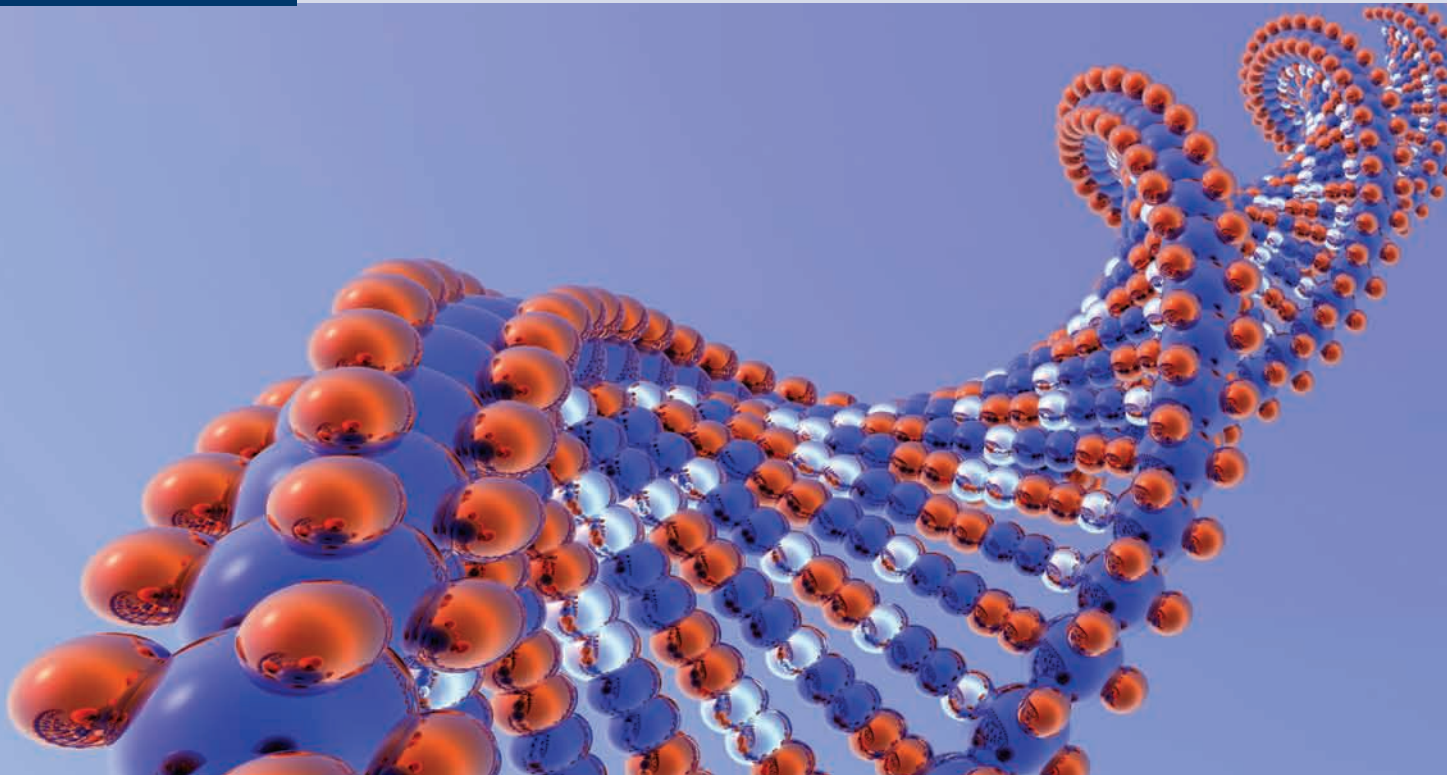


Genes and Chromosomes

The human male karyotype, or complete set of chromosomes, shown using chromosome painting. The "pseudocolors" are created by hybridizing fluorescent molecular probes to specific chromosomal regions. Photo courtesy of Steven M. Carr, Memorial University. Adapted from a photograph by Genetix Ltd. Used with permission.

CHAPTER 1	Genes Are DNA
CHAPTER 2	Genes Encode RNAs and Polypeptides
CHAPTER 3	Methods in Molecular Biology and Genetic Engineering
CHAPTER 4	The Interrupted Gene
CHAPTER 5	The Content of the Genome
CHAPTER 6	Genome Sequences and Gene Number
CHAPTER 7	Clusters and Repeats
CHAPTER 8	Genome Evolution
CHAPTER 9	Chromosomes
CHAPTER 10	Chromatin



A strand of DNA in blue and red. DNA is the genetic material of eukaryotic cells, bacteria, and many viruses.
© Artsilensecome/Shutterstock, Inc.

Genes Are DNA

CHAPTER OUTLINE

- 1.1** Introduction
- 1.2** DNA Is the Genetic Material of Bacteria, Viruses, and Eukaryotic Cells
Historical Perspectives: Determining That DNA Is the Genetic Material
- 1.3** Polynucleotide Chains Have Nitrogenous Bases Linked to a Sugar-Phosphate Backbone
- 1.4** DNA Is a Double Helix
- 1.5** Supercoiling Affects the Structure of DNA
- 1.6** DNA Replication Is Semiconservative
- 1.7** Polymerases Act on Separated DNA Strands at the Replication Fork
- 1.8** Genetic Information Can Be Provided by DNA or RNA
- 1.9** Nucleic Acids Hybridize by Base Pairing
- 1.10** Mutations Change the Sequence of DNA
- 1.11** Mutations May Affect Single Base Pairs or Longer Sequences
- 1.12** The Effects of Mutations Can Be Reversed
- 1.13** Mutations Are Concentrated at Hotspots
- 1.14** Some Hereditary Agents Are Extremely Small
- 1.15** Summary

1.1 Introduction

The hereditary basis of every living organism is its **genome**, a long sequence of DNA that provides the complete set of hereditary information carried by the organism's cells. The genome includes chromosomal DNA as well as DNA in plasmids and (in eukaryotes) organellar DNA as found in mitochondria and chloroplasts. We use the term *information* because the genome does not itself perform an active role in the development of the organism. It is the sequence of the individual subunits, or bases, of the DNA that determines development. By a complex series of interactions, the DNA sequence encodes all the RNAs and proteins of the organism that are to be produced at the appropriate time and place. Proteins serve a diverse series of roles in the development and functioning of an organism: they can form part of the structure of the organism, they have the capacity to build the structures, they perform the metabolic reactions necessary for life, and they participate in regulation as transcription factors, receptors, key players in signal transduction pathways, and other molecules. RNAs encoded by the genome, but that do not themselves encode proteins, also function in the expression of genes during development.

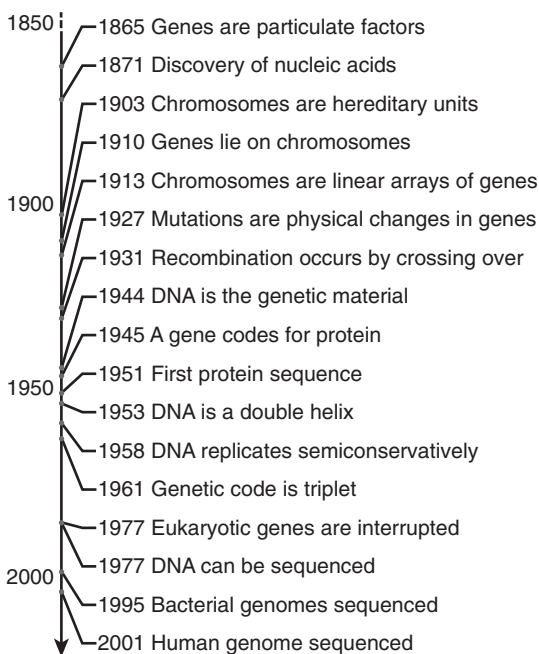
Physically, the genome may be divided into a number of different DNA molecules, or **chromosomes**. The ultimate definition of a genome is the sequence of the DNA of each chromosome. Functionally, the genome is divided into genes. Each gene is a sequence of DNA that encodes a single type of RNA or polypeptide (although it may encode multiple versions of its product). Each of the discrete chromosomes comprising the genome may contain a large number of genes. Genomes for living organisms may contain as few as ~500 genes (for a mycoplasma, a type of bacterium) to ~20,000 to 25,000 for a human being and as many as 50,000 in some plants.

In this chapter, we explore the gene in terms of its basic molecular construction. **FIGURE 1.1** summarizes the stages in the transition from the historical concept of the gene to the modern definition of the genome.

The first definition of the gene as a functional molecular unit followed from the discovery that individual genes are responsible for the production of specific proteins. The chemical differences between the DNA of the gene and its protein product led to the suggestion that a gene encodes a protein. This in turn led to the discovery of the complex apparatus by which the DNA sequence of a gene determines the amino acid sequence of a polypeptide.

Understanding the process by which a gene is expressed allows us to make a more rigorous definition of its nature. **FIGURE 1.2** shows the basic theme of this book. A gene is a sequence of DNA that directly produces a single strand of another nucleic acid, RNA, with a sequence that is identical to one of the two polynucleotide strands of DNA. In many cases, the RNA is in turn used to direct production of a polypeptide, whereas in other cases (such as rRNA, tRNA, and many other genes, the RNA transcribed from the gene is the functional end product. Thus a gene is a sequence of DNA that encodes an RNA, and in protein-coding (or **structural**) genes, the RNA in turn encodes a polypeptide.

From the demonstration that a gene consists of DNA and that a chromosome consists of a long stretch



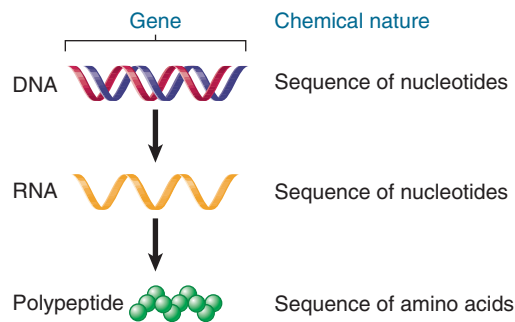
▶ **genome** The complete set of sequences in the genetic material of an organism. It includes the sequence of each chromosome plus any DNA in organelles.

▶ **chromosome** A discrete unit of the genome carrying many genes. Each consists of a very long molecule of duplex DNA and (in eukaryotes) an approximately equal mass of proteins. It is visible as a morphological entity only during cell division.

FIGURE 1.1 A brief history of genetics.

▶ **structural gene** A gene that encodes any RNA or polypeptide product other than a regulator.

FIGURE 1.2 A gene encodes an RNA, which may encode a polypeptide.



of DNA representing many genes, we will move to the overall organization of the genome. In *Chapter 4, The Interrupted Gene*, we take up in more detail the organization of the gene and its representation in proteins. In *Chapter 5, The Content of the Genome*, we consider the total number of genes, and in *Chapter 7, Clusters and Repeats*, we discuss other components of the genome and the maintenance of its organization.

CONCEPT AND REASONING CHECK

Why is it accurate to say that a genome has information for the development of an organism but does not directly participate in development?

1.2 DNA Is the Genetic Material of Bacteria, Viruses, and Eukaryotic Cells

► **transformation** In bacteria, it is the acquisition of new genetic material by incorporation of added DNA.

The idea that the genetic material is DNA has its roots in the discovery of **transformation** by Frederick Griffith in 1928 (see the accompanying box, *Historical Perspectives: Determining That DNA Is the Genetic Material*). Purification of the transforming principle from bacterial cells in 1944 by Avery, MacLeod, and McCarty showed that it is deoxyribonucleic acid (DNA).

