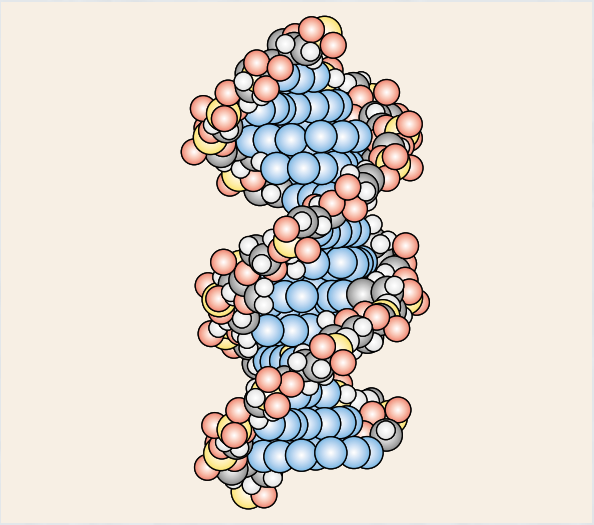
PART IV

Microbial Molecular Biology and Genetics

CHAPTER 11

Genes: Structure, Replication, and Mutation



This model illustrates double-stranded DNA. DNA is the genetic material for procaryotes and eucaryotes. Genetic information is contained in the sequence of base pairs that lie in the center of the helix.

Outline

Concepts

1. The two kinds of nucleic acid, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), differ from one another in chemical composition and structure. In procaryotic and eucaryotic cells, DNA serves as the repository for genetic information.
2. DNA is associated with basic proteins in the cell. In eucaryotes these are special histone proteins, whereas in procaryotes nonhistone proteins are complexed with DNA.
3. The flow of genetic information usually proceeds from DNA through RNA to protein. A protein's amino acid sequence reflects the nucleotide sequence of its mRNA. This messenger is a complementary copy of a portion of the DNA genome.
4. DNA replication is a very complex process involving a variety of proteins and a number of steps. It is designed to operate rapidly while minimizing errors and correcting those that arise when the DNA sequence is copied.
5. Genetic information is contained in the nucleotide sequence of DNA (and sometimes RNA). When a structural gene directs the synthesis of a polypeptide, each amino acid is specified by a triplet codon.
6. A gene is a nucleotide sequence that codes for a polypeptide, tRNA, or rRNA.
7. Most bacterial genes have at least four major parts, each with different functions: promoters, leaders, coding regions, and trailers.
8. Mutations are stable, heritable alterations in the gene sequence and usually, but not always, produce phenotypic changes. Nucleic acids are altered in several different ways, and these mutations may be either spontaneous or induced by chemical mutagens or radiation.
9. It is extremely important to keep the nucleotide sequence constant, and microorganisms have several repair mechanisms designed to detect alterations in the genetic material and restore it to its original state. Often more than one repair system can correct a particular type of mutation. Despite these efforts some alterations remain uncorrected and provide material and opportunity for evolutionary change.

***B**ut the most important qualification of bacteria for genetic studies is their extremely rapid rate of growth. . . . a single* E. coli *cell will grow overnight into a visible colony containing millions of cells, even under relatively poor growth conditions. Thus, genetic experiments on* E. coli *usually last one day, whereas experiments on corn, for example, take months. It is no wonder that we know so much more about the genetics of* E. coli *than about the genetics of corn, even though we have been studying corn much longer.*

—R.F. Weaver and P.W. Hedrick

The preceding chapters have introduced the essentials of microbial metabolism. We now turn to microbial genetics and molecular biology. This chapter reviews some of the most basic concepts of molecular genetics: how genetic information is stored and organized in the DNA molecule, the way in which DNA is replicated, the nature of the genetic code, gene structure, mutagenesis, and DNA repair. In addition, the use of microorganisms to identify potentially dangerous mutagenic agents in the fight against cancer is described. Much of this information will be familiar to those who have taken an introductory genetics course. Because of the importance of procaryotes, primary emphasis is placed on their genetics.

Based on the foundation provided by this chapter, chapter 12 will focus on gene expression and its regulation. Chapter 13 contains information on plasmids and the nature of genetic recombination in microorganisms. These three chapters provide the background needed for understanding the material in Part Five: recombinant DNA technology (chapter 14) and microbial genomics (chapter 15).

Geneticists, including microbial geneticists, use a specialized vocabulary because of the complexities of their discipline. Some knowledge of basic terminology is necessary at the beginning of this survey of general principles. The experimental material of the microbial geneticist is the **clone.** A clone is a population of cells that are derived asexually from a parental cell and are genetically identical. Sometimes a clone is called a pure culture. The term **genome** refers to all the genes present in a cell or virus. Procaryotes normally have one set of genes. That is, they are haploid (1N). Eucaryotic microorganisms usually have two sets of genes, or are diploid (2N). The genotype of an organism is the specific set of genes it possesses. In contrast, the phenotype is the collection of characteristics that are observable by the investigator. All genes are not expressed at the same time, and the environment profoundly influences phenotypic expression. Much genetics research has focused on the relationship between an organism's genotype and phenotype, and gene expression will be the focus of chapter 12.

Although genetic analysis began with the rediscovery of the work of Gregor Mendel in the early part of the twentieth century, subsequent elegant experimentation involving both bacteria and bacteriophages actually elucidated the nature of genetic information, gene structure, the genetic code, and mutations. We will first review a few of these early experiments and then summarize the view of DNA, RNA and protein relationships—sometimes called the Central Dogma—that has guided much of modern research.

11.1 DNA as Genetic Material

The early work of Fred Griffith in 1928 on the transfer of virulence in the pathogen *Streptococcus pneumoniae* (**figure 11.1**) set the stage for the research that first showed that DNA was the genetic material. Griffith found that if he boiled virulent bacteria and injected them into mice, the mice were not affected and no pneumococci could be recovered from the animals. When he injected a combination of killed virulent bacteria and a living nonvirulent strain, the mice died; moreover, he could recover living virulent bacteria from the dead mice. Griffith called this change of nonvirulent bacteria into virulent pathogens **transformation.**

Oswald T. Avery and his colleagues then set out to discover which constituent in the heat-killed virulent pneumococci was responsible for Griffith's transformation. These investigators selectively destroyed constituents in purified extracts of virulent pneumococci, using enzymes that would hydrolyze DNA, RNA, or protein. They then exposed nonvirulent pneumococcal strains to the treated extracts. Transformation of the nonvirulent bacteria was blocked only if the DNA was destroyed, suggesting that DNA was carrying the information required for transformation (**figure 11.2**). The publication of these studies by O. T. Avery, C. M. MacLeod, and M. J. McCarty in 1944 provided the first evidence that Griffith's transforming principle was DNA and therefore that DNA carried genetic information.

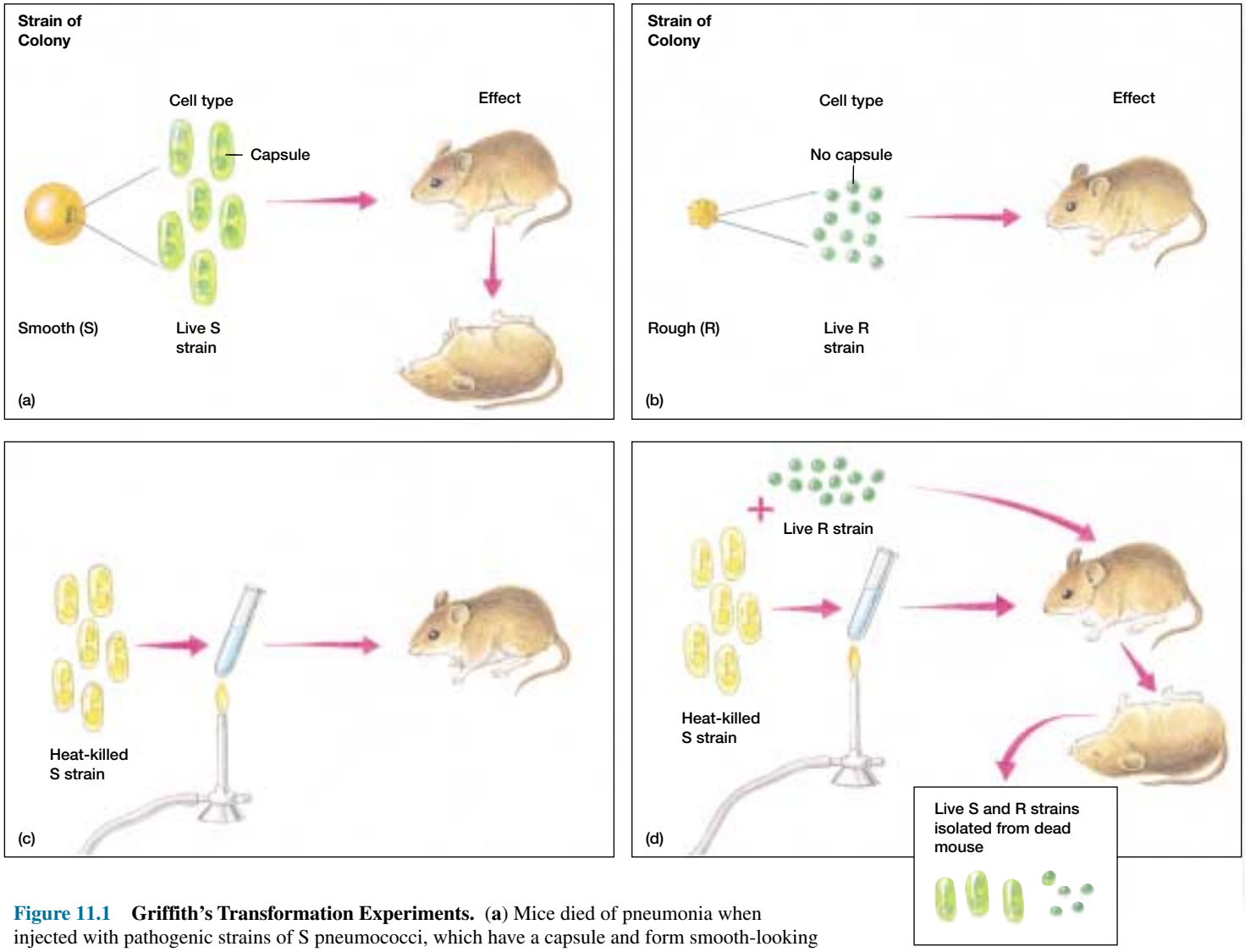


Figure 11.1 **Griffith's Transformation Experiments.** (**a**) Mice died of pneumonia when injected with pathogenic strains of S pneumococci, which have a capsule and form smooth-looking colonies. (**b**) Mice survived when injected with a nonpathogenic strain of R pneumococci, which lacks a capsule and forms rough colonies. (**c**) Injection with heat-killed strains of S pneumococci had no effect. (**d**) Injection with a live R strain and a heat-killed S strain gave the mice pneumonia, and live S strain pneumococci could be isolated from the dead mice.

R cells + purified S cell polysaccharide	→	R colonies
R cells + purified S cell protein	→	R colonies
R cells + purified S cell RNA	→	R colonies
R cells + purified S cell DNA	→	S colonies
S cell extract + protease + R cells	→	S colonies
S cell extract + RNase + R cells	→	S colonies

Figure 11.2 **Experiments on the Transforming Principle.** Summary of the experiments of Avery, MacLeod, and McCarty on the transforming principle. DNA alone changed R to S cells, and this effect was lost when the extract was treated with deoxyribonuclease. Thus DNA carried the genetic information required for the R to S conversion or transformation.

Some years later (1952), Alfred D. Hershey and Martha Chase performed several experiments that indicated that DNA was the genetic material in the T2 bacteriophage. Some luck was involved in their discovery, for the genetic material of many viruses is RNA and the researchers happened to select a DNA virus for their studies. Imagine the confusion if T2 had been an RNA virus! The controversy surrounding the nature of genetic information might have lasted considerably longer than it did. Hershey and Chase made the virus DNA radioactive with ^{32}P or labeled the viral protein coat with ^{35}S. They mixed radioactive bacteriophage with *E. coli* and incubated the mixture for a few minutes. The suspension was then agitated violently in a Waring blender to shear off any adsorbed bacteriophage particles

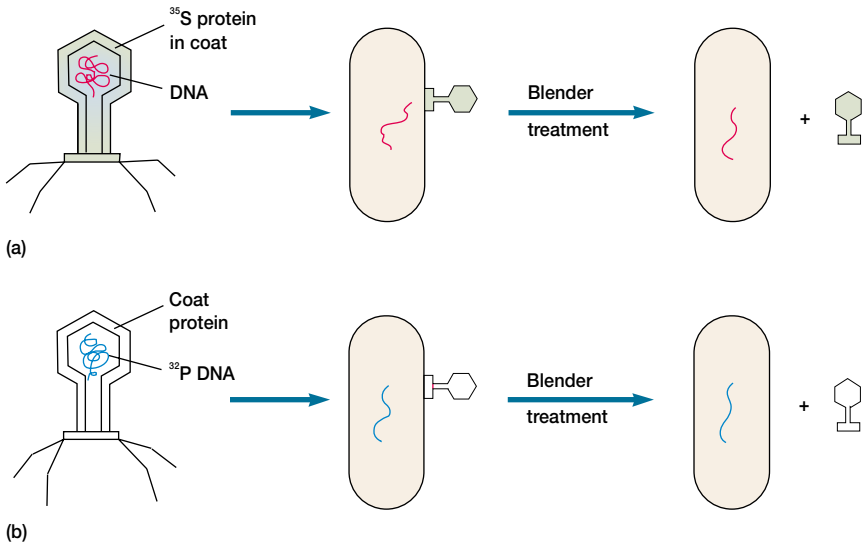


Figure 11.3 The Hershey-Chase Experiment. (**a**) When *E. coli* was infected with a T2 phage containing ^{35}S protein, most of the radioactivity remained outside the host cell. (**b**) When a T2 phage containing ^{32}P DNA was mixed with the host bacterium, the radioactive DNA was injected into the cell and phages were produced. Thus DNA was carrying the virus's genetic information.

(**figure 11.3**). After centrifugation, radioactivity in the supernatant and the bacterial pellet was determined. They found that most radioactive protein was released into the supernatant, whereas ^{32}P DNA remained within the bacteria. Since genetic material was injected and T2 progeny were produced, DNA must have been carrying the genetic information for T2. The biology of bacteriophages (chapter 17)

Subsequent studies on the genetics of viruses and bacteria were largely responsible for the rapid development of molecular genetics. Furthermore, much of the new recombinant DNA technology (*see chapter 14*) has arisen from recent progress in bacterial and viral genetics. Research in microbial genetics has had a profound impact on biology as a science and on the technology that affects everyday life.

Biologists have long recognized a relationship between DNA, RNA, and protein (**figure 11.4**), and this recognition has guided a vast amount of research over the past decades. DNA is precisely copied during its synthesis or **replication.** The expression of the information encoded in the base sequence of DNA begins with the synthesis of an RNA copy of the DNA sequence making up a gene. A gene is a DNA segment or sequence that codes for a polypeptide, an rRNA, or a tRNA. Although DNA has two complementary strands, only the template strand is copied at any particular point on DNA. If both strands of DNA were transcribed, two different mRNAs would result and cause genetic confusion. Thus the sequence corresponding to a gene is located only on one of the two complementary DNA strands. Different genes may be encoded on opposite strands. This process of DNA-directed RNA synthesis is called **transcription** because the DNA base sequence is being written into an RNA base sequence. The RNA that carries information from DNA and directs protein synthesis is **messenger RNA (mRNA).** The last phase of gene expression is **translation** or protein synthesis. The genetic information in the form of an mRNA nucleotide sequence is translated and governs the synthesis of protein. Thus the amino acid sequence of a protein is a direct reflection of the base sequence in mRNA. In turn the mRNA nucleotide sequence is a complementary copy of a portion of the DNA genome.

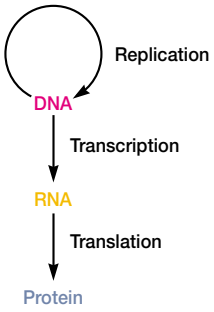


Figure 11.4 Relationships between DNA, RNA, and Protein Synthesis. This conceptual framework is sometimes called the Central Dogma.

1. Define clone, genome, genotype, and phenotype.
2. Briefly summarize the experiments of Griffith; Avery, MacLeod, and McCarty; and Hershey and Chase. What did each show, and why were these experiments important to the development of microbial genetics?
3. Describe the general relationship between DNA, RNA, and protein.

