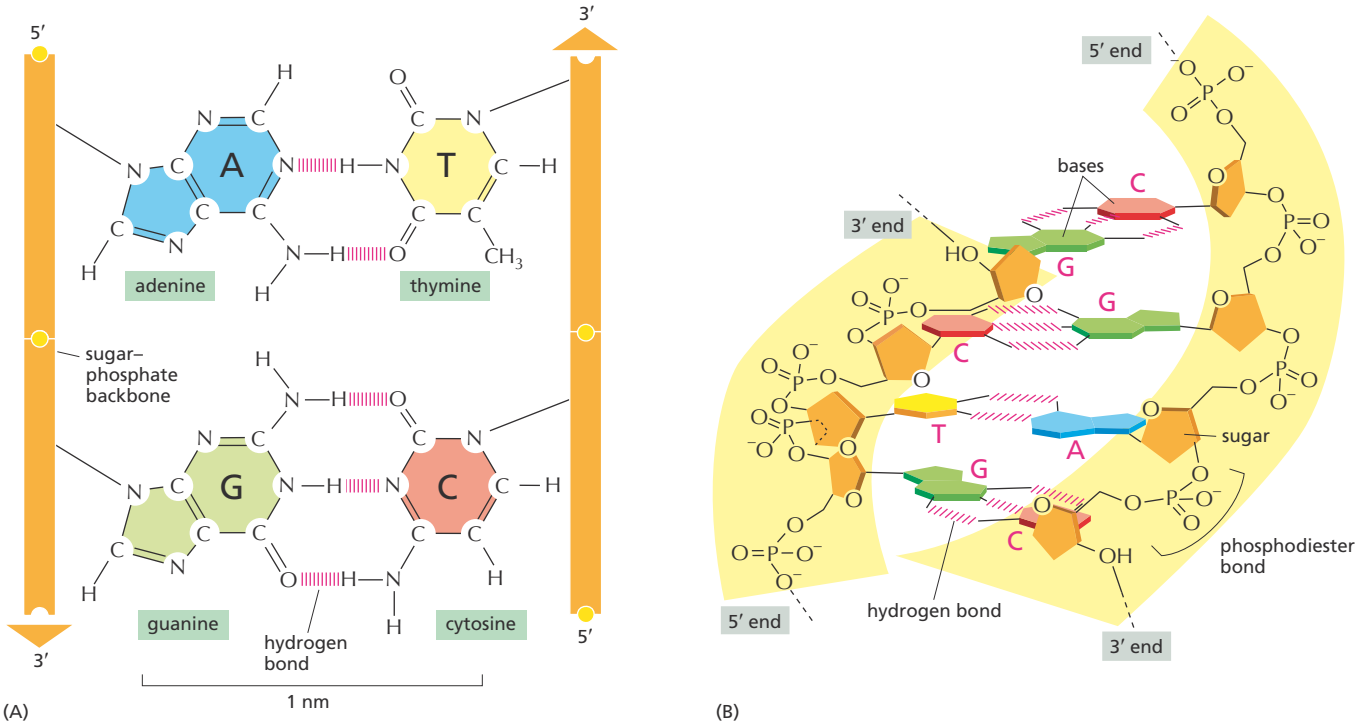
Figure 5–5 Hershey and Chase showed definitively that genes are made of DNA. (A) The researchers worked with T2 viruses, which are made entirely of protein and DNA. Each virus acts as a molecular syringe, injecting its genetic material into a bacterium; the empty viral capsule remains attached to the outside of the cell. (B) To determine whether the genetic material of the virus is protein or DNA, the researchers radioactively labeled the DNA in one batch of viruses with ^{32}P and the proteins in a second batch of viruses with ^{35}S. Because DNA lacks sulfur and the proteins lack phosphorus, these radioactive isotopes provided a handy way for the researchers to distinguish these two types of molecules. These labeled viruses were allowed to infect and replicate inside *E. coli*, and the mixture was then disrupted by brief pulsing in a Waring blender and separated to part the infected bacteria from the empty viral heads. When the researchers measured the radioactivity, they found that much of the ^{32}P-labeled DNA had entered the bacterial cells, while the vast majority of the ^{35}S-labeled proteins remained in solution with the spent viral particles.

As we saw in Chapter 2 (Panel 2–6, pp. 76–77), nucleotides are composed of a nitrogen-containing base and a five-carbon sugar, to which is attached one or more phosphate groups. For the nucleotides in DNA, the sugar is deoxyribose (hence the name deoxyribonucleic acid), and the base can be either *adenine* (*A*), *cytosine* (*C*), *guanine* (*G*), or *thymine* (*T*). The nucleotides are covalently linked together in a chain through the sugars and phosphates, which thus form a backbone of alternating sugar–phosphate–sugar–phosphate (see Figure 5–2B). Because it is only the base that differs in each of the four types of subunits, each polynucleotide chain in DNA can be thought of as a necklace: a sugar–phosphate backbone strung with four types of beads (the four bases A, C, G, and T). These same symbols (A, C, G, and T) are also commonly used to denote the four different nucleotides—that is, the bases with their attached sugar phosphates.

The way in which the nucleotide subunits are linked together gives a DNA strand a chemical polarity. If we imagine that each nucleotide has a knob (the phosphate) and a hole (see Figure 5–2A), each strand, formed by interlocking knobs with holes, will have all of its subunits lined up in the same orientation. Moreover, the two ends of the strand can be easily distinguished, as one will have a hole (the 3′ hydroxyl) and the other a knob (the 5′ phosphate). This polarity in a DNA strand is indicated by referring to one end as the *3′* end and the other as the *5′* end. This convention is based on the details of the chemical linkage between the nucleotide subunits.

The two polynucleotide chains in the DNA **double helix** are held together by hydrogen-bonding between the bases on the different strands. All the bases are therefore on the inside of the double helix, with the sugar–phosphate backbones on the outside (see Figure 5–2D). The bases do not pair at random, however: A always pairs with T, and G always pairs with C (**Figure 5–6**). In each case, a bulkier two-ring base (a purine, see Panel 2–6, pp. 76–77) is paired with a single-ring base (a pyrimidine). Each purine–pyrimidine pair is called a **base pair**, and this *complementary base-pairing* enables the base pairs to be packed in the energetically most favorable

Figure 5–6 The two strands of the DNA double helix are held together by hydrogen bonds between complementary base pairs. (A) The shapes and chemical structure of the bases allow hydrogen bonds to form efficiently only between A and T and between G and C, where atoms that are able to form hydrogen bonds (see Panel 2–2, pp. 68–69) can be brought close together without perturbing the double helix. Two hydrogen bonds form between A and T, whereas three form between G and C. The bases can pair in this way only if the two polynucleotide chains that contain them are antiparallel—that is, oriented in opposite directions. (B) A short section of the double helix viewed from its side. Four base pairs are shown. The nucleotides are linked together covalently by phosphodiester bonds through the 3′-hydroxyl (–OH) group of one sugar and the 5′-phosphate ($–OPO_3$) of the next (see Panel 2–6, pp. 76–77, to review how the carbon atoms in the sugar ring are numbered). This linkage gives each polynucleotide strand a chemical polarity; that is, its two ends are chemically different. The 3′ end carries an unlinked –OH group attached to the 3′ position on the sugar ring; the 5′ end carries a free phosphate group attached to the 5′ position on the sugar ring.



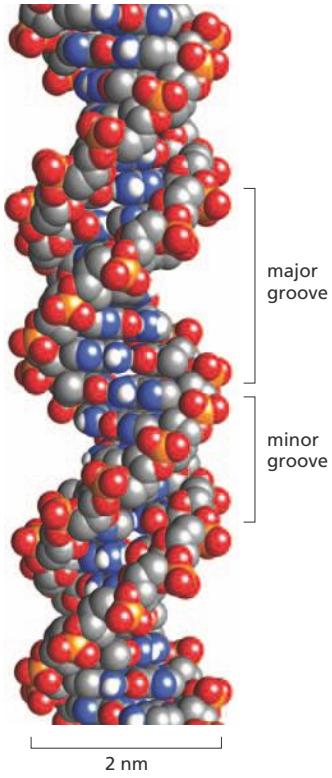


Figure 5–7 A space-filling model shows the conformation of the DNA double helix. The two DNA strands wind around each other to form a right-handed helix (see Figure 4–14) with 10 bases per turn. Shown here are 1.5 turns of the DNA double helix. The coiling of the two strands around each other creates two grooves in the double helix. The wider groove is called the major groove and the smaller one the minor groove. The colors of the atoms are: N, *blue*; O, *red*; P, *yellow*; and H, *white*.

arrangement in the interior of the double helix. In this arrangement, each base pair has a similar width, thus holding the sugar–phosphate backbones an equal distance apart along the DNA molecule. The members of each base pair can fit together within the double helix because the two strands of the helix run *antiparallel* to each other—that is, they are oriented with opposite polarities (see Figure 5–2C and D). The antiparallel sugar–phosphate strands then twist around each other to form a double helix containing 10 base pairs per helical turn (**Figure 5–7**). This twisting also contributes to the energetically favorable conformation of the DNA double helix.

A consequence of the base-pairing requirements is 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—an A always matches a T on the opposite strand, and a C always matches a G. This complementarity is of crucial importance when it comes to both copying and repairing the DNA, as we discuss in Chapter 6. An animated version of the DNA structure can be seen in **Movie 5.1**.

QUESTION 5–1

Which of the following statements are correct? Explain your answers.
A. A DNA strand has a polarity because its two ends contain different bases.
B. G-C base pairs are more stable than A-T base pairs.

The Structure of DNA Provides a Mechanism for Heredity

The need for genes to encode information that must be copied and transmitted accurately when a cell divides raised two fundamental questions: how can the information for specifying an organism be carried in chemical form, and how can the information be accurately copied? The discovery of the structure of the DNA double helix was a landmark in biology because it immediately suggested the answers—and thereby resolved the problem of heredity at the molecular level. In this chapter, we outline the answer to the first question; in the next chapter, we address in detail the answer to the second.

Information is encoded in the order, or sequence, of the nucleotides along each DNA strand. Each base—A, C, T, or G—can be considered a letter in a four-letter alphabet that is used to spell out biological messages (**Figure 5–8**). Organisms differ from one another because their respective DNA molecules have different *nucleotide sequences* and, consequently, carry different biological messages. But how is the nucleotide alphabet used to make up messages, and what do they spell out?

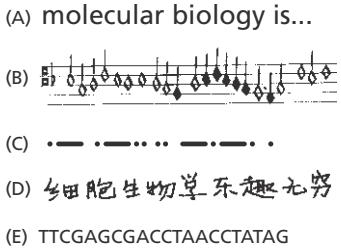


Figure 5–8 Linear messages come in many forms. The languages shown are (A) English, (B) a musical score, (C) Morse code, (D) Chinese, and (E) DNA.

It had already been established some time before the structure of DNA was determined that genes contain the instructions for producing proteins. DNA messages, therefore, must somehow be able to encode proteins. Consideration of the chemical character of proteins makes the problem easier to define. As discussed in Chapter 4, the function of a protein is determined by its three-dimensional structure, and this structure in turn is determined by the sequence of the amino acids in its polypeptide chain. The linear sequence of nucleotides in a gene must therefore be able to spell out the linear sequence of amino acids in a protein.

The exact correspondence between the 4-letter nucleotide alphabet of DNA and the 20-letter amino acid alphabet of proteins—the **genetic code**—is not obvious from the structure of the DNA molecule, and it took more than a decade after the discovery of the double helix to work it

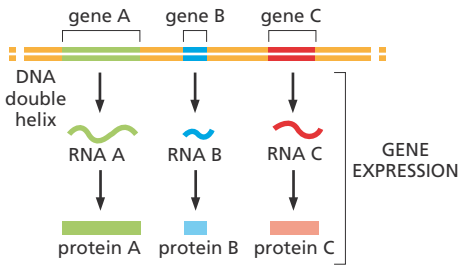


Figure 5–9 Most genes contain information to make proteins. As we discuss in Chapter 7, each protein-coding gene is used to produce RNA molecules, which then direct the production of the specific protein molecules.

out. In Chapter 7, we describe this code in detail when we discuss **gene expression**—the process by which the nucleotide sequence of a gene is *transcribed* into the nucleotide sequence of an RNA molecule, which, in most cases, is then *translated* into the amino acid sequence of a protein (**Figure 5–9**).

The amount of information in an organism's DNA is staggering: written out in the four-letter nucleotide alphabet, the nucleotide sequence of a very small protein-coding gene from humans occupies a quarter of a page of text, while the complete human DNA sequence would fill more than 1000 books the size of this one. Herein lies a problem that affects the architecture of all eukaryotic chromosomes: how can all this information be packed neatly into every cell nucleus? In the remainder of this chapter, we discuss the answer to this question.

THE STRUCTURE OF EUKARYOTIC CHROMOSOMES

Large amounts of DNA are required to encode all the information needed to make even a single-celled bacterium, and far more DNA is needed to encode the information to make a multicellular organism like you. Each human cell contains about 2 meters (m) of DNA; yet the cell nucleus is only 5–8 μm in diameter. Tucking all this material into such a small space is the equivalent of trying to fold 40 km (24 miles) of extremely fine thread into a tennis ball.

In eukaryotic cells, very long double-stranded DNA molecules are packaged into **chromosomes**. These DNA molecules not only fit readily inside the nucleus, but, after they are replicated, they can be easily apportioned between the two daughter cells at each cell division. The complex task of packaging DNA is accomplished by specialized proteins that bind to and fold the DNA, generating a series of coils and loops that provide increasingly higher levels of organization and prevent the DNA from becoming a tangled, unmanageable mess. Amazingly, the DNA is compacted in a way that allows it to remain accessible to all of the enzymes and other proteins that replicate it, repair it, and control the expression of its genes.

Bacteria typically carry their genes on a single, circular DNA molecule. This molecule is also associated with proteins that condense the DNA, but these proteins differ from the ones that package eukaryotic DNA. Although this prokaryotic DNA is called a bacterial "chromosome," it does not have the same structure as eukaryotic chromosomes, and less is known about how it is packaged. Our discussion of chromosome structure in this chapter will therefore focus entirely on eukaryotic chromosomes.

Eukaryotic DNA Is Packaged into Multiple Chromosomes

In eukaryotes, such as ourselves, the DNA in the nucleus is distributed among a set of different chromosomes. The DNA in a human nucleus, for example, contains approximately 3.2×10^9 nucleotides parceled out into 23 or 24 different types of chromosome (males, with their Y chromosome, have an extra type of chromosome that females do not have). Each chromosome consists of a single, enormously long, linear DNA molecule associated with proteins that fold and pack the fine thread of DNA into a more compact structure. The complex of DNA and protein is called *chromatin*. In addition to the proteins involved in packaging the DNA,

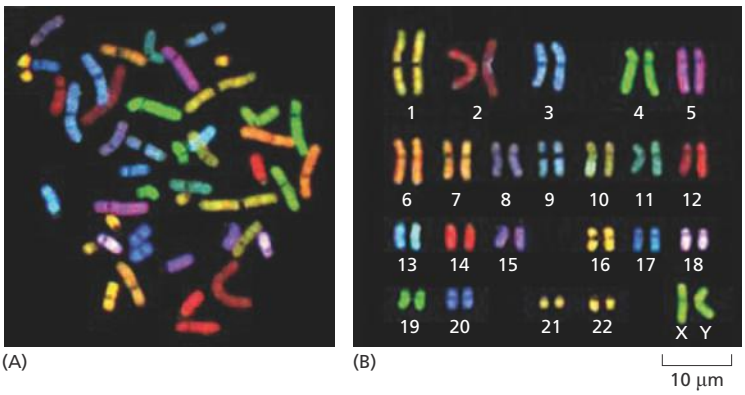


Figure 5–10 Each human chromosome can be "painted" a different color to allow its unambiguous identification. The chromosomes shown here were isolated from a cell undergoing nuclear division (mitosis) and are therefore in a highly compact (condensed) state. Chromosome painting is carried out by exposing the chromosomes to a collection of human DNA molecules that have been coupled to a combination of fluorescent dyes. For example, DNA molecules derived from Chromosome 1 are labeled with one specific dye combination, those from Chromosome 2 with another, and so on. Because the labeled DNA can form base pairs (hybridize) only to its chromosome of origin (discussed in Chapter 10), each chromosome is differently colored. For such experiments, the chromosomes are treated so that the individual strands of the double-helical DNA molecules partly separate to enable base-pairing with the labeled, single-stranded DNA, while keeping the chromosome structure relatively intact. (A) Micrograph shows the array of chromosomes as they originally spilled from the lysed cell. (B) The same chromosomes have been artificially lined up in order. In this so-called *karyotype*, the homologous chromosomes are numbered and arranged in pairs; the presence of a Y chromosome reveals that these chromosomes came from a male. (From E. Schröck et al., *Science* 273:494–497, 1996. With permission from the AAAS.)

chromosomes are also associated with many other proteins involved in DNA replication, DNA repair, and gene expression.

With the exception of the germ cells (sperm and eggs) and highly specialized cells that lack DNA entirely (such as mature red blood cells), human cells each contain two copies of each chromosome, one inherited from the mother and one from the father. The maternal and paternal chromosomes of a pair are called *homologous chromosomes* (*homologs*). The only nonhomologous chromosome pairs are the sex chromosomes in males, where a *Y chromosome* is inherited from the father and an *X chromosome* from the mother. (Females inherit one X chromosome from each parent and have no Y chromosome.)

In addition to being different sizes, the different human chromosomes can be distinguished from one another by a variety of techniques. Each chromosome can be "painted" a different color using sets of chromosome-specific DNA molecules coupled to different fluorescent dyes (**Figure 5–10**). This involves a technique called *DNA hybridization*, which takes advantage of complementary base-pairing, as we will describe in detail in Chapter 10. A more traditional way of distinguishing one chromosome from another is to stain the chromosomes with dyes that bind to certain types of DNA sequences. These dyes mainly distinguish between DNA that is rich in A-T nucleotide pairs and DNA that is G-C rich, and they produce a predictable pattern of bands along each type of chromosome. The patterns that result allow each chromosome to be identified and numbered.

An ordered display of the full set of 46 human chromosomes is called the human **karyotype** (see Figure 5–10). If parts of a chromosome are lost, or switched between chromosomes, these changes can be detected. Cytogeneticists analyze karyotypes to detect chromosomal abnormalities that are associated with some inherited defects (**Figure 5–11**) and with certain types of cancer.

Chromosomes Contain Long Strings of Genes

The most important function of chromosomes is to carry the genes—the functional units of heredity (**Figure 5–12**). A **gene** is often defined as a

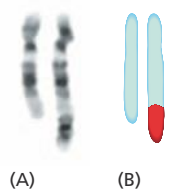


Figure 5–11 Abnormal chromosomes are associated with some inherited genetic defects. (A) A pair of Chromosomes 12 from a patient with inherited ataxia, a genetic disease of the brain characterized by progressive deterioration of motor skills. The patient has one normal Chromosome 12 (*left*) and one abnormally long Chromosome 12, which contains a piece of Chromosome 4 as identified by its banding pattern. (B) This interpretation was confirmed by chromosome painting, in which Chromosome 12 was painted *blue* and Chromosome 4 was painted *red*. (From E. Schröck et al., *Science* 273:494–497, 1996. With permission from the AAAS.)

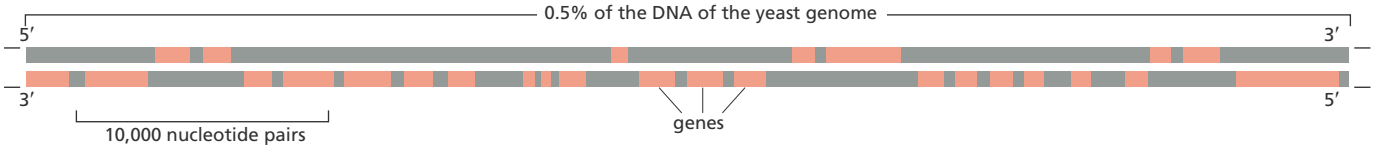


Figure 5–12 Genes are arranged along chromosomes. This figure shows a small region of the DNA double helix in one chromosome from the budding yeast *S. cerevisiae*. The *S. cerevisiae* genome contains about 12 million nucleotide pairs and 6600 genes—spread across 16 chromosomes. Note that, in each gene, only one of the two DNA strands actually encodes the information to make an RNA molecule, and this can be either strand, as indicated by the *light red* bars. However, a gene is generally denoted to contain both the "coding strand" and its complement, as in Figure 5–9. The high density of genes is characteristic of *S. cerevisiae*.

segment of DNA that contains the instructions for making a particular protein or RNA molecule. Most of the RNA molecules encoded by genes are subsequently used to produce a protein (see Figure 5–9). In some cases, however, the RNA molecule is the final product; like proteins, these RNA molecules have diverse functions in the cell, including structural, catalytic, and gene regulatory roles, as we discuss in later chapters.