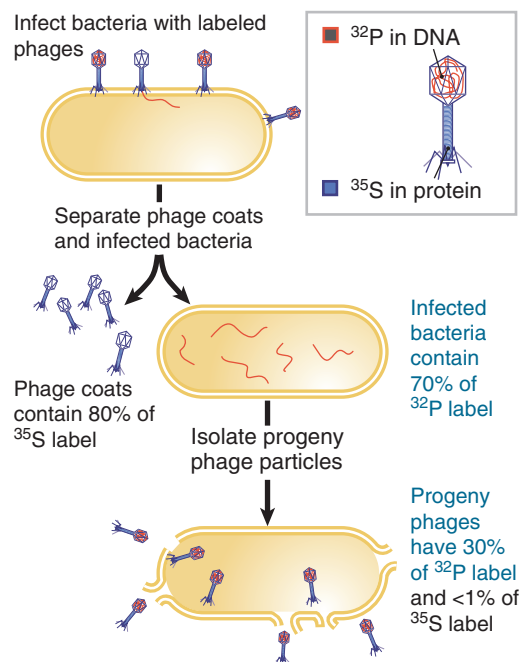
Having shown that DNA is the genetic material of bacteria, the next step was to demonstrate that DNA is the genetic material in a quite different system. Phage T2 is a bacteriophage virus (or phage) that infects the bacterium *Escherichia coli*. When phage particles are added to bacteria, they attach to the outside surface, some material enters the cell, and then ~20 minutes later each cell bursts open, or lyses, to release a large number of progeny phage.

FIGURE 1.3 illustrates the results of an experiment in 1952 by Alfred Hershey and Martha Chase in which bacteria were infected with T2 phages that had been radioactively labeled either in their DNA component (with ^{32}P) or in their protein component (with ^{35}S). The infected bacteria were agitated in a blender, and two fractions were separated by centrifugation. One fraction contained the empty phage “ghosts” that were released from the surface of the bacteria, and the other consisted of the infected bacteria themselves.

Previously, it had been shown that phage replication occurs intracellularly, so that the genetic material of the phage would have to enter the cell during infection.

Most of the ^{32}P label was present in the fraction containing infected bacteria. The progeny phage particles produced by the infection contained ~30% of the original ^{32}P label. The progeny received less than 1% of the protein contained in the original

FIGURE 1.3 The genetic material of phage T2 is DNA.



phage population. The phage ghosts consist of protein and therefore carried the ^{35}S radioactive label. This experiment directly showed that only the DNA of the parent phages enters the bacteria and becomes part of the progeny phages, which is exactly the pattern expected of genetic material.

A phage reproduces by commandeering the machinery of an infected host cell to manufacture more copies of itself. The phage possesses genetic material with properties analogous to those of cellular genomes: its traits are faithfully expressed and are subject to the same rules that govern inheritance of cellular traits. The case of T2 reinforces the general conclusion that DNA is genetic material of the genome of a cell or a virus.

When DNA is added to eukaryotic cells growing in culture, it enters the cells, and in some of them this results in the production of new proteins. When an isolated gene is used, its incorporation leads to the production of a particular protein, as depicted in **FIGURE 1.4**. Although for historical reasons these experiments are described as **transfection** when performed with animal cells, they are a direct counterpart to bacterial transformation. The DNA that is introduced into the recipient cell becomes part of its genome and is inherited with it, and expression of the new DNA results in a new trait. At first, these experiments were successful only with individual cells growing in culture, but in later experiments DNA was introduced into mouse eggs by microinjection and became a stable part of the genome of the mouse. Such experiments show directly that DNA is the genetic material in eukaryotes and that it can be transferred between different species and remain functional.

The genetic material of all known organisms and many viruses is DNA. However, some viruses use RNA as the genetic material. Therefore, the general nature of the genetic material is that it is always nucleic acid; specifically, it is DNA, except in the RNA viruses.

Cells that lack *TK* gene cannot produce thymidine kinase and die in absence of thymidine

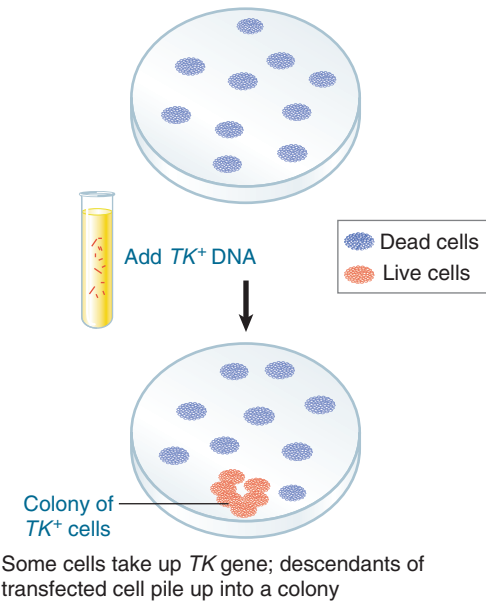


FIGURE 1.4 Eukaryotic cells can acquire a new phenotype as the result of transfection by added DNA.

► **transfection** In eukaryotic cells, the acquisition of new genetic material by incorporation of added DNA.

KEY CONCEPTS

- Bacterial transformation provided the first support that DNA is the genetic material of bacteria. Genetic properties can be transferred from one bacterial strain to another by extracting DNA from the first strain and adding it to the second strain.
- Phage infection showed that DNA is the genetic material of viruses. When the DNA and protein components of bacteriophages are labeled with different radioactive isotopes, only the DNA is transmitted to the progeny phages produced by infecting bacteria.
- DNA can be used to introduce new genetic traits into animal cells or whole animals.
- In some viruses, the genetic material is RNA.

CONCEPT AND REASONING CHECK

If Hershey and Chase had observed that nearly none of the DNA but a substantial fraction of the protein of phage T2 enters *E. coli* cells during infection, what would they have concluded?

HISTORICAL PERSPECTIVES

Determining That DNA Is the Genetic Material

Pneumonia was a leading cause of death in the early part of the 20th century, and much effort was put into understanding the structure and function of the bacteria known to cause it: the pneumococci. Scientists knew that several pneumococcal types existed and that these types could be distinguished by the molecules (capsular polysaccharides) displayed on their surface, but they did not know whether the different types represented individual and stable strains of bacteria or different developmental stages of a single strain. While conducting experiments to distinguish between these possibilities, Frederick Griffith set in motion a series of experiments that ultimately led to the identification of DNA as the genetic material of all cells.

Griffith studied S forms and R forms of pneumococcal bacteria (so-called for the smooth and rough appearance of the bacteria when grown in laboratory culture; see *Section 1.2, DNA Is the Genetic Material of Bacteria, Viruses, and Eukaryotic Cells*). The R form, which was generated in the laboratory from the S form, produced no capsular polysaccharides and did not cause pneumonia when injected into mice. The S form was lethal to mice but could be inactivated by exposing the bacteria to high heat before injection. Surprisingly, Griffith found that when the R form

was injected into mice alongside heat-killed S bacteria, the mice contracted pneumonia and died (**FIGURE B1.1**). What was more, live S-type bacteria—virulent and coated with capsular proteins—could be isolated from the dead mice. R-form bacteria had clearly been transformed into a stable S form, and the transformation process required some factor from the heat-killed S sample. Griffith suspected this factor, the so-called transforming principle, was a protein . . . and he was not alone.

Proteins are by far the most abundant macromolecules in living cells. Their essential components, the amino acids, come in twenty varieties and can be joined in

Pneumococcus types	Injection of cells	Result
Capsule smooth (S) appearance	Living S	Dies
	Heat-killed S	Lives
No capsule rough (R) appearance	Living R	Lives
	Heat-killed S + Living R	Dies

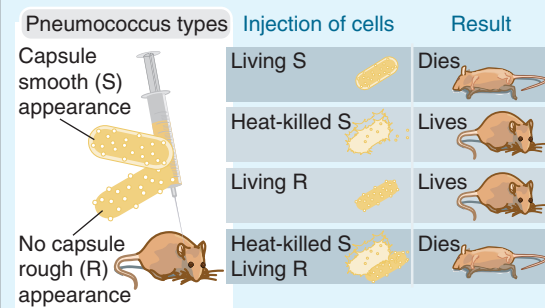


FIGURE B1.1 Neither heat-killed S-type nor live R-type bacteria can kill mice, but simultaneous injection of both can kill mice just as effectively as the live S-type.

1.3 Polynucleotide Chains Have Nitrogenous Bases Linked to a Sugar-Phosphate Backbone

- ▶ **purine** A double-ringed nitrogenous base, such as adenine or guanine.
- ▶ **pyrimidine** A single-ringed nitrogenous base, such as cytosine, thymine, or uracil.
- ▶ **nucleoside** A molecule consisting of a purine or pyrimidine base linked to the 1' carbon of a pentose sugar.
- ▶ **nucleotide** A molecule consisting of a purine or pyrimidine base linked to the 1' carbon of a pentose sugar and a phosphate group linked to either the 5' or 3' carbon of the sugar.
- ▶ **polynucleotide** A chain of nucleotides, such as DNA or RNA.

The essential component of nucleic acids (DNA and RNA) is the nucleotide, which has three parts:

- a nitrogenous base,
- a sugar, and
- one or more phosphates.

The nitrogenous base is a **purine** or **pyrimidine** ring. The base is linked to the 1' ("one prime") carbon on a pentose sugar by a glycosidic bond from the N₁ of pyrimidines or the N₉ of purines. The pentose sugar linked to a nitrogenous base is called a **nucleoside**. Nucleic acids are named for the type of sugar: DNA has 2'-deoxyribose, whereas RNA has 2'-ribose. The difference is that the sugar in RNA has a hydroxyl (–OH) group on the 2' carbon of the pentose ring. The sugar can be linked by its 5' or 3' carbon to a phosphate group. A nucleoside linked to a phosphate is a **nucleotide**.

A **polynucleotide** is a long chain of nucleotides. **FIGURE 1.5** shows that the backbone of the polynucleotide chain consists of an alternating series of pentose (sugar) and phosphate residues. The chain is formed by linking the 5' carbon of one pentose ring to the 3' carbon of the next pentose ring via a phosphate group, so the sugar-phosphate backbone is said to consist of 5'–3' phosphodiester linkages. The

seemingly endless combinations to result in molecules with a tremendous variety of size, shape, and function. By contrast, DNA is a relatively minor component of cells and was known at the time to comprise only four nucleotide types. These nucleotides were thought to be arranged in a specific, repetitive pattern, which contributed to the notion that deoxyribonucleic acids were molecules of “monotonous uniformity” that were unlikely to have any biological specificity. And, so, it was universally assumed that the transforming principle was a protein.

In 1944, Oswald Avery, Colin MacLeod, and Maclyn McCarty published the first report on the chemical nature of the transforming principle, and their results were startling. The main component of their active sample of purified transforming principle was not protein at all; it was DNA. To show that the transforming activity of this sample was due to the DNA and not to some minor but active contaminant, the authors treated it with enzymes that degrade protein, RNA, and carbohydrate; in all cases the sample retained the ability to stably transform R bacteria into S bacteria. The authors concluded that the transforming principle “consists principally, if not solely, of a highly polymerized, viscous form of deoxyribonucleic acid.”

Avery and his colleagues strengthened their claim that DNA was the transforming (or genetic) material by later showing that DNase I, an enzyme that specifically

degrades DNA, completely eliminated the transforming activity of their purified sample, but the scientific community remained skeptical. It was not until 1952, when A. D. Hershey and Martha Chase studied the independent roles of protein and DNA in bacteriophage T2 infection, that the notion of DNA as the genetic material finally gained favor.

Bacteriophage T2 is a virus that infects bacterial cells in discrete steps: the phage particle attaches itself to the bacterial cell wall, phage material is injected into the cell, the host cell is induced to produce new phage particles, and, finally, the bacterial cell bursts, releasing hundreds of new phage particles. To follow the movement of phage protein and phage DNA during this process, Hershey and Chase radioactively labeled each macromolecule separately: the protein with a radioactive form of sulfur and the DNA with a radioactive form of phosphorous. In their experiment, the majority of the phosphorous label entered the bacterial cell during infection, whereas the majority of the sulfur label stayed outside the cell. The authors went on to show that the new phage particles produced by the infected bacterial cell contained a large percentage of the labeled DNA but virtually none of the labeled protein (see Figure 1.3). These results indicated that it was DNA from the infecting phage, not protein, that entered the bacterial cell and induced genetic changes.

nitrogenous bases “stick out” from the backbone.

Each nucleic acid contains four types of nitrogenous base. The same two purines, adenine (A) and guanine (G), are present in both DNA and RNA. The two pyrimidines in DNA are cytosine (C) and thymine (T); in RNA uracil (U) is found instead of thymine. The only difference between uracil and thymine is the presence of a methyl group at position C₅.