11.2 Nucleic Acid Structure

The structure and synthesis of purine and pyrimidine nucleotides are introduced in chapter 10. These nucleotides can be combined to form nucleic acids of two kinds (**figure 11.5*a***). **Deoxyribonucleic acid (DNA)** contains the 2′-deoxyribonucleosides (figure 11.5*b*) of adenine, guanine, cytosine, and thymine. **Ribonucleic acid (RNA)** is composed of the ribonucleosides of adenine, guanine, cytosine, and uracil (instead of thymine). In both DNA and RNA, nucleosides are joined by phosphate groups to form long polynucleotide chains (figure 11.5*c*). The differences in chemical composition between the chains reside in their sugar and pyrimi-

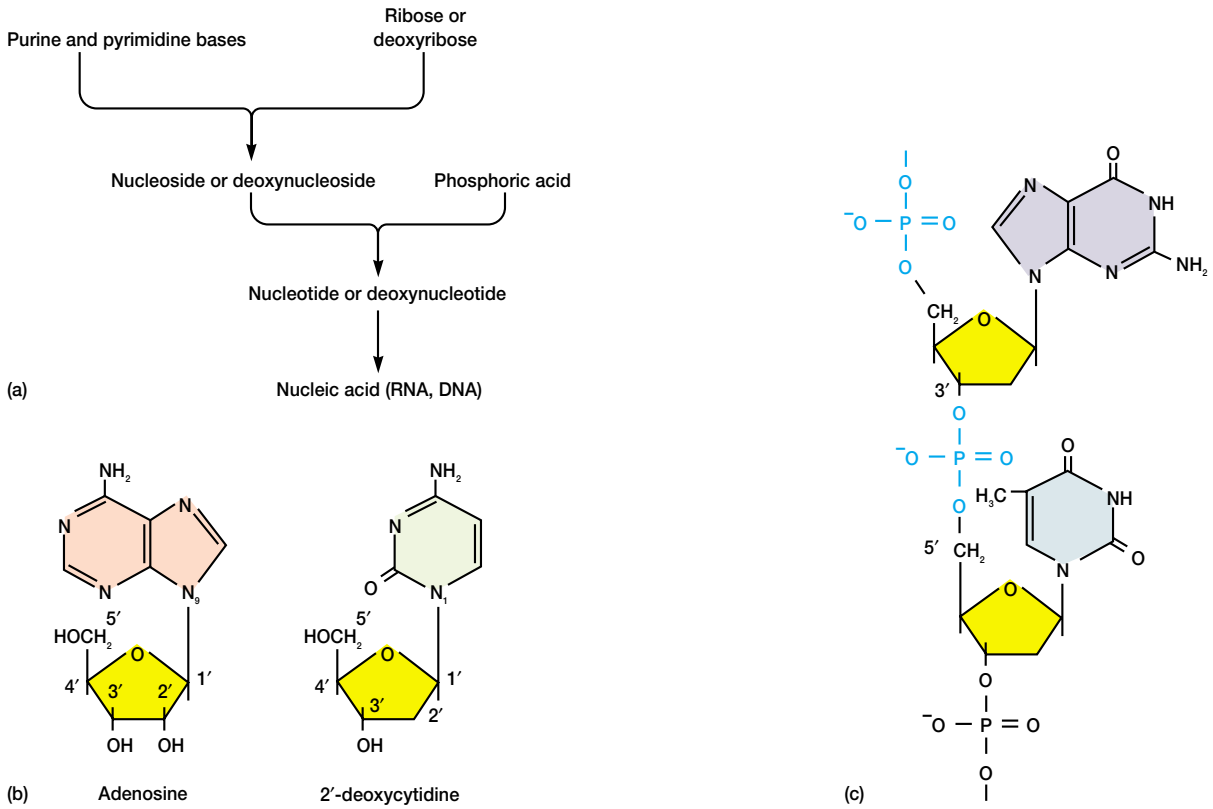


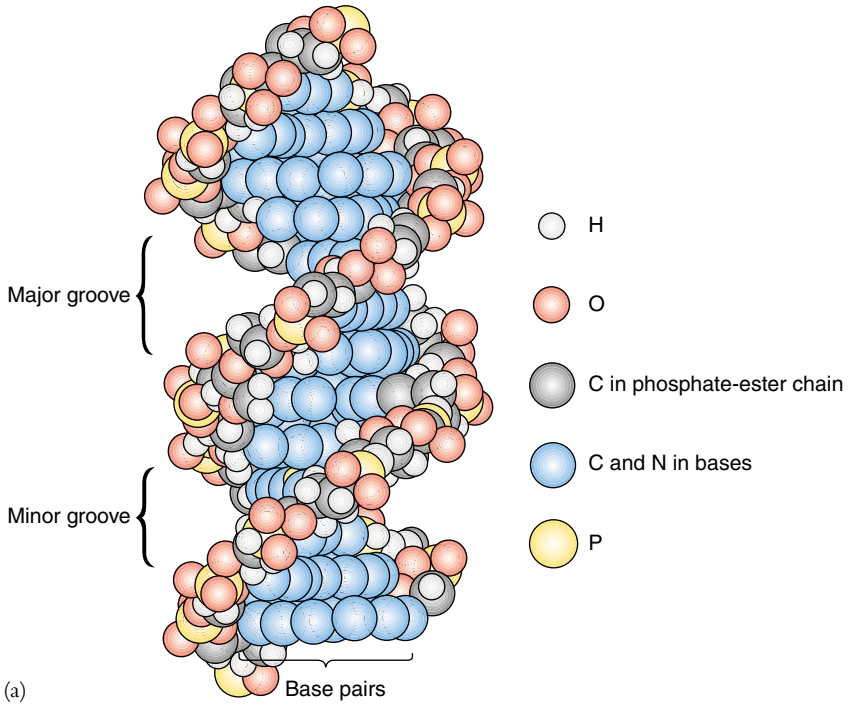
Figure 11.5 **The Composition of Nucleic Acids.** (**a**) A diagram showing the relationships of various nucleic acid components. Combination of a purine or pyrimidine base with ribose or deoxyribose gives a nucleoside (a ribonucleoside or deoxyribonucleoside). A nucleotide contains a nucleoside and one or more phosphoric acid molecules. Nucleic acids result when nucleotides are connected together in polynucleotide chains. (**b**) Examples of nucleosides—the purine nucleoside adenosine and the pyrimidine deoxynucleoside 2′-deoxycytidine. The carbons of nucleoside sugars are indicated by numbers with primes. (**c**) A segment of a polynucleotide chain showing two nucleosides, deoxyguanosine and thymidine, connected by a phosphodiester linkage between the 3′ and 5′-carbons of adjacent deoxyribose sugars.

dine bases: DNA has deoxyribose and thymine; RNA has ribose and uracil in place of thymine.

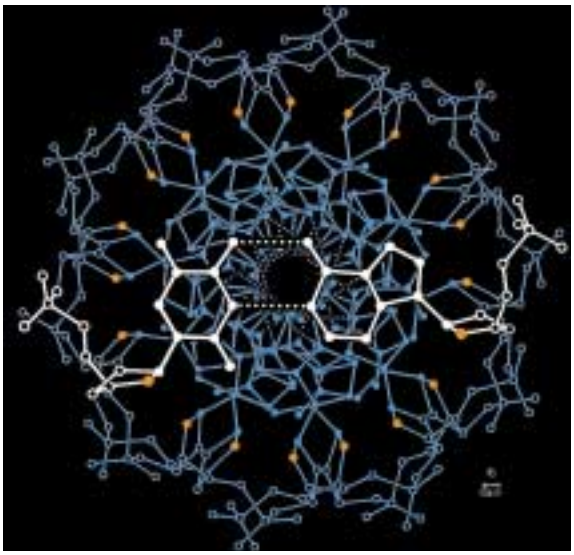
DNA Structure

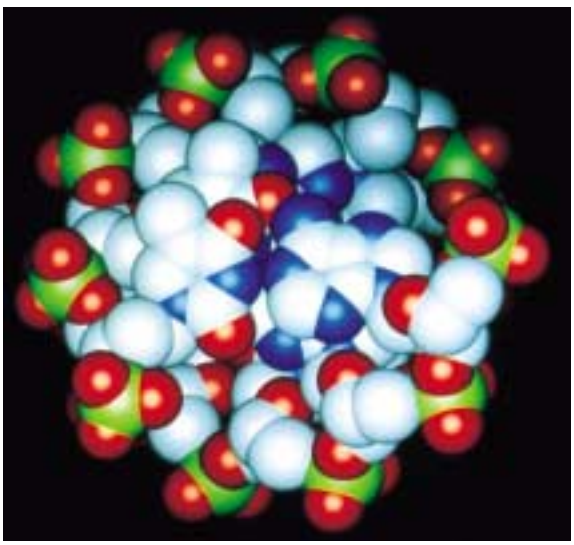
Deoxyribonucleic acids are very large molecules, usually composed of two polynucleotide chains coiled together to form a double helix 2.0 nm in diameter (**figure 11.6**). Each chain contains purine and pyrimidine deoxyribonucleosides joined by phosphodiester bridges (figure 11.5*c*). That is, two adjacent deoxyribose sugars are connected by a phosphoric acid molecule esterified to a 3′-hydroxyl of one sugar and a 5′-hydroxyl of the other. Purine and pyrimidine bases are attached to the 1′-carbon of the deoxyribose sugars and extend toward the middle of the cylinder formed by the two chains. They are stacked on top of each other in the center, one base pair every 0.34 nm. The purine adenine (A) is always paired with the pyrimidine thymine (T) by two hydrogen bonds. The purine guanine (G) pairs with cytosine (C) by three hydrogen bonds (**figure 11.7**). This AT and GC base pairing means that the two strands in a DNA double helix are **complementary.** That is, the bases in one strand match up with those of the other according to the base pairing rules. Because the sequences of bases in these strands encode genetic information, considerable effort has been devoted to determining the base sequences of DNA and RNA from many microorganisms (*see pp. 345–47).* *Nucleic acid sequence comparison and microbial taxonomy (chapter 19)*

The two polynucleotide strands fit together much like the pieces in a jigsaw puzzle because of complementary base pairing (**Box 11.1**). Inspection of figure 11.6*a,b,* depicting the B form of DNA (probably the most common form in cells), shows that the two strands are not positioned directly opposite one another in the helical cylinder. Therefore, when the strands twist about one another, a wide **major groove** and narrower **minor groove** are formed by the backbone. Each base pair rotates 36° around the cylinder with respect to adjacent pairs so that there are 10 base pairs per turn of the helical spiral. Each turn of the helix has a vertical length of 3.4 nm. The helix is right-handed—that is, the chains turn counterclockwise as they approach a viewer looking down the longitudinal axis. The two backbones are antiparallel or run in opposite directions with respect to the orientation of their

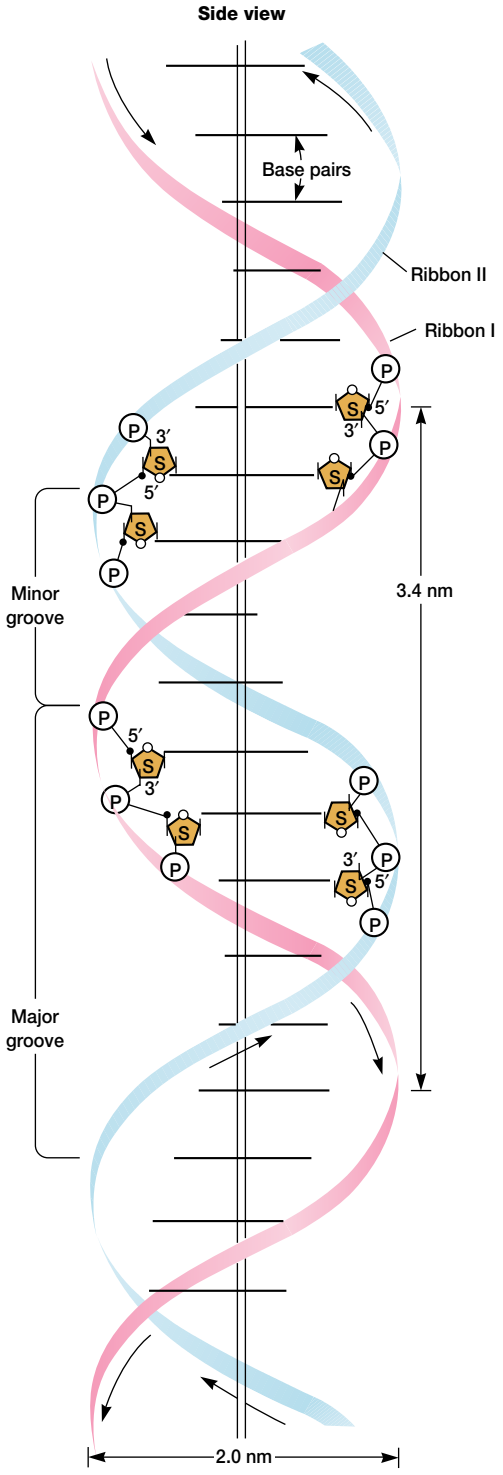


(a)





(c)



(b)

Figure 11.6 **The Structure of the DNA Double Helix.** (**a**) A space-filling model of the B form of DNA with the base pairs, major groove, and minor groove shown. The backbone phosphate groups, shown in color, spiral around the outside of the helix. (**b**) A diagrammatic representation of the double helix. The backbone consists of deoxyribose sugars (S) joined by phosphates (P) in phosphodiester bridges. The arrows at the top and bottom of the chains point in the 5′ to 3′ direction. The ribbons represent the sugar phosphate backbones. (**c**) An end view of the double helix showing the outer backbone and the bases stacked in the center of the cylinder. In the top drawing the ribose ring oxygens are red. The nearest base pair, an AT base pair, is highlighted in white.

Box 11.1

The Elucidation of DNA Structure

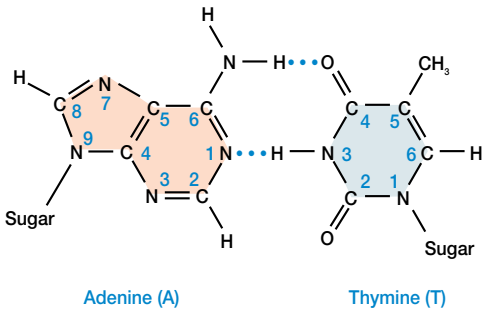
The basic chemical composition of nucleic acids was elucidated in the 1920s through the efforts of P. A. Levene. Despite his major contributions to nucleic acid chemistry, Levene mistakenly believed that DNA was a very small molecule, probably only four nucleotides long, composed of equal amounts of the four different nucleotides arranged in a fixed sequence. Partly because of his influence, biologists believed for many years that nucleic acids were too simple in structure to carry complex genetic information. They concluded that genetic information must be encoded in proteins because proteins were large molecules with complex amino sequences that could vary among different proteins.

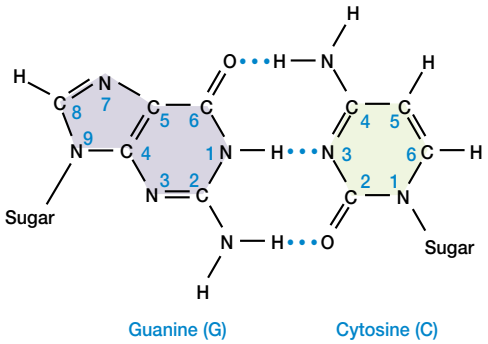
As so often happens, further advances in our understanding of DNA structure awaited the development of significant new analytical techniques in chemistry. One development was the invention of paper chromatography by Archer Martin and Richard Synge between 1941 and 1944. By 1948 the chemist Erwin Chargaff had begun using paper chromatography to analyze the base composition of DNA from a number of species. He soon found that the base composition of DNA from genetic material did indeed vary among species just as he expected. Furthermore, the total amount of purines always equaled the total amount of pyrimidines; and the adenine/thymine and guanine/cytosine ratios were always 1. These findings, known as Chargaff's rules, were a key to the understanding of DNA structure.

Another turning point in research on DNA structure was reached in 1951 when Rosalind Franklin arrived at King's College, London, and joined Maurice Wilkins in his efforts to prepare highly oriented DNA fibers and study them by X-ray crystallography. By the winter of 1952–1953, Franklin had obtained an excellent X-ray diffraction photograph of DNA.

The same year that Franklin began work at King's College, the American biologist James Watson went to Cambridge University and met Francis Crick. Although Crick was a physicist, he was very interested in the structure and function of DNA, and the two soon began to work on its structure. Their attempts were unsuccessful until Franklin's data provided them with the necessary clues. Her photograph of fibrous DNA contained a crossing pattern of dark spots, which showed that the molecule was helical. The dark regions at the top and bottom of the photograph showed that the purine and pyrimidine bases were stacked on top of each other and separated by 0.34 nm. Franklin had already concluded that the phosphate groups lay to the outside of the cylinder. Finally, the X-ray data and her determination of the density of DNA indicated that the helix contained two strands, not three or more as some had proposed.

Without actually doing any experiments themselves, Watson and Crick constructed their model by combining Chargaff's rules on base composition with Franklin's X-ray data and their predictions about how genetic material should behave. By building models, they found that a smooth, two-stranded helix of constant diameter could be constructed only when an adenine hydrogen bonded with thymine and when a guanine bonded with cytosine in the center of the helix. They immediately realized that the double helical structure provided a mechanism by which genetic material might be replicated. The two parental strands could unwind and direct the synthesis of complementary strands, thus forming two new identical DNA molecules (figure 11.10). Watson, Crick, and Wilkins received the Nobel Prize in 1962 for their discoveries. Franklin could not be considered for the prize because she had died of cancer in 1958 at the age of thirty-seven.





sugars. One end of each strand has an exposed 5′-hydroxyl group, often with phosphates attached, whereas the other end has a free 3′-hydroxyl group. If the end of a double helix is examined, the 5′ end of one strand and the 3′ end of the other are visible. In a given direction one strand is oriented 5′ to 3′ and the other, 3′ to 5′ (figure 11.6*b*).

RNA Structure

Besides differing chemically from DNA, ribonucleic acid is usually single stranded rather than double stranded like most DNA. An RNA strand can coil back on itself to form a hairpin-shaped structure with complementary base pairing and helical organization. Cells contain three different types of RNA—messenger RNA, ribosomal RNA, and transfer RNA—that differ from one another in function, site of synthesis in eucaryotic cells, and structure.

Figure 11.7 DNA Base Pairs. DNA complementary base pairing showing the hydrogen bonds (. . .).

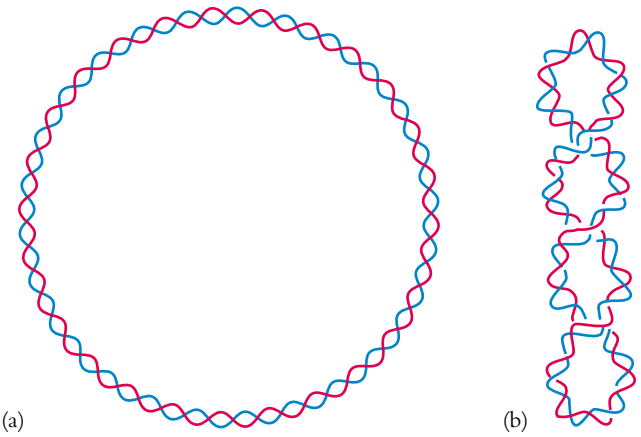
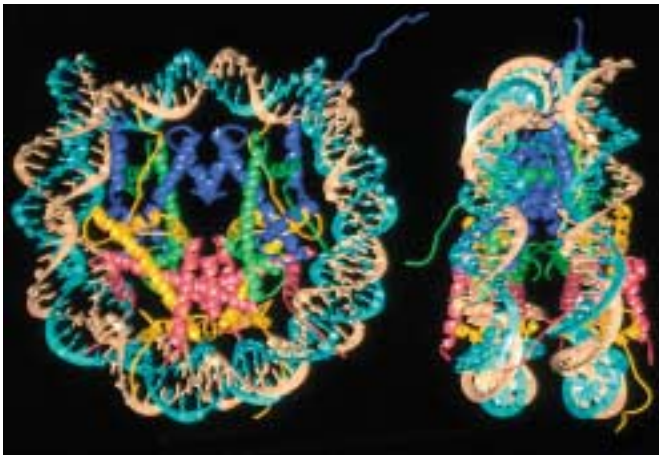


Figure 11.8 **DNA Forms.** (**a**) The DNA double helix of almost all bacteria is in the shape of a closed circle. (**b**) The circular DNA strands, already coiled in a double helix, are twisted a second time to produce supercoils.