Together, the total genetic information carried by all the chromosomes in a cell or organism constitutes its **genome**. Complete genome sequences have been determined for thousands of organisms, from *E. coli* to humans. As might be expected, some correlation exists between the complexity of an organism and the number of genes in its genome. For example, the total number of genes ranges from less than 500 for a simple bacterium to about 30,000 for humans. Bacteria and some single-celled eukaryotes, including *S. cerevisiae*, have especially compact genomes: the DNA molecules that make up their chromosomes are little more than strings of closely packed genes (see Figure 5–12). However, chromosomes from many eukaryotes—including humans—contain, in addition to genes and the specific nucleotide sequences required for normal gene expression, a large excess of interspersed DNA. This extra DNA is sometimes called "junk DNA," because the usefulness to the cell has not yet been demonstrated. Although the particular nucleotide sequence of most of this DNA might not be important, the DNA itself—acting as spacer material—may be crucial for the long-term evolution of the species and for the proper activity of the genes. In addition, comparisons of the genome sequences from many different species reveal that a portion of this extra DNA is highly conserved among related species, indicating that it serves an important function—although we don't yet know what that is.

In general, the more complex an organism, the larger is its genome. But this relationship does not always hold true. The human genome, for example, is 200 times larger than that of the yeast *S. cerevisiae*, but 30 times smaller than that of some plants and at least 60 times smaller than some species of amoeba (see Figure 1–40). Furthermore, how the DNA is apportioned over chromosomes also differs from one species to another. Humans have a total of 46 chromosomes (including both maternal and paternal sets), but a species of small deer has only 7, while some carp species have more than 100. Even closely related species with similar genome sizes can have very different numbers and sizes of chromosomes (**Figure 5–13**). Thus, although gene number is roughly correlated with species complexity, there is no simple relationship between gene number, chromosome number, and total genome size. The genomes and chromosomes of modern species have each been shaped by a unique history of seemingly random genetic events, acted on by specific selection pressures, as we discuss in Chapter 9.

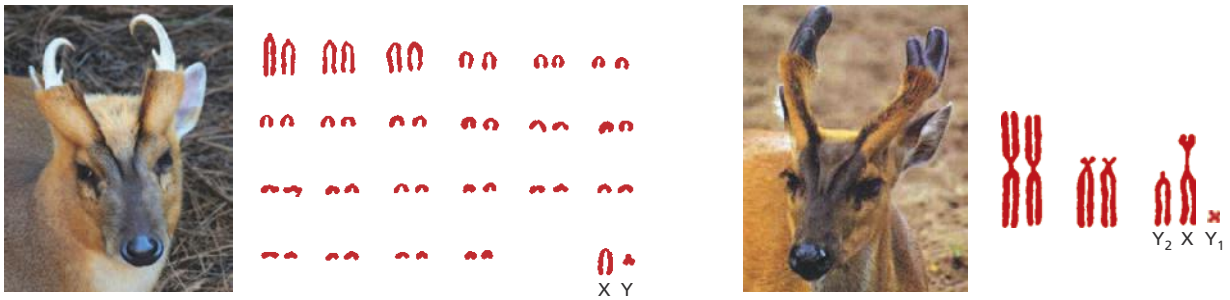


Figure 5–13 Two closely related species can have similar genome sizes but very different chromosome numbers. In the evolution of the Indian muntjac deer, chromosomes that were initially separate, and that remain separate in the Chinese species, fused without having a major effect on the number of genes—or the animal. (Courtesy of Deborah Carreno, Natural Wonders Photography.)

Specialized DNA Sequences Are Required for DNA Replication and Chromosome Segregation

To form a functional chromosome, a DNA molecule must do more than simply carry genes: it must be able to be replicated, and the replicated copies must be separated and partitioned equally and reliably into the two daughter cells at each cell division. These processes occur through an ordered series of events, known collectively as the **cell cycle**. This cycle of cell growth and division is briefly summarized in **Figure 5–14** and will be discussed in detail in Chapter 18. Only two broad stages of the cell cycle need concern us in this chapter: *interphase*, when chromosomes are duplicated, and *mitosis*, when they are distributed, or segregated, to the two daughter nuclei.

During interphase, the chromosomes are extended as long, thin, tangled threads of DNA in the nucleus and cannot be easily distinguished in the light microscope (see Figure 5–1). We refer to chromosomes in this extended state as *interphase chromosomes*. As we discuss in Chapter 6, specialized DNA sequences found in all eukaryotes ensure that DNA replication occurs efficiently during interphase. One type of nucleotide sequence acts as a **replication origin**, where replication of the DNA begins; eukaryotic chromosomes contain many replication origins to ensure that the long DNA molecules are replicated rapidly (**Figure 5–15**). Another DNA sequence forms the **telomeres** at each of the two ends of a chromosome. Telomeres contain repeated nucleotide sequences that are required for the ends of chromosomes to be replicated. They also cap the ends of the DNA molecule, preventing them from being mistaken by the cell as broken DNA in need of repair.

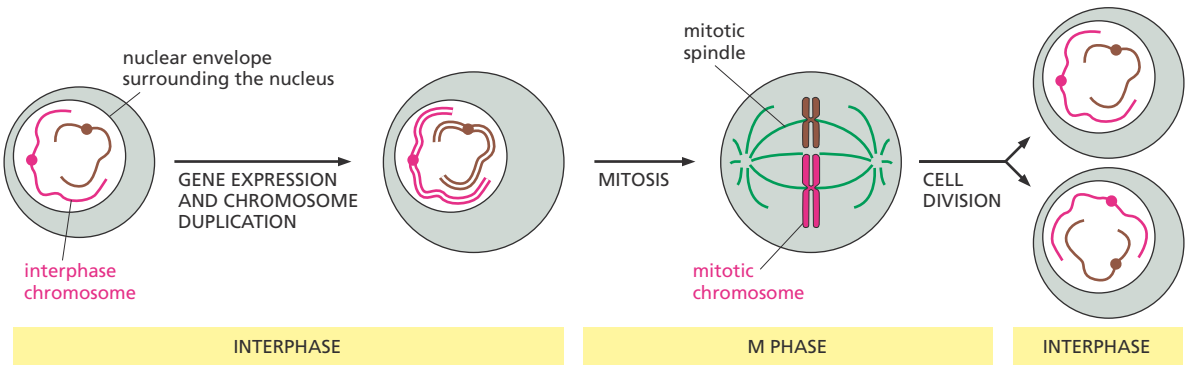


Figure 5–14 The duplication and segregation of chromosomes occurs through an ordered cell cycle in proliferating cells. During interphase, the cell expresses many of its genes, and—during part of this phase—it duplicates chromosomes. Once chromosome duplication is complete, the cell can enter *M phase*, during which nuclear division, or mitosis, occurs. In mitosis, the duplicated chromosomes condense, gene expression largely ceases, the nuclear envelope breaks down, and the mitotic spindle forms from microtubules and other proteins. The condensed chromosomes are then captured by the mitotic spindle, one complete set is pulled to each end of the cell, and a nuclear envelope forms around each chromosome set. In the final step of M phase, the cell divides to produce two daughter cells. Only two different chromosomes are shown here for simplicity.

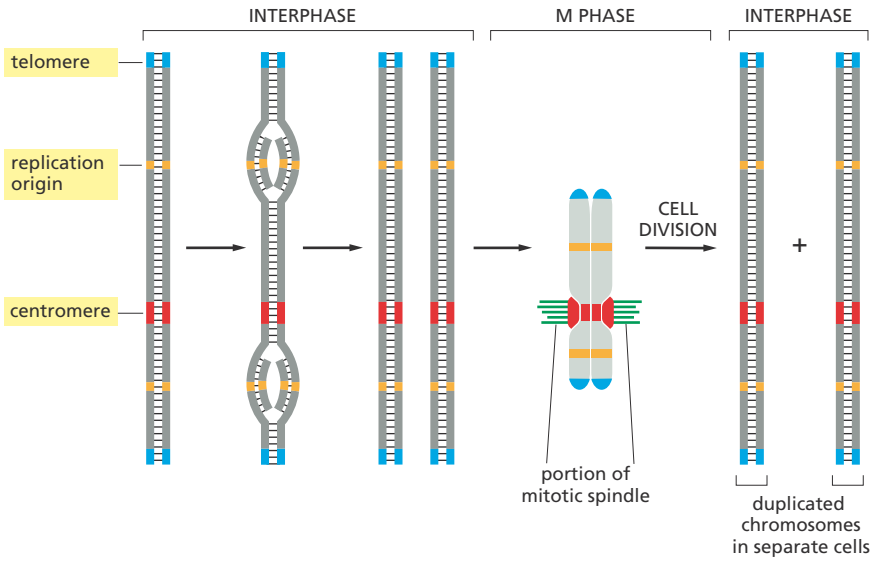


Figure 5–15 Three DNA sequence elements are needed to produce a eucaryotic chromosome that can be replicated and then segregated at mitosis. Each chromosome has multiple origins of replication, one centromere, and two telomeres. The sequence of events that a typical chromosome follows during the cell cycle is shown schematically. The DNA replicates in interphase, beginning at the origins of replication and proceeding bidirectionally from the origins across the chromosome. In M phase, the centromere attaches the duplicated chromosomes to the mitotic spindle so that one copy is distributed to each daughter cell when the cell divides. Prior to cell division, the centromere also helps to hold the compact, duplicated chromosomes together until they are ready to be pulled apart. Telomeres, which form special caps at the tips of each chromosome, aid in the replication of chromosome ends.

Eukaryotic chromosomes also contain a third type of specialized DNA sequence, called the **centromere**, that allows duplicated chromosomes to be separated during M phase (see Figure 5–15). During this stage of the cell cycle, the DNA coils up, adopting a more and more compact structure, ultimately forming highly compacted, or condensed, *mitotic chromosomes*. This is the state in which the duplicated chromosomes can be most easily visualized (**Figure 5–16** and see Figures 5–1 and 5–14). Once the chromosomes have condensed, the centromere attaches the mitotic spindle to each duplicated chromosome in a way that allows one copy of each chromosome to be segregated to each daughter cell (see Figure 5–15B). We describe the central role that centromeres play in cell division in Chapter 18.

Interphase Chromosomes Are Not Randomly Distributed Within the Nucleus

Inside the nucleus, the interphase chromosomes—although longer and finer than mitotic chromosomes—are nonetheless organized in various

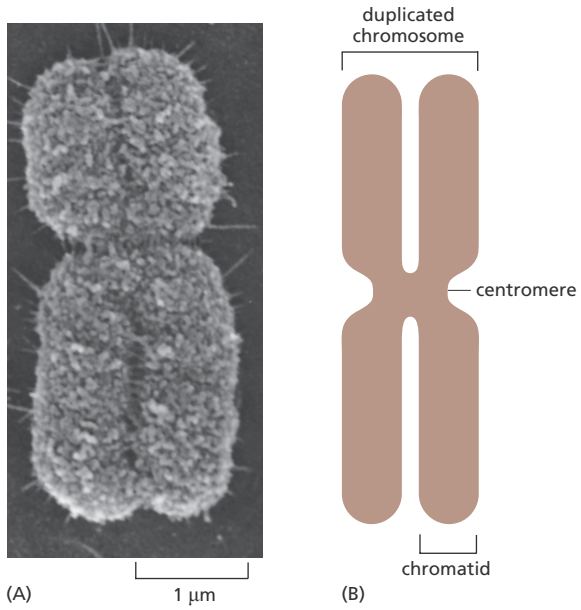


Figure 5–16 A typical duplicated mitotic chromosome is highly compact. Because DNA is replicated during interphase, each duplicated mitotic chromosome contains two identical daughter DNA molecules (see Figure 5–15A). Each of these very long DNA molecules, with its associated proteins, is called a *chromatid*; once the two sister chromatids separate, they are considered individual chromosomes. (A) A scanning electron micrograph of a mitotic chromosome. The two chromatids are tightly joined together. The constricted region reveals the position of the centromere. (B) A cartoon representation of a mitotic chromosome. (A, courtesy of Terry D. Allen.)

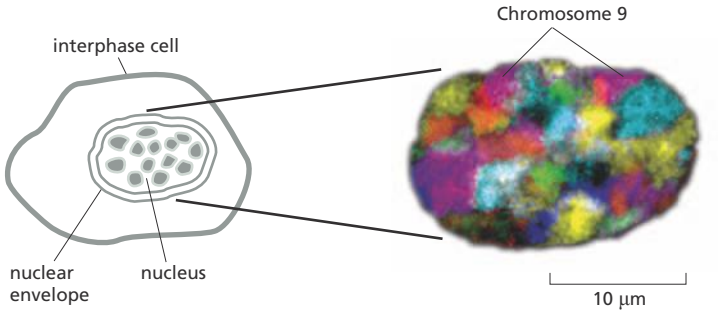


Figure 5–17 Interphase chromosomes occupy their own distinct territories within the nucleus. DNA probes coupled with different fluorescent markers were used to paint individual interphase chromosomes in a human cell. Viewed in a fluorescence microscope, each interphase chromosome is seen to occupy its own discrete territory within the nucleus, rather than being mixed with the other chromosomes like spaghetti in a bowl. Note that pairs of homologous chromosomes, such as the two copies of Chromosome 9 indicated, are not generally located in the same position. (From M.R. Speicher and N.P. Carter, *Nat. Rev. Genet.* 6:782–792, 2005. With permission from Macmillan Publishers Ltd.)

ways. First, each chromosome tends to occupy a particular region of the interphase nucleus, and so different chromosomes do not become extensively entangled with one another (**Figure 5–17**). In addition, some chromosomes are attached to particular sites on the *nuclear envelope*—the pair of concentric membranes that surround the nucleus—or to the underlying *nuclear lamina,* the protein meshwork that supports the envelope (discussed in Chapter 17).

The most obvious example of chromosome organization in the interphase nucleus is the **nucleolus** (**Figure 5–18**). The nucleolus is where the parts of the different chromosomes carrying genes that encode *ribosomal RNAs* cluster together. Here, ribosomal RNAs are synthesized and combine with proteins to form ribosomes, the cell's protein-synthesizing machines. As we discuss in Chapter 7, ribosomal RNAs play both structural and catalytic roles in the ribosome.

The DNA in Chromosomes Is Always Highly Condensed

As we have seen, all eukaryotic cells, whether in interphase or mitosis, package their DNA tightly into chromosomes. Human Chromosome 22, for example, contains about 48 million nucleotide pairs; stretched out end-to-end, its DNA would extend about 1.5 cm. Yet, during mitosis, Chromosome 22 measures only about 2 μm in length—that is, nearly 10,000 times more compact than the DNA would be if it were stretched to its full length. This remarkable feat of compression is performed by proteins that coil and fold the DNA into higher and higher levels of organization. The DNA of interphase chromosomes, although about 20 times less condensed than that of mitotic chromosomes (**Figure 5–19**), is still packed tightly.

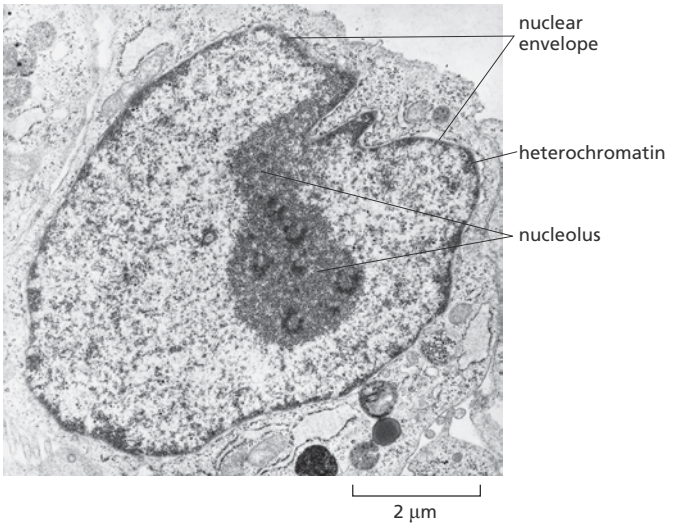


Figure 5–18 The nucleolus is the most prominent structure in the interphase nucleus. Electron micrograph of a thin section through the nucleus of a human fibroblast. The nucleus is surrounded by the nuclear envelope. Inside the nucleus, the chromatin appears as a diffuse speckled mass, with regions that are especially dense, called heterochromatin (dark staining). Heterochromatin contains few genes and is located mainly around the periphery of the nucleus, immediately under the nuclear envelope. The large dark region is the nucleolus, which contains the genes for ribosomal RNAs; these genes are located on multiple chromosomes but are clustered together in the nucleolus. (Courtesy of E.G. Jordan and J. McGovern.)

In the next sections, we introduce the specialized proteins that make this compression possible. Bear in mind, though, that chromosome structure is dynamic. Not only do chromosomes condense and decondense during the cell cycle, but chromosome packaging must be flexible enough to allow rapid, on-demand access to different regions of the interphase chromosome, unpacking enough to allow protein complexes access to specific, localized DNA sequences for replication, repair, or gene expression.

Nucleosomes Are the Basic Units of Eukaryotic Chromosome Structure

The proteins that bind to DNA to form eukaryotic chromosomes are traditionally divided into two general classes: the **histones** and the *nonhistone chromosomal proteins*. Histones are present in enormous quantities (more than 60 million molecules of several different types in each cell), and their total mass in chromosomes is about equal to that of the DNA itself. The complex of both classes of protein with nuclear DNA is called **chromatin**.

Histones are responsible for the first and most fundamental level of chromatin packing, the **nucleosome**, which was discovered in 1974. When interphase nuclei are broken open very gently and their contents examined with an electron microscope, much of the chromatin is in the form of *chromatin fibers* with a diameter of about 30 nm (**Figure 5–20A**). If this chromatin is subjected to treatments that cause it to unfold partially, it can then be seen in the electron microscope as a series of "beads on a string" (**Figure 5–20B**). The string is DNA, and each bead is a *nucleosome core particle*, which consists of DNA wound around a core of proteins formed from histones.

The structure of the nucleosome core particle was determined after first isolating nucleosomes by treating chromatin in its unfolded, "beads on a string" form with enzymes called nucleases, which break down DNA by cutting the phosphodiester bonds between nucleotides. After digestion for a short period, only the exposed DNA between the core particles—the *linker DNA*—is degraded, allowing the core particles to be isolated. An individual nucleosome core particle consists of a complex of eight histone proteins—two molecules each of histones H2A, H2B, H3, and H4—and a stretch of double-stranded DNA, 147 nucleotide pairs long, that winds around this *histone octamer* (**Figure 5–21**). The high-resolution structure of the nucleosome core particle was solved in 1997, revealing in atomic detail the disc-shaped histone octamer around which the DNA is tightly wrapped, making 1.7 turns in a left-handed coil (**Figure 5–22**).

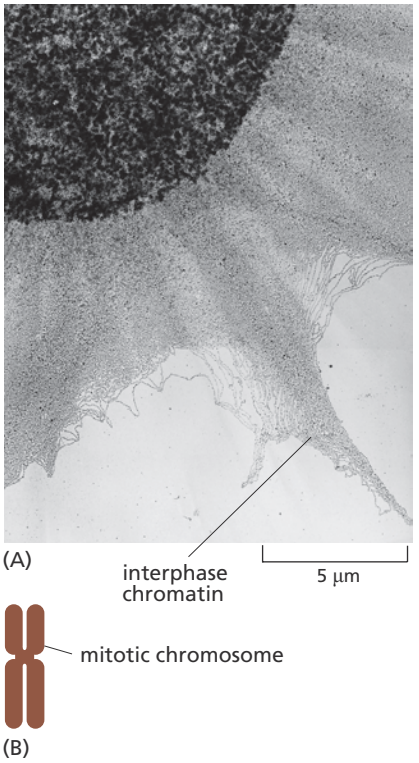


Figure 5–19 DNA in interphase chromosomes is less compact than in mitotic chromosomes. (A) An electron micrograph showing an enormous tangle of chromatin (DNA with its associated proteins) spilling out of a lysed interphase nucleus. (B) Schematic drawing of a human mitotic chromosome drawn to the same scale. (Courtesy of Victoria Foe.)