The terminal nucleotide at one end of the chain has a free 5′ phosphate group, whereas the terminal nucleotide at the other end has a free 3′ hydroxyl group. It is conventional to write nucleic acid sequences in the 5′ to 3′ direction—that is, from the 5′ terminus at the left to the 3′ terminus at the right.

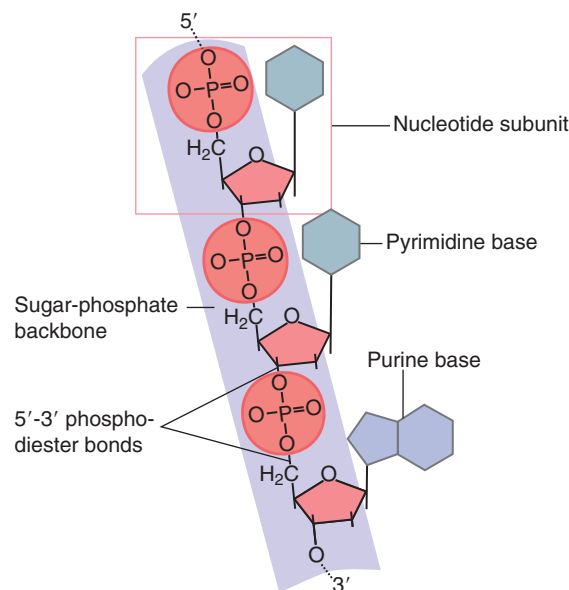


FIGURE 1.5 A polynucleotide chain consists of a series of 5′–3′ sugar-phosphate links that form a backbone from which the bases protrude.

KEY CONCEPTS

- A nucleoside consists of a purine or pyrimidine base linked to the 1' carbon of a pentose sugar.
- The difference between DNA and RNA is in the group at the 2' position of the sugar. DNA has a deoxyribose sugar (2'-H); RNA has a ribose sugar (2'-OH).
- A nucleotide consists of a nucleoside linked to a phosphate group on either the 5' or 3' carbon of the (deoxy)ribose.
- Successive (deoxy)ribose residues of a polynucleotide chain are joined by a phosphate group between the 3' carbon of one sugar and the 5' carbon of the next sugar.
- One end of the chain (conventionally written on the left) has a free 5' end and the other end of the chain has a free 3' end.
- DNA contains the four bases adenine, guanine, cytosine, and thymine; RNA normally has uracil instead of thymine.

CONCEPT AND REASONING CHECK

List the structural differences between DNA and RNA nucleotides.

1.4 DNA Is a Double Helix

By the 1950s, the observation by Erwin Chargaff that the nitrogenous bases are present in different amounts in the genomes of different species led to the concept that the sequence of bases is the form in which genetic information is carried. Given this concept, there were two remaining challenges: working out the structure of DNA and explaining how a sequence of bases in DNA could determine the sequence of amino acids in a protein.

Three pieces of evidence contributed to the construction of the double-helix model for DNA by James Watson and Francis Crick in 1953:

- X-ray diffraction data collected by Rosalind Franklin and Maurice Wilkins showed that the B-form of DNA (found in aqueous solution) is a regular helix, making a complete turn every 34 Å (3.4 nm), with a diameter of ~20 Å (2 nm). Since the distance between adjacent nucleotides is 3.4 Å (0.34 nm), there must be 10 nucleotides per turn.
- The density of DNA suggests that the helix must contain two polynucleotide chains. The constant diameter of the helix can be explained if the bases in each chain face inward and are restricted so that a purine is always paired with a pyrimidine, avoiding partnerships of purine-purine (which would be too wide) or pyrimidine-pyrimidine (which would be too narrow).
- Chargaff also observed that regardless of the absolute amounts of each base, the proportion of G is always the same as the proportion of C in DNA, and the proportion of A is always the same as that of T. Consequently, the composition of any DNA can be described by its G-C content, or the sum of the proportions of G and C bases. (The proportions of A and T bases can be determined by subtracting the G-C content from 1.) G-C content ranges from 0.26 to 0.74 for different species.

Watson and Crick proposed that the two polynucleotide chains in the double helix associate by hydrogen bonding between the nitrogenous bases. Normally, G can hydrogen bond specifically only with C, whereas A can bond specifically only with T. This hydrogen bonding between bases is described as *base pairing*, and the paired bases (G forming three hydrogen bonds with C, or A forming two hydrogen bonds with T) are said to be **complementary**. Base pairing occurs because of the complementary shapes of the complementary bases at the interfaces of where they pair, along with

► **complementary** Base pairs that match up in the pairing reactions in double helical nucleic acids (A with T in DNA or with U in RNA, and C with G).

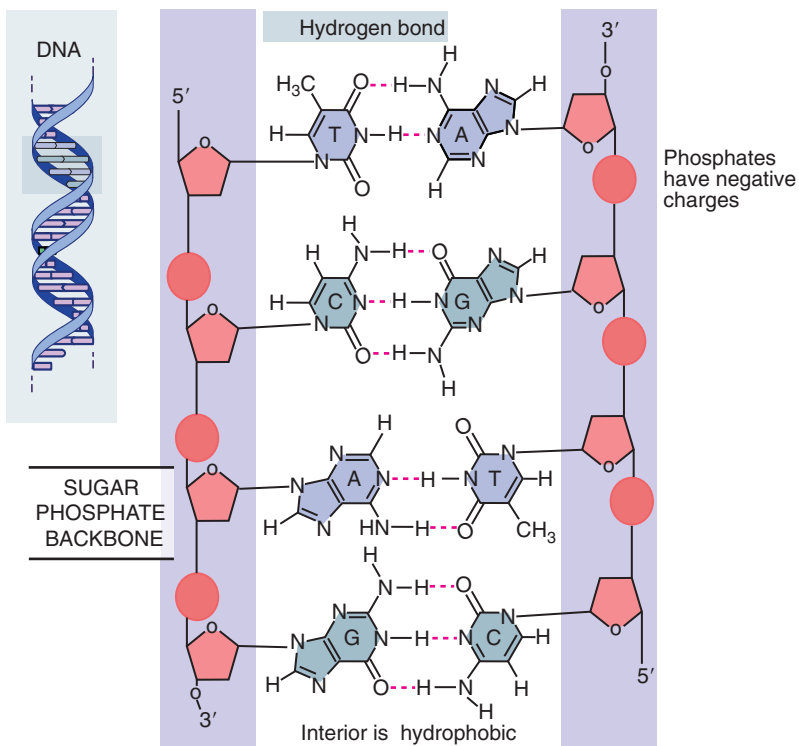


FIGURE 1.6 The DNA double helix maintains a constant width because purines always face pyrimidines in the complementary A-T and G-C base pairs. The sequence in the figure is T-A, C-G, A-T, G-C.

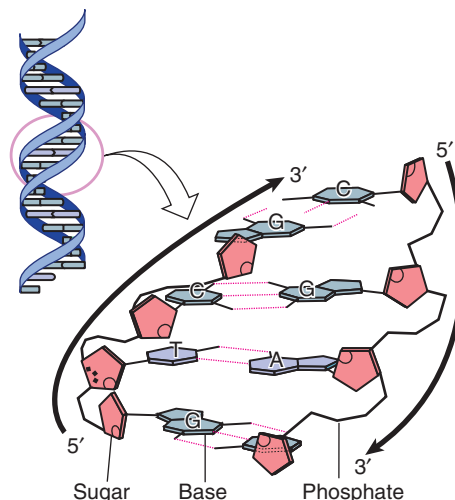
the location of just the right functional groups in just the right geometry along those interfaces so that hydrogen bonds can form.

The Watson-Crick model has the two polynucleotide chains running in opposite directions, so they are said to be **antiparallel**, as illustrated in **FIGURE 1.6**. Looking in one direction along the helix, one strand runs in the 5' to 3' direction, whereas its complement runs 3' to 5'. (Double-stranded DNA is often called *duplex DNA*.)

The sugar-phosphate backbones are on the outside of the double helix and carry negative charges on the phosphate groups. When DNA is in solution *in vitro*, the charges are neutralized by the binding of metal ions, typically sodium (Na^+). In the cell, positively charged proteins provide some of the neutralizing force. These proteins play important roles in determining the organization of DNA in the cell.

The base pairs (often abbreviated as bp) are on the inside of the double helix. They are flat and lie perpendicular to the axis of the helix. Using the analogy of the double helix as a spiral staircase, the base pairs form the steps, as illustrated schematically in **FIGURE 1.7**. Proceeding up the helix, bases are stacked above one another like a pile of plates.

Each base pair is rotated $\sim 36^\circ$ around the axis of the helix relative to the next base pair, so ~ 10 base pairs (in solution, 10.4 bp) make a complete turn of 360° . The twisting of the two strands around one another forms a double helix with a **minor groove** that is $\sim 12 \text{ \AA}$ (1.2 nm) across and a **major groove** that is $\sim 22 \text{ \AA}$ (2.2 nm) across, as can be seen from the scale model of **FIGURE 1.8**. In B-DNA, the double helix is said to be “right-handed”; the turns run clockwise as viewed along the helical axis.

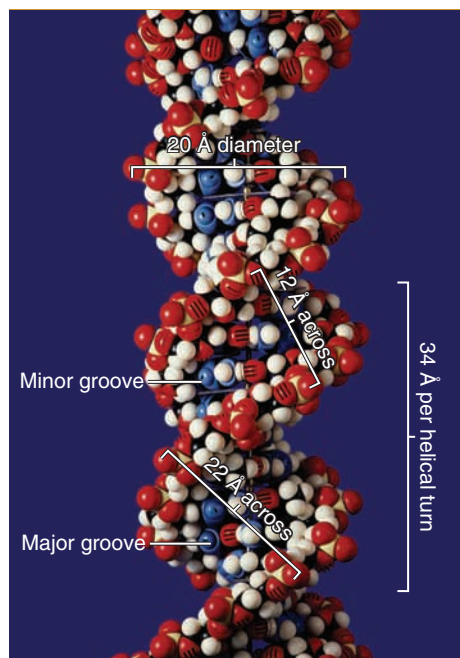


- ▶ **antiparallel** Strands of the DNA double helix organized in opposite orientation, so that the 5' end of one strand is aligned with the 3' end of the other strand.

FIGURE 1.7 Flat base pairs lie perpendicular to the sugar-phosphate backbone.

- ▶ **minor groove** A fissure running the length of the DNA double helix that is 12 \AA across.
- ▶ **major groove** A fissure running the length of the DNA double helix that is 22 \AA across.

FIGURE 1.8 The two strands of DNA form a double helix. Photo © Photodisc.



(The A-form of DNA, found in the absence of water, is also a right-handed helix that is shorter and thicker than the B-form. A third DNA structure, Z-DNA, is longer and narrower than the B-form and is a left-handed helix.)

It is important to realize that the Watson-Crick model of the B-form represents an average structure and that there can be local variations in the precise structure due to sequence. If the double helix has more base pairs per turn than the B-form, it is said to be **overwound**; if it has fewer base pairs per turn it is **underwound**. The degree of local winding can be affected by the overall conformation of the DNA double helix or by the binding of proteins to specific sites on the DNA.

- ▶ **overwound** B-form DNA that has more than 10.4 base pairs per turn of the helix.
- ▶ **underwound** B-form DNA that has fewer than 10.4 base pairs per turn of the helix.

KEY CONCEPTS

- The B-form of DNA is a double helix consisting of two polynucleotide chains that run antiparallel.
- The nitrogenous bases of each chain are flat purine or pyrimidine rings that face inward and pair with one another by hydrogen bonding to form only A-T or G-C pairs.
- The diameter of the double helix is 20 Å, and there is a complete turn every 34 Å, with ten base pairs per turn (~10.4 bp per turn in solution).
- The double helix has a major (wide) groove and a minor (narrow) groove.

CONCEPT AND REASONING CHECKS

1. Summarize the evidence that Watson and Crick used in proposing their model of the B-form of DNA.
2. If the G-C content of a DNA duplex is 0.44, what are the proportions of the four bases?

1.5 Supercoiling Affects the Structure of DNA

- ▶ **supercoiling** The coiling of a closed duplex DNA in space so that it crosses over its own axis.