The Organization of DNA in Cells

Although DNA exists as a double helix in both procaryotic and eucaryotic cells, its organization differs in the two cell types (*see table 4.2*). DNA is organized in the form of a closed circle in almost all procaryotes (the chromosome of *Borrelia* is a linear DNA molecule). This circular double helix is further twisted into supercoiled DNA (**figure 11.8**) and is associated with basic proteins but not with the histones found complexed with almost all eucaryotic DNA. These histonelike proteins do appear to help organize bacterial DNA into a coiled chromatinlike structure. The structure of the bacterial nucleoid (p. 54)

DNA is much more highly organized in eucaryotic chromatin (*see section 4.9*) and is associated with a variety of proteins, the most prominent of which are **histones.** These are small, basic proteins rich in the amino acids lysine and/or arginine. There are five types of histones in almost all eucaryotic cells studied: H1, H2A, H2B, H3, and H4. Eight histone molecules (two each of H2A, H2B, H3, and H4) form an ellipsoid about 11 nm long and 6.5 to 7 nm in diameter (**figure 11.9*a***). DNA coils around the surface of the ellipsoid approximately $1\frac{3}{4}$ turns or 166 base pairs before proceeding on to the next. This complex of histones plus DNA is called a



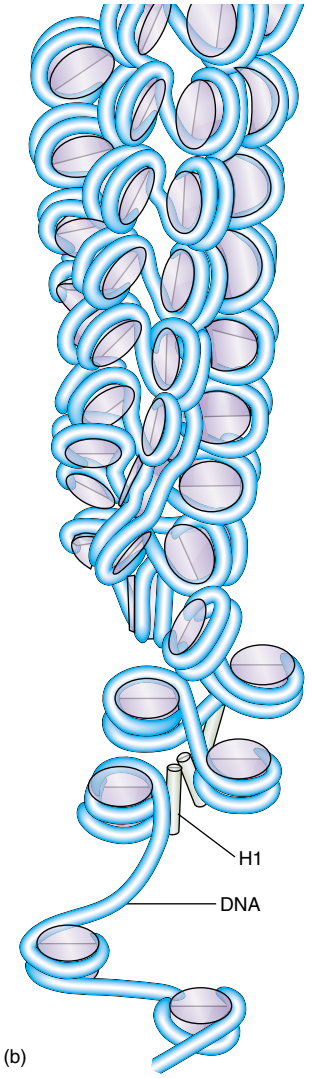


Figure 11.9 **Nucleosome Internal Organization and Function.** (**a**) The nucleosome core particle is a histone octamer surrounded by the 146 base pair DNA helix (brown and turquoise). The octamer is a disk-shaped structure composed of two H2A-H2B dimers and two H3-H4 dimers. The eight histone proteins are colored differently: blue, H3; green, H4; yellow, H2A; and red, H2B. Histone proteins interact with the backbone of the DNA minor groove. The DNA double helix circles the histone octamer in a left-handed helical path. (**b**) An illustration of how a string of nucleosomes, each associated with a histone H1, might be organized to form a highly supercoiled chromatin fiber. The nucleosomes are drawn as cylinders.

nucleosome. Thus DNA gently isolated from chromatin looks like a string of beads. The stretch of DNA between the beads or nucleosomes, the linker region, varies in length from 14 to over 100 base pairs. Histone H1 appears to associate with the linker regions to aid the folding of DNA into more complex chromatin structures (figure 11.9*b*). When folding reaches a maximum, the chromatin takes the shape of the visible chromosomes seen in eucaryotic cells during mitosis and meiosis (*see figure 4.20*).

1. What are nucleic acids? How do DNA and RNA differ in structure?
2. Describe in some detail the structure of the DNA double helix. What does it mean to say that the two strands are complementary and antiparallel?
3. What are histones and nucleosomes? Describe the way in which DNA is organized in the chromosomes of procaryotes and eucaryotes.

11.3 DNA Replication

The replication of DNA is an extraordinarily important and complex process, one upon which all life depends. We shall first discuss the overall pattern of DNA synthesis and then examine the mechanism of DNA replication in greater depth.

Patterns of DNA Synthesis

Watson and Crick published their description of DNA structure in April 1953. Almost exactly one month later, a second paper appeared in which they suggested how DNA might be replicated. They hypothesized that the two strands of the double helix unwind from one another and separate (**figure 11.10**). Free nucleotides now line up along the two parental strands through complementary base pairing—A with T, G with C (figure 11.7). When these nucleotides are linked together by one or more enzymes, two replicas result, each containing a parental DNA strand and a newly formed strand. Research in subsequent years has proved Watson and Crick's hypothesis correct.

Replication patterns are somewhat different in procaryotes and eucaryotes. For example, when the circular DNA chromosome of *E. coli* is copied, replication begins at a single point, the origin. Synthesis occurs at the **replication fork,** the place at which the DNA helix is unwound and individual strands are replicated. Two replication forks move outward from the origin until they have copied the whole **replicon,** that portion of the genome that contains an origin and is replicated as a unit. When the replication forks move around the circle, a structure shaped like the Greek letter theta (θ) is formed (**figure 11.11**). Finally, since the bacterial chromosome is a single replicon, the forks meet on the other side and two separate chromosomes are released.

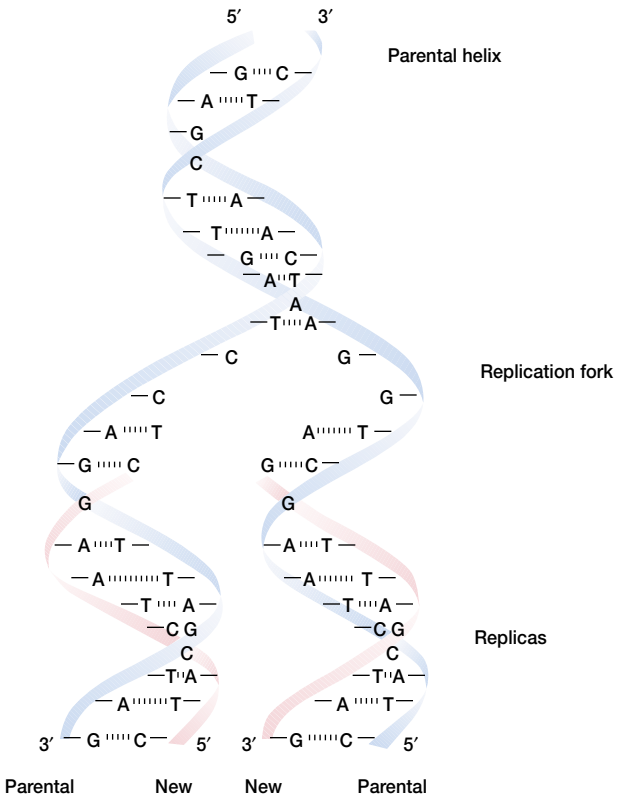


Figure 11.10 Semiconservative DNA Replication. The replication fork of DNA showing the synthesis of two progeny strands. Newly synthesized strands are in maroon. Each copy contains one new and one old strand. This process is called semiconservative replication.

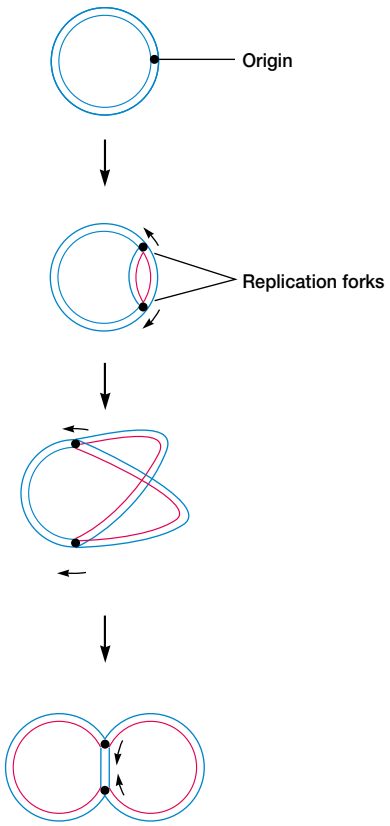


Figure 11.11 Bidirectional Replication. The replication of a circular bacterial genome. Two replication forks move around the DNA forming theta-shaped intermediates. Newly replicated DNA double helix is in red.

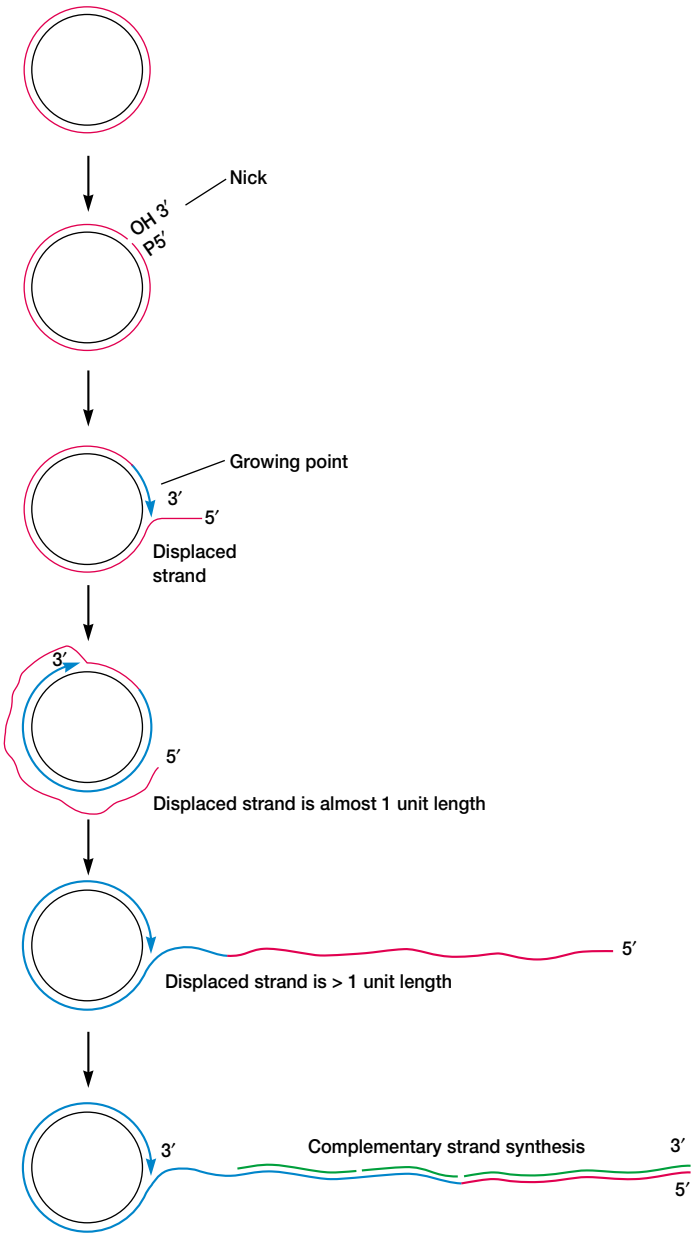


Figure 11.12 The Rolling-Circle Pattern of Replication. A single-stranded tail, often composed of more than one genome copy, is generated and can be converted to the double-stranded form by synthesis of a complementary strand. The "free end" of the rolling-circle strand is probably bound to the primosome.

A different pattern of DNA replication occurs during *E. coli* conjugation (*see section 13.4*) and the reproduction of viruses, such as phage lambda (*see section 17.5*). In the **rolling-circle mechanism** (**figure 11.12**), one strand is nicked and the free 3′-hydroxyl end is extended by replication enzymes. As the 3′ end is lengthened while the growing point rolls around the circular template, the 5′ end of the strand is displaced and forms an ever-lengthening tail. The single-stranded tail may be converted to the double-stranded form by complementary strand synthesis. This

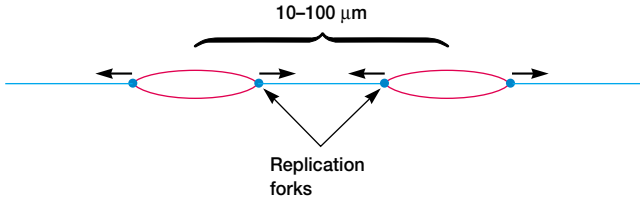


Figure 11.13 The Replication of Eucaryotic DNA. Replication is initiated every 10 to 100 μm and the replication forks travel away from the origin. Newly copied DNA is in red.

mechanism is particularly useful to viruses (*see p. 388*) because it allows the rapid, continuous production of many genome copies from a single initiation event.

Eucaryotic DNA is linear and much longer than procaryotic DNA; *E. coli* DNA is about 1,300 μm in length, whereas the 46 chromosomes in the human nucleus have a total length of 1.8 m (almost 1,400 times longer). Clearly many replication forks must copy eucaryotic DNA simultaneously so that the molecule can be duplicated in a relatively short period, and so many replicons are present that there is an origin about every 10 to 100 μm along the DNA. Replication forks move outward from these sites and eventually meet forks that have been copying the adjacent DNA stretch (**figure 11.13**). In this fashion a large molecule is copied quickly.

Mechanism of DNA Replication

Because DNA replication is so essential to organisms, a great deal of effort has been devoted to understanding its mechanism. The replication of *E. coli* DNA is probably best understood and is the focus of attention in this section. The process in eucaryotic cells is thought to be similar.

DNA replication is initiated at the *oriC* locus. The DnaA protein binds to *oriC* while hydrolyzing ATP. This leads to the initial unwinding of double-stranded DNA at the initiation site. Further unwinding occurs through the activity of the DnaB protein, a helicase (see below).

E. coli has three different **DNA polymerase** enzymes, each of which catalyzes the synthesis of DNA in the 5′ to 3′ direction while reading the DNA template in the 3′ to 5′ direction (**figures 11.14** and **11.15**). The polymerases require deoxyribonucleoside triphosphates (dATP, dGTP, dCTP, and dTTP) as substrates and a DNA template to copy. Nucleotides are added to the 3′ end of the growing chain when the free 3′-hydroxyl group on the deoxyribose attacks the first or alpha phosphate group of the substrate to release pyrophosphate (figure 11.14). DNA polymerase III plays the major role in replication, although it is probably assisted by polymerase I. It is thought that polymerases I and II participate in the repair of damaged DNA (p. 254).

During replication the DNA double helix must be unwound to generate separate single strands. Unwinding occurs very quickly; the fork may rotate as rapidly as 75 to 100 revolutions per second. **Helicases** are responsible for DNA unwinding. These enzymes use energy from ATP to unwind short stretches of helix just ahead of the replication fork. Once the strands have separated, they are kept sin-

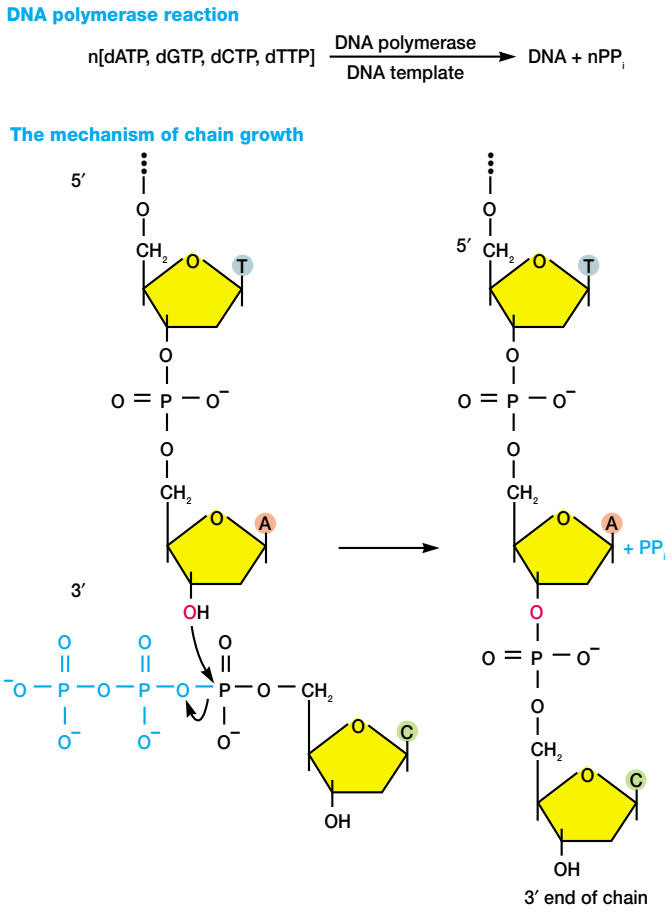


Figure 11.14 The DNA Polymerase Reaction and Its Mechanism. The mechanism involves a nucleophilic attack by the hydroxyl of the 3′ terminal deoxyribose on the alpha phosphate group of the nucleotide substrate (in this example, adenosine attacks cytidine triphosphate).

gle through specific binding with **single-stranded DNA binding proteins (SSBs)** as shown in figure 11.15. Rapid unwinding can lead to tension and formation of supercoils or supertwists in the helix, just as rapid separation of two strands of a rope can lead to knotting or coiling of the rope. The tension generated by unwinding is relieved, and the unwinding process is promoted by enzymes known as **topoisomerases.** These enzymes change the structure of DNA by transiently breaking one or two strands in such a way that it remains unaltered as its shape is changed (e.g., a topoisomerase might tie or untie a knot in a DNA strand). **DNA gyrase** is an *E. coli* topoisomerase that removes the supertwists produced during replication (*see figure 35.6*).

After the double helix has been unwound, successful replication requires the solution of two problems. First, DNA polymerase only synthesizes a new copy of DNA while moving in the 5′ to 3′ direction. Inspection of figure 11.15 shows that synthesis of the leading strand copy is relatively simple because the new strand can be extended continuously at its 3′ end as the DNA unwinds. In contrast, the lagging strand cannot be extended in the same direction because this would require 3′ to 5′ synthesis, which is not possible. As a result, the lagging strand copy is synthesized discontinuously in the 5′ to 3′ direction as a series of fragments; then the fragments are joined to form a complete copy. The second problem arises because DNA polymerase cannot start a new copy from scratch, but must build on an already existing strand. In figure 11.15, the leading strand copy already exists; however, the lagging strand fragments must be synthesized without a DNA strand to build upon. In this case, a special RNA primer is first synthesized and then a DNA copy can be built on the primer.

The details of DNA replication are outlined in a diagram of the replication fork (**figure 11.16**). The replication process takes place in four stages.

1. Helicases unwind the helix with the aid of topoisomerases like the DNA gyrase (figure 11.16, step 1). It appears that the DnaB protein is the helicase most actively involved in

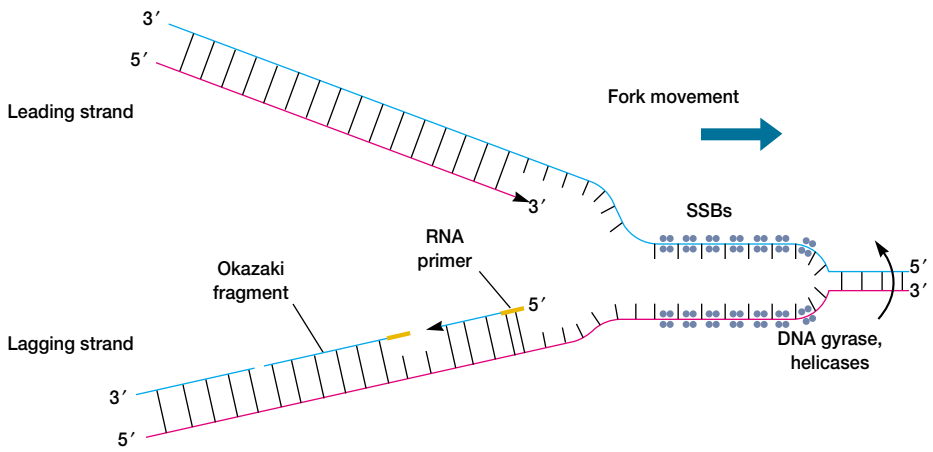


Figure 11.15 Bacterial DNA Replication. A general diagram of the synthesis of DNA in *E. coli* at the replication fork. Bases and base pairs are represented by lines extending outward from the strands. The RNA primer is in gold. See text for details.