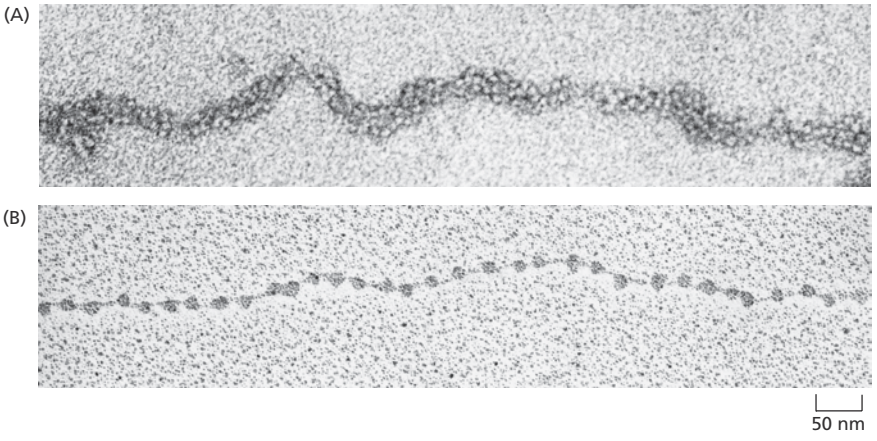


Figure 5–20 Nucleosomes can be seen in the electron microscope. (A) Chromatin isolated directly from an interphase nucleus appears in the electron microscope as a chromatin fiber about 30-nm thick; a part of one such fiber is shown here. (B) This electron micrograph shows a length of a chromatin fiber that has been experimentally unpacked, or decondensed, after isolation to show the "beads-on-a-string" appearance of the nucleosomes. (A, courtesy of Barbara Hamkalo; B, courtesy of Victoria Foe.)

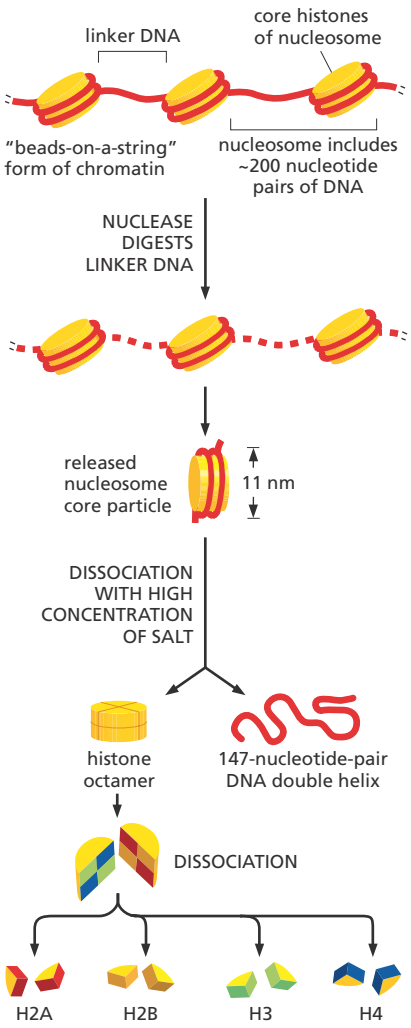


Figure 5–21 Nucleosomes contain DNA wrapped around a protein core of eight histone molecules. In a test tube, the nucleosome core particle can be released from chromatin by digestion of the linker DNA with a nuclease, which degrades the exposed DNA but not the DNA wound tightly around the nucleosome core. The DNA around each isolated nucleosome core particle can then be released and its length determined. With 147 nucleotide pairs in each fragment, the DNA wraps almost twice around each histone octamer.

The linker DNA between each nucleosome core particle can vary in length from a few nucleotide pairs up to about 80. (The term nucleosome technically refers to a nucleosome core particle plus one of its adjacent DNA linkers, as shown in Figure 5–21, but it is often used to refer to the nucleosome core particle itself.) The formation of nucleosomes converts a DNA molecule into a chromatin thread that is approximately one-third the length of the initial piece of DNA, and it provides the first level of DNA packing.

All four of the histones that make up the octamer are relatively small proteins, with a high proportion of positively charged amino acids (lysine and arginine). The positive charges help the histones bind tightly to the negatively charged sugar–phosphate backbone of DNA. These numerous electrostatic interactions explain in part why DNA of virtually any sequence can bind to a histone octamer. Each of the histones in the

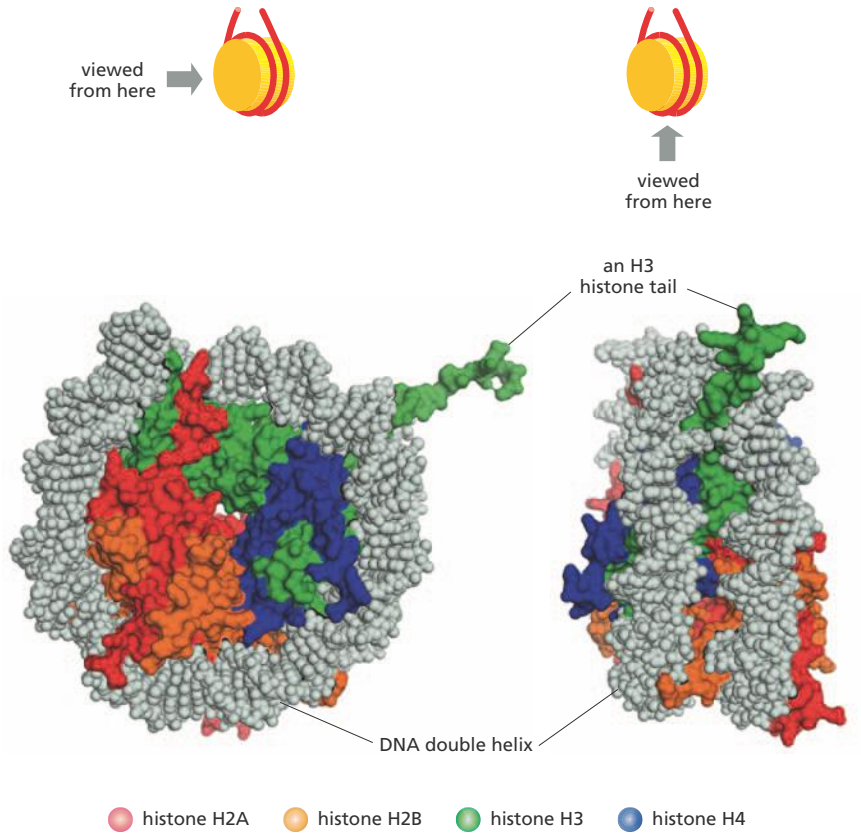


Figure 5–22 The structure of the nucleosome core particle, as determined by X-ray diffraction analysis, reveals how DNA is tightly wrapped around a disc-shaped histone octamer. Two views of a nucleosome core particle are shown here. The two strands of the DNA double helix are shown in *gray*. A portion of an H3 histone tail (*green*) can be seen extending from the nucleosome core particle, but the tails of the other histones have been truncated. (Reprinted by permission from K. Luger et al., *Nature* 389:251–260, 1997. With permission from Macmillan Publishers Ltd.)

octamer also has a long, unstructured N-terminal amino acid "tail" that extends out from the nucleosome core particle (see Figure 5–22). These histone tails are subject to several types of reversible, covalent chemical modifications that control many aspects of chromatin structure.

The histones that form the nucleosome core are among the most highly conserved of all known eukaryotic proteins: there are only two differences between the amino acid sequences of histone H4 from peas and cows, for example. This extreme evolutionary conservation reflects the vital role of histones in controlling eukaryotic chromosome structure.

Chromosome Packing Occurs on Multiple Levels

Although long strings of nucleosomes form on most chromosomal DNA, chromatin in the living cell rarely adopts the extended beads-on-a-string form seen in Figure 5–20B. Instead, the nucleosomes are further packed on top of one another to generate a more compact structure, such as the chromatin fiber shown in Figure 5–20A and **Movie 5.2**. This additional packing of nucleosomes into a chromatin fiber depends on a fifth histone called histone H1, which is thought to pull adjacent nucleosomes together into a regular repeating array. This "linker" histone changes the path the DNA takes as it exits the nucleosome core, allowing it to form a more condensed chromatin fiber (**Figure 5–23**).

We saw earlier that during mitosis chromatin becomes so highly condensed that individual chromosomes can be seen in the light microscope. How is a chromatin fiber folded to produce mitotic chromosomes? The answer is not yet known in detail, but it is known that the chromatin fiber is folded into a series of loops, and that these loops are further condensed to produce the interphase chromosome; finally, this compact string of loops is thought to undergo at least one more level of packing to form the mitotic chromosome (**Figure 5–24** and **Figure 5–25**).

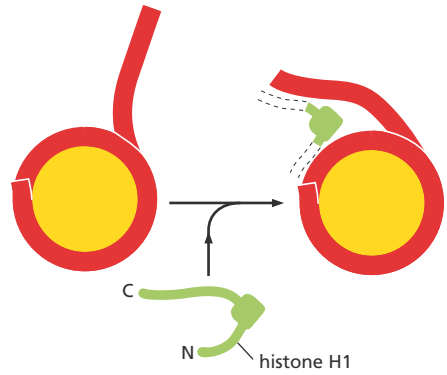


Figure 5–23 A linker histone helps to pull nucleosomes together and pack them into a more compact chromatin fiber. Histone H1 consists of a globular region plus a pair of long tails at its C-terminal and N-terminal ends. The globular region constrains an additional 20 base pairs of the DNA where it exits from the nucleosome core, an activity that is thought to be important for the formation of the chromatin fiber. The long C-terminal tail is required for H1 to bind to chromatin. The positions of the C-terminal and N-terminal tails in the nucleosome are not known.

QUESTION 5–2

Assuming that the histone octamer (shown in Figure 5–21) forms a cylinder 9 nm in diameter and 5 nm in height and that the human genome forms 32 million nucleosomes, what volume of the nucleus (6 μm in diameter) is occupied by histone octamers? (Volume of a cylinder is $\pi r^2 h$; volume of a sphere is $4/3\ \pi r^3$.) What fraction of the total volume of the nucleus do the histone octamers occupy? How does this compare with the volume of the nucleus occupied by human DNA?

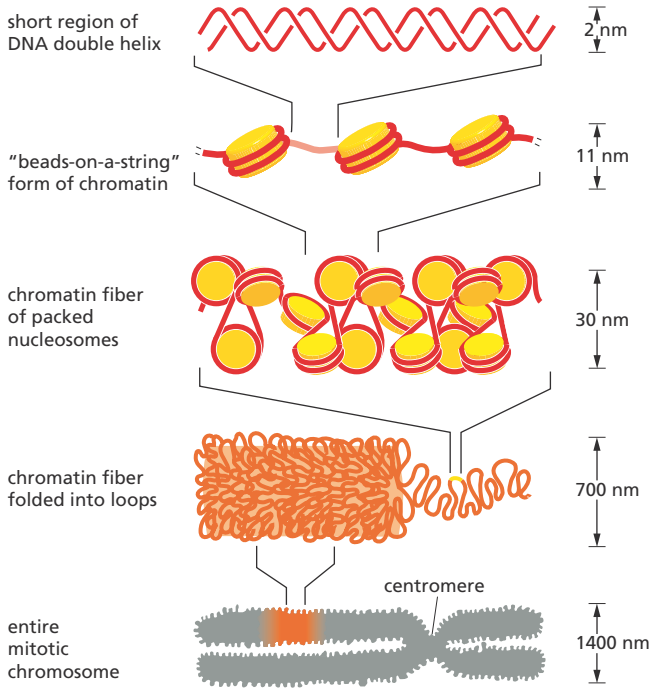


Figure 5–24 DNA packing occurs on several levels in chromosomes. This schematic drawing shows some of the levels thought to give rise to the highly condensed mitotic chromosome. The actual structures are still uncertain.

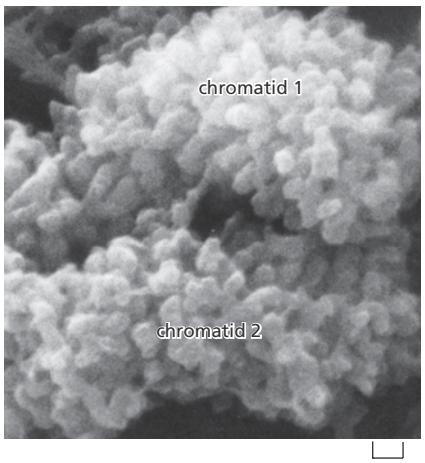


Figure 5–25 The mitotic chromosome contains chromatin that is packed especially tightly. This scanning electron micrograph shows a region near one end of a typical mitotic chromosome. Each knoblike projection is believed to represent the tip of a separate loop of chromatin. The chromosome has duplicated, forming two sister chromatids that are still held close together (see Figure 5–16). The ends of the two chromatids can be distinguished on the right of the photo. (From M.P. Marsden and U.K. Laemmli, *Cell* 17:849–858, 1989. With permission from Elsevier.)

THE REGULATION OF CHROMOSOME STRUCTURE

So far, we have discussed how DNA is packed tightly into chromatin. We now turn to the question of how this packaging can be regulated to allow rapid access to the underlying DNA. The DNA in cells carries enormous amounts of coded information, and cells must be able to get to this information as needed.

In this section, we discuss how a cell can alter its chromatin structure to expose localized regions of DNA and allow access to specific proteins and protein complexes, particularly those involved in gene expression and in DNA replication and repair. We then discuss how chromatin structure is established and maintained—and how a cell can pass on some forms of this structure to its descendants. The regulation and inheritance of chromatin structure play crucial parts in the development of eukaryotic organisms.

QUESTION 5–3

Histone proteins are among the most highly conserved proteins in eukaryotes. Histone H4 proteins from a pea and a cow, for example, differ in only 2 of 102 amino acids. Comparison of the gene sequences shows many more differences, but only two change the amino acid sequence. These observations indicate that mutations that change amino acids must have been selected against during evolution. Why do you suppose that amino-acid-altering mutations in histone genes are deleterious?

Changes in Nucleosome Structure Allow Access to DNA

Eukaryotic cells have several ways to adjust the local structure of their chromatin rapidly. One way takes advantage of **chromatin-remodeling complexes**, protein machines that use the energy of ATP hydrolysis to change the position of the DNA wrapped around nucleosomes (**Figure 5–26A**). The complexes, which attach to both the histone octamer and the DNA wrapped around it, can locally alter the arrangement of nucleosomes on the DNA, making the DNA either more accessible (**Figure 5–26B**) or less accessible to other proteins in the cell. During mitosis, many of the chromatin-remodeling complexes are inactivated, which may help mitotic chromosomes maintain their tightly packed structure.

Another way of altering chromatin structure relies on the reversible chemical modification of the histones. The tails of all four of the core histones are particularly subject to these covalent modifications (**Figure 5–27A**). For example, acetyl, phosphate, or methyl groups can be added to and removed from the tails by enzymes that reside in the nucleus (**Figure 5–27B**). These and other modifications can have important consequences for the stability of the chromatin fiber. Acetylation of lysines, for instance, can reduce the affinity of the tails for adjacent nucleosomes, thereby loosening chromatin structure and allowing access to particular nuclear proteins.

Most importantly, however, these modifications can serve as docking sites on the histone tails for a variety of regulatory proteins. Different patterns of modifications attract different proteins to particular stretches of chromatin. Some of these proteins promote chromatin condensation, whereas others decondense chromatin and facilitate access to the DNA. Specific combinations of tail modifications and the proteins that bind to them have different meanings for the cell: one pattern, for example, indicates that a particular stretch of chromatin has been newly replicated;

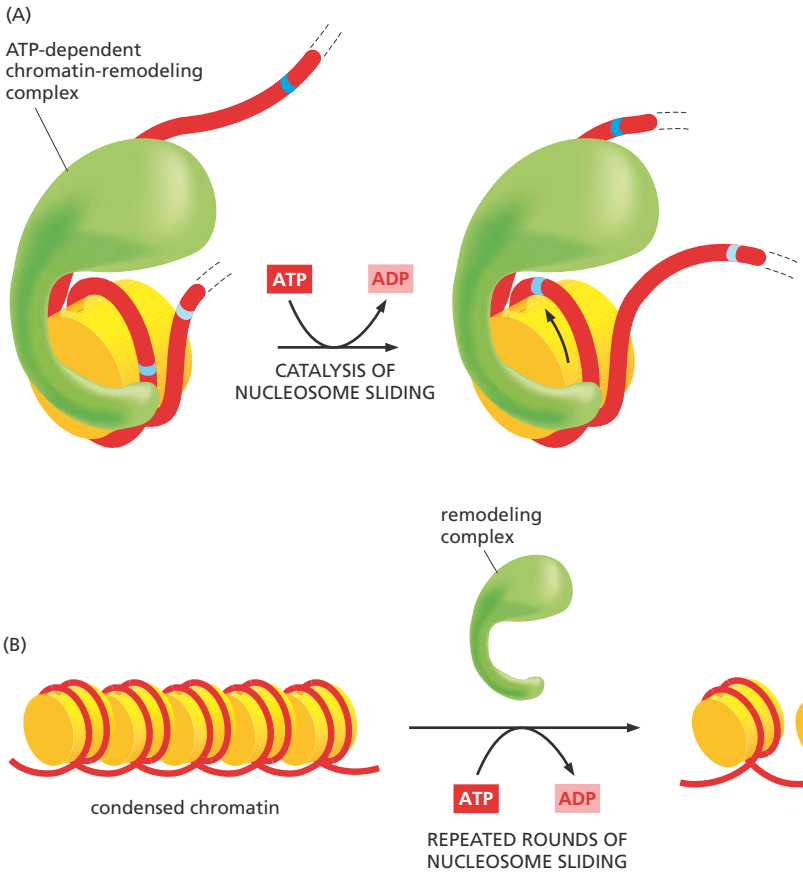


Figure 5–26 Chromatin-remodeling complexes locally reposition the DNA wrapped around nucleosomes. (A) The complexes use energy derived from ATP hydrolysis to loosen the nucleosomal DNA and push it along the histone octamer, thereby exposing the DNA to other DNA-binding proteins. The *blue stripes* have been added to show how the nucleosome moves along the DNA. Many cycles of ATP hydrolysis are required to produce such a shift. (B) In the case shown, the repositioning of nucleosomes decondenses the chromatin in a particular chromosomal region; in other cases, it condenses the chromatin.

another indicates that the genes in that stretch of chromatin should be expressed; still others indicate that the nearby genes should be silenced (**Figure 5–27C**).

Like the chromatin-remodeling complexes, the enzymes that modify histone tails are tightly regulated. They are brought to particular chromatin regions mainly by interactions with proteins that bind to specific

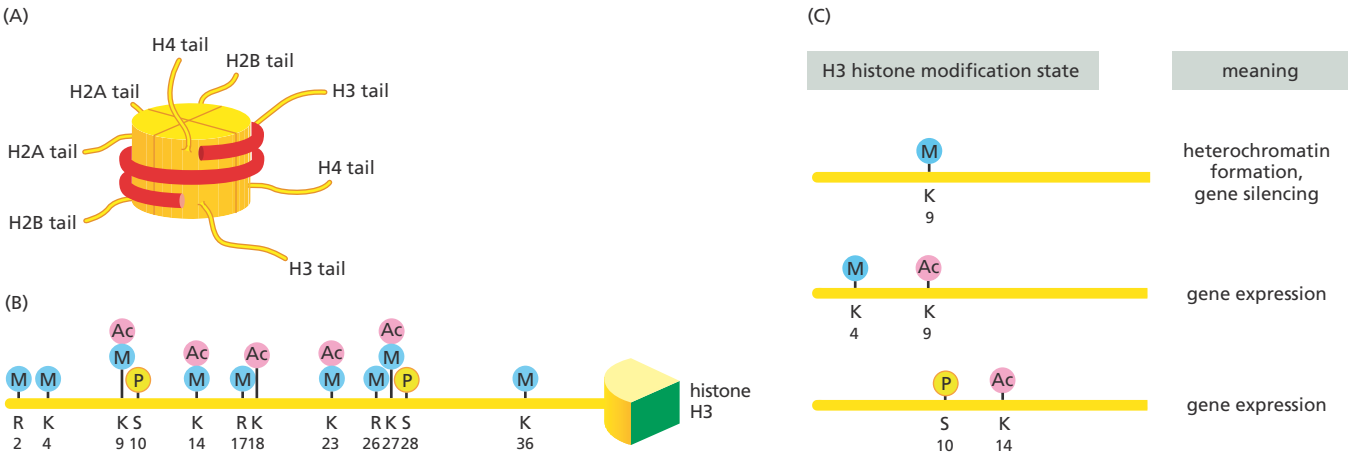


Figure 5–27 The pattern of modification of histone tails can dictate how a stretch of chromatin is treated by the cell. (A) Schematic drawing showing the positions of the histone tails that extend from each nucleosome. (B) Each histone can be modified by the covalent attachment of a number of different chemical groups, mainly to the tails. Histone H3, for example, can receive an acetyl group (Ac), a methyl group (M), or a phosphate group (P). The numbers denote the positions of the modified amino acids in the protein chain, with each amino acid designated by its one-letter code. Note that some positions, such as lysines (K) 9, 14, 23, and 27, can be modified in more than one way. Moreover, lysines can be modified with either one, two, or three methyl groups (not shown). Note that histone H3 contains 135 amino acids, most of which are in its globular portion (*green*), and that most modifications are on its N-terminal tail (*orange*). (C) Different combinations of histone tail modifications can confer a specific meaning on the stretch of chromatin on which they occur, as indicated. Only a few of these "meanings" are known.

sequences in DNA (we discuss these proteins in Chapter 8). The histone-modifying enzymes work in concert with the chromatin-remodeling complexes to condense or decondense stretches of chromatin, allowing local chromatin structure to change rapidly according to the needs of the cell.