The two strands of DNA are wound around each other to form the double helical structure; the double helix can also wind around itself to change the overall conformation, or *topology*, of the DNA molecule in space. This is called **supercoiling**. The effect can be imagined like a rubber band twisted around itself. Supercoiling creates tension in the DNA and, therefore, can occur only if the DNA has no free ends (otherwise the free ends can rotate to relieve the tension) or in linear DNA if it is anchored to a protein scaffold, as in eukaryotic chromosomes. The simplest example of a DNA with no free ends is a circular molecule. The effect of supercoiling can be seen by comparing the nonsupercoiled circular DNA lying flat (**FIGURE 1.9**, center) with the supercoiled circular molecule that forms a twisted (and, therefore, more condensed) shape (**FIGURE 1.9**, bottom).

The consequences of supercoiling depend on whether the DNA is twisted around itself in the same direction as the two strands within the double helix (clockwise) or in the opposite direction. Twisting in the same direction produces *positive supercoiling*, which overwinds the DNA so that there are fewer base pairs per turn than in the

B-form. Twisting in the opposite direction produces *negative supercoiling*, or underwinding, so there are more base pairs per turn than in the B-form. Both types of supercoiling of the double helix in space are tensions in the DNA (which is why DNA molecules with no supercoiling are said to be “relaxed”). Negative supercoiling can be thought of as creating tension in the DNA that is relieved by the unwinding of the double helix. The effect of severe negative supercoiling is to generate a region in which the two strands of DNA have separated (technically, zero base pairs per turn).

Topological manipulation of DNA is a central aspect of all its functional activities (recombination, replication, and transcription) as well as of the organization of its higher-order structure. All synthetic activities involving double-stranded DNA require the strands to separate. The strands do not simply lie side by side, though; they are intertwined. Their separation therefore requires the strands to rotate about each other in space. Some possibilities for the unwinding reaction are illustrated in **FIGURE 1.10**.

Unwinding a short linear DNA presents no problems, as the DNA ends are free to spin around the axis of the double helix to relieve any tension. However, DNA in a typical chromosome is not only extremely long but is also coated with proteins that serve to anchor the DNA at numerous points. Therefore, even a linear eukaryotic chromosome does not functionally possess free ends.

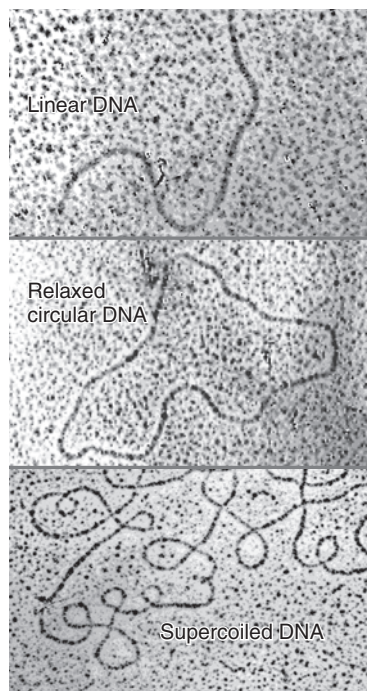


FIGURE 1.9 Linear DNA is extended (top); a circular DNA remains extended if it is relaxed (nonsupercoiled) (center); but a supercoiled DNA has a twisted and condensed form (bottom). Photos courtesy of Nirupam Roy Choudhury, International Centre for Genetic Engineering and Biotechnology (ICGEB).

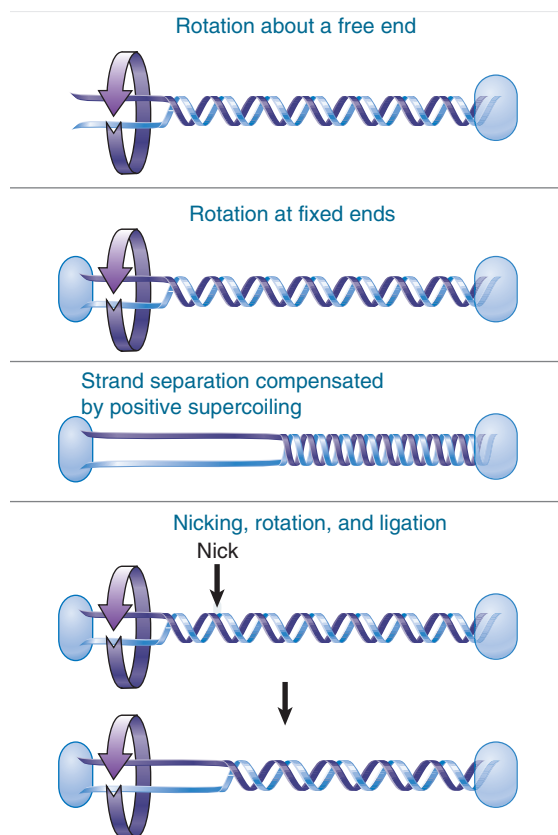


FIGURE 1.10 Separation of the strands of a DNA double helix can be achieved in several ways.

Consider the effects of separating the two strands in a molecule whose ends are not free to rotate. When two intertwined strands are pulled apart from one end, the result is to *increase* their winding about each other farther along the molecule, resulting in positive supercoiling elsewhere in the molecule to balance the underwinding generated in the single-stranded region. The problem can be overcome by introducing a transient nick in one strand. An internal free end allows the nicked strand to rotate about the intact strand, after which the nick can be sealed. Each repetition of the nicking and sealing reaction releases one superhelical turn. The topoisomerase enzymes that perform these reactions to control supercoiling in the cell will be discussed in *Section 15.7, Topoisomerases Relax or Introduce Supercoils in DNA*.

KEY CONCEPTS

- Supercoiling occurs only in “closed” DNA with no free ends.
- Closed DNA is either circular DNA or linear DNA in which the ends are anchored so that they are not free to rotate.

CONCEPT AND REASONING CHECK

Why does negative supercoiling facilitate unwinding of DNA but positive supercoiling inhibit unwinding?

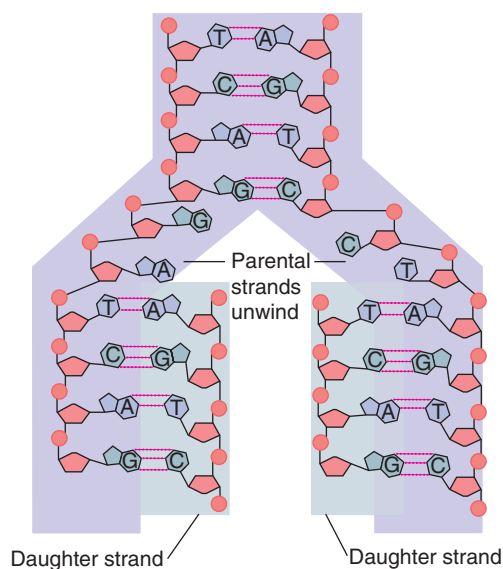
1.6 DNA Replication Is Semiconservative

It is crucial that DNA be reproduced accurately. The two polynucleotide strands are joined only by hydrogen bonds, so they are able to separate without the breakage of covalent bonds. The specificity of base pairing suggests that both of the separated parental strands could act as template strands for the synthesis of complementary daughter strands. **FIGURE 1.11** shows the principle that a new daughter strand is assembled from each parental strand. The sequence of the daughter strand is determined by the parental strand: an A in the parental strand causes a T to be placed in the daughter strand, a parental G directs incorporation of a daughter C, and so on.

The top part of Figure 1.11 shows an unreplicated parental duplex with the original two parental strands. The lower part shows the two daughter duplexes produced by complementary base pairing. Each of the daughter duplexes is identical in

FIGURE 1.11 Base pairing provides the mechanism for replicating DNA.

► **semiconservative replication** DNA replication accomplished by separation of the strands of a parental duplex, each strand then acting as a template for synthesis of a complementary strand.



sequence to the original parent duplex, containing one parental strand and one newly synthesized strand. The structure of DNA carries the information needed for its own replication. The consequences of this mode of replication, called **semi-conservative replication**, are illustrated in **FIGURE 1.12**. The unit conserved from one generation to the next is one of the two individual strands comprising the parental duplex.

Figure 1.12 illustrates a prediction of this model. If the parental DNA carries a “heavy” density label because the organism has been grown in medium containing a suitable isotope (such as ^{15}N), its strands can be distinguished from those that are synthesized when the organism is transferred to a medium

containing “light” isotopes. The parental DNA is a duplex of two “heavy” strands (red). After one generation of growth in “light” medium, the duplex DNA is *hybrid* in density—it consists of one heavy parental strand (red) and one light daughter strand (blue). After a second generation, the two strands of each hybrid duplex have separated. Each strand gains a light partner, so that now one half of the duplex DNA remains hybrid and the other half is entirely light (both strands are blue).

The individual strands of these duplexes are entirely heavy or entirely light. This pattern was confirmed experimentally by Matthew Meselson and Franklin Stahl in 1958. Meselson and Stahl followed the semiconservative replication of DNA through three generations of growth of *E. coli*. When DNA was extracted from bacteria and separated in a density gradient by centrifugation, the DNA formed bands corresponding to its density—heavy for parental, hybrid for the first generation, and half hybrid bands and half light bands in the second generation, indicating that a single parental strand is retained in the daughter molecule. (See the box in *Chapter 12* entitled *Historical Perspectives: The Meselson-Stahl Experiment* for more detail on this experiment.)

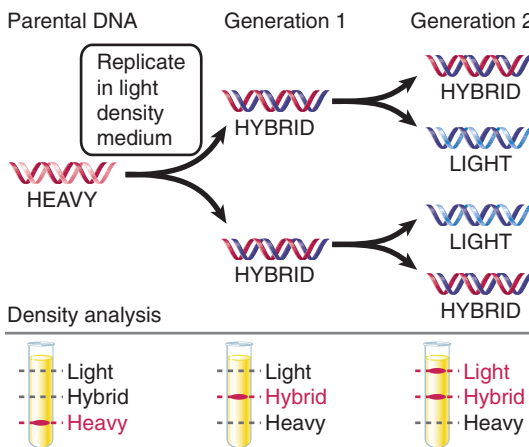


FIGURE 1.12 Replication of DNA is semiconservative.

KEY CONCEPTS

- The Meselson-Stahl experiment used “heavy” isotope labeling to show that the single polynucleotide strand is the unit of DNA that is conserved during replication.
- Each strand of a DNA duplex acts as a template for synthesis of a daughter strand.
- The sequences of the daughter strands are determined by complementary base pairing with the separated parental strands.

CONCEPT AND REASONING CHECK

What is the expected result of an experiment similar to that of Meselson and Stahl, but beginning with “light” DNA and culturing cells in a “heavy” medium?

1.7 Polymerases Act on Separated DNA Strands at the Replication Fork

Replication requires the two strands of the parental duplex to undergo separation, or **denaturation**. The disruption of the duplex, however, is only transient and is reversed, or undergoes **renaturation**, as the daughter duplex is formed. Only a small stretch of the duplex DNA is denatured at any moment during replication.

The helical structure of a molecule of DNA during replication is illustrated in **FIGURE 1.13**. The unreplicated region consists of the parental duplex, opening into the replicated region where the two daughter duplexes have formed. The duplex is disrupted at the junction between the two regions, which is called the **replication fork**. Replication involves movement of the replication fork along the parental DNA, so that there is continuous denaturation of the parental strands and formation of daughter duplexes.

- ▶ **denaturation** A molecule’s conversion from the physiological conformation to some other (inactive) conformation. In DNA, this involves the separation of the two strands due to breaking of hydrogen bonds between bases.
- ▶ **renaturation** The reassociation of denatured complementary single strands of a DNA double helix.
- ▶ **replication fork** The point at which strands of parental duplex DNA are separated so that replication can proceed. A complex of proteins, including DNA polymerase, is found there.

FIGURE 1.13 The replication fork is the region of DNA in which there is a transition from the unwound parental duplex to the newly replicated daughter duplexes.

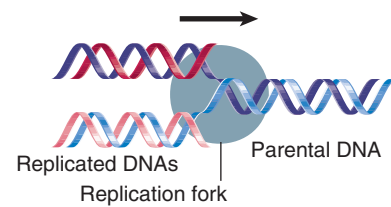


FIGURE 1.14 An endonuclease cleaves a bond within a nucleic acid. This example shows an enzyme that attacks one strand of a DNA duplex.