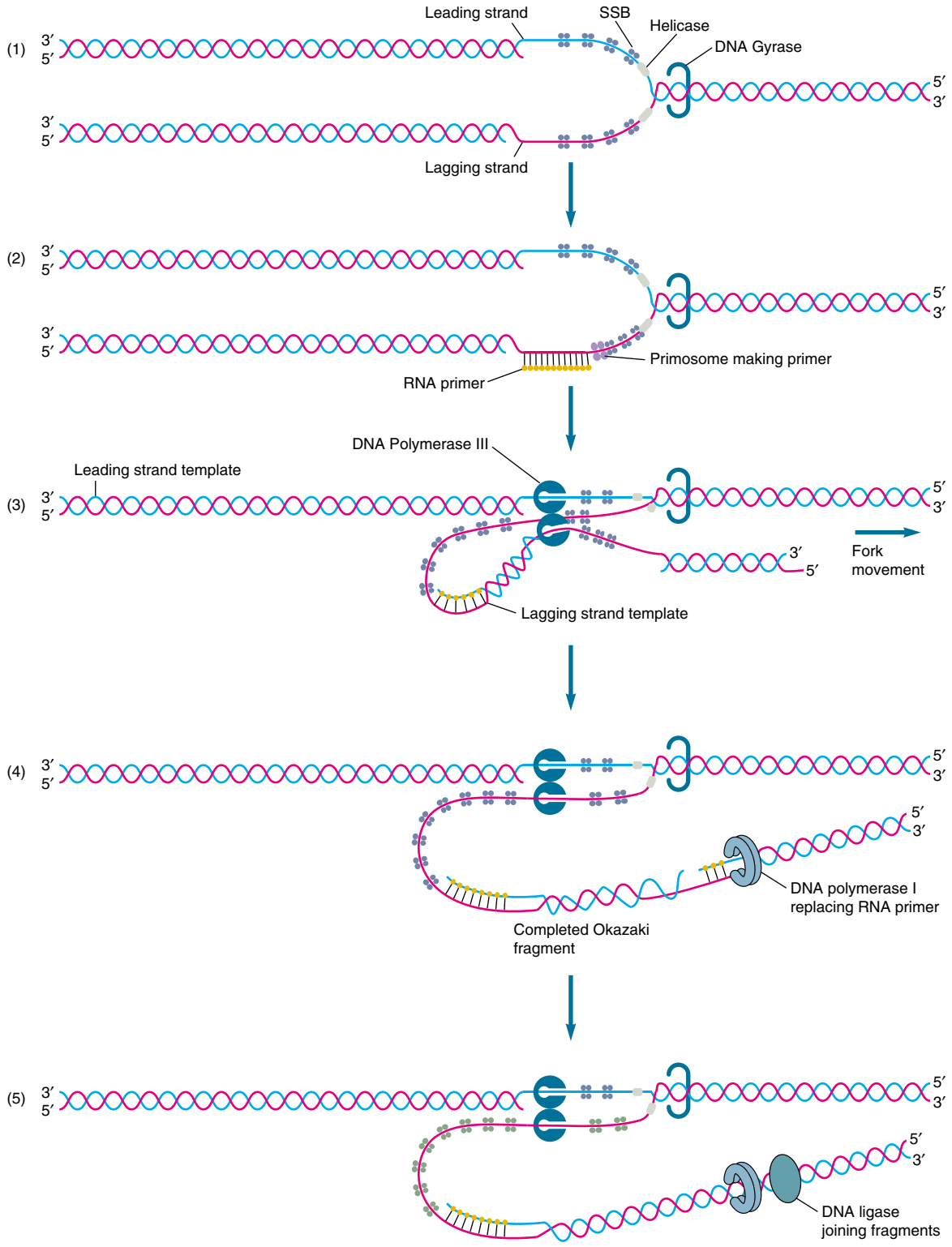


Figure 11.16 A Hypothetical Model for Activity at the Replication Fork. The overall process is pictured in five stages with only one cycle of replication shown for sake of clarity. In practice, all these enzymes are functioning simultaneously and more than one round of replication can occur simultaneously; for example, new primer RNA can be synthesized at the same time as DNA is being replicated. (*1*) DNA gyrase, helicases, and single-stranded DNA binding proteins (SSBs) unwind DNA to produce a single-stranded stretch. (*2*) The primosome synthesizes an RNA primer. (*3*) The replisome has two DNA polymerase III complexes. One polymerase continuously copies the leading strand. The lagging strand loops around the other polymerase so that both strands can be replicated simultaneously. When DNA polymerase III encounters a completed Okazaki fragment, it releases the lagging strand. (*4*) DNA polymerase I removes the RNA primer and fills in the gap with complementary DNA. (*5*) DNA ligase seals the nick and joins the two fragments.

replication, but the n′ protein also may participate in unwinding. The single strands are kept separate by the DNA binding proteins (SSBs).

2. DNA is probably replicated continuously by DNA polymerase III when the leading strand is copied. Lagging strand replication is discontinuous, and the fragments are synthesized in the 5′ to 3′ direction just as in leading strand synthesis. First, a special RNA polymerase called a **primase** synthesizes a short RNA primer, usually around 10 nucleotides long, complementary to the DNA (figure 11.16, step 2). It appears that the primase requires the assistance of several other proteins, and the complex of the primase with its accessory proteins is called the **primosome.** DNA polymerase III holoenzyme then synthesizes complementary DNA beginning at the 3′ end of the RNA primer. Both leading and lagging strand synthesis probably occur concurrently on a single multiprotein complex with two catalytic sites, the replisome. If this is the case, the lagging strand template must be looped around the complex (figure 11.16, step 3). The final fragments are around 1,000 to 2,000 nucleotides long in bacteria and approximately 100 nucleotides long in eucaryotic cells. They are called **Okazaki fragments** after their discoverer, Reiji Okazaki.
3. After most of the lagging strand has been duplicated by the formation of Okazaki fragments, DNA polymerase I or RNase H removes the RNA primer. Polymerase I synthesizes complementary DNA to fill the gap resulting from RNA deletion (figure 11.16, step 4). The polymerase appears to remove one primer nucleotide at a time and replace it with the appropriate complementary deoxyribonucleotide. Polymerase III holoenzyme also may be able to fill in the gap.
4. Finally, the fragments are joined by the enzyme **DNA ligase,** which forms a phosphodiester bond between the 3′-hydroxyl of the growing strand and the 5′-phosphate of an Okazaki fragment (figure 11.16, step 5, and **figure 11.17**). Bacterial ligases use the pyrophosphate bond of NAD+ as an energy source; many other ligases employ ATP.

DNA polymerase III holoenzyme, the enzyme complex that synthesizes most of the DNA copy, is a very large entity containing DNA polymerase III and several other proteins. The γδ complex and β subunits of the holoenzyme bind it to the DNA template and primer. The α subunit carries out the actual polymerization reaction. It appears that most or all of the replication proteins form a huge complex or replication factory, sometimes called a replisome, that is relatively stationary and probably bound to the plasma membrane. The DNA moves through this factory and is copied, emerging as two daughter chromosomes. In slowly growing bacteria there seem to be two factories located at or close to the center of the cell. Rapidly growing cells might have four or more factories.

DNA replication stops when the polymerase complex reaches a termination site on the DNA in *E. coli.* The Tus protein binds to these *Ter* sites and halts replication. In many procaryotes, replication stops randomly when the forks meet.

DNA replication is an extraordinarily complex process. At least 30 proteins are required to replicate the *E. coli* chromosome.

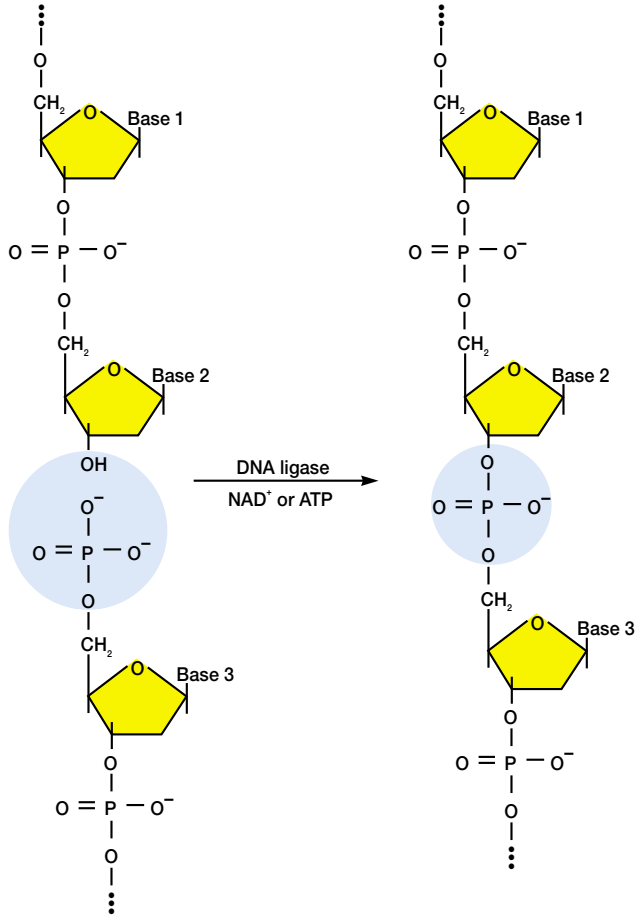


Figure 11.17 **The DNA Ligase Reaction.** The groups being altered are shaded in blue.

Presumably much of the complexity is necessary for accuracy in copying DNA. It would be very dangerous for any organism to make many errors during replication because a large number of mutations would certainly be lethal. In fact, *E. coli* makes errors with a frequency of only 10^{-9} or 10^{-10} per base pair replicated (or about 10^{-6} per gene per generation). Part of this precision results from the low error rate of the copying process itself. However, DNA polymerase III (and DNA polymerase I) also can proofread the newly formed DNA. As polymerase III moves along synthesizing a new DNA strand, it recognizes any errors resulting in improper base pairing and hydrolytically removes the wrong nucleotide through a special 3′ to 5′ exonuclease activity (which is found in the ϵ subunit). The enzyme then backs up and adds the proper nucleotide in its place. Polymerases delete errors by acting much like correcting typewriters. DNA repair (pp. 254–56)

Despite its complexity and accuracy, replication occurs very rapidly. In procaryotes replication rates approach 750 to 1,000 base pairs per second. Eucaryotic replication is much slower, about 50 to 100 base pairs per second. This is not surprising because eucaryotic replication also involves operations like unwinding the DNA from nucleosomes.

1. Define the following terms: replication, transcription, messenger RNA, translation, replicon, replication fork, primosome, and replisome.
2. Be familiar with the nature and functions of the following replication components and intermediates: DNA polymerases I and III, topoisomerase, DNA gyrase, helicase, single-stranded DNA binding protein, Okazaki fragment, DNA ligase, leading strand, and lagging strand.

11.4 The Genetic Code

The realization that DNA is the genetic material triggered efforts to understand how genetic instructions are stored and organized in the DNA molecule. Early studies on the nature of the genetic code showed that the DNA base sequence corresponds to the amino acid sequence of the polypeptide specified by the gene. That is, the nucleotide and amino acid sequences are colinear. It also became evident that many mutations are the result of changes of single amino acids in a polypeptide chain. However, the exact nature of the code was still unclear.

Establishment of the Genetic Code

Since only 20 amino acids normally are present in proteins, there must be at least 20 different code words in a linear, single strand of DNA. The code must be contained in some sequence of the four nucleotides commonly found in the linear DNA sequence. There are only 16 possible combinations (4^2) of the four nucleotides if only nucleotide pairs are considered, not enough to code for all 20 amino acids. Therefore a code word, or **codon,** must involve at least nucleotide triplets even though this would give 64 possible combinations (4^3), many more than the minimum of 20 needed to specify the common amino acids.

The actual codons were discovered in the early 1960s through the experiments carried out by Marshall Nirenberg, Heinrich Matthaei, Philip Leder, and Har Gobind Khorana. In 1968 Nirenberg and Khorana shared the Nobel prize with Robert W. Holley, the first person to sequence a nucleic acid (phenylalanyl-tRNA).

Organization of the Code

The genetic code, presented in RNA form, is summarized in **table 11.1.** Note that there is **code degeneracy.** That is, there are up to six different codons for a given amino acid. Only 61 codons, the **sense**

Table 11.1 The Genetic Code

First Position (5′ End)[a]	*Second Position* U	*Second Position* C	*Second Position* A	*Second Position* G	*Third Position (3′ End)*
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	**U**
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	**C**
	UUA Leu	UCA Ser	UAA STOP	UGA STOP	**A**
	UUG Leu	UCG Ser	UAG STOP	UGG Trp	**G**
C	CUU Leu	CCU Pro	CAU His	CGU Arg	**U**
	CUC Leu	CCC Pro	CAC His	CGC Arg	**C**
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	**A**
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	**G**
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	**U**
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	**C**
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	**A**
	AUG Met	ACG Thr	AAG Lys	AGG Arg	**G**
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	**U**
	GUC Val	GCC Ala	GAC Asp	GGC Gly	**C**
	GUA Val	GCA Ala	GAA Glu	GGA Gly	**A**
	GUG Val	GCG Ala	GAG Glu	GGG Gly	**G**

[a]The code is presented in the RNA form. Codons run in the 5′ to 3′ direction. See text for details.

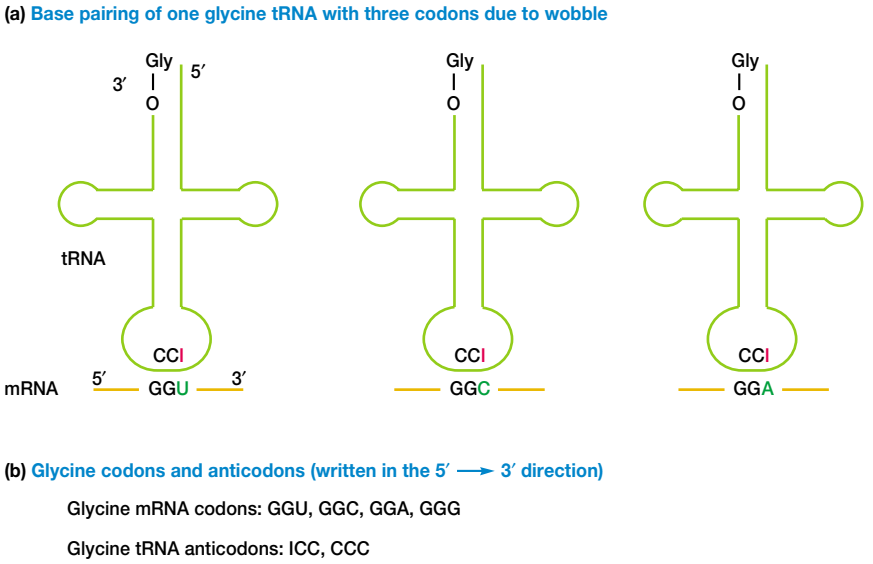


Figure 11.18 Wobble and Coding. The use of wobble in coding for the amino acid glycine. (**a**) Inosine (I) is a wobble nucleoside that can base pair with uracil (U), cytosine (C), or adenine (A). Thus ICC base pairs with GGU, GGC, and GGA in the mRNA. (**b**) Because of the wobble produced by inosine, two tRNA anticodons can recognize the four glycine (Gly) codons. ICC recognizes GGU, GGC, and GGA; CCC recognizes GGG.

codons, direct amino acid incorporation into protein. The remaining three codons (UGA, UAG, and UAA) are involved in the termination of translation and are called **stop** or **nonsense codons.** Despite the existence of 61 sense codons, there are not 61 different tRNAs, one for each codon. The 5′ nucleotide in the anticodon can vary, but generally, if the nucleotides in the second and third anticodon positions complement the first two bases of the mRNA codon, an aminoacyl-tRNA with the proper amino acid will bind to the mRNA-ribosome complex. This pattern is evident on inspection of changes in the amino acid specified with variation in the third position (table 11.1). This somewhat loose base pairing is known as **wobble** and relieves cells of the need to synthesize so many tRNAs (**figure 11.18**). Wobble also decreases the effects of DNA mutations. The mechanism of protein synthesis and tRNA function (pp. 265–71)

1. Why must a codon contain at least three nucleotides?
2. Define the following: code degeneracy, sense codon, stop or nonsense codon, and wobble.

11.5 Gene Structure

The **gene** has been defined in several ways. Initially geneticists considered it to be the entity responsible for conferring traits on the organism and the entity that could undergo recombination. Recombination involves exchange of DNA from one source with that from another (*see section 13.1*) and is responsible for generating much of the genetic variability found in viruses and living organisms. Genes were typically named for some mutant or altered phenotype. With the discovery and characterization of DNA, the gene was defined more precisely as a linear sequence of nucleotides or codons (this term can be used for RNA as well as DNA) with a fixed start point and end point.

At first, it was thought that a gene contained information for the synthesis of one enzyme, the one gene–one enzyme hypothesis. This has been modified to the one gene–one polypeptide hypothesis because of the existence of enzymes and other proteins composed of two or more different polypeptide chains coded for by separate genes. The segment that codes for a single polypeptide is sometimes also called a **cistron.** More recent results show that even this description is oversimplified. Not all genes are involved in protein synthesis; some code instead for rRNA and tRNA. Thus a gene might be defined as a polynucleotide sequence that codes for a polypeptide, tRNA, or rRNA. Some geneticists think of it as a segment of nucleic acid that is transcribed to give an RNA product. Most genes consist of discrete sequences of codons that are "read" only one way to produce a single product. That is, the code is not overlapping and there is a single starting point with one **reading frame** or way in which nucleotides are grouped into codons (**figure 11.19**). Chromosomes therefore usually consist of gene sequences that do not overlap one another (**figure 11.20*a***). However, there are exceptions to the rule. Some viruses such as the phage ϕX174 do have overlapping genes (figure 11.20*b*), and parts of genes overlap in some bacterial genomes.