Interphase Chromosomes Contain Both Condensed and More Extended Forms of Chromatin

The localized alteration of chromatin packing by remodeling complexes and histone modification has important effects on the large-scale structure of interphase chromosomes. Interphase chromatin is not uniformly packed. Instead, regions of the chromosome that contain genes that are being expressed are generally more extended, while those that contain silent genes are more condensed. Thus, the detailed structure of an interphase chromosome can differ from one cell type to the next, helping to determine which genes are expressed. Most cell types express about 20 to 30 % of the genes they contain.

The most highly condensed form of interphase chromatin is called **heterochromatin** (from the Greek *heteros*, "different," plus chromatin). It was first observed in the light microscope in the 1930s as discrete, strongly staining regions within the mass of chromatin. Heterochromatin typically makes up about 10% of an interphase chromosome, and in mammalian chromosomes, it is concentrated around the centromere region and in the telomeres at the ends of the chromosomes (see Figure 5–15).

The rest of the interphase chromatin is called **euchromatin** (from the Greek *eu*, "true" or "normal," plus chromatin). Although we use the term euchromatin to refer to chromatin that exists in a more decondensed state than heterochromatin, it is now clear that both euchromatin and heterochromatin are composed of mixtures of different chromatin structures (**Figure 5–28**).

Each type of chromatin structure is established and maintained by different sets of histone tail modifications that attract distinct sets of non-histone proteins. The modifications that direct the formation of the most common type of heterochromatin, for example, include the methylation of lysine 9 in histone H3 (see Figure 5–27). Once it has been established, heterochromatin can spread because these histone tail modifications attract a set of heterochromatin-specific proteins, including histone-modifying enzymes, which then create the same histone tail modifications on adjacent nucleosomes. These modifications in turn recruit more of the heterochromatin-specific proteins, causing a wave of condensed chromatin to propagate along the chromosome. This heterochromatin will continue to spread until it encounters a barrier DNA sequence that stops the propagation (**Figure 5–29**). In this manner, extended regions of heterochromatin can be established along the DNA.

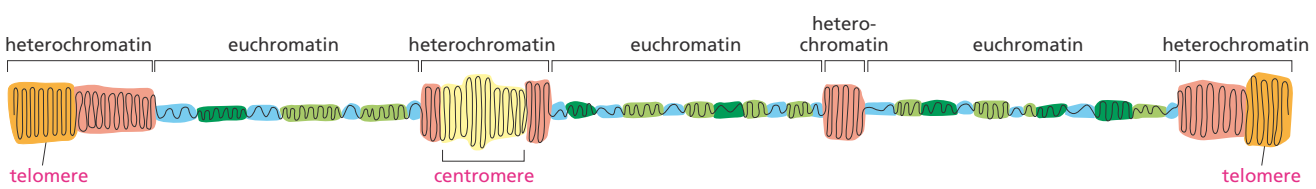


Figure 5–28 The structure of chromatin varies along a single interphase chromosome. As schematically indicated by different colors (and the path of the DNA molecule represented by the central *black line*), heterochromatin and euchromatin each represent a set of different chromatin structures with different degrees of condensation. Overall, heterochromatin is more condensed than euchromatin.

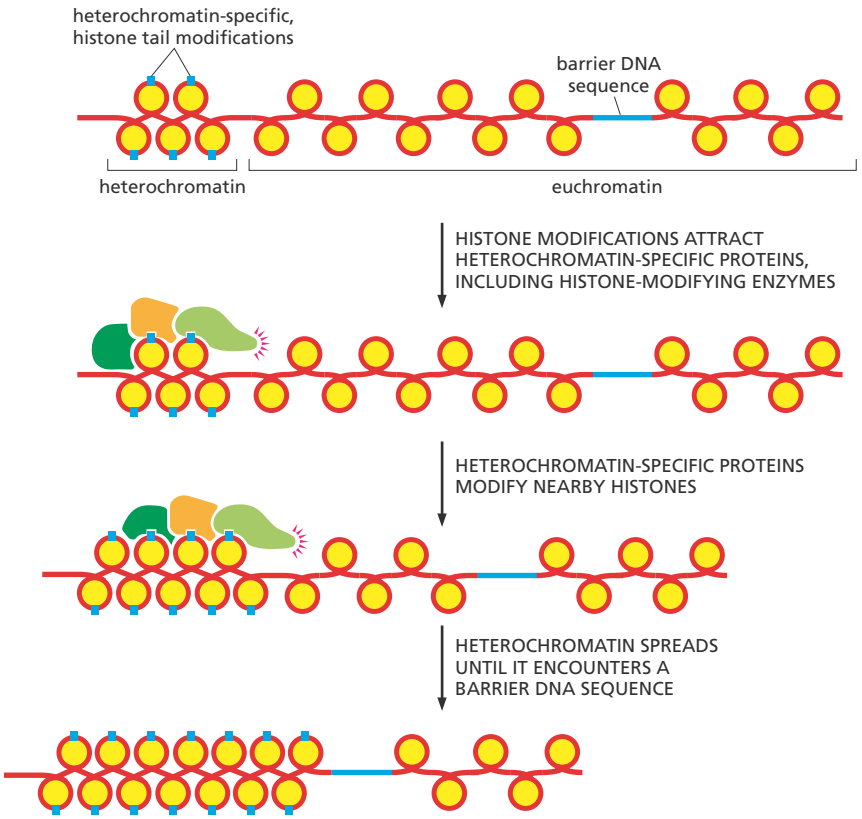


Figure 5–29 Heterochromatin-specific modifications allow heterochromatin to form and to spread. These modifications attract heterochromatin-specific proteins that reproduce the same modifications on neighboring histones. In this manner, heterochromatin can spread until it encounters a barrier DNA sequence that blocks its propagation into regions of euchromatin.

Most DNA that is permanently folded into heterochromatin in the cell does not contain genes. Because heterochromatin is so compact, genes that accidentally become packaged into heterochromatin usually fail to be expressed. Such inappropriate packaging of genes in heterochromatin can cause disease: in humans, the gene that encodes β-globin—which forms part of the oxygen-carrying hemoglobin molecule—is situated next to a region of heterochromatin. If, because of an inherited DNA deletion, that heterochromatin spreads, the β-globin gene is poorly expressed and the person develops a severe form of anemia.

Perhaps the most striking example of the use of heterochromatin to keep genes shut down, or *silenced*, is found in the interphase X chromosomes of female mammals. In mammals, female cells contain two X chromosomes, whereas male cells contain one X and one Y. Because a double dose of X-chromosome products would be lethal, female mammals have evolved a mechanism for permanently inactivating one of the two X chromosomes in each cell. At random, one or other of the two X chromosomes in each cell becomes highly condensed into heterochromatin early in embryonic development. Thereafter, the condensed and inactive state of that X chromosome is inherited in all of the many descendants of those cells (**Figure 5–30**).

When a cell divides, it generally passes on its histone modifications, chromatin structure, and gene expression patterns to the two daughter cells. Such "cell memory" is critical for the establishment and maintenance of different cell types during the development of a complex multicellular organism. We discuss the mechanisms involved in cell memory in Chapter 8, where we consider the control of gene expression.

QUESTION 5–4

Mutations in a particular gene on the X chromosome result in color blindness in men. By contrast, most women carrying the mutation have proper color vision but see colored objects with reduced resolution, as though functional cone cells (the photoreceptor cells responsible for color vision) are spaced farther apart than normal in the retina. Can you give a plausible explanation for this observation? If a woman is color-blind, what could you say about her father? About her mother? Explain your answers.

Figure 5–30 One of the two X chromosomes is inactivated in the cells of mammalian females by heterochromatin formation. Each female cell contains two X chromosomes, one from the mother (X_m) and the other from the father (X_p). At an early stage in embryonic development, one of these two chromosomes becomes condensed into heterochromatin in each cell, apparently at random. At each cell division, the same X chromosome becomes condensed (and inactivated) in all the descendants of that original cell. Thus, all mammalian females end up as mixtures (mosaics) of cells bearing maternal or paternal inactivated X chromosomes. In most of their tissues and organs, about half the cells will be of one type, and the other half will be of the other.

ESSENTIAL CONCEPTS

- Life depends on the stable storage and inheritance of genetic information.
- Genetic information is carried by very long DNA molecules and is encoded in the linear sequence of four nucleotides: A, T, G, and C.
- Each molecule of DNA is a double helix composed of a pair of antiparallel, complementary DNA strands, which are held together by hydrogen bonds between G-C and A-T base pairs.
- The genetic material of a eukaryotic cell is contained in a set of chromosomes, each formed from a single, enormously long DNA molecule that contains many genes.
- When a gene is expressed, part of its nucleotide sequence is transcribed into RNA molecules, many of which are translated into protein.
- The DNA that forms each eukaryotic chromosome contains, in addition to genes, many replication origins, one centromere, and two telomeres. These special DNA sequences ensure that, before cell division, each chromosome can be duplicated efficiently, and that the resulting daughter chromosomes are parceled out equally to the two daughter cells.
- In eukaryotic chromosomes, the DNA is tightly folded by binding to a set of histone and nonhistone proteins. This complex of DNA and protein is called chromatin.
- Histones pack the DNA into a repeating array of DNA–protein particles called nucleosomes, which further fold up into even more compact chromatin structures.

- A cell can regulate its chromatin structure—temporarily decondensing or condensing particular regions of its chromosomes—using chromatin-remodeling complexes and enzymes that covalently modify histone tails in various ways.
- The loosening of chromatin to a more decondensed state allows proteins involved in gene expression, DNA replication, and DNA repair to gain access to the necessary DNA sequences.
- Some forms of chromatin have a pattern of histone tail modification that causes the DNA to become so highly condensed that its genes cannot be expressed to produce RNA; such condensation occurs on all chromosomes during mitosis and in the heterochromatin of interphase chromosomes.

KEY TERMS

base pair
cell cycle
centromere
chromatin
chromatin-remodeling complex
chromosome
complementary
deoxyribonucleic acid (DNA)
double helix
euchromatin
gene
gene expression
genetic code
genome
heterochromatin
histone
karyotype
nucleolus
nucleolus
replication origin
telomere gene

QUESTIONS

QUESTION 5–5

A. The nucleotide sequence of one DNA strand of a DNA double helix is

5′-GGATTTTTGTCCACAATCA-3′.

What is the sequence of the complementary strand?

B. In the DNA of certain bacterial cells, 13% of the nucleotides are adenine. What are the percentages of the other nucleotides?

C. How many possible nucleotide sequences are there for a stretch of DNA that is *N* nucleotides long, if it is (a) single-stranded or (b) double-stranded?

D. Suppose you had a method of cutting DNA at specific sequences of nucleotides. How many nucleotides long (on average) would such a sequence have to be in order to make just one cut in a bacterial genome of 3×10^6 nucleotide pairs? How would the answer differ for the genome of an animal cell that contains 3×10^9 nucleotide pairs?

QUESTION 5–6

An A-T base pair is stabilized by only two hydrogen bonds. Hydrogen-bonding schemes of very similar strengths can also be drawn between other base combinations that normally do not occur in DNA molecules, such as the A-C and the A-G pairs shown in **Figure Q5–6**.

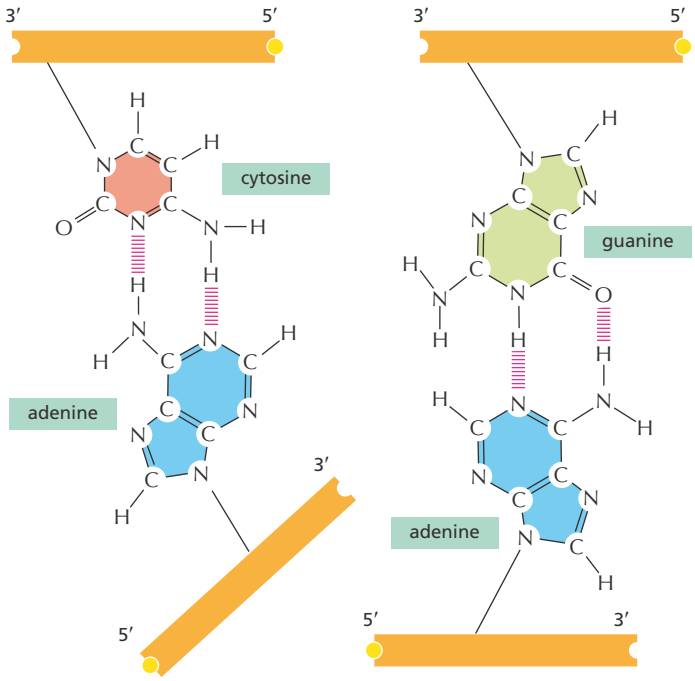


Figure Q5–6

What would happen if these pairs formed during DNA replication and the inappropriate bases were incorporated? Discuss why this does not often happen. (Hint: see Figure 5–6.)

QUESTION 5–7

A. A macromolecule isolated from an extraterrestrial source superficially resembles DNA, but closer analysis reveals that the bases have quite different structures (**Figure Q5–7**). Bases V, W, X, and Y have replaced bases A, T, G, and C. Look at these structures closely. Could these DNA-like molecules have been derived from a living organism that uses principles of genetic inheritance similar to those used by organisms on Earth?

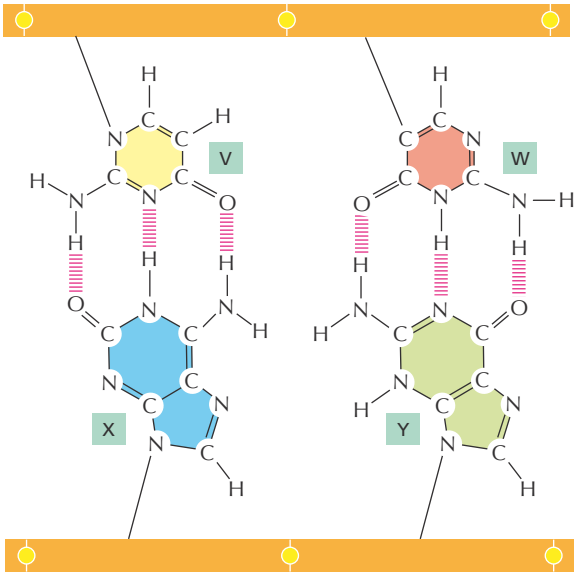


Figure Q5–7

B. Simply judged by their potential for hydrogen-bonding, could any of these extraterrestrial bases replace terrestrial A, T, G, or C in terrestrial DNA? Explain your answer.

QUESTION 5–8

The two strands of a DNA double helix can be separated by heating. If you raised the temperature of a solution containing the following three DNA molecules, in what order do you suppose they would "melt"? Explain your answer.

A. 5′-GCGGGCCAGCCCGAGTGGGTAGCCCAGG-3′
3′-CGCCCGGTCGGGCTCACCCATCGGGTCC-5′

B. 5′-ATTATAAAATATTTAGATACTATATTTACAA-3′
3′-TAATATTTTATAAATCTATGATATAAATGTT-5′

C. 5′-AGAGCTAGATCGAT-3′
3′-TCTCGATCTAGCTA-5′

QUESTION 5–9

The total length of DNA in the human genome is about 1 m, and the diameter of the double helix is about 2 nm. Nucleotides in a DNA double helix are stacked (see Figure 5–6B) at an interval of 0.34 nm. If the DNA were enlarged so that its diameter equaled that of an electrical extension cord (5 mm), how long would the extension cord be from one end to the other (assuming that it is completely stretched out)? How close would the bases be to each other? How long would a gene of 1000 nucleotide pairs be?

QUESTION 5–10

A compact disc (CD) stores about 4.8×10^9 bits of information in a 96 cm^2 area. This information is stored as a binary code—that is, every bit is either a 0 or a 1.

A. How many bits would it take to specify each nucleotide pair in a DNA sequence?

B. How many CDs would it take to store the information contained in the human genome?

QUESTION 5–11

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

A. Each eukaryotic chromosome must contain the following DNA sequence elements: multiple origins of replication, two telomeres, and one centromere.

B. Nucleosome core particles are 30 nm in diameter.

QUESTION 5–12

Define the following terms and their relationships to one another:

A. Interphase chromosome

B. Mitotic chromosome

C. Chromatin

D. Heterochromatin

E. Histones

F. Nucleosome

QUESTION 5–13

Carefully consider the result shown in **Figure Q5–13**. Each of the two colonies shown on the *left* is a clump of approximately 100,000 yeast cells that has grown up from a single cell, which is now somewhere in the middle of the colony. The two yeast colonies are genetically different, as shown by the chromosomal maps on the right.

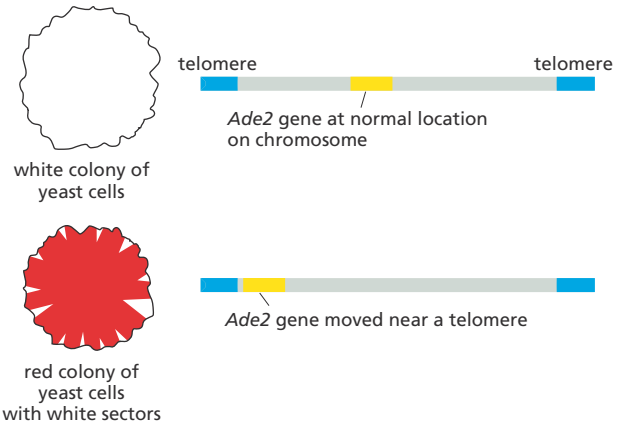
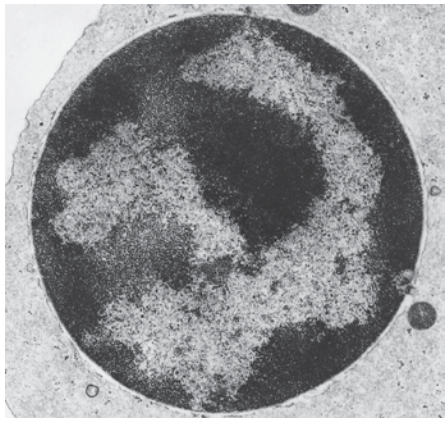


Figure Q5–13

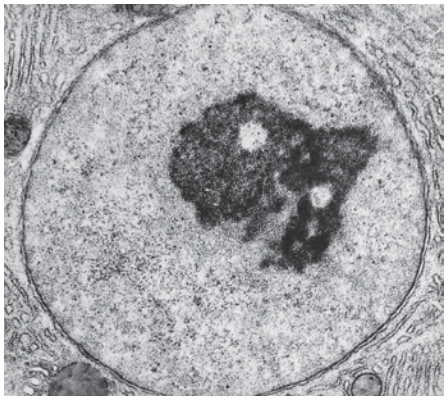
The yeast *Ade2* gene encodes one of the enzymes required for adenine biosynthesis, and the absence of the *Ade2* gene product leads to the accumulation of a red pigment. At its normal chromosome location, *Ade2* is expressed in all cells. When it is positioned near the telomere, which is highly condensed, *Ade2* is no longer expressed. How do you think the white sectors arise? What can you conclude about the propagation of the transcriptional state of the *Ade2* gene from mother to daughter cells?

QUESTION 5–14

The two electron micrographs in **Figure Q5–14** show nuclei of two different cell types. Can you tell from these pictures which of the two cells is transcribing more of its genes? Explain how you arrived at your answer. (Micrographs courtesy of Don W. Fawcett.)