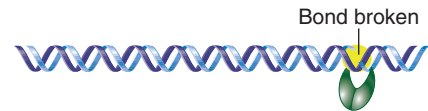
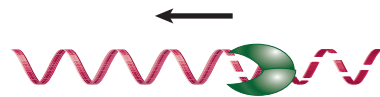


FIGURE 1.15 An exonuclease removes bases one at a time by cleaving the last bond in a polynucleotide chain.



- ▶ **DNA polymerase** An enzyme that synthesizes a daughter strand(s) of DNA (under direction from a DNA template). Any particular enzyme may be involved in repair or replication (or both).
- ▶ **DNase** An enzyme that degrades DNA.
- ▶ **RNase** An enzyme that degrades RNA.
- ▶ **exonuclease** An enzyme that cleaves nucleotides one at a time from the end of a polynucleotide chain; it may be specific for either the 5' or 3' end of DNA or RNA.
- ▶ **endonuclease** An enzyme that cleaves bonds within a nucleic acid chain; it may be specific for RNA or for single-stranded or double-stranded DNA.

The synthesis of DNA is aided by specific enzymes, **DNA polymerases**, that recognize the template strand and catalyze the addition of nucleotide subunits to the polynucleotide chain that is being synthesized. They are accompanied in DNA replication by ancillary enzymes such as helicases that unwind the DNA duplex, a primase that synthesizes an RNA primer required by DNA polymerase, and DNA ligase that connects discontinuous DNA fragments. Degradation of nucleic acids also requires specific enzymes: deoxyribonucleases (**DNases**) degrade DNA, and ribonucleases (**RNases**) degrade RNA. The nucleases (see *Chapter 3, Methods in Molecular Biology and Genetic Engineering*) fall into the general classes of **exonucleases** and **endonucleases**:

- Endonucleases break individual phosphodiester linkages within RNA or DNA molecules, generating discrete fragments. Some DNases cleave both strands of a duplex DNA at the target site, whereas others cleave only one of the two strands. Endonucleases are involved in cutting reactions, as shown in **FIGURE 1.14**.
- Exonucleases remove nucleotide residues one at a time from the end of the molecule, generating mononucleotides. They always function on a single nucleic acid strand, and each exonuclease proceeds in a specific direction, that is, starting either at a 5' or at a 3' end and proceeding toward the other end. They are involved in trimming reactions, as shown in **FIGURE 1.15**.

KEY CONCEPTS

- Replication of DNA is undertaken by a complex of enzymes that separate the parental strands and synthesize the daughter strands.
- The replication fork is the point at which the parental strands are separated.
- The enzymes that synthesize DNA are called DNA polymerases.
- Nucleases are enzymes that degrade nucleic acids; they include DNases and RNases and can be categorized as endonucleases or exonucleases.

CONCEPT AND REASONING CHECK

What are the functions of DNA polymerases, DNases, and RNases in living cells?

1.8 Genetic Information Can Be Provided by DNA or RNA

The **central dogma** is the dominant paradigm of molecular biology. Structural genes exist as sequences of nucleic acid, but they function by being expressed in the form of polypeptides. Replication makes possible the inheritance of genetic information, while transcription and translation are responsible for its expression to another form.

The central dogma includes several observations about the processes of replication, transcription, and translation (see Figure 1.29):

- Transcription of DNA by a DNA-dependent **RNA polymerase** generates RNA molecules. Messenger RNAs (mRNAs) are translated to polypeptides. Other types of RNA, such as rRNAs and tRNAs, are functional themselves and are not translated.
- A genetic system may involve either DNA or RNA as the genetic material. Cells use only DNA. Some viruses use RNA, and replication of viral RNA by an RNA-dependent RNA polymerase occurs in the infected cell.
- The expression of cellular genetic information is usually unidirectional. Transcription of DNA generates RNA molecules; the exception is the reverse transcription of retroviral RNA to DNA that occurs when retroviruses infect cells (see below). Generally polypeptides cannot be retrieved for use as genetic information; translation of RNA into polypeptide is always irreversible.

These mechanisms are equally effective for the cellular genetic information of prokaryotes or eukaryotes and for the information carried by viruses. The genomes of all living organisms consist of duplex DNA. Viruses have genomes that consist of DNA or RNA, and there are examples of each type that are double-stranded (dsDNA or dsRNA) or single-stranded (ssDNA or ssRNA). Details of the mechanism used to replicate the nucleic acid vary among viruses, but the principle of replication via synthesis of complementary strands remains the same, as illustrated in **FIGURE 1.16**.

The restriction of a unidirectional transfer of information from DNA to RNA in cells is not absolute. It is broken by the retroviruses, which have genomes consisting of a single-stranded RNA molecule. During the retroviral cycle of infection, the RNA is converted into a single-stranded DNA by the process of **reverse transcription**, which is accomplished by the enzyme *reverse transcriptase*, an RNA-dependent DNA polymerase. The resulting ssDNA is in turn converted into dsDNA. This duplex DNA becomes part of the genome of the host cell and is inherited like any other gene. So reverse transcription allows a sequence of RNA to be retrieved and used as DNA in a cell.

The existence of RNA replication and reverse transcription establishes the general principle that information in the form of either type of nucleic acid sequence can be converted into the other type. In the usual course of events, however, the cell relies on the processes of DNA replication, transcription, and translation. But on rare occasions (possibly mediated by an RNA virus), information from a cellular RNA is converted into DNA and inserted into the genome. Although retroviral reverse transcription is not necessary for the regular operations of the cell, it becomes a mechanism

▶ **central dogma** Information cannot be transferred from polypeptide to polypeptide or polypeptide to nucleic acid, but can be transferred between nucleic acids and from nucleic acid to polypeptide.

▶ **RNA polymerase** An enzyme that synthesizes RNA using a DNA template (formally described as DNA-dependent RNA polymerases).

▶ **reverse transcription** Synthesis of DNA from a template of RNA by the enzyme reverse transcriptase.

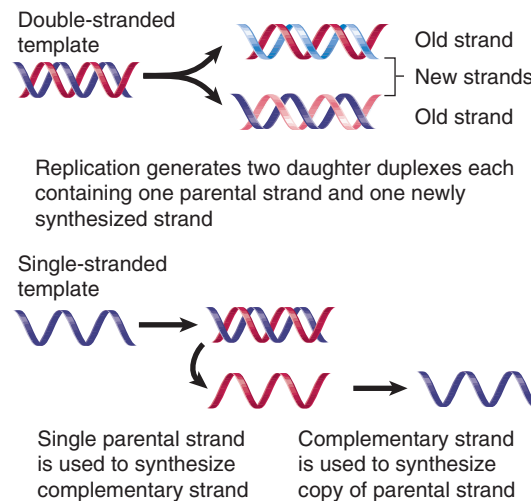


FIGURE 1.16 Double-stranded and single-stranded nucleic acids both replicate by synthesis of complementary strands governed by the rules of base pairing.

FIGURE 1.17 The size of the genome varies over an enormous range.

Genome	Gene Number	Base Pairs
Organisms		
Plants	<50,000	<10 ¹¹
Mammals	30,000	~3 × 10 ⁹
Worms	14,000	~10 ⁸
Flies	12,000	1.6 × 10 ⁸
Fungi	6,000	1.3 × 10 ⁷
Bacteria	2–4,000	<10 ⁷
Mycoplasma	500	<10 ⁶
dsDNA Viruses		
Vaccinia	<300	187,000
Papova (SV40)	~6	5,226
Phage T4	~200	165,000
ssDNA Viruses		
Parvovirus	5	5,000
Phage φX174	11	5,387
dsRNA Viruses		
Reovirus	22	23,000
ssRNA Viruses		
Coronavirus	7	20,000
Influenza	12	13,500
TMV	4	6,400
Phage MS2	4	3,569
STNV	1	1,300
Viroids		
PSTV RNA	0	359

of potential importance when we consider the evolution of the genome (see *Chapter 8, Genome Evolution*).

The same principles for the perpetuation of genetic information apply to the massive genomes of plants or amphibians as well as the tiny genomes of mycoplasma and the even smaller genomes of DNA or RNA viruses. **FIGURE 1.17** presents some examples that illustrate the range of genome types and sizes. The reasons for such variation in genome size and gene number will be explored in *Chapters 5 and 6*.

Among the various living organisms, with genomes varying in size over a 100,000-fold range, a common principle prevails: the DNA encodes all of the polypeptides that the cell(s) of the organism must synthesize, and the polypeptides in turn (directly or indirectly) provide the functions needed for survival. A similar principle describes the function of the genetic information of viruses, whether DNA or RNA: the nucleic acid codes for the polypeptide(s) needed to

package the genome and for any other functions in addition to those provided by the host cell that are needed to reproduce the virus. (The smallest virus—the satellite tobacco necrosis virus [STNV]—cannot replicate independently. It requires the presence of a “helper” virus—the tobacco necrosis virus [TNV], which is itself a normally infectious virus.)

KEY CONCEPTS

- Cellular genes are DNA, but viruses may have genomes of RNA.
- DNA is converted into RNA by transcription, and RNA may be converted into DNA by reverse transcription.
- The translation of RNA into protein is unidirectional.

CONCEPT AND REASONING CHECK

What types of enzymes would be necessary to replicate ssDNA, ssRNA, dsDNA, and dsRNA genomes to produce exact copies of the same type of nucleic acid?

1.9 Nucleic Acids Hybridize by Base Pairing

A crucial property of the double helix is the capacity to separate the two strands without disrupting the covalent bonds that form the polynucleotides and at the (very rapid) rates needed to sustain genetic functions. The specificity of the processes of denaturation and renaturation is determined by complementary base pairing.

The concept of base pairing is central to all processes involving nucleic acids. Disruption of the base pairs is crucial to the function of a double-stranded nucleic acid, whereas the ability to form base pairs is essential for the activity of a single-stranded nucleic acid. **FIGURE 1.18** shows that base pairing enables complementary single-stranded nucleic acids to form a duplex.

- An intramolecular duplex region can form by base pairing between two complementary sequences that are part of a single-stranded nucleic acid.
- A single-stranded nucleic acid may base pair with an independent, complementary single-stranded nucleic acid to form an intermolecular duplex.

Formation of duplex regions from single-stranded nucleic acids is most prevalent in RNA but is also important for single-stranded viral DNA genomes. Base pairing between independent complementary single strands is not restricted to DNA–DNA or RNA–RNA but can also occur between DNA and RNA.

The lack of covalent bonds between complementary strands makes it possible to manipulate DNA *in vitro*. The hydrogen bonds that stabilize the double helix are disrupted by heating or low salt concentration. The two strands of a double helix separate entirely when all the hydrogen bonds between them are broken.

Denaturation of DNA occurs over a narrow temperature range and results in striking changes in many of its physical properties. The midpoint of the temperature range over which the strands of DNA separate is called the **melting temperature** (T_m) and depends on the G-C content of the duplex. Because each G-C base pair has three hydrogen bonds, it is more stable than an A-T base pair, which has only two hydrogen bonds. The more G-C base pairs in a DNA, the greater the energy that is needed to separate the two strands. In solution under physiological conditions, a DNA that is 40% G-C (a value typical of mammalian genomes) denatures with a T_m of about 87°C, so duplex DNA is stable at the temperature of the cell.

The denaturation of DNA is reversible under appropriate conditions. Renaturation depends on specific base pairing between the complementary strands. **FIGURE 1.19** shows that the reaction takes place in two stages. First, single strands of DNA in the solution encounter one another by chance; if their sequences are complementary, the two strands base pair to generate a short double-stranded region. This region of base pairing then extends along the molecule, much like a zipper, to form a lengthy duplex.

Complete renaturation restores the properties of the original double helix. The property of renaturation applies to any two complementary nucleic acid sequences. This is sometimes called **annealing**, but the reaction is more generally called **hybridization** whenever nucleic acids from different sources are involved, as in the case when DNA hybridizes to RNA. The ability of two nucleic acids to hybridize constitutes a precise test for their complementarity because only complementary sequences can form a duplex (although under certain conditions imperfect matches can be tolerated).