Procaryotic and viral gene structure differs greatly from that of eucaryotes. In bacterial and viral systems, the coding information within a cistron normally is continuous (some bacterial genes do contain introns); however, in eucaryotic organisms, many genes contain coding information (exons) interrupted periodically by noncoding sequences (introns). An interesting exception to this rule is eucaryotic histone genes, which lack introns. Because procaryotic and viral systems are the best characterized, the more detailed description of gene structure that follows will focus on *E. coli* genes. Exons and introns in eucaryotic genes (p. 263)

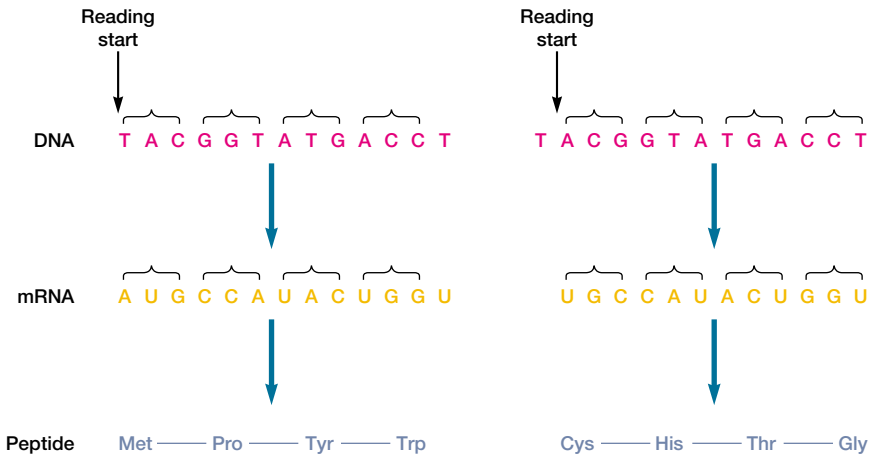
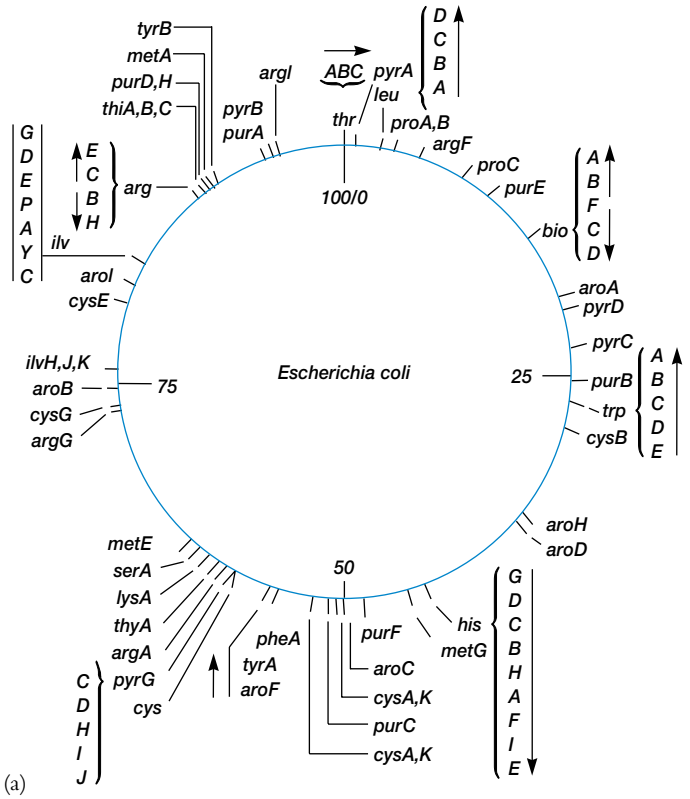


Figure 11.19 Reading Frames and Their Importance. The place at which DNA sequence reading begins determines the way nucleotides are grouped together in clusters of three (outlined with brackets), and this specifies the mRNA codons and the peptide product. In the example, a change in the reading frame by one nucleotide yields a quite different mRNA and final peptide.



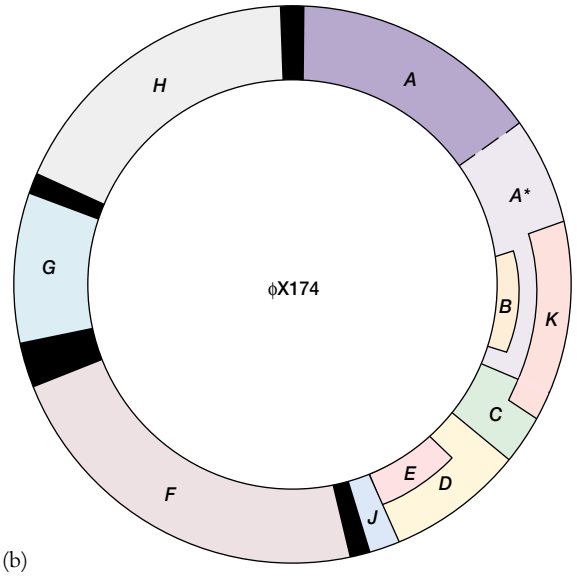


Figure 11.20 Chromosomal Organization in Bacteria and Viruses. (**a**) Simplified genetic map of *E. coli.* The *E. coli* map is divided into 100 minutes. (**b**) The map of phage ϕX174 shows the overlap of gene *B* with *A*, *K* with *A* and *C*, and *E* with *D*. The solid regions are spaces lying between genes. Protein A* consists of the last part of protein A and arises from reinitiation of transcription within gene *A*.

Genes That Code for Proteins

Recall from the discussion of transcription that although DNA is double stranded, only one strand contains coded information and directs RNA synthesis. This strand is called the **template strand,** and the complementing strand is known as the nontemplate strand (**figure 11.21**). Because the mRNA is made from the 5′ to the 3′ end, the polarity of the DNA template strand is therefore 3′ to 5′. Therefore the beginning of the gene is at the 3′ end of the template strand (also the 5′ end of the nontemplate strand). An RNA polymerase recognition/binding and regulatory site known as the **promoter** is located at the start of the gene. The mechanism of transcription (pp. 261–64)

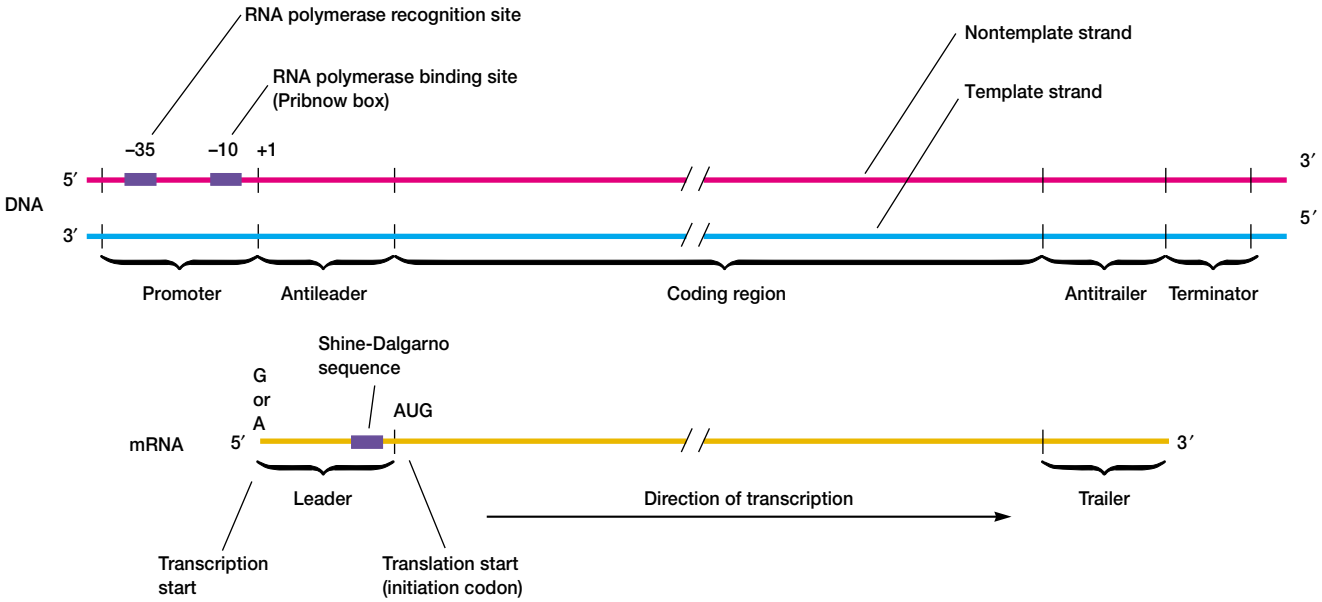


Figure 11.21 **A Bacterial Structural Gene.** The organization of a typical structural gene in bacteria. Leader and trailer sequences are included even though some genes lack one or both. Transcription begins at the +1 position in DNA, and the first nucleotide incorporated into mRNA is usually GTP or ATP. Translation of the mRNA begins with the AUG initiation codon. Regulatory sites are not shown.

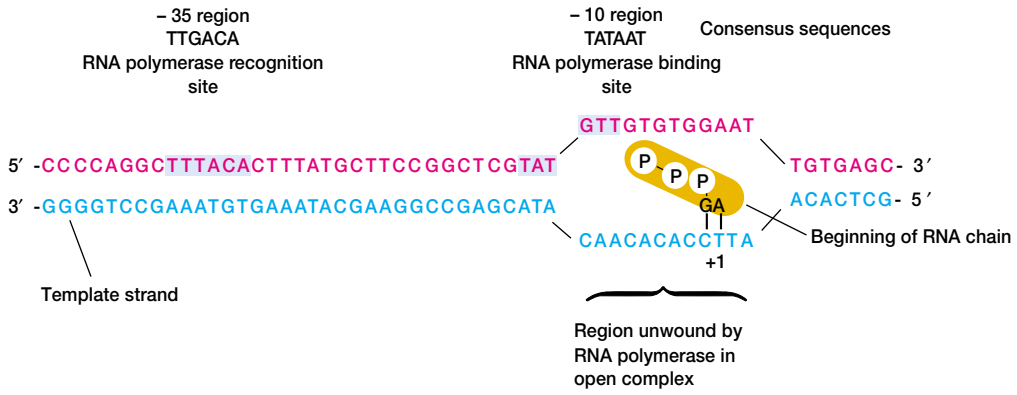


Figure 11.22 **A Bacterial Promoter.** The lactose operon promoter and its consensus sequences. The start point for RNA synthesis is labeled +1. The region around −35 is the site at which the RNA polymerase first attaches to the promoter. RNA polymerase binds and begins to unwind the DNA helix at the Pribnow box or RNA polymerase binding site, which is located in the −10 region.

Promoters are sequences of DNA that are usually upstream from the actual coding or transcribed region (figure 11.21)—that is, the promoter is located before or upstream from the coding region in relationship to the direction of transcription (the direction of transcription is referred to as downstream). Different genes have different promoters, and promoters also vary in sequence between bacteria. In *E. coli* the promoter has two important functions, and these relate to two specific segments within the promoter (figure 11.21 and **figure 11.22**). Although these two segments do vary slightly in sequence between bacterial strains and different genes, they are fairly constant and may be represented by consensus sequences. These are idealized sequences composed of the bases most often found at each position when the sequences from different bacteria are compared. The RNA polymerase recognition site, with a consensus sequence of 5′TTGACA3′ on the nontemplate strand in *E. coli,* is centered about 35 base pairs before (the −35 region) the transcriptional start point (labeled as +1) of RNA synthesis. This sequence seems to be the site of the initial association of the RNA polymerase with the DNA. The **RNA polymerase binding site,** also

known as the **Pribnow box,** is centered at the −10 region and has a consensus sequence 5′TATAAT3′ in *E. coli* (a sequence that favors the localized unwinding of DNA). This is where the RNA polymerase begins to unwind the DNA for eventual transcription. The initially transcribed portion of the gene is not necessarily coding material. Rather, a **leader sequence** may be synthesized first. The leader is usually a nontranslated sequence that is important in the initiation of translation and sometimes is involved in regulation of transcription.

The leader (figure 11.21) in procaryotes generally contains a consensus sequence known as the **Shine-Dalgarno sequence,** 5′AGGA3′, the transcript of which complements a sequence on the 16S rRNA in the small subunit of the ribosome. The binding of mRNA leader with 16S rRNA properly orients the mRNA on the ribosome. The leader also sometimes regulates transcription by attenuation (*see section 12.4*). Downstream and next to the leader is the most important part of the structural gene, the coding region.

The **coding region** (figure 11.21) of genes that direct the synthesis of proteins typically begins with the template DNA sequence 3′TAC5′. This produces the RNA translation initiation codon 5′AUG3′, which codes for *N*-formylmethionine. This modified form of methionine is the first amino acid incorporated in most procaryotic proteins. The remainder of the gene coding region consists of a sequence of codons that specifies the sequence of amino acids for that particular protein. Transcription does not stop at the translation stop codon but rather at a **terminator sequence** (*see section 12.1*). The terminator often lies after a nontranslated **trailer sequence** located downstream from the coding region (figure 11.21). The trailer sequence, like the leader, is needed for the proper expression of the coding region of the gene.

Besides the basic components described above—the promoter, leader, coding region, trailer, and terminator—many procaryotic genes have a variety of regulatory sites. These are locations where DNA-recognizing regulatory proteins bind to stimulate or prevent gene expression. Regulatory sites often are associated with promoter function, and some consider them to be parts of special promoters. Two such sites, the operator and the CAP binding site, are discussed in *section 12.3.* Certainly everything is not known about genes and their structure. With the ready availability of purified cloned genes and DNA sequencing technology, major discoveries continue to be made in this area. The operon and transcription regulation (pp. 275–78)

Genes That Code for tRNA and rRNA

The DNA segments that code for tRNA and rRNA also are considered genes, although they give rise to structurally important RNA rather than protein. In *E. coli* the genes for tRNA are fairly typical, consisting of a promoter and transcribed leader and trailer sequences that are removed during the maturation process (**figure 11.23*a***). The precise function of the leader is not clear; however, the trailer is required for termination. Genes coding for tRNA may code for more than a single tRNA molecule or type of tRNA (figure 11.23*a*). The segments coding for tRNAs are separated by short spacer sequences that are removed after transcription by special ribonucleases, at least one of which contains catalytic RNA. As mentioned in chapter 12 (*see p. 266*), mature tRNAs contain unusual nucleosides. Modified nucleosides such as inosine, ribothymidine, and pseudouridine almost always are formed after the tRNA has been synthesized. Special tRNA-modifying enzymes are responsible. RNA splicing and ribozymes (pp. 264–65)

The genes for rRNA also are similar in organization to genes coding for proteins and have promoters, trailers, and terminators (figure 11.23*b*). Interestingly all the rRNAs are transcribed as a single, large precursor molecule that is cut up by ribonucleases after transcription to yield the final rRNA products. *E. coli* pre-rRNA spacer and trailer regions even contain tRNA genes. Thus the synthesis of tRNA and rRNA involve posttranscriptional modification, a relatively rare process in procaryotes.

1. Define or describe the following: gene, template and nontemplate strands, promoter, consensus sequence, RNA polymerase recognition and binding sites, Pribnow box, leader, Shine-Dalgarno sequence, coding region, reading frame, trailer, and terminator.
2. How do the genes of procaryotes and eucaryotes usually differ from each other?
3. Briefly discuss the general organization of tRNA and rRNA genes. How does their expression differ from that of structural genes with respect to posttranscriptional modification of the gene product?

11.6 Mutations and Their Chemical Basis

Considerable information is embedded in the precise order of nucleotides in DNA. For life to exist with stability, it is essential that the nucleotide sequence of genes is not disturbed to any great extent. However, sequence changes do occur and often result in altered phenotypes. These changes are largely detrimental but are important in generating new variability and contribute to the process of evolution. Microbial mutation rates also can be increased, and these genetic changes have been put to many important uses in the laboratory and industry.

Mutations [Latin *mutare,* to change] were initially characterized as altered phenotypes or phenotypic expressions. Long before the existence of direct proof that a mutation is a stable, heritable change in the nucleotide sequence of DNA, geneticists predicted that several basic types of transmitted mutations could exist. They believed that mutations could arise from the alteration of single pairs of nucleotides and from the addition or deletion of one or two nucleotide pairs in the coding regions of a gene. Clearly, mutations may be characterized according to either the kind of genotypic change that has occurred or their phenotypic consequences. In this section the molecular basis of mutations and mutagenesis is first considered. Then the phenotypic effects of mutations, the detection of mutations, and the use of mutations in carcinogenicity testing are discussed.

Mutations and Mutagenesis

Mutations can alter the phenotype of a microorganism in several different ways. Morphological mutations change the microorganism's colonial or cellular morphology. Lethal mutations, when expressed,

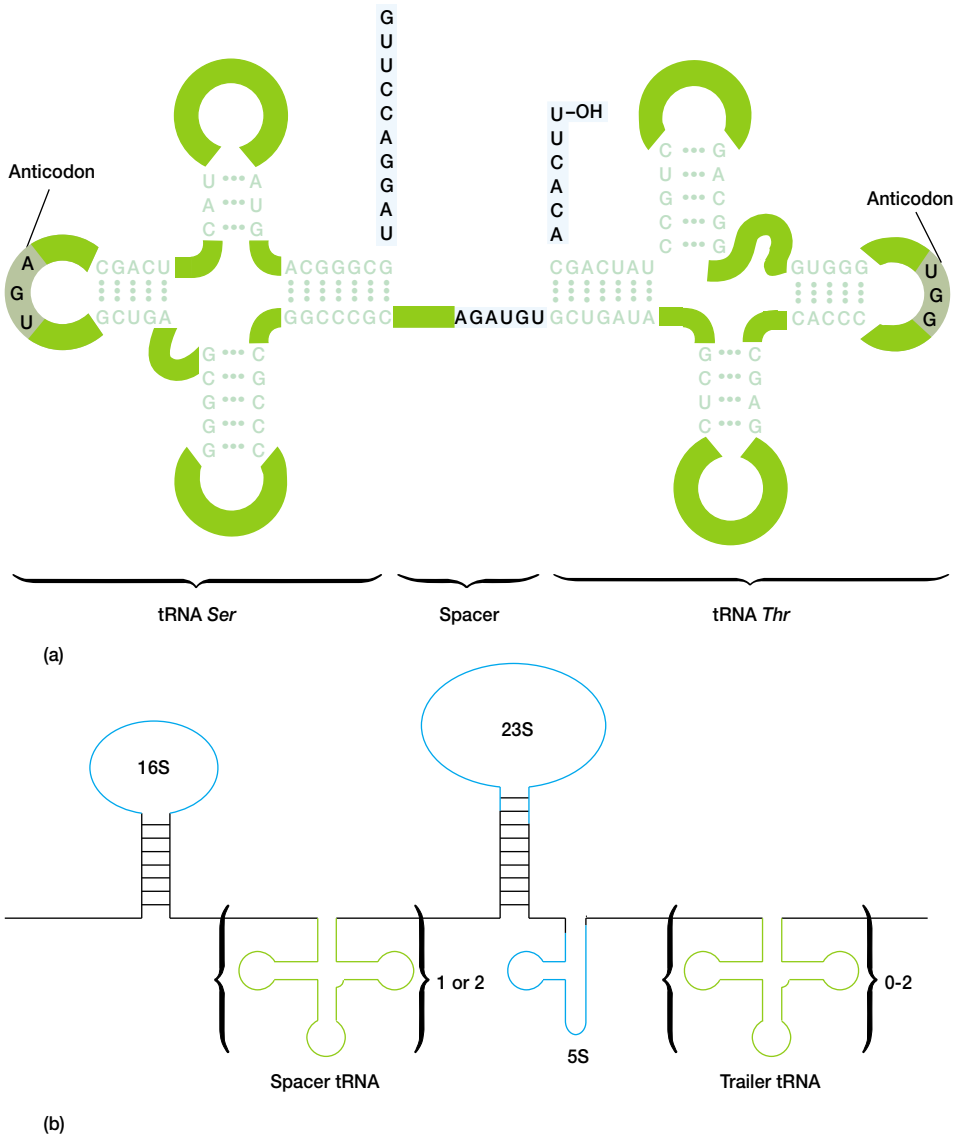


Figure 11.23 **tRNA and rRNA Genes.** (**a**) A tRNA precursor from *E. coli* that contains two tRNA molecules. The spacer and extra nucleotides at both ends are removed during processing. (**b**) The *E. coli* ribosomal RNA gene codes for a large transcription product that is cleaved into three rRNAs and one to three tRNAs. The 16S, 23S, and 5S rRNA segments are represented by blue lines, and tRNA sequences are placed in brackets. The seven copies of this gene vary in the number and kind of tRNA sequences.

result in the death of the microorganism. Since the microorganism must be able to grow in order to be isolated and studied, lethal mutations are recovered only if they are recessive in diploid organisms or conditional (see the following) in haploid organisms.