(A)



(B)

Figure Q5–14

QUESTION 5–15

DNA forms a right-handed helix. Pick out the right-handed helix from those shown in **Figure Q5–15**.

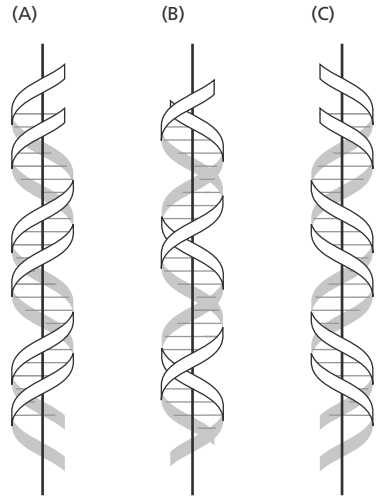


Figure Q5–15

QUESTION 5–16

A single nucleosome core particle is 11 nm in diameter and contains 147 bp of DNA (the DNA double helix measures 0.34 nm/bp). What packing ratio (ratio of DNA length to nucleosome diameter) has been achieved by wrapping DNA around the histone octamer? Assuming that there are an additional 54 bp of extended DNA in the linker between nucleosomes, how condensed is "beads-on-a-string" DNA relative to fully extended DNA? What fraction of the 10,000-fold condensation that occurs at mitosis does this first level of packing represent?

Page left intentionally blank

CHAPTER **SIX** 6

DNA Replication, Repair, and Recombination

The ability of a cell to survive and proliferate in a chaotic environment depends on the accurate duplication of the vast quantity of genetic information carried in its DNA. This duplication process, called *DNA replication*, must occur before a cell can divide to produce two genetically identical daughter cells. Maintaining order in a cell also requires the continual surveillance and repair of its genetic information, as DNA is subjected to unavoidable damage by chemicals and radiation in the environment and by reactive molecules that are generated inside the cell. In this chapter, we describe the protein machines that replicate and repair the cell's DNA. These machines catalyze some of the most rapid and accurate processes that take place within cells, and the strategies they have evolved to achieve this feat are marvels of elegance and efficiency.

Despite these systems for protecting 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**); when allowed to accumulate over millions of years, such changes provide the variety in genetic material that makes one species distinct from another, as we discuss in Chapter 9.

But, mutations are much more likely to be detrimental than beneficial: in humans, they are responsible for thousands of genetic diseases, including cancer. The survival of a cell or organism, therefore, depends on keeping

Figure 6–1 Genetic information is passed from one generation to the next. Differences in DNA can produce the variations that underlie the differences between individuals of the same species—or, over time, the differences between one species and another. In this family photo, the children resemble one another and their parents more closely than they resemble other people because they inherit their genes from their parents. The cat shares many features with humans, but during the millions of years of evolution that have separated humans and cats, both have accumulated many changes in DNA that now make the two species different. The chicken is an even more distant relative.

changes in its DNA to a minimum. Without the protein machines that are continually monitoring and repairing damage to DNA, it is questionable whether life could exist at all.

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 duplicating 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. But the task is awe-inspiring, as it can involve copying billions of nucleotide pairs every time a cell divides. The copying must be carried out with incredible speed and accuracy: in about 8 hours, a dividing animal cell will copy the equivalent of 1000 books like this one 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*.

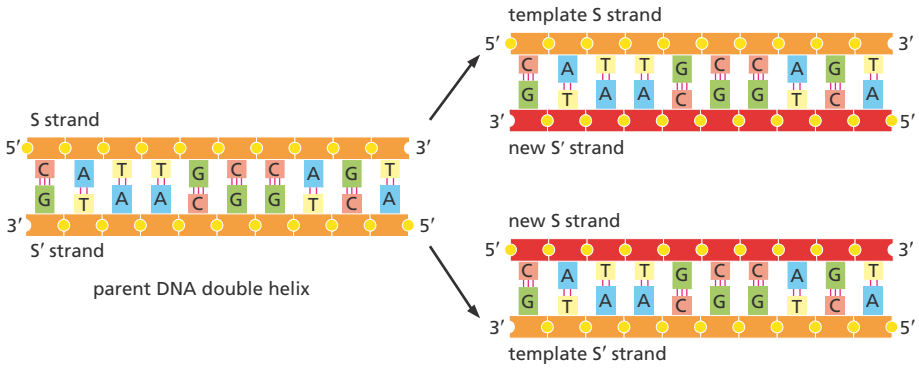


Figure 6–2 DNA acts as a template for its own duplication. Because the nucleotide A will successfully pair only with T, and G with C, each strand of a DNA double helix—labeled here as the S strand and its complementary S′ strand—can serve as a template to specify the sequence of nucleotides in its complementary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

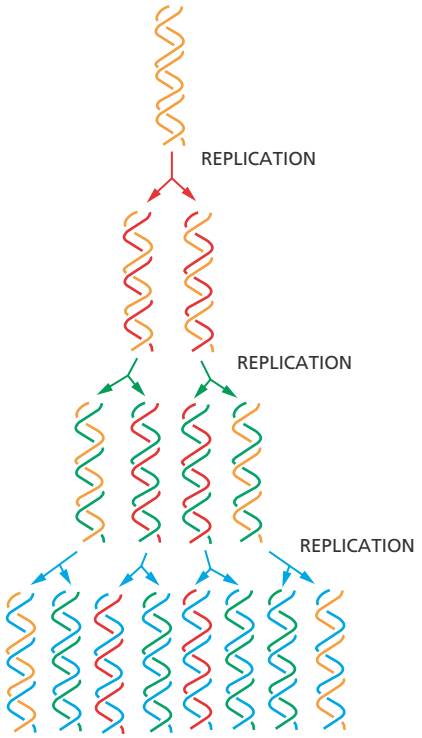


Figure 6–3 In each round of DNA replication, each of the two strands of DNA is used as a template for the formation of a new, complementary strand. DNA replication is "semiconservative" because each daughter DNA double helix is composed of one conserved strand and one newly synthesized strand.

DNA replication produces two complete double helices from the original DNA molecule, with each new DNA helix being identical (except for rare copying errors) in nucleotide sequence 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**). In **How We Know**, pp. 200–202, we discuss the experiments that first demonstrated that DNA is replicated in this way.

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 unpaired bases. How does this 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 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 relatively easy 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.

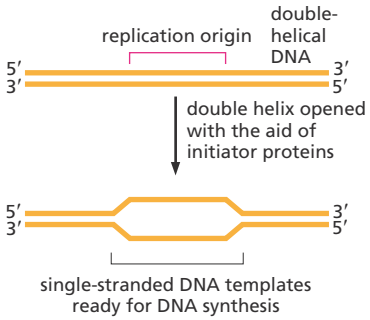


Figure 6–4 A DNA double helix is opened at replication origins. DNA sequences at replication origins are recognized by initiator proteins (not shown), which locally pry apart the two strands of the double helix. The exposed single strands can then serve as templates for copying the DNA.

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

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 (see Figure 5–2). In it, 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. 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 duplicated. 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.

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 model, it

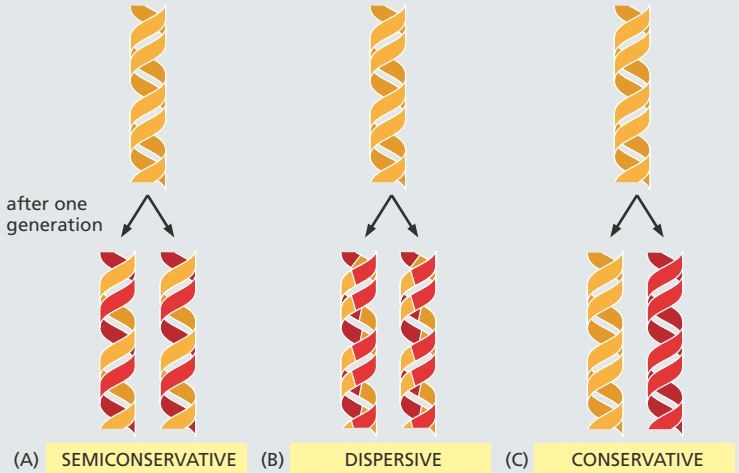


Figure 6–5 Three models for DNA replication make different predictions. (A) In the semiconservative model, each parent strand serves as a template for the synthesis of a new daughter strand. The first round of replication would produce two hybrid molecules, each containing one strand from the original parent in addition to 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 daughter DNA will contain a mixture 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.

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 *E. 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 whose DNA was heavy, the other whose DNA 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 surrounding salt solution (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**).

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 the 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

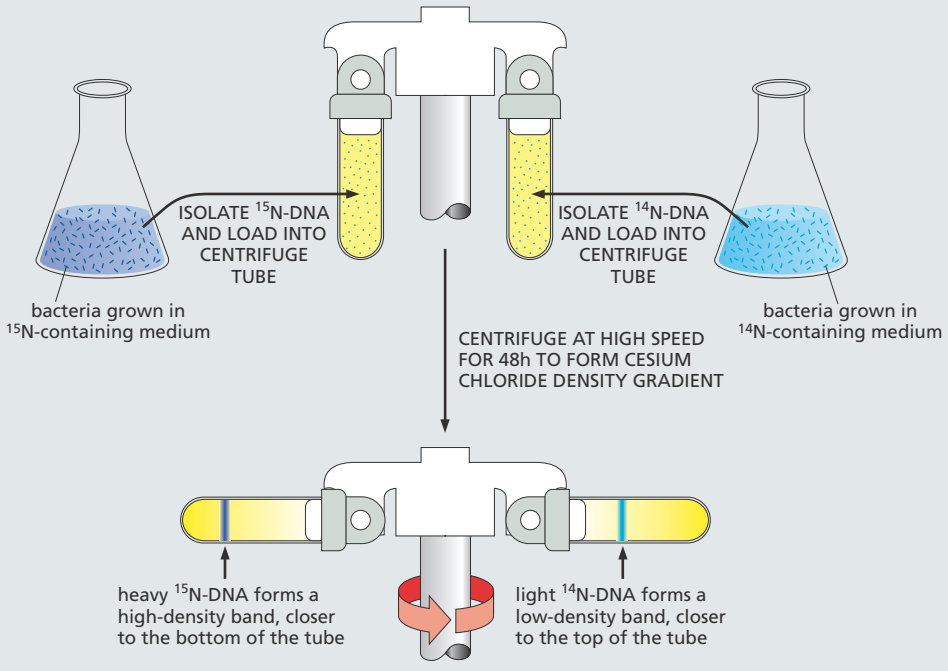


Figure 6–6 Centrifugation in a cesium chloride gradient allows the separation of heavy and light DNA. Bacteria are grown for several generations in a medium containing either ^{15}N (the heavy isotope) or ^{14}N (the light isotope) to label their DNA. The cells are then broken open, and the DNA is loaded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

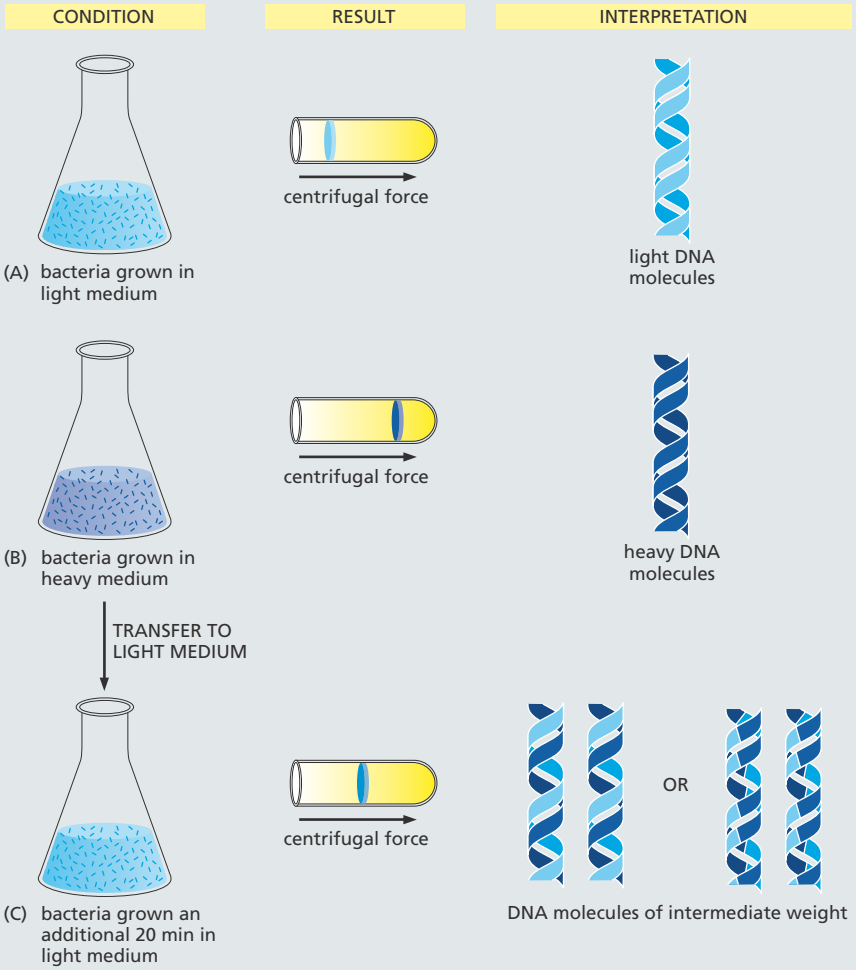


Figure 6–7 The first part of the Meselson–Stahl experiment ruled out the conservative model of DNA replication. (A) Bacteria grown in light medium (containing ^{14}N) yield DNA that forms a band high up in the centrifuge tube, whereas bacteria grown in ^{15}N-containing heavy medium (B) produce DNA that migrates further down the tube. When bacteria grown in a heavy medium are transferred to a light medium and allowed to continue dividing, they produce a band whose position falls somewhere between that of the parent bands (C). 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 hybrid daughter DNA molecules.

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 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 nice, clean results.

the parental DNA would remain entirely heavy, while the daughter DNA would be entirely light (see Figure 6–5C). The data matched with 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 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 their hybrid molecules before centrifuging, they discovered that one strand of the DNA was heavy, whereas the other was light. This observation supported only the semiconservative model; if the dispersive 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.

each 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 replicating the DNA as they go (**Figure 6–9**). DNA replication in 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.

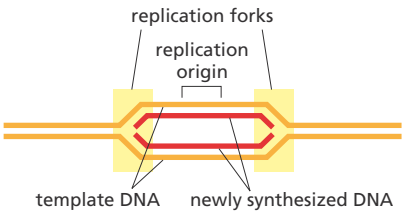


Figure 6–8 DNA synthesis occurs at Y-shaped junctions called replication forks. Two replication forks are formed at each replication origin.

DNA Polymerase Synthesizes DNA Using a Parental Strand as 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

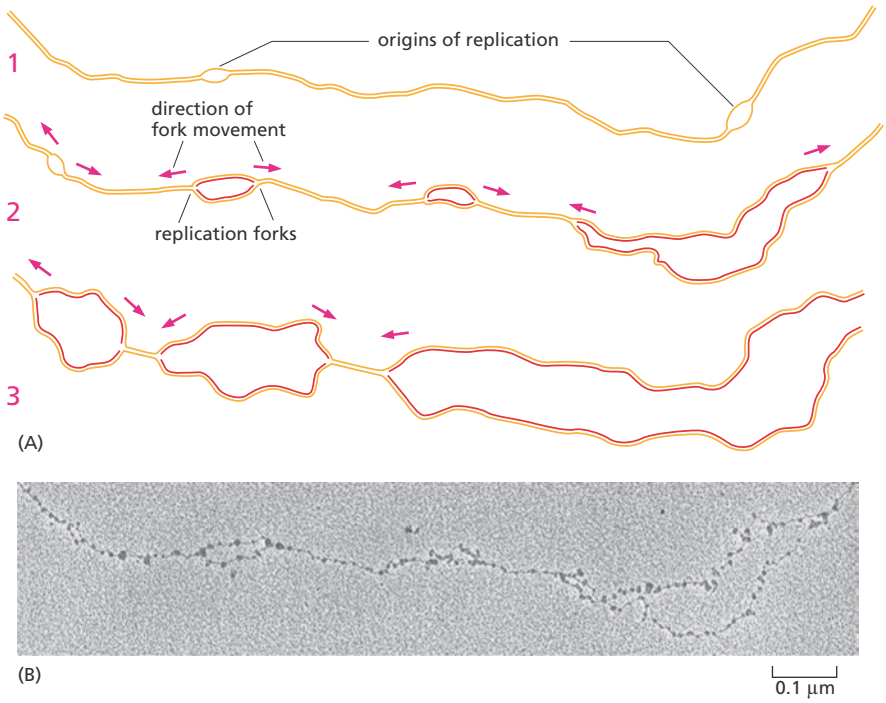


Figure 6–9 The two replication forks move away in opposite directions at each replication origin. (A) These drawings represent the same portion of a DNA molecule as it might appear at different times during replication. The *orange* lines represent the two parental DNA strands; the *red* lines represent 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 protein complexes around which the DNA is wrapped (discussed in Chapter 5). The chromosome in this micrograph is the one that was redrawn in sketch (2) above. (Electron micrograph courtesy of Victoria Foe.)

QUESTION 6–1

Look carefully at the micrograph and drawing 2 in Figure 6–9.

A. Using the scale bar, estimate the lengths of the DNA strands 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?

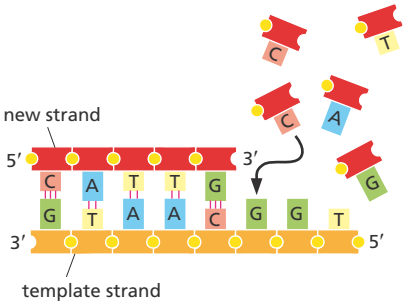


Figure 6–10 A new DNA strand is synthesized in the 5′–to–3′ direction. At each step, the appropriate incoming nucleotide is selected by forming base pairs with the next nucleotide in the template strand: A with T, T with A, C with G, and G with C. Each is added to the 3′ end of the growing new strand, 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 3–41).

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 the 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′) (**Figure 6–12**). The replication fork is therefore asymmetrical. Looking at Figure 6–9A, however, it appears that both of the new DNA strands are growing in the same direction; that is, the direction in which the replication fork is moving. That observation suggests that one strand is being synthesized in the 5′-to-3′ direction and the other in the 3′-to-5′ direction.

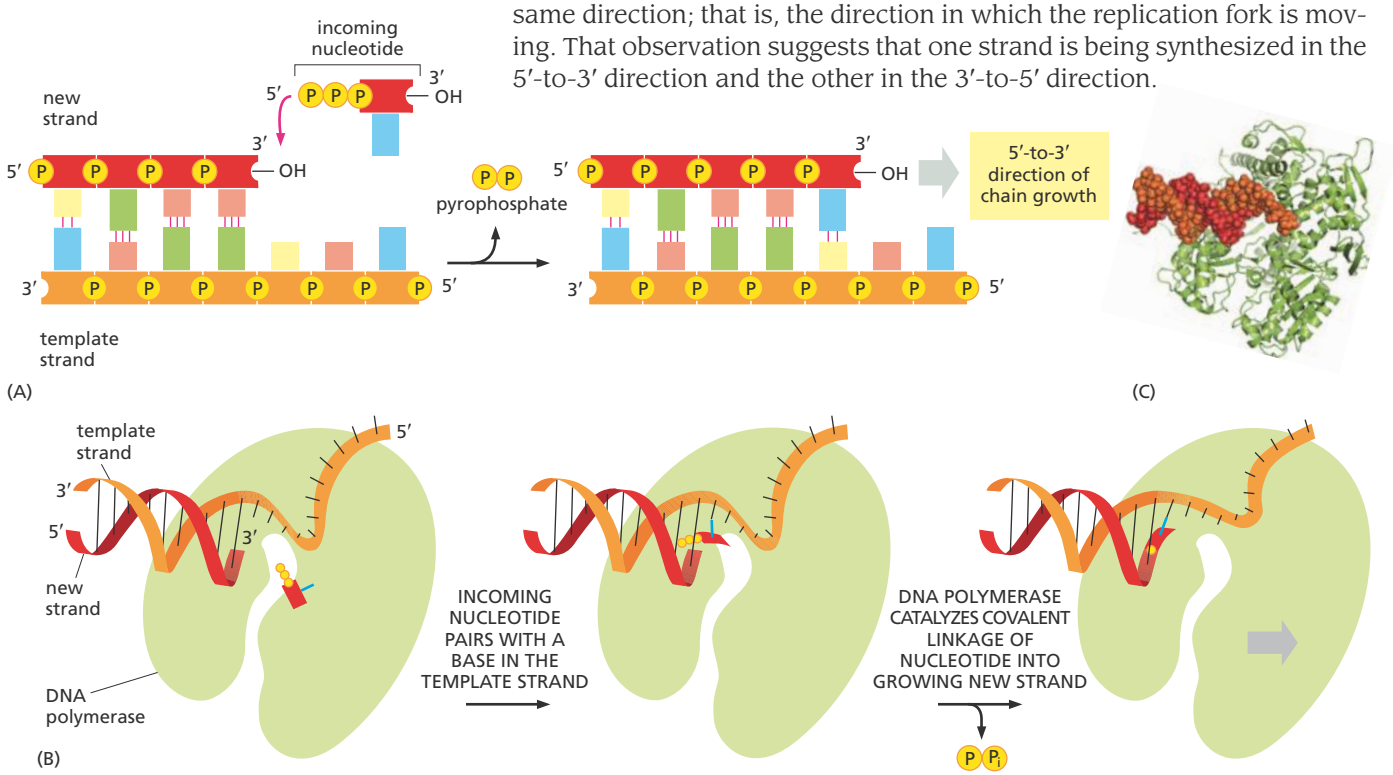


Figure 6–11 DNA polymerase adds a deoxyribonucleotide to the 3′ end of a growing DNA chain. (A) Nucleotides enter the reaction as deoxyribonucleoside triphosphates. This incoming nucleotide forms a base pair with its partner in the template strand. It is then linked to the free 3′ hydroxyl on the growing DNA strand. The new DNA strand is therefore synthesized in the 5′-to-3′ direction. Breakage of a high-energy phosphate bond in the incoming nucleoside triphosphate—accompanied by the release of pyrophosphate—provides the energy for the polymerization reaction. (B) The reaction is catalyzed by the enzyme DNA polymerase (*light green*). The polymerase guides the incoming nucleotide to the template strand and positions it such that its 5′ terminal phosphate 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, which shows the positioning of the DNA double helix. The template strand is the longer of the two DNA strands (Movie 6.1).