The principle of the hybridization reaction is to combine two single-stranded nucleic acids in solution and then to measure the amount of double-stranded material that forms. **FIGURE 1.20** illustrates a procedure in which a DNA preparation is denatured and the single strands are attached to a filter. Then a second denatured DNA (or RNA) preparation is added. The filter is treated so that the second preparation can attach to it only if it is able to base pair with the DNA that

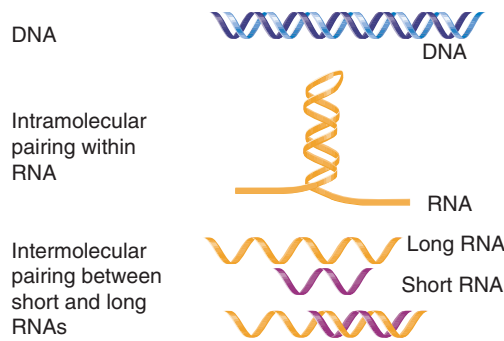


FIGURE 1.18 Base pairing occurs in duplex DNA and also in intra- and intermolecular interactions in single-stranded RNA (or DNA).

► **melting temperature** The midpoint of the temperature range over which the strands of DNA separate.

► **annealing** The renaturation of a duplex structure from single strands that were obtained by denaturing duplex DNA.

► **hybridization** The pairing of complementary RNA and DNA strands to give an RNA–DNA hybrid.

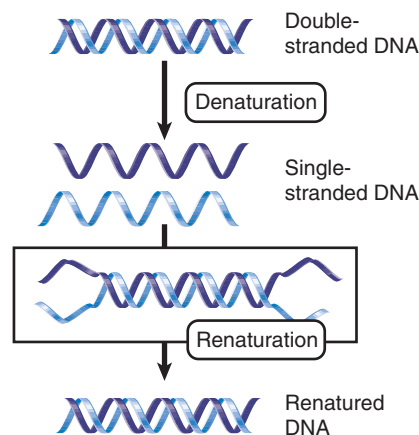
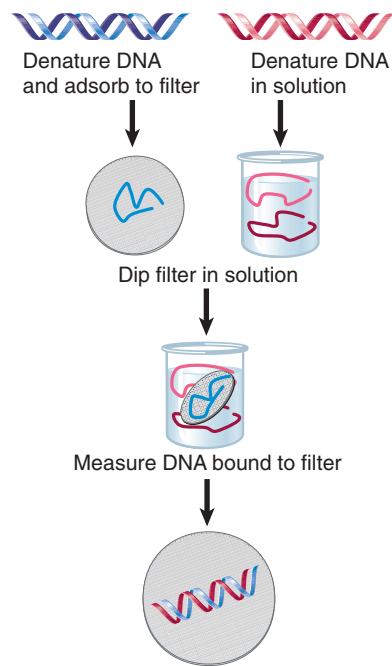


FIGURE 1.19 Denatured single strands of DNA can renature to give the duplex form.

FIGURE 1.20 Filter hybridization establishes whether a solution of denatured DNA (or RNA) contains sequences complementary to the strands immobilized on the filter.



was originally attached. Usually the second preparation is labeled so that the hybridization reaction can be measured as the amount of label retained by the filter. Alternatively, hybridization in solution can be measured as the change in UV-absorbance of a nucleic acid solution at 260 nm as detected via spectrophotometry. As DNA denatures to single strands with increasing temperature, UV-absorbance of the DNA solution increases; UV-absorbance consequently decreases as ssDNA hybridizes to complementary DNA or RNA with decreasing temperature.

Two sequences need not be perfectly complementary to hybridize. If they are similar but not identical, an imperfect duplex is formed in which base pairing is interrupted at positions where the two single strands are not complementary.

KEY CONCEPTS

- Heating causes the two strands of a DNA duplex to separate.
- T_m is the midpoint of the temperature range for denaturation.
- Complementary single strands can renature when the temperature is reduced.
- Denaturation and renaturation/hybridization can occur with DNA–DNA, DNA–RNA, or RNA–RNA combinations and can be intermolecular or intramolecular.
- The ability of two single-stranded nucleic acids to hybridize is a measure of their complementarity.

CONCEPT AND REASONING CHECK

Describe how measures of hybridization between DNA from different species can be used to estimate the evolutionary relationships between those species.

1.10 Mutations Change the Sequence of DNA

Mutations provide decisive evidence that DNA is the genetic material. When a change in the sequence of DNA causes an alteration in a polypeptide, we may conclude that the DNA encodes that polypeptide. Furthermore, a corresponding change in the phenotype of the organism may allow us to identify the function of that polypeptide. The existence of many mutations in a gene may allow many variant forms of a polypeptide to be compared, and a detailed analysis can be used to identify regions of the polypeptide responsible for individual enzymatic or other functions.

All organisms suffer a certain number of mutations as the result of normal cellular operations or random interactions with the environment. These are called **spontaneous mutations**, and the rate at which they occur (the “background level”) is characteristic for any particular organism. Mutations are rare events, and of course those that have deleterious effects are selected against during evolution. It is, therefore, difficult to observe large numbers of spontaneous mutants from natural populations.

► **spontaneous mutations** Mutations that occur in the absence of any added reagent to increase the mutation rate as the result of errors in replication (or other events involved in the reproduction of DNA) or by random changes to the chemical structure of bases.

The occurrence of mutations can be increased by treatment with certain compounds. These are called **mutagens**, and the changes they cause are called **induced mutations**. Most mutagens either modify a particular base of DNA or become incorporated into the nucleic acid. The potency of a mutagen is judged by how much it increases the rate of mutation above background. By using mutagens, it becomes possible to induce many changes in any gene.

Mutation rates can be measured at several levels of resolution: mutation across the whole genome (as the rate per genome per generation), mutation in a gene (as the rate per locus per generation), or mutation at a specific nucleotide site (as the rate per base pair per generation). These rates correspondingly decrease as smaller units are observed.

Spontaneous mutations that inactivate gene function occur in bacteriophages and bacteria at a relatively constant rate of 3 to 4×10^{-3} per genome per generation. Given the large variation in genome sizes between bacteriophages and bacteria, this corresponds to great differences in the mutation rate per base pair. This suggests that the overall rate of mutation has been subject to selective forces that have balanced the deleterious effects of most mutations against the advantageous effects of some mutations. This conclusion is strengthened by the observation that an archaean that lives under harsh conditions of high temperature and acidity (which are expected to damage DNA) does not show an elevated mutation rate but in fact has an overall mutation rate just below the average range. **FIGURE 1.21** shows that in bacteria, the mutation rate corresponds to $\sim 10^{-6}$ events per locus per generation or to an average rate of change per base pair of 10^{-9} to 10^{-10} per generation. The rate at individual base pairs varies very widely, over a 10,000-fold range. We have no accurate measurement of the rate of mutation in eukaryotes, although usually it is thought to be somewhat similar to that of bacteria on a per locus per generation basis. One reason that mutation rates vary across species is that the activity and efficacy of DNA repair systems also vary. DNA repair systems will be discussed in *Chapter 16, Repair Systems*.

KEY CONCEPTS

- All mutations are changes in the sequence of DNA.
- Mutations may occur spontaneously or may be induced by mutagens.

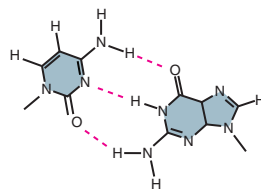
CONCEPT AND REASONING CHECK

What are the advantages of maintaining a nonzero mutation rate?

1.11 Mutations May Affect Single Base Pairs or Longer Sequences

Any base pair of DNA can be mutated. A **point mutation** changes only a single base pair and can be caused by either of two types of event:

- Chemical modification of DNA directly changes one base into a different base.
- An error during the replication of DNA causes the wrong base to be inserted into a polynucleotide.



Mutation rate

Any base pair
1 in 10^9 – 10^{10}
generations



Any gene
1 in 10^5 – 10^6
generations



The genome
1 in 300
generations

FIGURE 1.21 A base pair is mutated at a rate of 10^{-9} – 10^{-10} per generation, a gene of 1000 bp is mutated at $\sim 10^{-6}$ per generation, and a bacterial genome is mutated at 3×10^{-3} per generation.

- ▶ **mutagens** Substances that increase the rate of mutation by inducing changes in DNA sequence, directly or indirectly.
- ▶ **induced mutations** Mutations that result from the action of a mutagen. The mutagen may act directly on the bases in DNA, or it may act indirectly to trigger a pathway that leads to a change in DNA sequence.

FIGURE 1.22 Mutations can be induced by chemical modification of a base.

- ▶ **transition** A mutation in which one pyrimidine is replaced by the other or in which one purine is replaced by the other.
- ▶ **transversion** A mutation in which a purine is replaced by a pyrimidine, or vice versa.

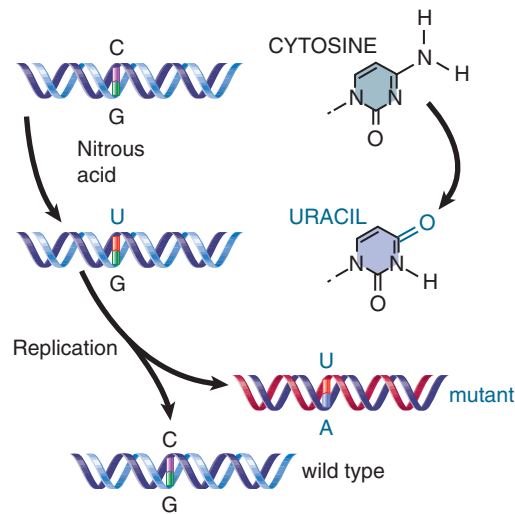
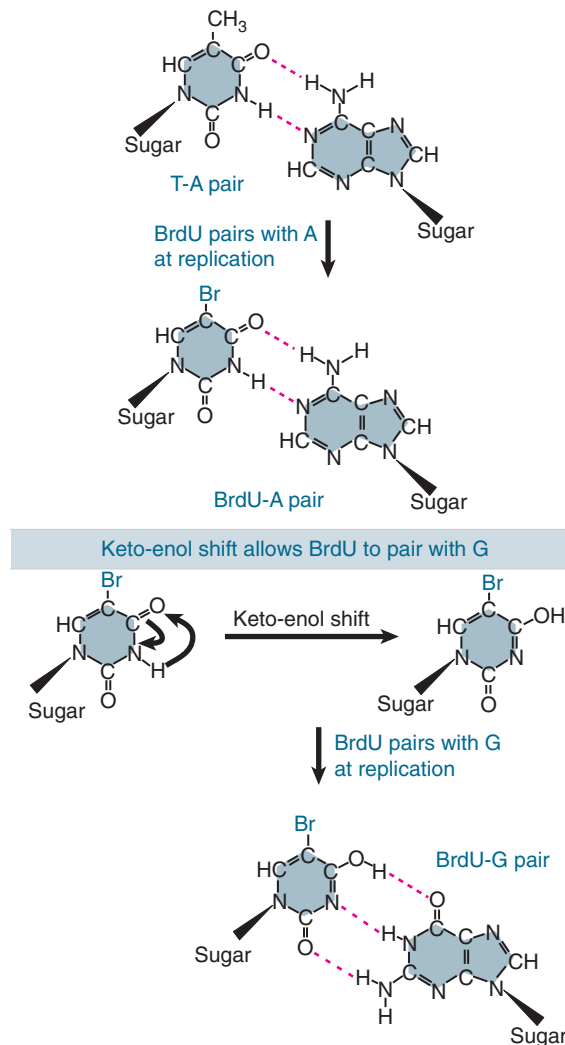


FIGURE 1.23 Mutations can be induced by the incorporation of base analogs into DNA.



position within a normal base and results in an anomalous, but more stable, base pairing. For example, while the common keto form of guanine pairs stably with cytosine, the rare enol form of guanine pairs stably with thymine.

Point mutations can be divided into two types, depending on the nature of the base substitution:

- The most common class is the **transition**, resulting from the substitution of one pyrimidine by the other or of one purine by the other. This replaces a G-C pair with an A-T pair, or vice versa.
- The less common class is the **transversion**, in which a purine is replaced by a pyrimidine, or vice versa, so that (for example) an A-T pair becomes a T-A or C-G pair.

As shown in **FIGURE 1.22**, the mutagen nitrous acid performs an oxidative deamination that converts cytosine into uracil, resulting in a transition. In the replication cycle following the transition, the U pairs with an A instead of the G with which the original C would have paired. So the C-G pair is replaced by a T-A pair when the A pairs with the T in the next replication cycle. (Nitrous acid can also deaminate adenine, causing the reverse transition from A-T to G-C.)