Conditional mutations are those that are expressed only under certain environmental conditions. For example, a conditional lethal mutation in *E. coli* might not be expressed under permissive conditions such as low temperature but would be expressed under restrictive conditions such as high temperature. Thus the hypothetical mutant would grow normally at the permissive temperature but would die at high temperatures.

Biochemical mutations are those causing a change in the biochemistry of the cell. Since these mutations often inactivate a biosynthetic pathway, they frequently make a microorganism unable to grow on a medium lacking an adequate supply of the pathway's end product. That is, the mutant cannot grow on minimal medium and requires nutrient supplements. Such mutants are called **auxotrophs,** whereas microbial strains that can grow on minimal medium are **prototrophs.** Analysis of auxotrophy has been quite important in microbial genetics due to the ease of auxotroph selection and the relative abundance of this mutational type. Mutant detection and replica plating (pp. 251–52); Nutrient requirements and nutritional types (pp. 96–98)

A resistant mutant is a particular type of biochemical mutant that acquires resistance to some pathogen, chemical, or antibiotic. Such mutants also are easy to select for and very useful in microbial genetics. Mechanisms of drug resistance (pp. 818–19)

Mutations occur in one of two ways. (1) Spontaneous mutations arise occasionally in all cells and develop in the absence of any added agent. (2) Induced mutations, on the other hand, are the result of exposure of the organism to some physical or chemical agent called a **mutagen.**

Although most geneticists believe that spontaneous mutations occur randomly in the absence of an external agent and are then selected, observations by some microbiologists have led to a new and controversial hypothesis. John Cairns and his collaborators have reported that a mutant *E. coli* strain, which is unable to use lactose as a carbon and energy source, regains the ability to do so more rapidly when lactose is added to the culture medium as the only carbon source. Lactose appears to induce mutations that allow *E. coli* to use the sugar again. It has been claimed that these and similar observations on different mutations are examples of **directed or adaptive mutation**—that is, some bacteria seem able to choose which mutations occur so that they can better adapt to their surroundings. Many explanations have been offered to account for this phenomenon without depending on bacterial selection of particular mutations. One of the most interesting is the proposal that **hypermutation** can produce such results. Some starving bacteria might rapidly generate multiple mutations through activation of special mutator genes. This would produce many mutant bacterial cells. In such a random process, the rate of production of favorable mutants would increase, with many of these mutants surviving to be counted. There would appear to be directed or adaptive mutation because many of the unfavorable mutants would die. There is support for this hypothesis. Mutator genes have been discovered and do cause hypermutation under nutritional stress. Even if the directed mutation hypothesis is incorrect, it has stimulated much valuable research and led to the discovery of new phenomena. Some of these issues will be discussed in chapter 42 in the context of evolutionary biotechnology.

Spontaneous Mutations

Spontaneous mutations arise without exposure to external agents. This class of mutations may result from errors in DNA replication, or even from the action of transposons (*see section 13.3*). A few of the more prevalent mechanisms are described in the following paragraphs.

Generally replication errors occur when the base of a template nucleotide takes on a rare tautomeric form. Tautomerism is the relationship between two structural isomers that are in chemical equilibrium and readily change into one another. Bases typically exist in the keto form. However, they can at times take on either an imino or enol form (**figure 11.24*a***). These tautomeric shifts change the hydrogen-bonding characteristics of the bases, allowing purine for purine or pyrimidine for pyrimidine substitutions that can eventually lead to a stable alteration of the nucleotide sequence (figure 11.24*b*). Such substitutions are known as **transition mutations** and are relatively common, although most of them are repaired by various proofreading functions (*see pp. 239 and 254*). In **transversion mutations,** a purine is substituted for a pyrimidine, or a pyrimidine for a purine. These mutations are rarer due to the steric problems of pairing purines with purines and pyrimidines with pyrimidines.

Spontaneous mutations also arise from **frameshifts,** usually caused by the deletion of DNA segments resulting in an altered codon reading frame. These mutations generally occur where there is a short stretch of the same nucleotide. In such a location, the pairing of template and new strand can be displaced by the distance of the repeated sequence leading to additions or deletions of bases in the new strand (**figure 11.25**).

Spontaneous mutations originate from lesions in DNA as well as from replication errors. For example, it is possible for purine nucleotides to be depurinated—that is, to lose their base. This results in the formation of an **apurinic site,** which will not base pair normally and may cause a transition type mutation after the next round of replication. Cytosine can be deaminated to uracil, which is then removed to form an **apyrimidinic site.** Reactive forms of oxygen such as oxygen free radicals and peroxides are produced by aerobic metabolism (*see p. 128*). These may alter DNA bases and cause mutations. For example, guanine can be converted to 8-oxo-7,8-dihydrodeoxyguanine, which often pairs with adenine rather than cytosine during replication.

Finally, spontaneous mutations can result from the insertion of DNA segments into genes. This results from the movement of insertion sequences and transposons (*see pp. 298–302*), and usually inactivates the gene. Insertion mutations are very frequent in *E. coli* and many other bacteria.

Induced Mutations

Virtually any agent that directly damages DNA, alters its chemistry, or interferes with repair mechanisms (pp. 254–56) will induce mutations. Mutagens can be conveniently classified according to their mechanism of action. Four common modes of mutagen action are incorporation of base analogs, specific mispairing, intercalation, and bypass of replication.

Base analogs are structurally similar to normal nitrogenous bases and can be incorporated into the growing polynucleotide chain during replication. Once in place, these compounds typically exhibit base pairing properties different from the bases they replace and can eventually cause a stable mutation. A widely used base analog is 5-bromouracil (5-BU), an analog of thymine. It undergoes a tautomeric shift from the normal keto form to an enol much more frequently than does a normal base. The enol forms hydrogen bonds like cytosine and directs the incorporation of guanine rather than adenine (**figure 11.26**). The mechanism of action of other base analogs is similar to that of 5-bromouracil.

Specific mispairing is caused when a mutagen changes a base's structure and therefore alters its base pairing characteristics. Some mutagens in this category are fairly selective; they preferentially react with some bases and produce a specific kind of DNA damage. An example of this type of mutagen is

Figure 11.24 Transition and Transversion Mutations. Errors in replication due to base tautomerization. (**a**) Normally AT and GC pairs are formed when keto groups participate in hydrogen bonds. In contrast, enol tautomers produce AC and GT base pairs. (**b**) Mutation as a consequence of tautomerization during DNA replication. The temporary enolization of guanine leads to the formation of an AT base pair in the mutant, and a GC to AT transition mutation occurs. The process requires two replication cycles. The mutation only occurs if the abnormal first-generation GT base pair is missed by repair mechanisms.

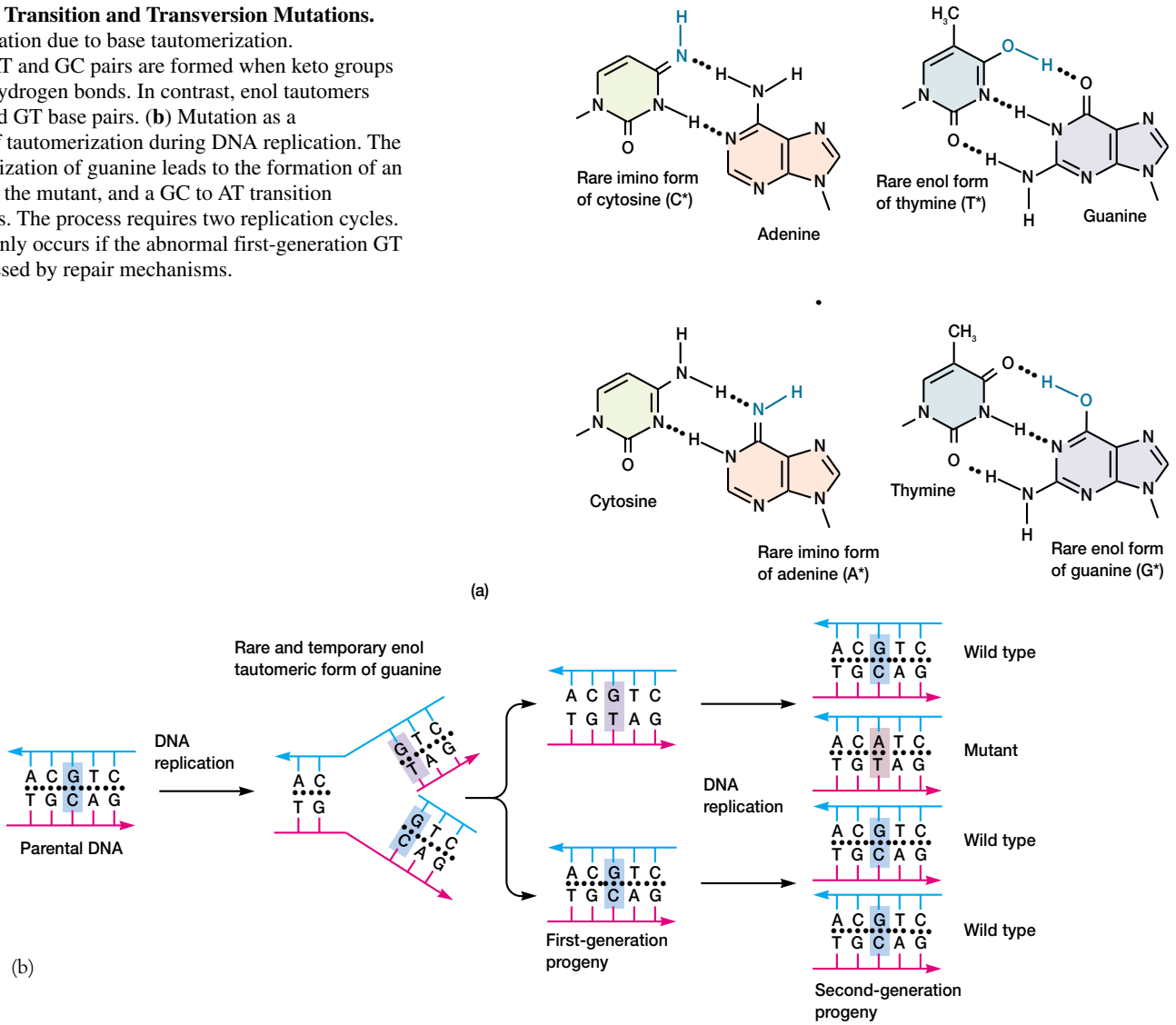
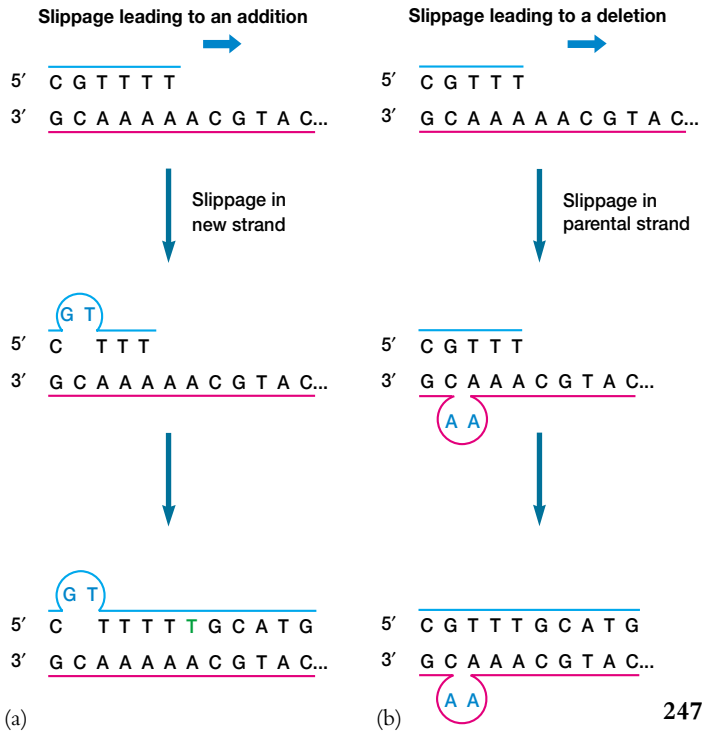
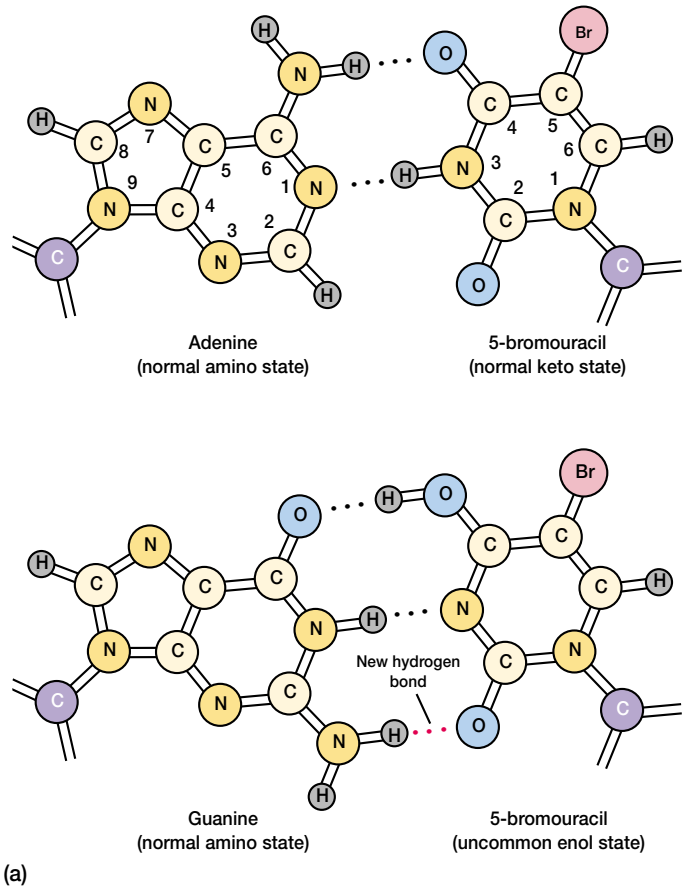


Figure 11.25 Additions and Deletions. A hypothetical mechanism for the generation of additions and deletions during replication. The direction of replication is indicated by the large arrow. In each case there is strand slippage resulting in the formation of a small loop that is stabilized by the hydrogen bonding in the repetitive sequence, the AT stretch in this example. DNA synthesis proceeds to the right in this figure. (**a**) If the new strand slips, an addition of one T results. (**b**) Slippage of the parental strand yields a deletion (in this case, a loss of two Ts).





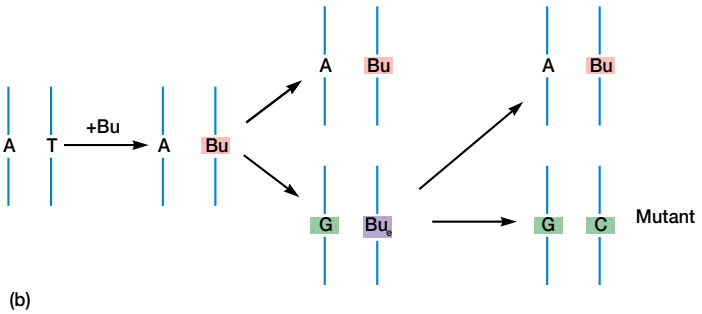


Figure 11.26 **Mutagenesis by the Base Analog 5-Bromouracil.** (**a**) Base pairing of the normal keto form of 5-BU is shown in the top illustration. The enol form of 5-BU (bottom illustration) base pairs with guanine rather than with adenine as might be expected for a thymine analog. (**b**) If the keto form of 5-BU is incorporated in place of thymine, its occasional tautomerization to the enol form (BU_e) will produce an AT to GC transition mutation.

methyl-nitrosoguanidine, an alkylating agent that adds methyl groups to guanine, causing it to mispair with thymine (**figure 11.27**). A subsequent round of replication could then result in a GC-AT transition. DNA damage also stimulates error-prone repair mechanisms. Other examples of mutagens with this mode of action are the alkylating agents ethylmethanesulfonate and hydroxylamine. Hydroxylamine hydroxylates the C-4 nitrogen of cytosine, causing it to base pair like thymine. There are many other DNA modifying agents that can cause mispairing.

Intercalating agents distort DNA to induce single nucleotide pair insertions and deletions. These mutagens are planar and insert themselves (intercalate) between the stacked bases of the helix. This results in a mutation, possibly through the formation of a loop in DNA. Intercalating agents include acridines such as proflavin and acridine orange.

Many mutagens, and indeed many carcinogens, directly damage bases so severely that hydrogen bonding between base pairs is impaired or prevented and the damaged DNA can no longer act as a template. For instance, UV radiation generates cyclobutane type dimers, usually thymine dimers, between adjacent pyrimidines (**figure 11.28**). Other examples are ionizing radiation and carcinogens such as aflatoxin B1 and other benzo(a)pyrene derivatives. Such damage to DNA would generally be lethal but may trigger a repair mechanism that restores much of the damaged genetic material, although with considerable error incorporation (pp. 254–56).

Retention of proper base pairing is essential in the prevention of mutations. Often the damage can be repaired before a mutation is permanently established. If a complete DNA replication cycle takes place before the initial lesion is repaired, the mutation frequently becomes stable and inheritable.

The Expression of Mutations

The expression of a mutation will only be readily noticed if it produces a detectable, altered phenotype. A mutation from the most prevalent gene form, the **wild type,** to a mutant form is called a **forward mutation.** Later, a second mutation may make the mutant appear to be a wild-type organism again. Such a mutation is called a **reversion mutation** because the organism seems to have reverted back to its original phenotype. A true **back mutation** converts the mutant nucleotide sequence back to the wild-type sequence. The wild-type phenotype also can be regained by a second mutation in a different gene, a **suppressor mutation,** which overcomes the effect of the first mutation (**table 11.2**). If the second mutation is within the same gene, the change may be called a second site reversion or intragenic suppression. Thus, although revertant phenotypes appear to be wild types, the original DNA sequence may not be restored. In practice, a mutation is visibly expressed when a protein that is in some way responsible for the phenotype is altered sufficiently to produce a new phenotype.