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 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 backstitching 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 provides cells with an important advantage, 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 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 nucleotide and the template strand. Only when the match is correct does DNA polymerase catalyze the nucleotide-addition reaction. Second,

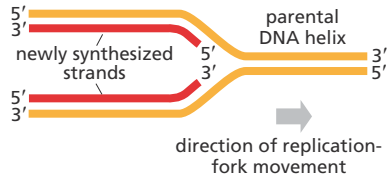


Figure 6–12 At a replication fork, the two newly synthesized DNA strands are of opposite polarities. This is because the two template strands are oriented in opposite directions.

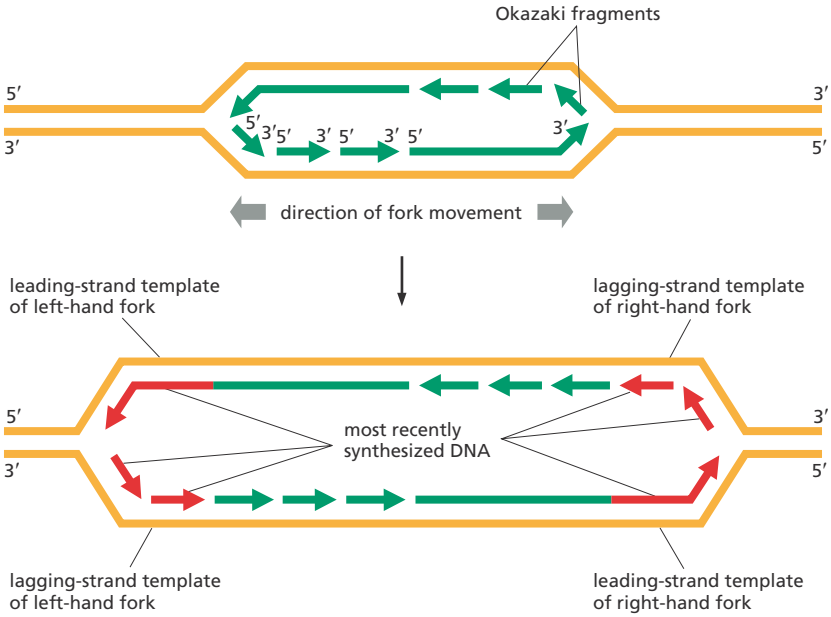


Figure 6–13 At each replication fork, the lagging DNA strand is synthesized in pieces. Because both of the new strands at a replication fork are synthesized in the 5′-to-3′ direction, the lagging strand of DNA must be made initially as a series of short DNA strands, which are later joined together. The 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.

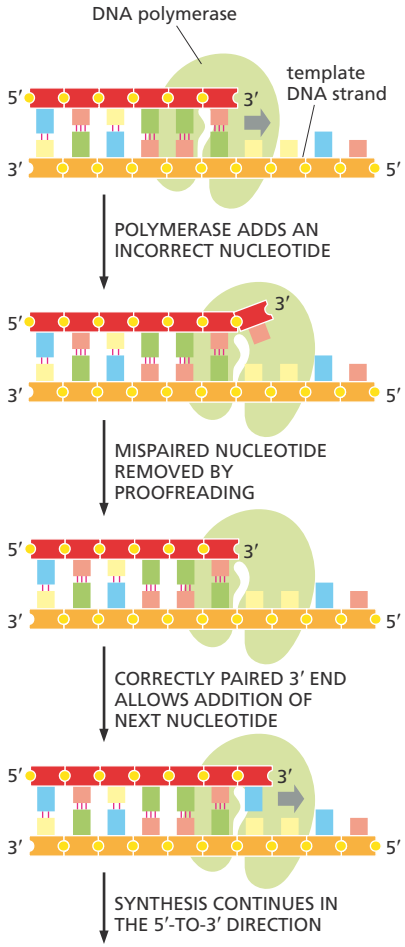


Figure 6–14 During DNA synthesis, DNA polymerase proofreads its own work. If an incorrect nucleotide is added to a growing strand, the DNA polymerase cleaves it from the strand and replaces it with the correct nucleotide before continuing.

when DNA polymerase makes 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**). This proofreading is carried out by a nuclease that cleaves the phosphodiester backbone. 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 explains why DNA polymerases synthesize DNA only in the 5′-to-3′ direction, despite the need that this imposes for a cumbersome backstitching mechanism at the replication fork (see Figure 6–13). A hypothetical DNA polymerase that synthesized in the 3′-to-5′ direction (and would thereby circumvent the need for backstitching) would be unable to proofread: if it removed an incorrectly paired nucleotide, the polymerase would create a chemical dead end—a chain that could no longer be elongated. 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.

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. It thus serves as a *primer* for DNA synthesis, and the enzyme that synthesizes the RNA primer is known as *primase*.

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 2–6, pp. 76–77). 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 (**Figure 6–16**).

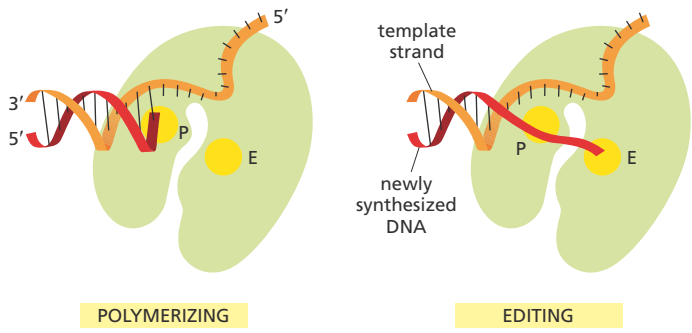


Figure 6–15 DNA polymerase contains separate sites for DNA synthesis and proofreading. The diagrams are based on the structure of an *E. coli* DNA polymerase molecule, as determined by X-ray crystallography. DNA polymerase is shown with the replicating DNA molecule and the polymerase in the polymerizing mode (*left*) and in the proofreading mode (*right*). The catalytic sites for the polymerization activity (P) and error-correcting proofreading 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 growing 3′ end moves into the error-correcting catalytic site (E) to be removed.

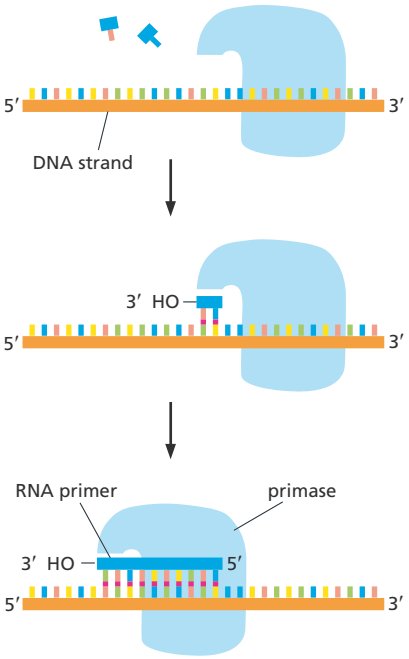


Figure 6–16 RNA primers are synthesized by an RNA polymerase called primase, which uses a DNA strand as a template. Like DNA polymerase, primase works in the 5′-to-3′ direction. Unlike DNA polymerase, however, primase can start a new polynucleotide chain by joining together two nucleoside triphosphates without the need for a base-paired 3′ end as a starting point. (In this case, ribonucleoside triphosphates, rather than deoxyribonucleoside triphosphates, provide the incoming nucleotides.)

For the leading strand, an RNA primer is needed only to start replication at a replication origin; once a replication fork has been established, the DNA polymerase is continuously presented with a base-paired 3′ end as it tracks along the template strand. But on the lagging strand, where DNA synthesis is discontinuous, new primers are 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 are laid down at intervals along the newly exposed, single-stranded stretch. DNA polymerase adds a deoxyribonucleotide to the 3′ end of each primer to start a new Okazaki fragment, and it will continue to elongate this fragment until it runs into the next RNA primer (**Figure 6–17**).

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 DNA fragments together. Thus, a nuclease degrades the RNA primer, a DNA polymerase called a *repair polymerase* then replaces this RNA with DNA (using the end of the adjacent Okazaki fragment as a 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 6–18**).

Primase can begin new polynucleotide chains, but this activity is possible because the enzyme 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 DNA polymerases that make this DNA, like the replicative polymerases, proofread as they synthesize. In this way, the cell's replication machinery is able to begin new DNA chains 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 open up the double helix and synthesize new DNA. These proteins form part of a remarkably complex replication machine. The first problem faced by the replication machine is accessing the

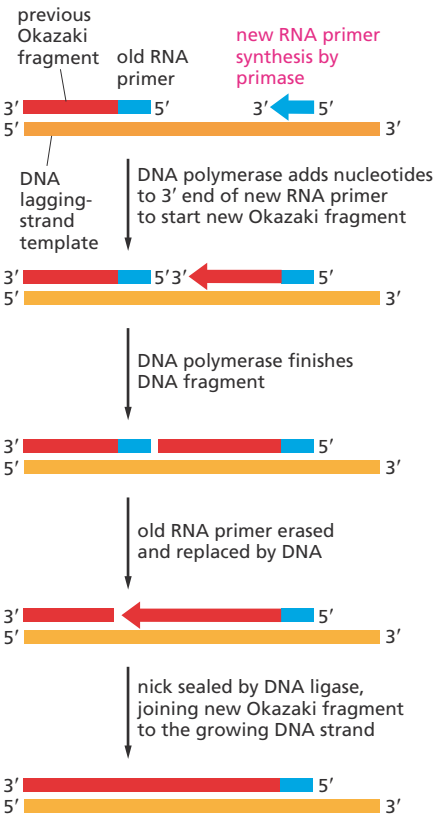


Figure 6–17 Multiple enzymes are required to synthesize Okazaki fragments on the lagging DNA strand. In eukaryotes, RNA primers are made at intervals of about 200 nucleotides on the lagging strand, and each RNA primer is approximately 10 nucleotides long. Primers are removed by nucleases that recognize an RNA strand in an RNA/DNA 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 fragments are finally joined together by an enzyme called DNA ligase, which catalyzes the formation of a phosphodiester bond between the 3′-OH 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 (not shown; see Figure 6–18).

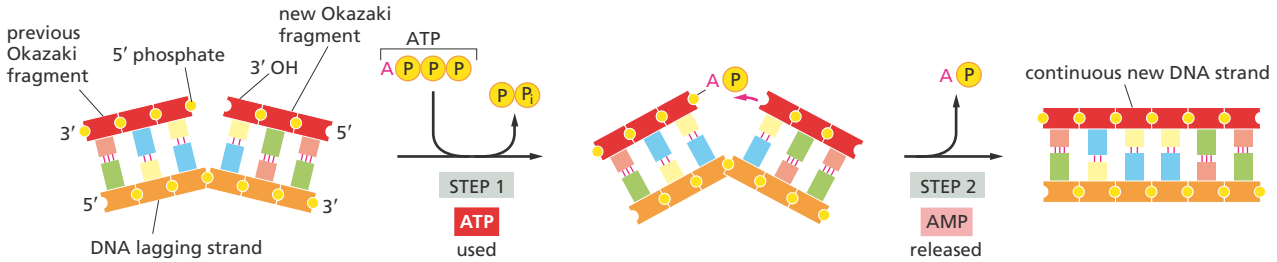


Figure 6–18 DNA ligase joins together Okazaki fragments on the lagging strand during DNA synthesis. The ligase enzyme uses a molecule of ATP to activate the 5′ end of one fragment (step 1) before forming a new bond with the 3′ end of the other fragment (step 2).

nucleotides that lie at the center of the helix. For DNA replication to occur, the double helix must be unzipped ahead of the replication fork so that the incoming nucleoside triphosphates can form base pairs with each template strand. Two types of replication proteins—*DNA helicases* and *single-strand DNA-binding proteins*—cooperate to carry out this task. The 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 6–19A** and **Movie 6.2**). Single-strand DNA-binding proteins cling to the single-stranded DNA exposed by the helicase, transiently preventing the strands from re-forming base pairs and keeping them in an elongated form so that they can serve as efficient templates.

This localized unwinding of the DNA double helix itself presents a problem. As the helicase pries open the DNA within the replication fork, the

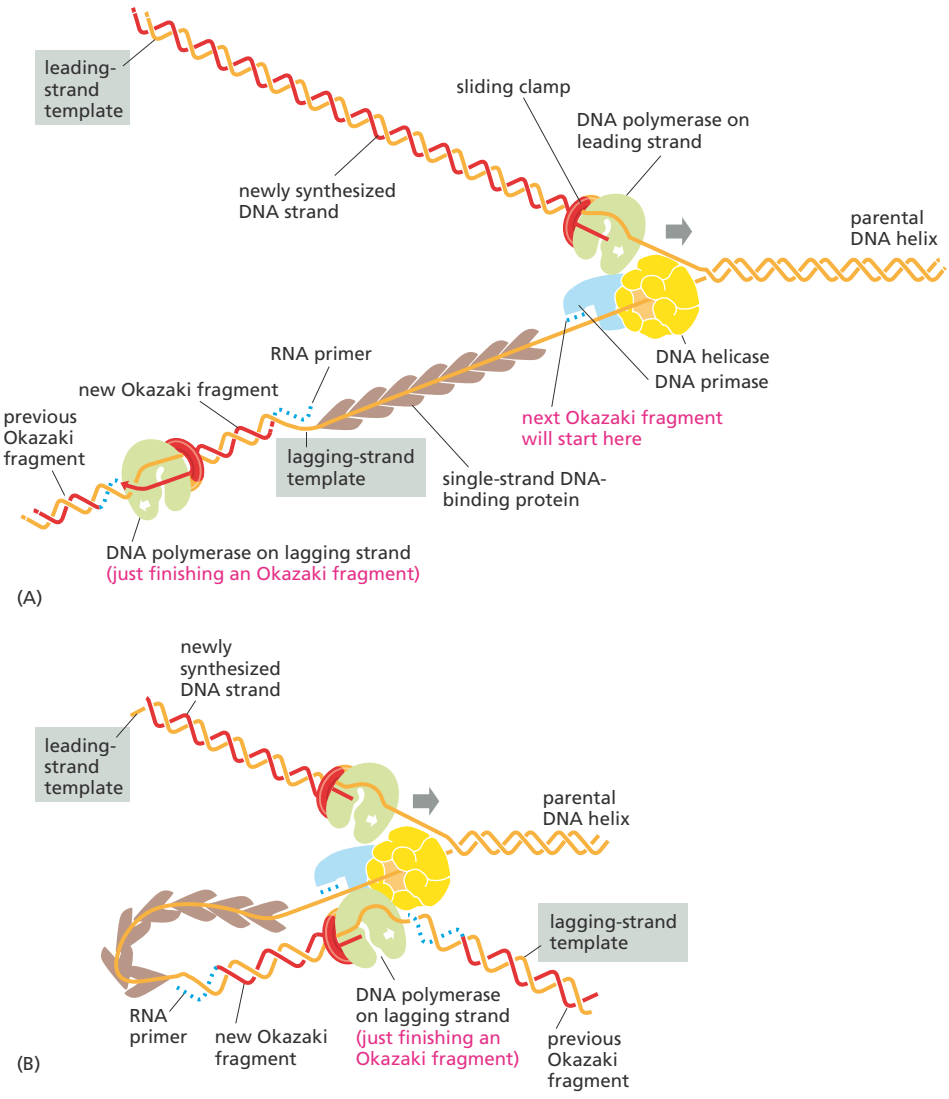


Figure 6–19 DNA synthesis is carried out by a group of proteins that act together as a replication machine. (A) DNA polymerases are held on the leading and lagging strands by circular protein clamps that allow the polymerases to slide. On the lagging-strand template, the clamp detaches each time the polymerase completes an Okazaki fragment. A clamp loader (not shown) is required to attach a sliding clamp each time a new Okazaki fragment is begun. 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, it 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 RNA primer that is being synthesized by the nearby primase. To watch the replication complex in action, see **Movies 6.4** and **6.5**.

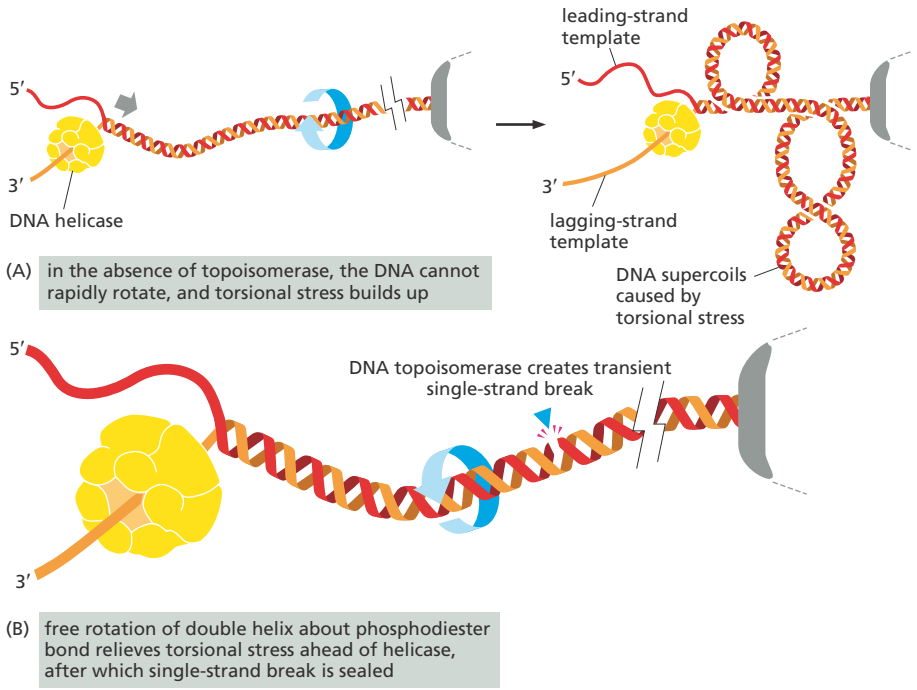


Figure 6–20 DNA topoisomerases relieve the tension that builds up in front of a replication fork. (A) As DNA helicase unwinds the DNA double helix, it generates a section of overwound DNA. Tension builds up because the chromosome is too large to rotate fast enough to relieve the buildup of torsional stress. The broken bars in the left-hand panel represent approximately 20 turns of DNA. (B) DNA topoisomerases relieve this stress by generating temporary nicks in the DNA.

DNA on the other side 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 impedes the forward movement of the replication machinery (**Figure 6–20A**). Cells use proteins called *DNA topoisomerases* to relieve this tension. These enzymes produce transient nicks in the DNA backbone, which temporarily release the tension; they then reseal the nick before falling off the DNA (**Figure 6–20B**).