Transitions are also caused by base mispairing, when noncomplementary bases pair instead of the usual Watson-Crick pairs. Base mispairing usually occurs as an aberration resulting from the incorporation into DNA of an abnormal base that has flexible pairing properties.

FIGURE 1.23 shows the example of the mutagen bromouracil (BrdU), an analog of thymine that contains a bromine atom in place of thymine's methyl group and can be incorporated into DNA in place of thymine. However, BrdU has flexible pairing properties because the presence of the bromine atom allows a *tautomeric shift* from a keto (=O) form to an enol (-OH) form. The enol form of BrdU can pair with guanine, which after replication leads to substitution of the original A-T pair by a G-C pair. Tautomeric shifts can also occur when a proton shifts

Transversions are rarer than transitions, since they require a temporary purine-purine or pyrimidine-pyrimidine pairing that would alter the diameter of the DNA duplex. However, one cause of transversions is a proton shift followed by a 180° rotation of the base around the glycosidic bond. For example, a rotated synadenine (produced by a proton shift in adenine) will pair stably with a normal adenine.

Point mutations were thought for a long time to be the principal means of change in individual genes. However, we now know that insertions and deletions (“indels”) of short sequences are quite frequent. Some mutagens, such as intercalating agents, can cause the insertion or deletion of a single base pair. An intercalating agent will insert itself between two adjacent base pairs in a DNA duplex, so that the duplex is distorted and a DNA polymerase can either skip or add a base during DNA replication. If this occurs in the coding sequence of a gene, a frameshift mutation will result (see Section 2.7, *The Genetic Code Is Triplet*). Often, the insertions are the result of transposable elements, which are sequences of DNA with the ability to move from one site to another (see Chapter 17, *Transposable Elements and Retroviruses*). An insertion within a coding region usually abolishes the activity of the gene. However, both insertions and deletions of short sequences can occur by other mechanisms—for example, those involving errors during replication or recombination. In addition, mutagens belonging to a class called acridines introduce very small insertions and deletions.

KEY CONCEPTS

- A point mutation changes a single base pair.
- Point mutations can be caused by the chemical conversion of one base into another or by errors that occur during replication.
- A transition replaces a G-C base pair with an A-T base pair, or vice versa.
- A transversion replaces a purine with a pyrimidine, such as changing A-T to T-A.
- Insertions can result from the movement of transposable elements.

CONCEPT AND REASONING CHECK

Why are transitions more common than transversions? Consider how the DNA repair mechanisms might recognize errors and the effects of these mutations on DNA structure in your answer.

1.12 The Effects of Mutations Can Be Reversed

FIGURE 1.24 shows that the possibility of reversion mutations, or **revertants**, is an important characteristic that distinguishes point mutations and insertions from deletions. Mutations that inactivate a gene are called **forward mutations**. Their effects are reversed by **back mutations**, which are of two types: true reversions and second-site reversions.

- A point mutation can revert either by a true reversion or a second-site reversion.
- An insertion can revert by deletion of the inserted sequence.
- A deletion of a sequence cannot revert in the absence of some mechanism to restore the lost sequence.

An exact reversal of the original mutation is called a **true reversion**. For example, if an A-T pair was replaced by a G-C pair in the original mutation, another mutation to restore the A-T pair will exactly regenerate the original sequence. The exact removal of a transposable element following its insertion is another example of a true reversion.

The second type of back mutation, **second-site reversion**, may occur elsewhere in the gene, and its effects compensate for the first mutation. For example, one amino acid change in a protein may abolish its function, but a second alteration may

- ▶ **revertants** Reversions of a mutant cell or organism to the wild-type phenotype.
- ▶ **forward mutation** A mutation that inactivates a functional gene.
- ▶ **back mutation** A mutation that reverses the effect of a mutation that had inactivated a gene; thus it restores the original sequence or function of the gene product.
- ▶ **true reversion** A mutation that restores the original sequence of the DNA.
- ▶ **second-site reversion** A second mutation suppressing the effect of a first mutation.

FIGURE 1.24 Point mutations and insertions can revert, but deletions cannot revert.



► **suppression mutation** A second event eliminates the effects of a mutation without reversing the original change in DNA.

compensate for the first and restore protein activity.

A forward mutation results from any change that alters the function of a gene product, whereas a back mutation must restore the original function to the altered gene product. Therefore, the possibilities for back mutations are much more restricted than those for forward mutations. The rate of back mutations is correspondingly lower than that of forward mutations, typically by a factor of ~10.

Mutations in other genes can also occur to circumvent the effects of mutation in the original gene. This is called a **suppression mutation**. A locus in which a mutation suppresses the effect of a mutation in another locus is called a *suppressor*. For example, a point mutation may cause an amino acid substitution in a polypeptide, whereas a second mutation in a tRNA gene may cause it to recognize the mutated codon and, as a result, insert the original amino acid during translation. (Note that this suppresses the original mutation but causes errors during translation of other mRNAs.)

KEY CONCEPTS

- Forward mutations alter the function of a gene, and back mutations (or revertants) reverse their effects.
- Insertions can revert by deletion of the inserted material, but deletions cannot revert.
- Suppression occurs when a mutation in a second gene bypasses the effect of mutation in the first gene.

CONCEPT AND REASONING CHECK

Transposable elements were originally identified because the rate of reversion of mutations caused by them is much higher than the reversion rate for point mutations. Explain why the reversion rate is so high.

1.13 Mutations Are Concentrated at Hotspots

So far we have presented mutations in terms of individual changes in the sequence of DNA that influence the activity of the DNA in which they occur. When we consider mutations in terms of the alteration of function of the gene, most genes within a species show more or less similar rates of mutation relative to their size. This suggests that the gene can be regarded as a target for mutation and that damage to any part of it can alter its function. As a result, susceptibility to mutation is roughly proportional to the size of the gene. But are all base pairs in a gene equally susceptible, or are some more likely to be mutated than others?

What happens when we isolate a large number of independent mutations in the same gene? Each is the result of an individual mutational event. Most mutations

will occur at different sites, but some will occur at the same position. Two independently isolated mutations at the same site may constitute exactly the same change in DNA (in which case the same mutation has happened more than once), or they may constitute different changes (three different point mutations are possible at each base pair).

The histogram of **FIGURE 1.25** shows the frequency with which mutations are found at each base pair in the *lacI* gene of *E. coli*. The statistical probability that more than one mutation occurs at a particular site is given by random-hit kinetics (as seen in the Poisson distribution). Some sites will gain one, two, or three mutations, whereas others will not gain any. Some sites gain far more than the number of mutations expected from a random distribution; they may have 10× or even 100× more mutations than predicted by random hits. These sites are called **hotspots**. Spontaneous mutations may occur at hotspots, and different mutagens may have different hotspots.

A major cause of spontaneous mutation is the presence of an unusual base in the DNA. In addition to the four standard bases of DNA, *modified bases* are sometimes found. The name reflects their origin; they are produced by chemical modification of one of the four standard bases. The most common modified base is 5-methylcytosine, which is generated when a methylase enzyme adds a methyl group to cytosine residues at specific sites in the DNA. Sites containing 5-methylcytosine are hotspots for spontaneous point mutation in *E. coli*. In each case, the mutation is a G-C to A-T transition. The hotspots are not found in mutant strains of *E. coli* that cannot methylate cytosine.

The reason for the existence of these hotspots is that cytosine bases suffer a higher frequency of spontaneous deamination. In this reaction, the amino group is replaced by a keto group. Recall that deamination of cytosine generates uracil (see Figure 1.22). **FIGURE 1.26** compares this reaction with the deamination of 5-methylcytosine where deamination generates thymine. The effect is to generate the mismatched base pairs G-U and G-T, respectively.

FIGURE 1.27 shows that the consequences of deamination are different for 5-methylcytosine and cytosine. Deaminating the (rare) 5-methylcytosine causes a mutation, whereas deaminating cytosine does not have this effect. This happens because the DNA repair systems are much more effective in recognizing G-U than G-T and always correct the U (which normally should be present only in RNA, not in DNA).

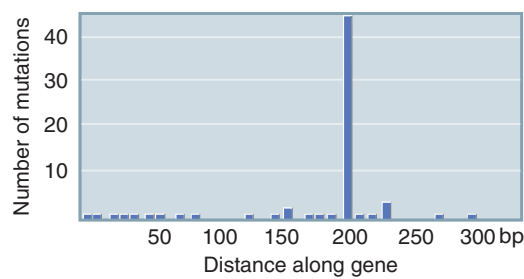
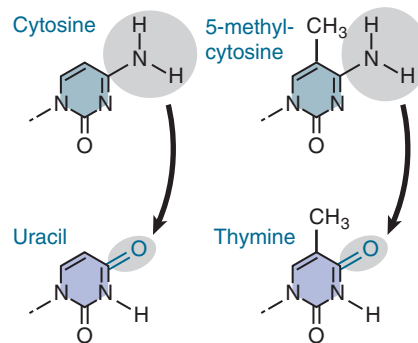


FIGURE 1.25 Spontaneous mutations occur throughout the *lacI* gene of *E. coli* but are concentrated at a hotspot.



► **hotspots** A site in the genome at which the frequency of mutation (or recombination) is very much increased, usually by at least an order of magnitude relative to neighboring sites.

FIGURE 1.26 Deamination of cytosine produces uracil, whereas deamination of 5-methylcytosine produces thymine.

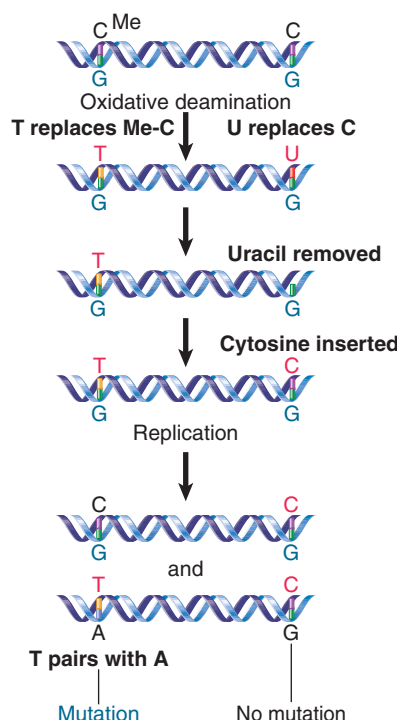


FIGURE 1.27 The deamination of 5-methylcytosine produces thymine (by C-G to T-A transitions), whereas the deamination of cytosine produces uracil (which usually is removed and then replaced by cytosine).

E. coli contain an enzyme, uracil-DNA-glycosidase, that removes uracil residues from DNA (see Section 16.4, *Base Excision Repair Systems Require Glycosylases*). This action leaves an unpaired G residue, and a repair system then inserts a complementary C base. The net result of these reactions is to restore the original sequence of the DNA. Thus, this system protects DNA against the consequences of spontaneous deamination of cytosine. (This system is not, however, efficient enough to prevent the effects of the increased deamination caused by nitrous acid; see Figure 1.22.)

Note that the deamination of 5-methylcytosine creates thymine and results in a mismatched base pair, G-T. If the mismatch is not corrected before the next replication cycle, a mutation results; the bases in the mispaired G-T separate, and then they pair with the correct complements to produce the original G-C and the mutant A-T.

Deamination of 5-methylcytosine is the most common cause of mismatched G-T pairs in DNA. Repair systems that act on G-T mismatches have a bias toward replacing the T with a C (rather than the alternative of replacing the G with an A), which helps to reduce the rate of mutation (see Section 16.6, *Controlling the Direction of Mismatch Repair*). However, these systems are not as effective as those that remove U from G-U mismatches. As a result, deamination of 5-methylcytosine leads to mutation much more often than does deamination of cytosine.

5-methylcytosine also creates hotspots in eukaryotic DNA. It is common at CpG dinucleotides that are concentrated in regions called *CpG islands* (see Section 29.6, *CpG Islands Are Subject to Methylation*). Although 5-methylcytosine accounts for ~1% of the bases in human DNA, sites containing the modified base account for ~30% of all point mutations.