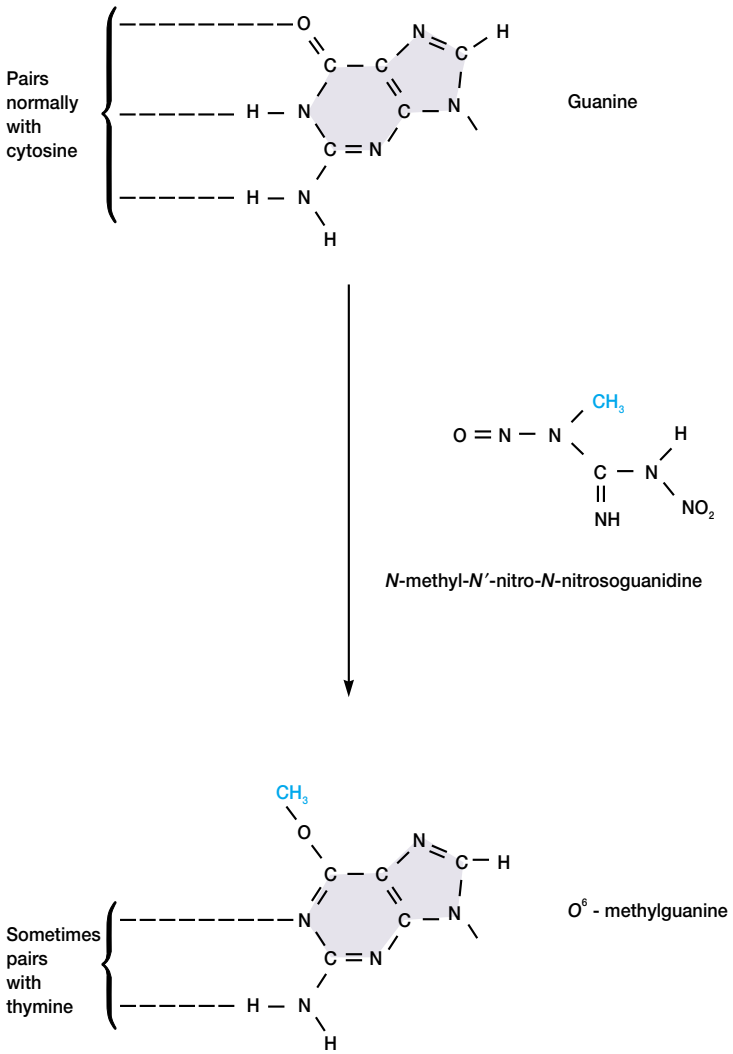


Figure 11.27 **Methyl-Nitrosoguanidine Mutagenesis.** Mutagenesis by methyl-nitrosoguanidine due to the methylation of guanine.

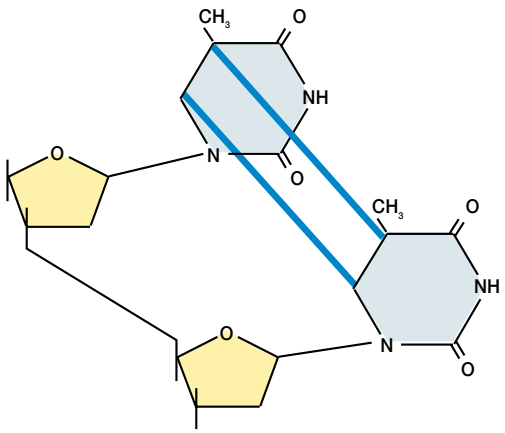


Figure 11.28 **Thymine Dimer.** Thymine dimers are formed by ultraviolet radiation. The enzyme photolyase cleaves the two colored bonds during photoreactivation.

However, mutations may occur and not alter the phenotype for a variety of reasons.

Although very large deletion and insertion mutations exist, most mutations affect only one base pair in a given location and therefore are called **point mutations.** There are several types of point mutations (table 11.2).

One kind of point mutation that could not be detected until the advent of nucleic acid sequencing techniques is the **silent mutation.** If a mutation is an alteration of the nucleotide sequence of DNA, mutations can occur and have no visible effect because of code degeneracy. When there is more than one codon for a given amino acid, a single base substitution could result in the formation of a new codon for the same amino acid. For example, if the codon CGU were changed to CGC, it usually would still code for arginine even though a mutation had occurred. The expression of this mutation often would not be detected except at the level of the DNA or mRNA. When there is no change in the protein or its concentration, there will be no change in the phenotype of the organism.

Table 11.2 Summary of Some Molecular Changes from Gene Mutations

Type of Mutation	Result and Example
Forward Mutations	
Single Nucleotide-Pair (Base-Pair) Substitutions	
At DNA Level	
Transition	Purine replaced by a different purine, or pyrimidine replaced by a different pyrimidine (e.g., AT ⟶ GC).
Transversion	Purine replaced by a pyrimidine, or pyrimidine replaced by a purine (e.g., AT ⟶ CG).
At Protein Level	
Silent mutation	Triplet codes for same amino acid: AGG ⟶ **C**GG (both code for Arg)
Neutral mutation	Triplet codes for different but functionally equivalent amino acid: AAA (Lys) ⟶ A**G**A (Arg)
Missense mutation	Triplet codes for a different amino acid.
Nonsense mutation	Triplet codes for chain termination: CAG (Gln) ⟶ **U**AG (stop)
Single Nucleotide-Pair Addition or Deletion: Frameshift Mutation	Any addition or deletion of base pairs that is not a multiple of three results in a frameshift in reading the DNA segments that code for proteins.
Intragenic Addition or Deletion of Several to Many Nucleotide Pairs	
Reverse Mutations	
True Reversion	AAA (Lys) (wild type) —forward→ GAA (Glu) (mutant) —reverse→ AAA (Lys) (wild type)
Equivalent Reversion	UCC (Ser) (wild type) —forward→ UGC (Cys) (mutant) —reverse→ AGC (Ser) (wild type) CGC (Arg, basic) (wild type) —forward→ CCC (Pro, not basic) (mutant) —reverse→ CAC (His, basic) (pseudo-wild type)
Suppressor Mutations	
Intragenic Suppressor Mutations Frameshift of opposite sign at site within gene. Addition of X to the base sequence shifts the reading frame from the CAT codon to XCA followed by TCA codons. The subsequent deletion of a C base shifts the reading frame back to CAT.	CATCATCATCATCATCAT (+) (–) CATXCATATCATCATCAT y x z y y y
Extragenic Suppressor Mutations	
Nonsense suppressors	Gene (e.g., for tyrosine tRNA) undergoes mutational event in its anticodon region that enables it to recognize and align with a mutant nonsense codon (e.g., UAG) to insert an amino acid (tyrosine) and permit completion of the translation.
Physiological suppressors	A defect in one chemical pathway is circumvented by another mutation—for example, one that opens up another chemical pathway to the same product, or one that permits more efficient uptake of a compound produced in small quantities because of the original mutation.

A second type of point mutation is the **missense mutation.** This mutation involves a single base substitution in the DNA that changes a codon for one amino acid into a codon for another. For example, the codon GAG, which specifies glutamic acid, could be changed to GUG, which codes for valine. The expression of missense mutations can vary. Certainly the mutation is expressed at the level of protein structure. However, at the level of protein function, the effect may range from complete loss of activity to no change at all.

Mutations also occur in the regulatory sequences responsible for the control of gene expression and in other noncoding portions of structural genes. Constitutive lactose operon mutants in *E. coli* are excellent examples. These mutations map in the operator site and produce altered operator sequences that are not recognized by the repressor protein, and therefore the operon is continuously active in transcription. If a mutation renders the promoter sequence nonfunctional, the coding region of the struc-

tural gene will be completely normal, but a mutant phenotype will result due to the absence of a product. RNA polymerase rarely transcribes a gene correctly without a fully functional promoter. The lac operon and gene regulation (pp. 275–78)

Mutations also occur in rRNA and tRNA genes and can alter the phenotype through disruption of protein synthesis. In fact, these mutants often are initially identified because of their slow growth. One type of suppressor mutation is a base substitution in the anticodon region of a tRNA that allows the insertion of the correct amino acid at a mutant codon (table 11.2).

1. Define or describe the following: mutation, conditional mutation, auxotroph and prototroph, spontaneous and induced mutations, mutagen, transition and transversion mutations, frameshift, apurinic site, base analog, specific mispairing, intercalating agent, thymine dimer, wild type, forward and reverse mutations, suppressor mutation, point mutation, silent mutation, missense and nonsense mutations, directed or adaptive mutation and hypermutation, and frameshift mutation.
2. Give four ways in which spontaneous mutations might arise.
3. How do the mutagens 5-bromouracil, methyl-nitrosoguanidine, proflavin, and UV radiation induce mutations?
4. Give examples of intragenic and extragenic suppressor mutations.

11.7 Detection and Isolation of Mutants

In order to study microbial mutants, one must be able to detect them readily, even when there are few, and then efficiently isolate them from the parent organism and other mutants. Fortunately this often is easy to do. This section describes some techniques used in mutant detection, selection, and isolation.

Mutant Detection

When collecting mutants of a particular organism, one must know the normal or wild-type characteristics so as to recognize an altered phenotype. A suitable detection system for the mutant phenotype under study also is needed. Since mutations are generally rare, about one per 10^7 to 10^{11} cells, it is important to have a very sensitive detection system so that these events will not be missed. Geneticists often induce mutations to increase the probability of obtaining specific changes at high frequency (about one in 10^3 to 10^6); even so, mutations are rare.

Many proteins are still functional after the substitution of a single amino acid, but this depends on the type and location of the amino acid. For instance, replacement of a nonpolar amino acid in the protein's interior with a polar amino acid probably will drastically alter the protein's three-dimensional structure and therefore its function. Similarly the replacement of a critical amino acid at the active site of an enzyme will destroy its activity. However, the replacement of one polar amino acid with another at the protein surface may have little or no effect. Missense mutations may actually play a very important role in providing new variability to drive evolution because they often are not lethal and therefore remain in the gene pool. Protein structure (appendix I)

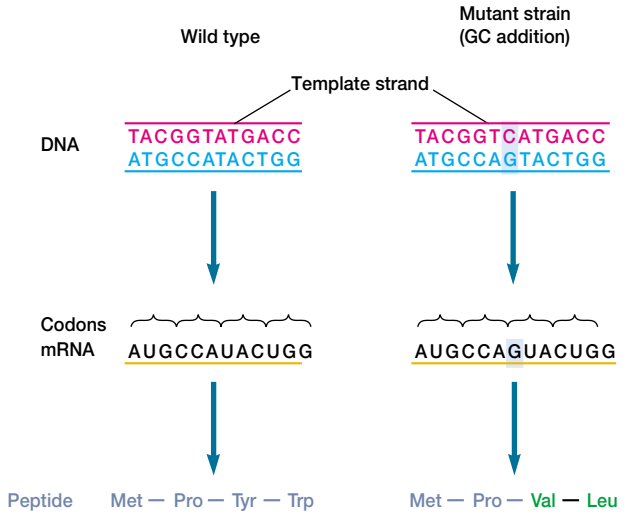


Figure 11.29 **Frameshift Mutation.** A frameshift mutation resulting from the insertion of a GC base pair. The reading frameshift produces a different peptide after the addition.

A third type of point mutation causes the early termination of translation and therefore results in a shortened polypeptide. Such mutations are called **nonsense mutations** because they involve the conversion of a sense codon to a nonsense or stop codon. Depending on the relative location of the mutation, the phenotypic expression may be more or less severely affected. Most proteins retain some function if they are shortened by only one or two amino acids; complete loss of normal function will almost certainly result if the mutation occurs closer to the middle of the gene.

The **frameshift mutation** is a fourth type of point mutation and was briefly mentioned earlier. Frameshift mutations arise from the insertion or deletion of one or two base pairs within the coding region of the gene. Since the code consists of a precise sequence of triplet codons, the addition or deletion of fewer than three base pairs will cause the reading frame to be shifted for all codons downstream (figure 11.19). **Figure 11.29** shows the effect of a frameshift mutation on a short section of mRNA and the amino acid sequence it codes for.

Frameshift mutations usually are very deleterious and yield mutant phenotypes resulting from the synthesis of nonfunctional proteins. The reading frameshift often eventually produces a nonsense or stop codon so that the peptide product is shorter as well as different in sequence. Of course if the frameshift occurred near the end of the gene, or if there were a second frameshift shortly downstream from the first that restored the reading frame, the phenotypic effect might not be as drastic. A second nearby frameshift that restores the proper reading frame is a good example of an intragenic suppressor mutation.

Detection systems in bacteria and other haploid organisms are straightforward because any new allele should be seen immediately, even if it is a recessive mutation. Sometimes detection of

mutants is direct. If albino mutants of a normally pigmented bacterium are being studied, detection simply requires visual observation of colony color. Other direct detection systems are more complex. For example, the **replica plating** technique is used to detect auxotrophic mutants. It distinguishes between mutants and the wild-type strain based on their ability to grow in the absence of a particular biosynthetic end product (**figure 11.30**). A lysine auxotroph, for instance, will grow on lysine-supplemented media but not on a medium lacking an adequate supply of lysine because it cannot synthesize this amino acid.

Once a detection method is established, mutants are collected. Since a specific mutation is a rare event, it is necessary to look at perhaps thousands to millions of colonies or clones. Using direct detection methods, this could become quite a task, even with microorganisms. Consider a search for the albino mutants mentioned previously. If the mutation rate were around one in a million, on the average a million or more organisms would have to be tested to find one albino mutant. This probably would require several thousand plates. The task of isolating auxotrophic mutants in this way would be even more taxing with the added labor of replica plating. This difficulty can be partly overcome by using mutagens to increase the mutation rate, thus reducing the number of colonies to be examined. However, it is more efficient to use a selection system employing some environmental factor to separate mutants from wild-type microorganisms.

Mutant Selection

An effective selection technique uses incubation conditions under which the mutant will grow, because of properties given it by the mutation, whereas the wild type will not. Selection methods often involve reversion mutations or the development of resistance to an environmental stress. For example, if the intent is to isolate revertants from a lysine auxotroph (Lys^-), the approach is quite easy. A large population of lysine auxotrophs is plated on minimal medium lacking lysine, incubated, and examined for colony formation. Only cells that have mutated to restore the ability to manufacture lysine will grow on minimal medium (**figure 11.31**). Thus several million cells can be plated on a single petri dish, and many cells can be tested for mutations by scanning a few petri dishes for growth. This is because the auxotrophs will not grow on minimal medium and confuse the results; only the phenotypic revertants will form colonies. This method has proven very useful in determining the relative mutagenicity of many substances.

Resistance selection methods follow a similar approach. Often wild-type cells are not resistant to virus attack or antibiotic treatment, so it is possible to grow the bacterium in the presence of the agent and look for surviving organisms. Consider the example of a phage-sensitive wild-type bacterium. When the organism is cultured in medium lacking the virus and then plated out on selective medium containing phages, any colonies that form will be resistant to phage attack and very likely will be mutants in this regard. Resistance selection can be used together with virtually any environmental parameter; resistance to bacteriophages, antibiotics, or temperature are most commonly employed.

Substrate utilization mutations also are employed in bacterial selection. Many bacteria use only a few primary carbon sources.

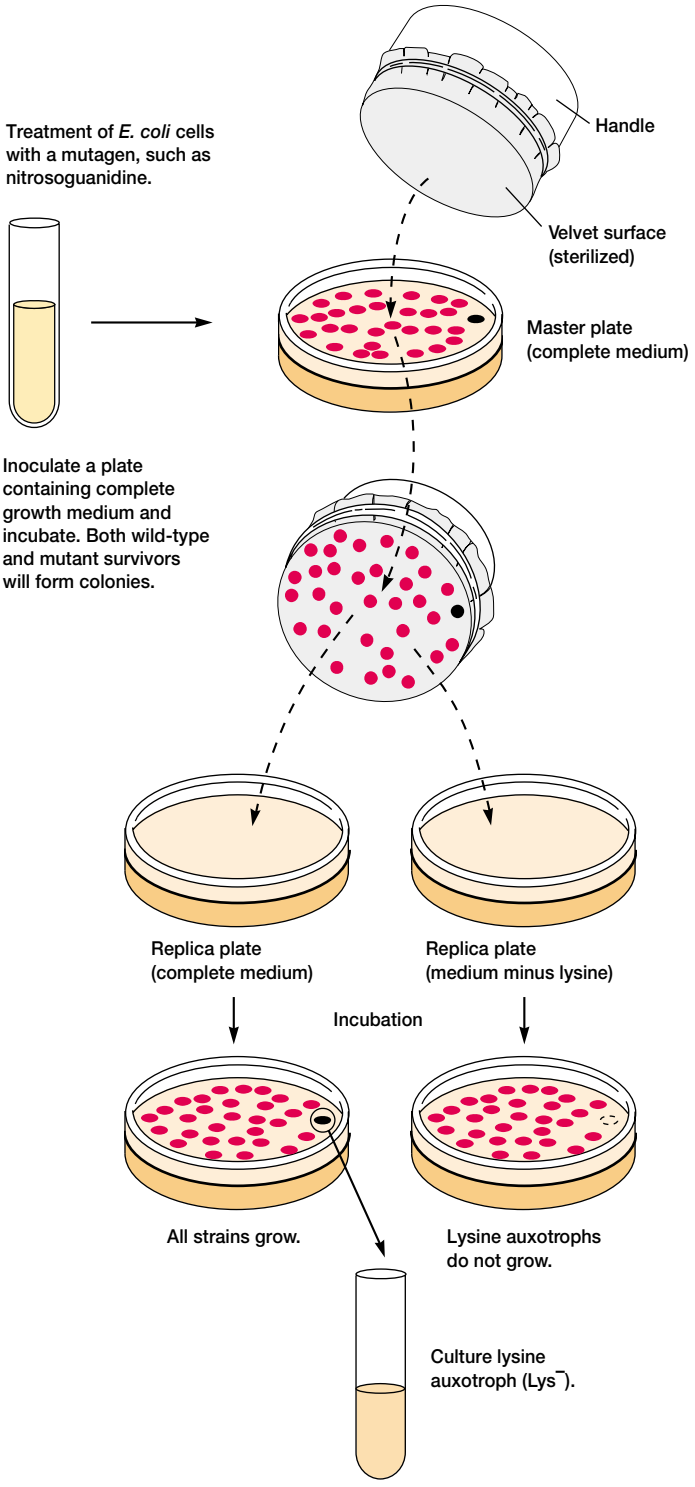


Figure 11.30 Replica Plating. The use of replica plating in isolating a lysine auxotroph. Mutants are generated by treating a culture with a mutagen. The culture containing wild type and auxotrophs is plated on complete medium. After the colonies have developed, a piece of sterile velveteen is pressed on the plate surface to pick up bacteria from each colony. Then the velvet is pressed to the surface of other plates and organisms are transferred to the same position as on the master plate. After determining the location of Lys^- colonies growing on the replica with complete medium, the auxotrophs can be isolated and cultured.

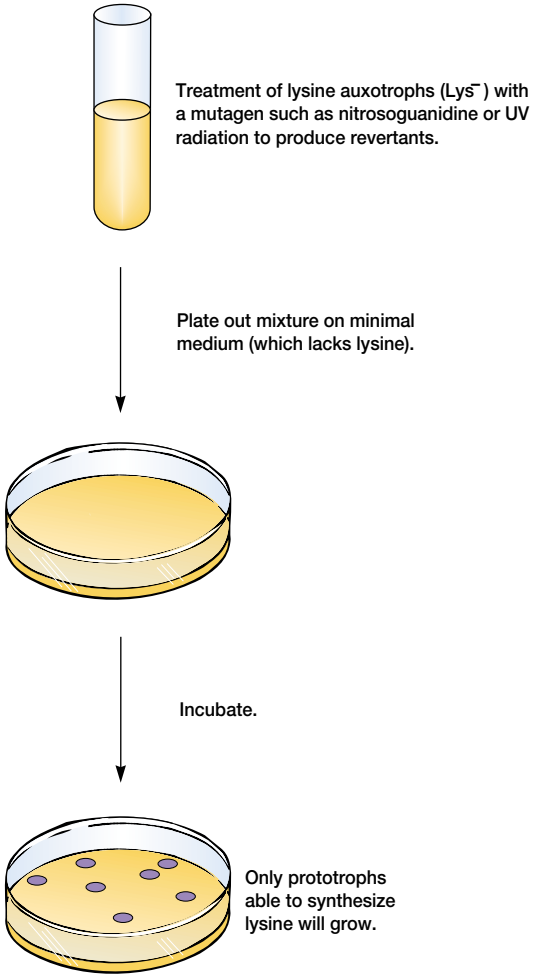


Figure 11.31 **Mutant Selection.** The production and direct selection of auxotroph revertants. In this example, lysine revertants will be selected after treatment of a lysine auxotroph culture because the agar contains minimal medium that will not support auxotroph growth.