An additional replication 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 6–19A and **Movie 6.3**).

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.

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 6–19B and **Movies 6.4** and **6.5**).

Telomerase Replicates the Ends of Eukaryotic Chromosomes

Having discussed how DNA replication begins at origins and how movement of a replication fork proceeds, we now turn to the special problem

QUESTION 6–2

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

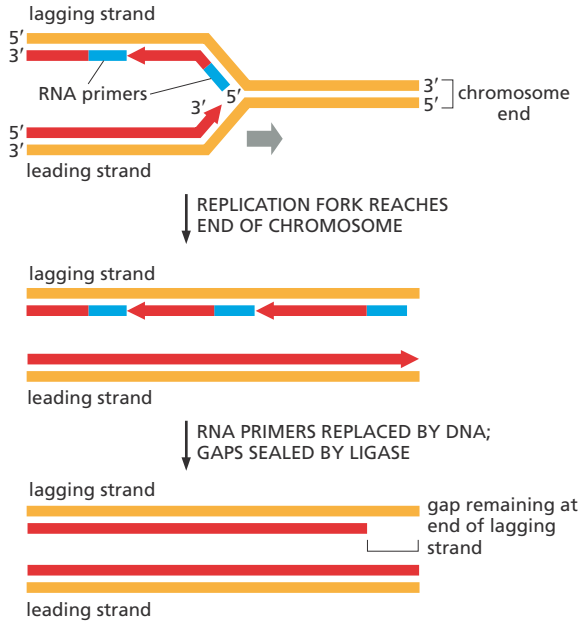


Figure 6–21 Without a special mechanism to replicate the ends of linear chromosomes, DNA would be lost during each round of cell division. DNA synthesis begins at origins of replication and continues until the replication machinery reaches the ends of the chromosome. The leading strand is reproduced 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 way to replace it with DNA. These gaps at the ends of the lagging strand must be filled in by a special mechanism to keep the chromosome ends from shrinking with each cell division.

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 for DNA polymerase
D. Nuclease that removes RNA primers
E. DNA helicase
F. Primase

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 has to be synthesized in the form of discontinuous DNA fragments, each of which is primed with an RNA primer laid down by a primase (see Figure 6–17). 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 way to replace it (**Figure 6–21**). Without a strategy to deal with this problem, the lagging strand would become shorter with each round of DNA replication; after repeated cell divisions, chromosomes would shrink—and eventually lose valuable genetic information.

Bacteria solve this "end-replication" problem by having circular DNA molecules as chromosomes. Eukaryotes solve it by having long, repetitive nucleotide sequences at the ends of their chromosomes which are incorporated into structures called **telomeres**. These telomeric DNA sequences attract an enzyme called **telomerase** to the chromosome ends. Using an RNA template that is part of the enzyme itself, telomerase extends the ends of the replicating lagging strand by adding multiple copies of the same short DNA sequence to the template strand. This extended template allows replication of the lagging strand to be completed by conventional DNA replication (**Figure 6–22**).

In addition to allowing replication of chromosome ends, telomeres form structures that mark the true ends of a chromosome. This allows the cell to distinguish unambiguously between the natural ends of chromosomes and the double-strand DNA breaks that sometimes occur accidentally in

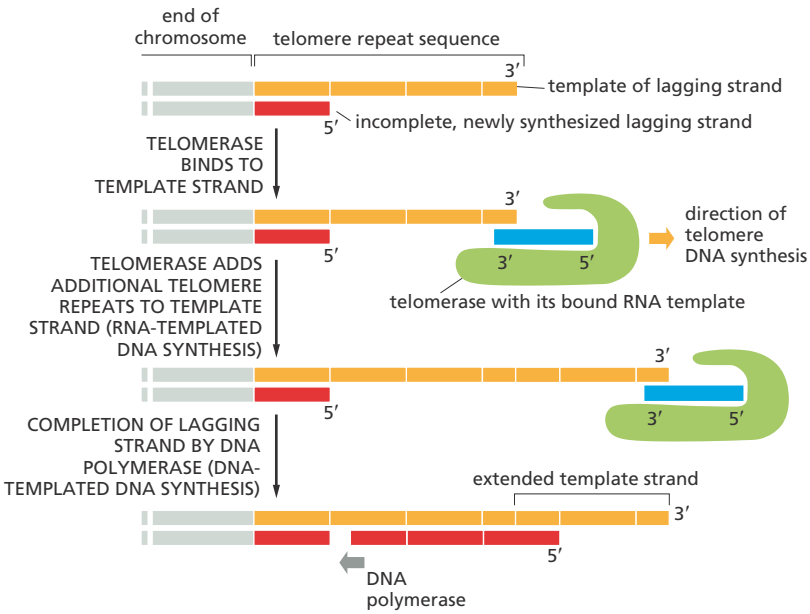


Figure 6–22 Telomeres and telomerase prevent linear eukaryotic chromosomes from shortening with each cell division. For clarity, only the template DNA (*orange*) and newly synthesized DNA (*red*) of the lagging strand are shown (see bottom of Figure 6–21). To complete the replication of the lagging strand at the ends of a chromosome, the template strand is first extended beyond the DNA that is to be copied. To achieve this, the enzyme telomerase adds more repeats to the telomere repeat sequences at the 3′ end of the template strand, which then allows the lagging strand to be completed by DNA polymerase, as shown. The telomerase enzyme 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, as shown. To see telomerase in action, view **Movie 6.6**.

the middle of chromosomes. These breaks are dangerous and must be immediately repaired, as we see in the next section.

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 millions of years. Some of these changes allow organisms to adapt to changing conditions and to thrive in new habitats. However, in the short term, and from the perspective of an individual organism, genetic alterations can be detrimental. In a multicellular organism, such permanent changes in the DNA—called mutations—can upset the organism's extremely complex and finely tuned development and physiology.

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 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 accumulation of DNA damage in cells that are 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 discuss how the fidelity of DNA replication and repair are reflected in our genome.

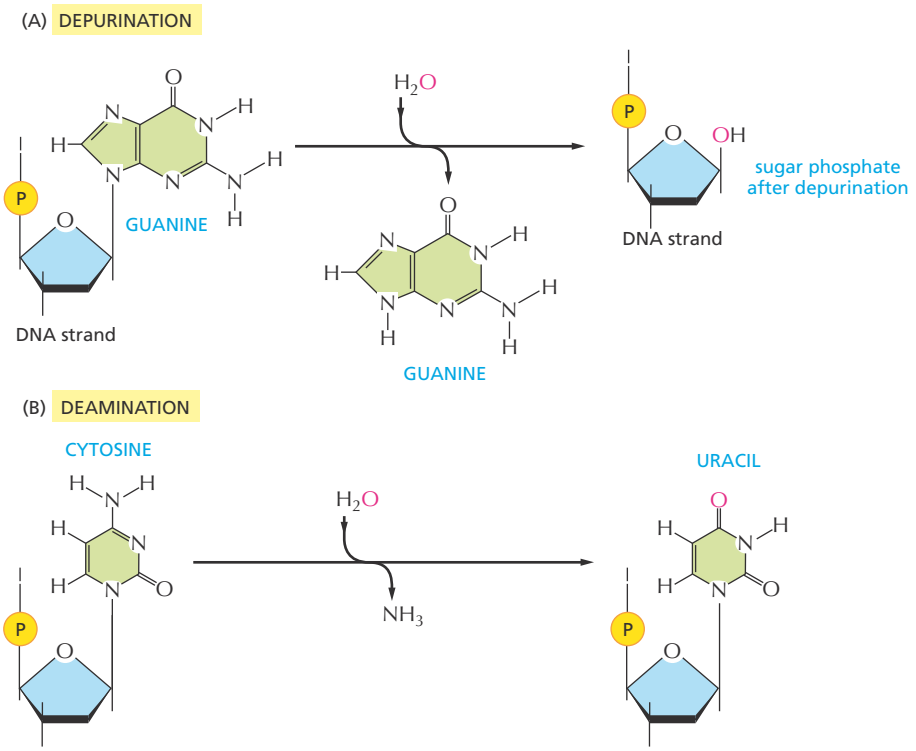


Figure 6–23 Depurination and deamination are the most frequent chemical reactions known to create serious DNA damage in cells. (A) Depurination can remove guanine (or adenine) from DNA. (B) The major type of deamination reaction converts cytosine to an altered DNA base, uracil; however, deamination can also occur on other bases as well. Both depurination and deamination take place on double-helical DNA, and neither break the phosphodiester backbone.

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."

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, during 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 6–23A**). 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 6–25B). Another common reaction is the spontaneous loss of an amino group (*deamination*) from a cytosine in DNA to produce the base uracil (**Figure 6–23B**). Some chemically reactive by-products of cell metabolism also occasionally react with the bases in DNA, altering them in such a way that their base-pairing properties are changed.

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 6–24**. It is the failure to repair thymine dimers that spells trouble for individuals with the disease xeroderma pigmentosum.

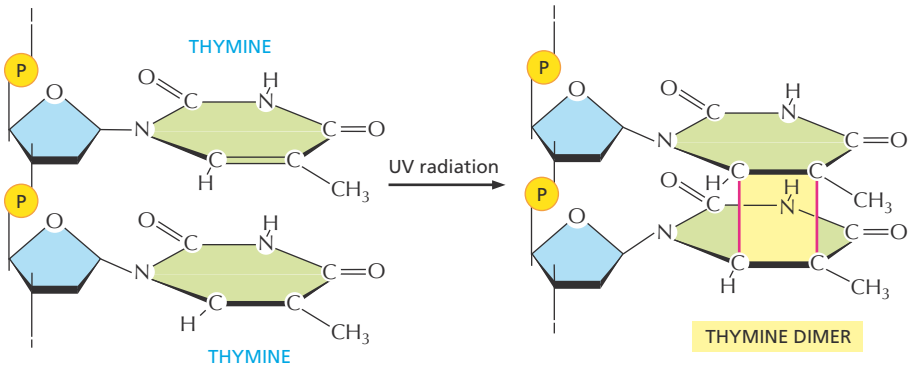


Figure 6–24 The ultraviolet radiation in sunlight can cause the formation of thymine dimers. Two adjacent thymine bases have become covalently attached to each other to form a thymine dimer. Skin cells that are exposed to sunlight are especially susceptible to this type of DNA damage.

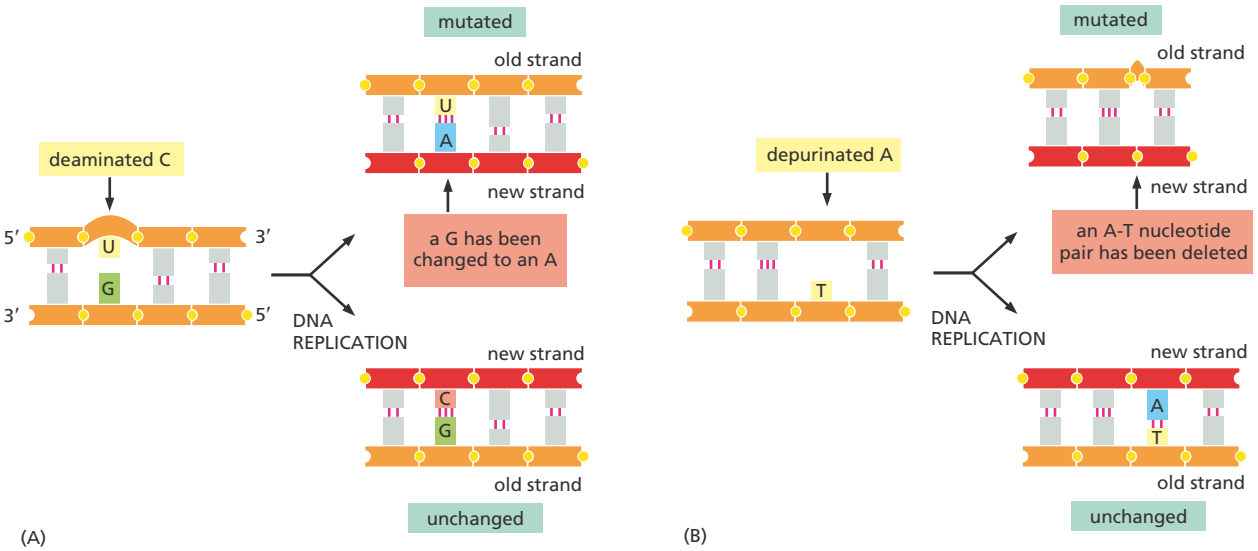


Figure 6–25 Chemical modifications of nucleotides, if left unrepaired, produce mutations. (A) Deamination of cytosine, if uncorrected, results in the substitution of one base for another when the DNA is replicated. As shown in Figure 6–23B, deamination of cytosine produces uracil. Uracil differs from 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 (not shown), the replication machinery places an incorrect nucleotide across from the missing base, again resulting in a mutation.

These are only a few of many chemical changes that can occur in our DNA. If left unrepaired, many of them would lead either to the substitution of one nucleotide pair for another as a result of incorrect base-pairing during replication (**Figure 6–25A**) or to deletion of one or more nucleotide pairs in the daughter DNA strand after DNA replication (**Figure 6–25B**). 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—quite rarely—incorporate an incorrect nucleotide that it fails to correct via proofreading (see Figure 6–14).

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

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 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 6–26**, 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 in the region.
2. A *repair DNA polymerase* binds to the 3′-hydroxyl end of the cut DNA strand. It then fills in the gap by making a complementary copy of the information stored in the undamaged strand. Although

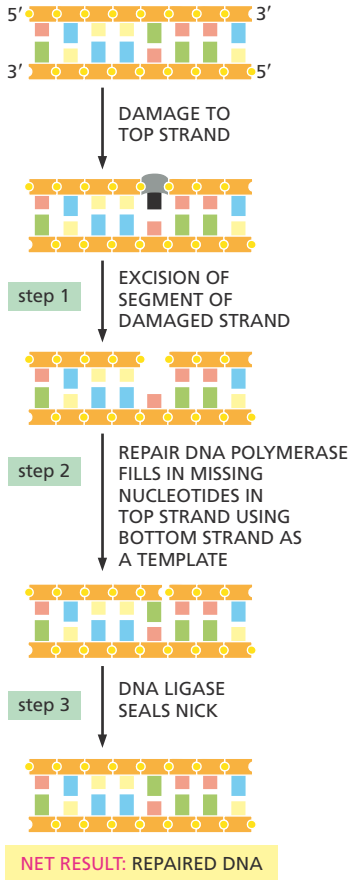


Figure 6–26 The basic mechanism of DNA repair involves three steps. In step 1 (excision), the damage is cut out by one of a series of nucleases, each specialized for a type of DNA damage. In step 2 (resynthesis), the original DNA sequence is restored by a repair DNA polymerase, which fills in the gap 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 6–18).

different 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, this is the same enzyme that fills in the gap left after the RNA primers are removed during the normal DNA replication process (see Figure 6–17).

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.

Steps 2 and 3 are nearly the same for most types of DNA damage, including the rare errors that arise during DNA replication. However, step 1 uses a series of different enzymes, each specialized for removing different types of DNA damage. Humans produce hundreds of different proteins that function in DNA repair.

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**—which is dedicated to correcting these errors. The replication machine makes approximately one mistake per 10^7 nucleotides copied; DNA mismatch repair corrects 99% of these replication errors, increasing the overall accuracy to one mistake in 10^9 nucleotides copied. This level of accuracy is much, much higher than that generally encountered in our day-to-day lives (**Table 6–1**).

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 6–27**). A complex of mismatch repair proteins recognizes such a DNA mismatch, removes a portion of the DNA strand containing the error, and then resynthesizes the missing DNA. This repair mechanism restores the correct sequence (**Figure 6–28**).

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 of DNA that contains the correct sequence would only

TABLE 6–1 ERROR RATES

US Postal Service on-time delivery of local first-class mail	13 late deliveries per 100 parcels
Airline luggage system	1 lost bag per 150
A professional typist typing at 120 words per minute	1 mistake per 250 characters
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

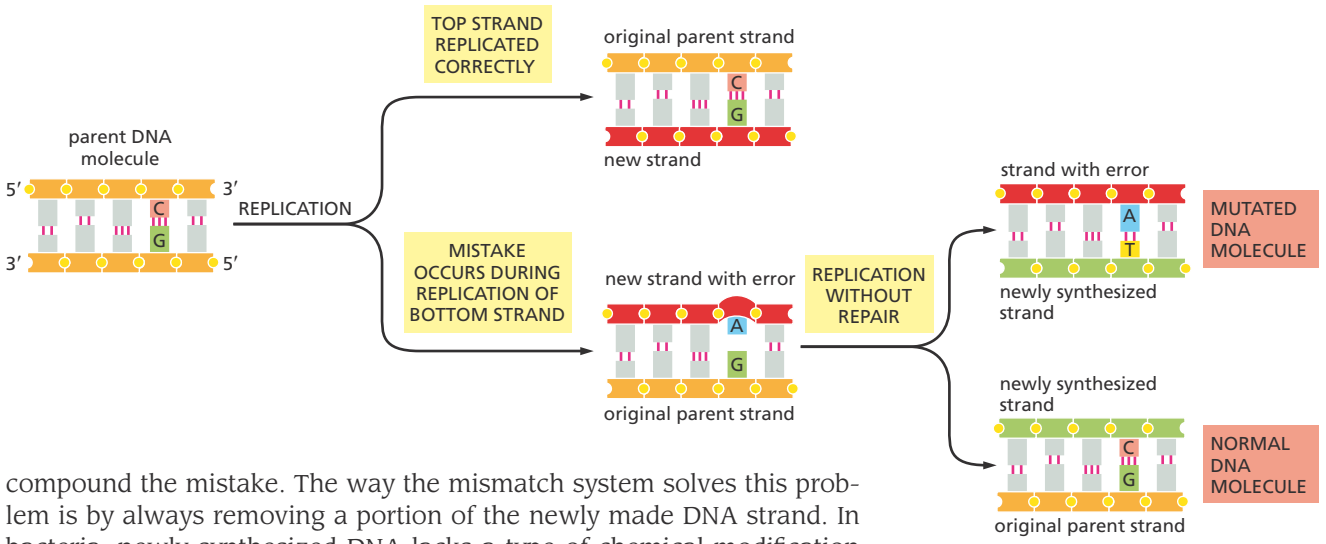


Figure 6–27 Errors made during DNA replication must be corrected to avoid mutations. If uncorrected, a mismatch will lead to a permanent mutation in one of the two DNA molecules produced by the next round of DNA replication.

compound the mistake. The way the mismatch system solves this problem is by always removing a portion of the newly made DNA strand. In bacteria, newly synthesized DNA lacks a type of chemical modification that is present on the preexisting parent DNA. Other cells use other strategies for distinguishing their parent DNA from a newly replicated strand.

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. Humans inherit two copies of these genes (one from each parent), and individuals who inherit one damaged mismatch 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 damaged mismatch repair gene strongly predisposes an individual to cancer.

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.

But what happens when both strands of the double helix are damaged at the same time? Radiation, mishaps at the replication fork, and various chemical assaults can all fracture the backbone of DNA, creating a

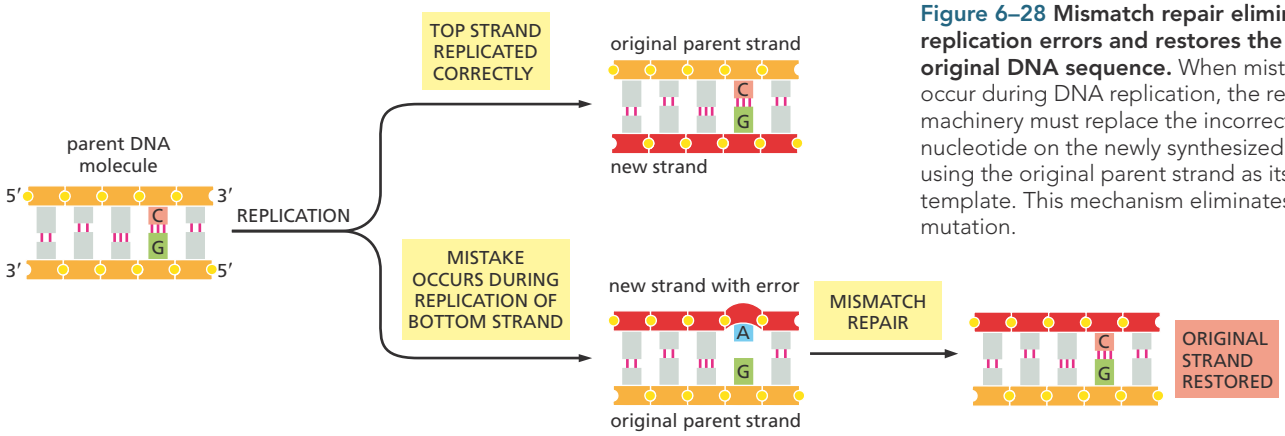


Figure 6–28 Mismatch repair eliminates replication errors and restores the original DNA sequence. When mistakes occur during DNA replication, the repair machinery must replace the incorrect nucleotide on the newly synthesized strand, using the original parent strand as its template. This mechanism eliminates the mutation.