The importance of repair systems in reducing the rate of mutation is emphasized by the effects of eliminating the mouse enzyme MBD4, a glycosylase that can remove T (or U) from mismatches with G. The result is to increase the mutation rate at CpG sites by a factor of 3. (The reason the effect is not greater is that MBD4 is only one of several systems that act on G-T mismatches; probably the elimination of all the systems would increase the mutation rate much more.)

Another type of hotspot, though not often found in coding regions, is the “slippery sequence”—a homopolymer run, or region where a very short sequence (one or a few nucleotides) is repeated many times in tandem. During replication, a DNA polymerase may skip one repeat or replicate the same repeat twice, leading to a decrease or increase in repeat number.

KEY CONCEPTS

- The frequency of mutation at any particular base pair is statistically equivalent, except for hotspots, where the frequency is increased by at least an order of magnitude.
- A common cause of hotspots is the modified base 5-methylcytosine, which is spontaneously deaminated to thymine.
- A hotspot can result from the high frequency of change in copy number of a short tandemly repeated sequence.

CONCEPT AND REASONING CHECK

Suggest several possible reasons that a particular base pair can be a mutational hotspot.

1.14 Some Hereditary Agents Are Extremely Small

► **viroid** A small infectious nucleic acid that does not have a protein coat.

Viroids (or subviral pathogens) are infectious agents that cause diseases in higher plants. They are very small circular molecules of RNA. Unlike viruses—for which the infectious agent consists of a *virion*, a genome encapsulated in a protein coat—the viroid RNA is itself the infectious agent. The viroid consists solely of the RNA molecule, which is extensively folded by imperfect base pairing, forming a characteristic

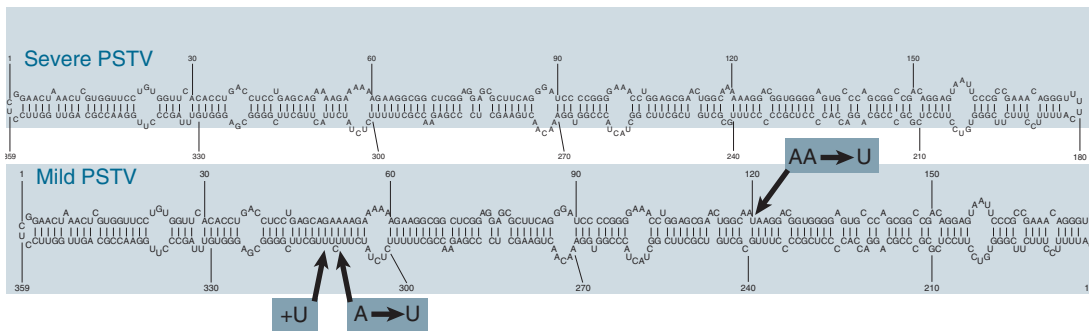


FIGURE 1.28 PSTV RNA is a circular molecule that forms an extensive double-stranded structure, interrupted by many interior loops. The severe and mild forms of PSTV have RNAs that differ at three sites.

rod as shown in **FIGURE 1.28**. Mutations that interfere with the structure of this rod reduce the infectivity of the viroid.

A viroid RNA consists of a single molecule that is replicated autonomously and accurately in infected cells. Viroids are categorized into several groups. A given viroid is assigned to a group according to sequence similarity with other members of the group. For example, four viroids in the PSTV (potato spindle tuber viroid) group have 70%–83% sequence similarity with PSTV. Different isolates of a particular viroid strain vary from one another in sequence, which may result in phenotypic differences among infected cells. For example, the “mild” and “severe” strains of PSTV differ by three nucleotide substitutions.

Viroids are similar to viruses in having heritable nucleic acid genomes but differ from viruses in both structure and function. Viroid RNA does not appear to be translated into polypeptide, so it cannot itself encode the functions needed for its survival. This situation poses two as yet unanswered questions: How does viroid RNA replicate, and how does it affect the phenotype of the infected plant cell?

Replication must be carried out by enzymes of the host cell. The heritability of the viroid sequence indicates that viroid RNA is the template for replication. Thus, the modern version of the central dogma, as presented in **FIGURE 1.29**, includes replication of RNA as a component in some systems.

Viroids are presumably pathogenic because they interfere with normal cellular processes. They might do this in a relatively random way; for example, they may take control of an essential enzyme for their own replication or interfere with the production of necessary cellular RNAs. Alternatively, they might behave as abnormal regulatory molecules, with particular effects upon the expression of individual genes.

An even more unusual agent is the cause of scrapie, a degenerative neurological disease of sheep and goats. The disease is similar to the human diseases of kuru and Creutzfeldt-Jakob syndrome, which affect brain function. The infectious agent of scrapie does not contain nucleic acid. This extraordinary agent is called a **prion** (proteinaceous infectious agent). It is a 28-kD hydrophobic glycoprotein, PrP. PrP is encoded by a cellular gene (conserved among the mammals) that is expressed in normal brain cells. The protein exists in two forms: the version found in normal brain cells is called PrP^c and is entirely degraded by proteases during normal protein turnover. The version found in infected brains is called PrP^{sc} and is extremely resistant to degradation by proteases. PrP^c is converted to PrP^{sc} by a conformational change that confers protease resistance.

As the infectious agent of scrapie, PrP^{sc} must in some way modify

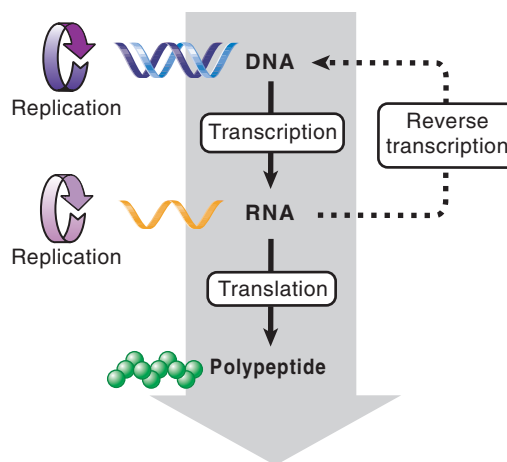


FIGURE 1.29 The central dogma states that information in nucleic acid can be perpetuated or transferred, but the transfer of information into a polypeptide (protein) is irreversible.

the synthesis of its normal cellular counterpart so that it becomes infectious instead of harmless (see *Section 29.9, Prions Cause Diseases in Mammals*). Mice that lack a PrP gene cannot develop scrapie, which demonstrates that PrP is essential for development of the disease.

KEY CONCEPT

- Some very small hereditary agents do not encode polypeptide but consist of RNA or protein with heritable properties.

CONCEPT AND REASONING CHECK

How would you distinguish whether a newly discovered infectious agent is an organism, a virus, a viroid, or a prion?

1.15 Summary

Two classic experiments provided strong evidence that DNA is the genetic material of bacteria, eukaryotic cells, and many viruses. DNA isolated from one strain of *Pneumococcus* bacteria can confer properties of that strain upon another strain. In addition, DNA is the only component that is inherited by progeny phages from parental phages. DNA can be used to transfect new properties into eukaryotic cells.

DNA is a double helix consisting of antiparallel strands in which the nucleotide units are linked by 5' to 3' phosphodiester bonds. The backbone is on the exterior; purine and pyrimidine bases are stacked in the interior in pairs in which A is complementary to T and G is complementary to C. In semiconservative replication, the two strands separate and daughter strands are assembled by complementary base pairing. Complementary base pairing is also used to transcribe an RNA from one strand of a DNA duplex.

A mutation consists of a change in the sequence of A-T and G-C base pairs in DNA. A mutation in a coding sequence may change the sequence of amino acids in the corresponding polypeptide. A point mutation changes only the amino acid represented by the codon in which the mutation occurs. Point mutations may be reverted by back mutation of the original mutation. Insertions may revert by loss of the inserted material, but deletions cannot revert. Mutations may also be suppressed indirectly when a mutation in a different gene counters the original defect.

The natural incidence of mutations is increased by mutagens. Mutations may be concentrated at hotspots. A type of hotspot responsible for some point mutations is caused by deamination of the modified base 5-methylcytosine. Forward mutations occur at a rate of $\sim 10^{-6}$ per locus per generation; back mutations are rarer.

Although all genetic information in cells is carried by DNA, viruses have genomes of double-stranded or single-stranded DNA or RNA. Viroids are subviral pathogens that consist solely of small molecules of RNA with no protective packaging. The RNA does not encode protein and its mode of perpetuation and of pathogenesis is unknown. Scrapie results from a proteinaceous infectious agent, or prion.

CHAPTER QUESTIONS

1. Which strain of *S. pneumoniae* was found to be avirulent (nonlethal) when mice were infected with it?
 - A. the smooth strain
 - B. the rough strain
 - C. both strains
 - D. none were virulent

2. What isotope can be used to specifically label protein?
 - A. ^{14}C
 - B. ^3H
 - C. ^{32}P
 - D. ^{35}S
3. The difference between RNA and DNA is the presence of a:
 - A. 2'- PO_4 group on the ribose sugar in RNA.
 - B. 3'- PO_4 group on the ribose sugar in RNA.
 - C. 2'-OH group on the ribose sugar in RNA.
 - D. 3'-OH group on the ribose sugar in RNA.
4. Which pair of scientists determined that DNA replication is semiconservative?
 - A. Meselson and Stahl
 - B. Watson and Crick
 - C. Okazaki and Okazaki
 - D. Griffith and Avery
5. In the density labeling experiment to study DNA replication, parental DNA in the cell was labeled with a high density isotope and, after one or more generations of growth, subjected to density gradient centrifugation. Which of the following populations was not detected after the second generation?
 - A. the light density population
 - B. the hybrid density population
 - C. the heavy density population
 - D. both the light and heavy density populations
6. Which class of infectious agent is inserted into the host genome as a double-stranded DNA segment?
 - A. retroviruses
 - B. double-stranded RNA viruses
 - C. double-stranded DNA viruses
 - D. viroids
7. The mutation rate in bacteria is about:
 - A. 10^{-6} per locus per generation.
 - B. 10^{-7} per locus per generation.
 - C. 10^{-8} per locus per generation.
 - D. 10^{-9} per locus per generation.
8. About 30% of human point mutations are associated with which of the following modified bases?
 - A. 5-methylguanine
 - B. 5-methyladenine
 - C. 5-methylthymine
 - D. 5-methylcytosine
9. The presence of the modified base in the previous question often leads to:
 - A. transitions.
 - B. transversions.
 - C. deletions.
 - D. insertions.
10. The reversal of an original base pair that was changed from A-T to G-C, then back to A-T, is an example of a:
 - A. true reversion.
 - B. second-site reversion.
 - C. forward mutation.
 - D. suppression.

KEY TERMS

annealing	genome	polynucleotide	spontaneous mutations
antiparallel	hotspots	prion	structural gene
back mutation	hybridization	purine	supercoiling
central dogma	induced mutations	pyrimidine	suppression mutation
chromosome	major groove	renaturation	transfection
complementary	melting temperature	replication fork	transformation
denaturation	minor groove	reverse transcription	transforming principle
DNA polymerase	mutagens	revertants	transition
DNase	nucleoside	RNA polymerase	transversion
endonuclease	nucleotide	RNase	true reversion
exonuclease	overwound	second-site reversion	underwound
forward mutation	point mutation	semiconservative replication	viroid

FURTHER READING

- Holmes, F. (2001). *Meselson, Stahl, and the Replication of DNA: A History of the Most Beautiful Experiment in Biology*. Yale University Press, New Haven, CT. An account of Meselson and Stahl's scientific partnership with a unique look into the daily business of "doing science."
- Maki, H. (2002). Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annu. Rev. Genet.* **36**, 279–303. A review of causes and effects of spontaneous mutations and mutational hotspots.
- Prusiner, S. B. (1998). Prions. *Proc. Natl. Acad. Sci. USA* **95**, 13363–13383. An edited version of Prusiner's Nobel lecture, including an overview of the biology of prions and an account of their discovery.
- Watson, J. D. (1981). *The Double Helix: A Personal Account of the Discovery of the Structure of DNA* (Norton Critical Editions). W. W. Norton, New York, NY. Watson's 1968 best-selling personal account of the discovery of the double helix along with reprints of original publications and additional commentary.