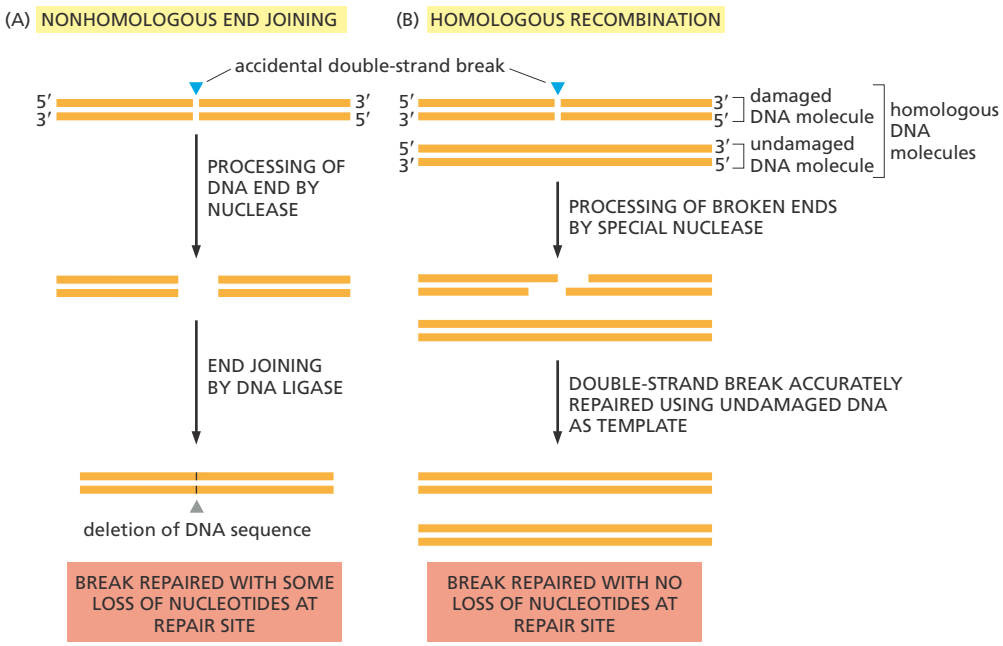


Figure 6–29 Cells can repair double-strand breaks in one of two ways. (A) In nonhomologous end joining, the break is first "cleaned" by a nuclease that chews back the broken ends to produce flush ends. The flush ends are then stitched together by a DNA ligase. Some nucleotides are lost in the repair process, as indicated by the black lines in the repaired DNA. (B) If a double-strand break occurs in one of two daughter DNA double helices after DNA replication has occurred, but before the daughter chromosomes have been separated, the undamaged double helix can be readily used as a template to repair the damaged double helix by homologous recombination. This is a more involved process than non-homologous end joining, but it accurately restores the original DNA sequence at the site of the break. The detailed mechanism is presented in Figure 6–30.

double-strand break. Such lesions are particularly dangerous, because they can lead to the fragmentation of chromosomes and the subsequent loss of genes.

This type of damage is especially difficult to repair. Each chromosome contains unique information; if a chromosome undergoes 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 rapidly 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 repairs the damage, 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 6–29A**).

In most cases, this emergency repair mechanism mends the damage without creating any additional problems. But if the 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. So cells have an alternative, error-free strategy for repairing double-strand breaks, called homologous recombination (**Figure 6–29B**), as we discuss next.

Homologous Recombination Can Flawlessly Repair DNA Double-Strand Breaks

The problem with repairing a double-strand break, as we mentioned, is finding an intact template to guide the repair. However, if a double-strand break occurs in one double helix shortly after a stretch of DNA has been replicated, the undamaged double helix can readily serve as a template to guide the repair of the broken DNA: information on the undamaged strand of the intact double helix is used to repair the complementary broken strand in the other. Because the two DNA molecules

are homologous—they have 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 6–29B).

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 6–30A**). To initiate the repair, a nuclease chews back the 5′ ends of the two broken strands at the break (**Figure 6–30B**). Then, with the help of specialized enzymes, one of the broken 3′ ends "invades" the unbroken homologous DNA duplex and searches for a complementary sequence through base-pairing (**Figure 6–30C**). Once an extensive, accurate match is found, the invading strand is elongated by a repair DNA polymerase, using the complementary strand as a template (**Figure 6–30D**). After the repair polymerase has passed the point where the break occurred, the newly repaired strand rejoins its original partner, forming base pairs that hold the two strands of the broken double helix together (**Figure 6–30E**). Repair is then completed by additional DNA synthesis at the 3′ ends of both strands of the broken double helix (**Figure 6–30F**), followed by DNA ligation (**Figure 6–30G**).

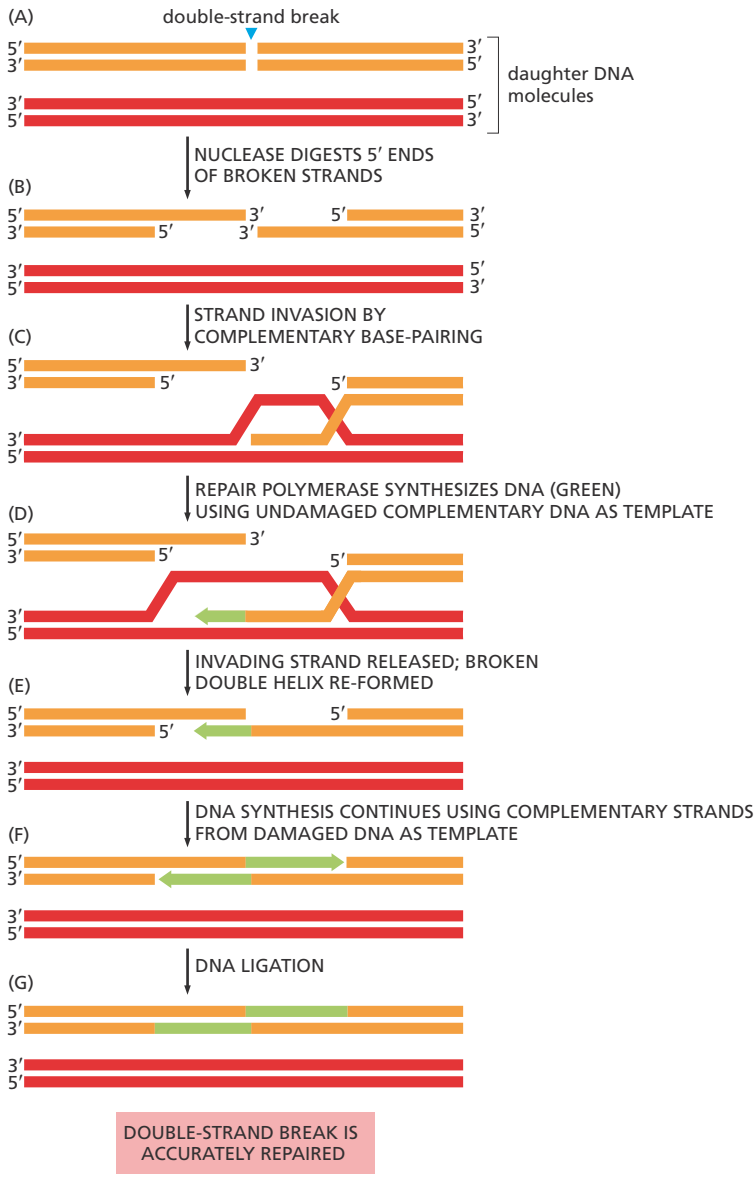


Figure 6–30 Homologous recombination allows the flawless repair of DNA double-strand breaks. This is the preferred method for repairing double-strand breaks that arise shortly after the DNA has been replicated but before the cell has divided. See text for details. (Adapted from M. McVey et al., *Proc. Natl. Acad. Sci. USA* 101:15694–15699, 2004. With permission from the National Academy of Sciences.)

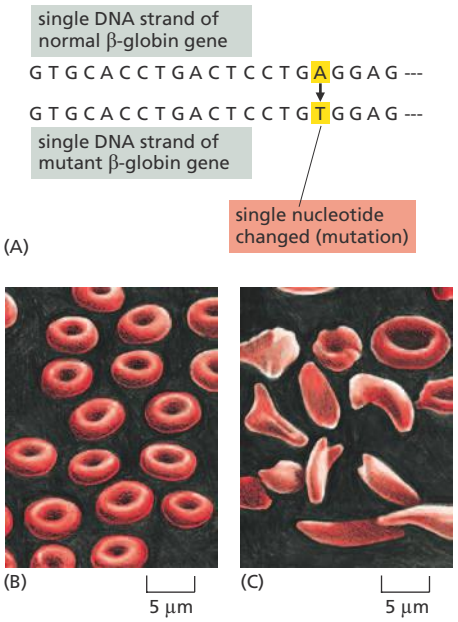


Figure 6–31 A single nucleotide change causes the disease sickle-cell anemia. (A) β-globin is one of the two types of protein subunits that form hemoglobin (see Figure 4–24). A single nucleotide change (mutation) in the β-globin gene produces a β-globin subunit that differs from normal β-globin only by a change from glutamic acid to valine at the sixth amino acid position. (Only a small portion of the gene is shown here; the β-globin subunit contains a total of 146 amino acids.) 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. 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 mutation responsible 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 net result is two intact DNA helices, where 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 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 has a crucial role in the exchange of genetic information during the formation of the germ cells—sperm and eggs. This specialized process, 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 give rise to a mutation. This permanent change in the DNA sequence can have profound consequences. A mutation that affects just a single nucleotide pair can severely compromise an organism's fitness if the change occurs in a vital position in the DNA sequence. Because the structure and activity of each protein depend on its amino acid sequence, a protein with an altered sequence may function poorly or not at all. For example, humans use the protein hemoglobin to transport oxygen in the blood (see Figure 4–24). A permanent change in a single nucleotide in a hemoglobin gene can cause cells to make hemoglobin with an incorrect sequence of amino acids. One such mutation causes the disease *sickle-cell anemia*. The sickle-cell hemoglobin is less soluble than normal hemoglobin and forms fibrous intracellular precipitates, which produce the characteristic sickle shape of affected red blood cells (**Figure 6–31**). 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. This anemia can cause weakness, dizziness, headaches, and breathlessness. Moreover, the abnormal red blood cells 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. Over the course of evolution, many other mutations in the hemoglobin gene have arisen, but only those that do not completely destroy the protein remain in the population.

The example of sickle-cell anemia, which is an inherited disease, illustrates the importance of protecting reproductive cells (*germ cells*) against mutation. A mutation in a germ cell will be passed on to all the cells in the body of the multicellular organism that develops from it, including the germ cells responsible for the production of the next generation.

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 an 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 largely by a gradual accumulation of random mutations in a somatic cell and its

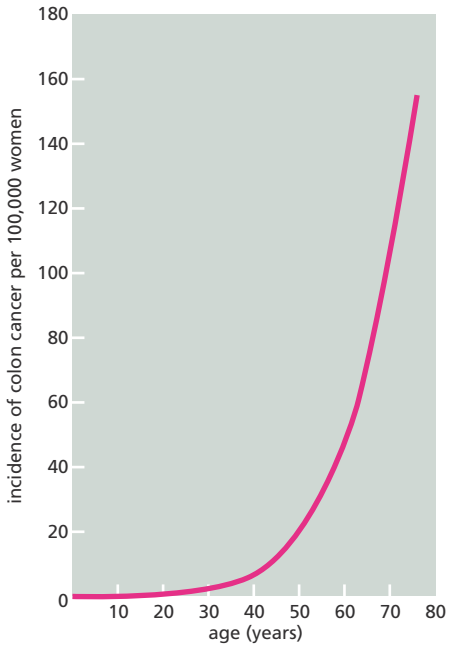


Figure 6–32 Cancer incidence increases dramatically with age. The number of newly diagnosed cases of cancer of the colon in women in England and Wales in one year is plotted as a function of age at diagnosis. Colon cancer, like most human cancers, is caused by the accumulation of multiple mutations. 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.)

descendants (**Figure 6–32**). 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 reproductive 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.

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 harmful consequences are usually eliminated from the population through natural selection; individuals carrying the altered DNA may die or experience decreased fertility, in which case these changes will be lost. 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, and many proteins have amino acid sequences that are almost identical (**Figure 6–33**). 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.

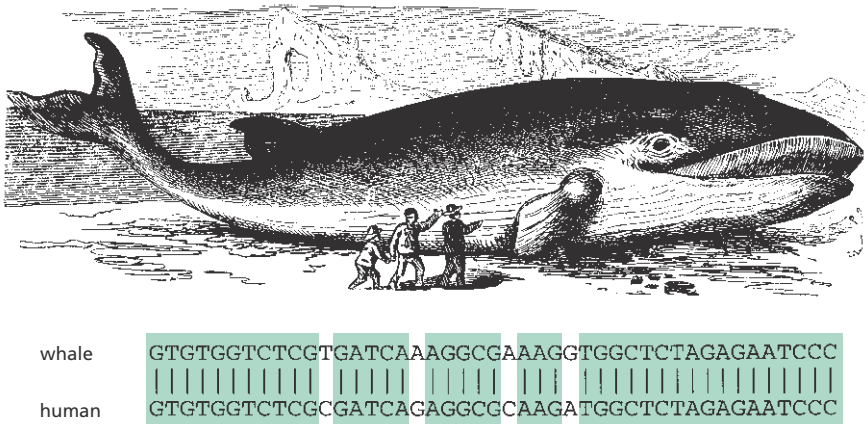


Figure 6–33 The sex-determination genes from humans and whales are unmistakably similar. Although their body plans are strikingly different, humans and whales are built from the same proteins. Despite the many millions of years that have passed since humans and whales diverged, the nucleotide sequences of many of their genes are closely similar. The DNA sequences of a part of the gene that determines maleness in humans and in whales are shown, one above the other; the positions where the two are identical are shaded in *green*.

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 only one 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 chain 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 erased and replaced with DNA.
- DNA replication requires the cooperation of many proteins that form a multienzyme replication machine that copies both DNA strands as it moves along the double helix.
- In eukaryotes, a special enzyme called telomerase replicates the DNA at the ends of the chromosomes.
- 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 lost in the process, altering the DNA sequence at the repair site.
- Homologous recombination can flawlessly repair double-strand breaks 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? 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? What is wrong with this argument?

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 are unstable.

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**.

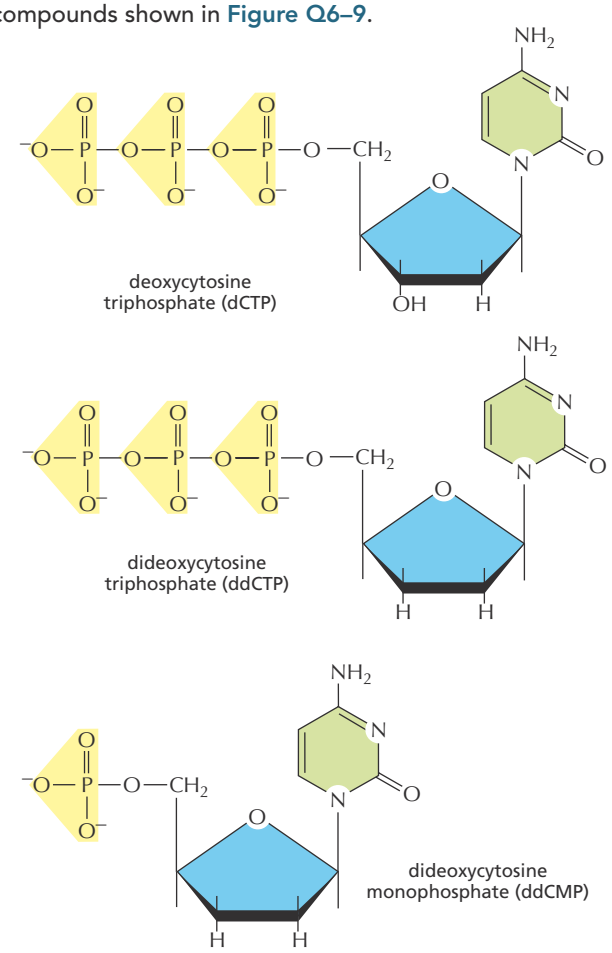


Figure Q6–9

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

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.

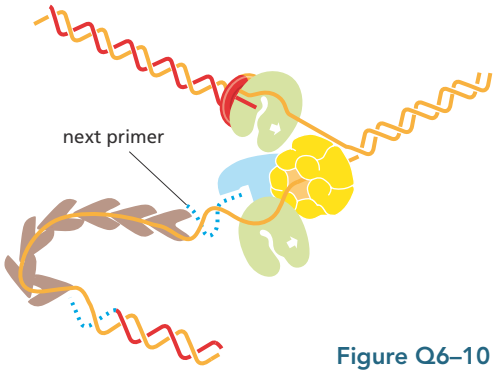


Figure Q6–10

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.

QUESTION 6–13

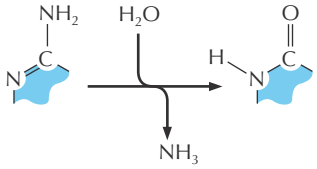


Figure Q6–13

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 by 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 cannot contain uracil?

QUESTION 6–14

A. Explain why telomeres and telomerase are needed for replication of eukaryotic chromosomes but not for replication of a circular bacterial chromosome. 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 one or both telomeres

C. Had no 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.

QUESTION 6–16

Because DNA polymerase proceeds only in the 5′-to-3′ direction, the enzyme is able to correct its own polymerization errors as it moves along the DNA (**Figure Q6–16**). A hypothetical DNA polymerase that synthesized in the 3′-to-5′ direction would be unable to proofread. Given what you know about nucleic acid chemistry and DNA synthesis, draw a sketch similar to Figure Q6–16 that shows what would happen if a DNA polymerase operating in the 3′-to-5′ direction were to remove an incorrect nucleotide from a growing DNA strand. Why would the edited strand be unable to be elongated?

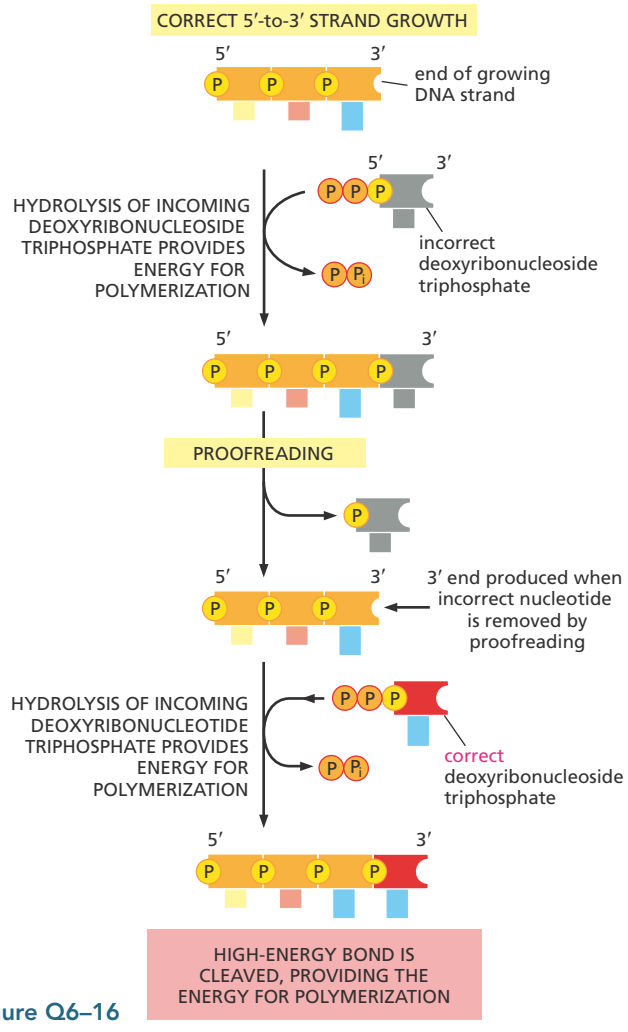


Figure Q6–16