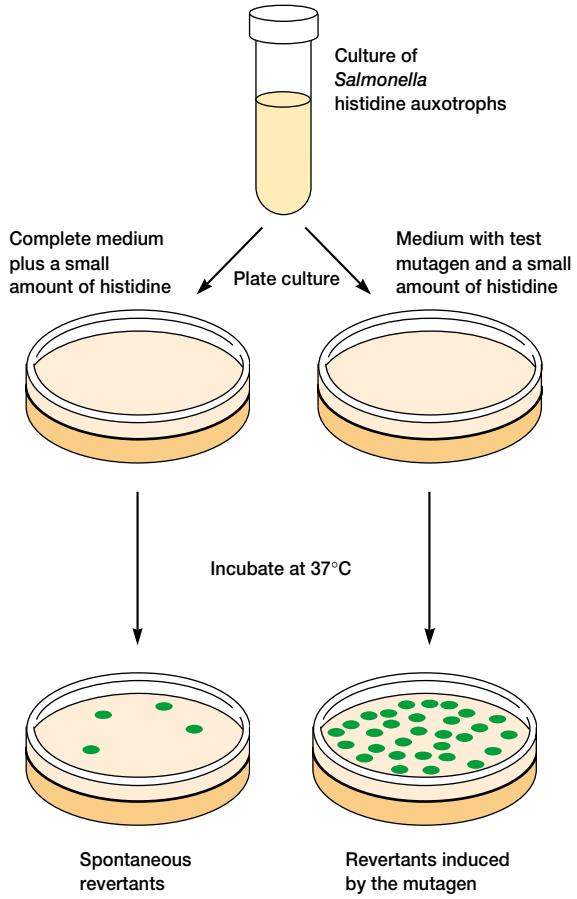


Figure 11.32 **The Ames Test for Mutagenicity.** See text for details.

With such bacteria, it is possible to select mutants by a method similar to that employed in resistance selection. The culture is plated on medium containing an alternate carbon source. Any colonies that appear can use the substrate and are probably mutants.

Mutant detection and selection methods are used for purposes other than understanding more about the nature of genes or the biochemistry of a particular microorganism. One very important role of mutant selection and detection techniques is in the study of carcinogens. The next section briefly describes one of the first and perhaps best known of the carcinogen testing systems.

Carcinogenicity Testing

An increased understanding of the mechanisms of mutation and cancer induction has stimulated efforts to identify environmental carcinogens so that they can be avoided. The observation that many carcinogenic agents also are mutagenic is the basis for detecting potential carcinogens by testing for mutagenicity while taking advantage of bacterial selection techniques and short generation times. The **Ames test,** developed by Bruce Ames in the 1970s, has been widely used to test for carcinogens. The Ames test is a mutational reversion assay employing several special strains of *Salmonella typhimurium,* each of which has a different mutation in the histidine biosynthesis operon. The bacteria also have mutational alterations of their cell walls that make them more permeable to test substances. To further increase assay sensitivity, the strains are defective in the ability to carry out repair of DNA and have plasmid genes that enhance error-prone DNA repair.

In the Ames test these special tester strains of *Salmonella* are plated with the substance being tested and the appearance of visible colonies followed (**figure 11.32**). To ensure that DNA replication can take place in the presence of the potential mutagen, the bacteria and test substance are mixed in dilute molten top agar to which a trace of histidine has been added. This molten mix is then poured on top of minimal agar plates and incubated for 2 to 3 days at 37°C. All of the histidine auxotrophs will grow for the first few hours in the presence of the test compound until the histidine is depleted. Once the histidine supply is exhausted, only revertants that have mutationally

regained the ability to synthesize histidine will grow. The visible colonies need only be counted and compared to controls in order to estimate the relative mutagenicity of the compound: the more colonies, the greater the mutagenicity.

A mammalian liver extract is also often added to the molten top agar prior to plating. The extract converts potential carcinogens into electrophilic derivatives that will readily react with DNA. This process occurs naturally when foreign substances are metabolized in the liver. Since bacteria do not have this activation system, liver extract often is added to the test system to promote the transformations that occur in mammals. Many potential carcinogens, such as the aflatoxins (*see pp. 967–68),* are not actually carcinogenic until they are modified in the liver. The addition of extract shows which compounds have intrinsic mutagenicity and which need activation after uptake. Despite the use of liver extracts, only about half the potential animal carcinogens are detected by the Ames test.

1. Describe how replica plating is used to detect and isolate auxotrophic mutants.
2. Why are mutant selection techniques generally preferable to the direct detection and isolation of mutants?
3. Briefly discuss how reversion mutations, resistance to an environmental factor, and the ability to use a particular nutrient can be employed in mutant selection.
4. What is the Ames test and how is it carried out? What assumption concerning mutagenicity and carcinogenicity is it based upon?

11.8 DNA Repair

Since replication errors and a variety of mutagens can alter the nucleotide sequence, a microorganism must be able to repair changes in the sequence that might be fatal. DNA is repaired by several different mechanisms besides **proofreading** by replication enzymes (DNA polymerases can remove an incorrect nucleotide immediately after its addition to the growing end of the chain). Repair in *E. coli* is best understood and is briefly described in this section. DNA replication and proofreading (pp. 235–39)

Excision Repair

Excision repair is a general repair system that corrects damage that causes distortions in the double helix. A repair endonuclease or uvrABC endonuclease removes the damaged bases along with some bases on either side of the lesion (**figure 11.33**). The resulting single-stranded gap, about 12 nucleotides long, is filled by DNA polymerase I, and DNA ligase joins the fragments (p. 239). This system can remove thymine dimers (figure 11.28) and repair almost any other injury that produces a detectable distortion in DNA.

Besides this general excision repair system, specialized versions of the system excise specific sites on the DNA where the sugar phosphate backbone is intact but the bases have been removed to form apurinic or apyrimidinic sites (AP sites). Special endonucleases called AP endonucleases recognize these locations and nick the backbone at the site. Excision repair then commences, beginning with the excision of a short stretch of nucleotides.

Another type of excision repair employs DNA glycosylases. These enzymes remove damaged or unnatural bases yielding AP sites that are then repaired as above. Not all types of damaged bases are repaired in this way, but new glycosylases are being discovered and the process may be of more general importance than first thought.

Removal of Lesions

Thymine dimers and alkylated bases often are directly repaired. **Photoreactivation** is the repair of thymine dimers by splitting them apart into separate thymines with the help of visible light in a photochemical reaction catalyzed by the enzyme photolyase. Because this repair mechanism does not remove and replace nucleotides, it is error free.

Sometimes damage caused by alkylation is repaired directly as well. Methyls and some other alkyl groups that have been added to the O–6 position of guanine can be removed with the help of an enzyme known as alkyltransferase or methylguanine methyltransferase. Thus damage to guanine from mutagens such as methyl-nitrosoguanidine (figure 11.27) can be repaired directly.

Postreplication Repair

Despite the accuracy of DNA polymerase action and continual proofreading, errors still are made during DNA replication. Remaining mismatched bases and other errors are usually detected and repaired by the **mismatch repair system** in *E. coli.* The mismatch correction enzyme scans the newly replicated DNA for mismatched pairs and removes a stretch of newly synthesized DNA around the mismatch. A DNA polymerase then replaces the excised nucleotides, and the resulting nick is sealed with a ligase. Postreplication repair is a type of excision repair.

Successful postreplication repair depends on the ability of enzymes to distinguish between old and newly replicated DNA strands. This distinction is possible because newly replicated DNA strands lack methyl groups on their bases, whereas older DNA has methyl groups on the bases of both strands. **DNA methylation** is catalyzed by DNA methyltransferases and results in three different products: *N*6-methyladenine, 5-methylcytosine, and *N*4-methylcytosine. After strand synthesis, the *E. coli* DNA adenine methyltransferase (DAM) methylates adenine bases in d(GATC) sequences to form *N*6-methyladenine. For a short time after the replication fork has passed, the new strand lacks methyl groups while the template strand is methylated. The repair system cuts out the mismatch from the unmethylated strand.

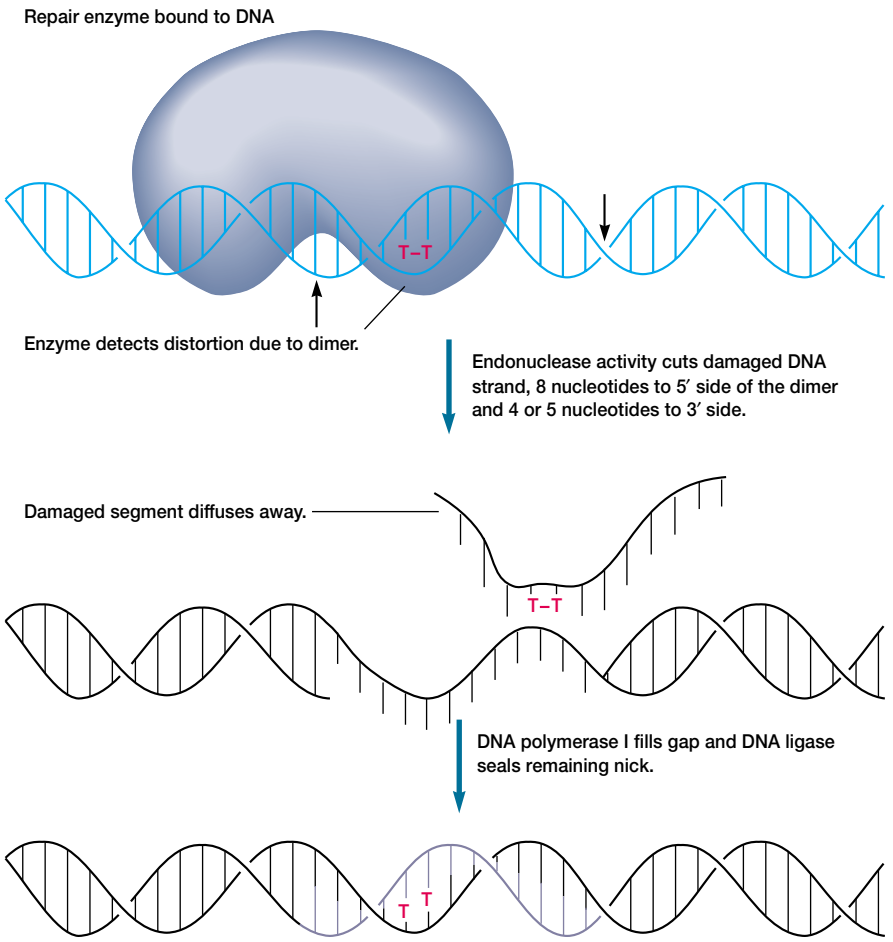


Figure 11.33 **Excision Repair.** Excision repair of a thymine dimer that has distorted the double helix. The repair endonuclease or uvrABC endonuclease is coded for by the *uvrA, B,* and *C* genes.

Recombination Repair

In **recombination repair,** damaged DNA for which there is no remaining template is restored. This situation arises if both bases of a pair are missing or damaged, or if there is a gap opposite a lesion. In this type of repair the **recA protein** cuts a piece of template DNA from a sister molecule and puts it into the gap or uses it to replace a damaged strand. Although bacteria are haploid, another copy of the damaged segment often is available because either it has recently been replicated or the cell is growing rapidly and has more than one copy of its chromosome. Once the template is in place, the remaining damage can be corrected by another repair system.

The recA protein also participates in a type of inducible repair known as **SOS repair.** In this instance the DNA damage is so great that synthesis stops completely, leaving many large gaps. RecA will bind to the gaps and initiate strand exchange. Simultaneously it takes on a proteolytic function that destroys the lexA repressor protein, which regulates the function of many genes involved in DNA repair and synthesis (**figure 11.34**). As a result many more copies of these enzymes are produced, accelerating the replication and repair processes. The system can quickly repair extensive damage caused by agents such as UV radiation, but it is error prone and does produce mutations. However, it is certainly better to have a few mutations than no DNA replication at all.

1. Define the following: proofreading, excision repair, photoreactivation, methylguanine methyltransferase, mismatch repair, DNA methylation, recombination repair, recA protein, SOS repair, and lexA repressor.
2. Describe in general terms the mechanisms of the following repair processes: excision repair, recombination repair, and SOS repair.

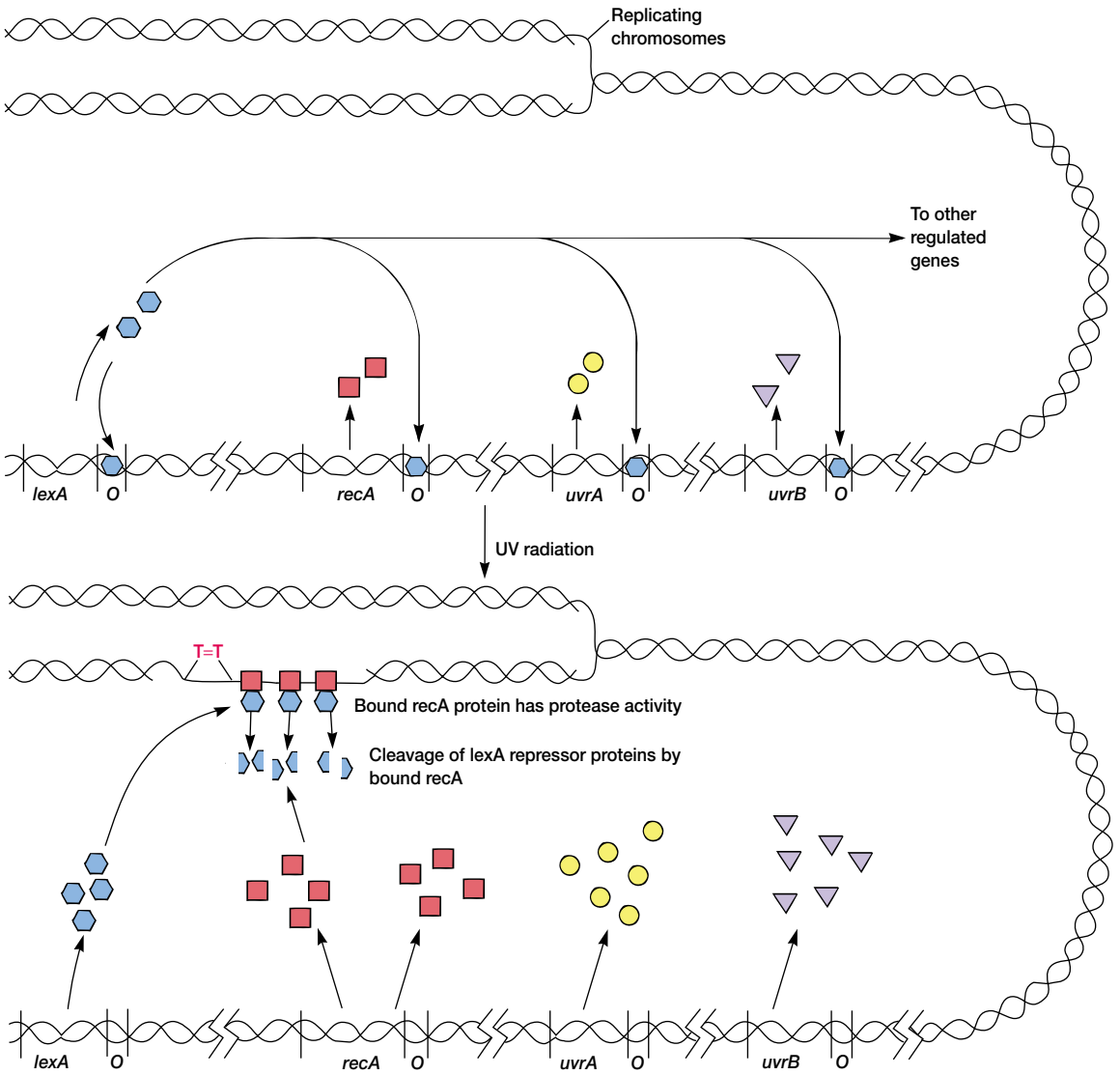


Figure 11.34 **The SOS Repair Process.** In the absence of damage, repair genes are expressed in *E. coli* at low levels due to binding of the lexA repressor protein at their operators (*O*). When the recA protein binds to a damaged region—for example, a thymine dimer created by UV radiation—it destroys lexA and the repair genes are expressed more actively. The *uvr* genes code for the repair endonuclease or uvrABC endonuclease responsible for excision repair.

Summary

1. The knowledge that DNA is the genetic material for cells came from studies on transformation by Griffith and Avery and from experiments on T2 phage reproduction by Hershey and Chase.
2. DNA differs in composition from RNA in having deoxyribose and thymine rather than ribose and uracil.
3. DNA is double stranded, with complementary AT and GC base pairing between the strands. The strands run antiparallel and are twisted into a right-handed double helix (**figure 11.6**).
4. RNA is normally single stranded, although it can coil upon itself and base pair to form hairpin structures.
5. In almost all procaryotes DNA exists as a closed circle that is twisted into supercoils and associated with histonelike proteins.
6. Eucaryotic DNA is associated with five types of histone proteins. Eight histones associate to form ellipsoidal octamers around which the DNA is coiled to produce the nucleosome (**figure 11.9**).
7. DNA synthesis is called replication. Transcription is the synthesis of an RNA copy of DNA and produces three types of RNA: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA).
8. The synthesis of protein under the direction of mRNA is called translation.
9. Most circular procaryotic DNAs are copied by two replication forks moving around the circle to form a theta-shaped (θ) figure. Sometimes a rolling-circle mechanism is employed instead.
10. Eucaryotic DNA has many replicons and replication origins located every 10 to 100 μm along the DNA.

11. DNA polymerase enzymes catalyze the synthesis of DNA in the 5′ to 3′ direction while reading the DNA template in the 3′ to 5′ direction.
12. The double helix is unwound by helicases with the aid of topoisomerases like the DNA gyrase. DNA binding proteins keep the strands separate.
13. DNA polymerase III holoenzyme synthesizes a complementary DNA copy beginning with a short RNA primer made by a primase enzyme.
14. The leading strand is probably replicated continuously, whereas DNA synthesis on the lagging strand is discontinuous and forms Okazaki fragments (**figures 11.15** and **11.16**).
15. DNA polymerase I excises the RNA primer and fills in the resulting gap. DNA ligase then joins the fragments together.
16. Genetic information is carried in the form of 64 nucleotide triplets called codons (**table 11.1**); sense codons direct amino acid incorporation, and stop or nonsense codons terminate translation.
17. The code is degenerate—that is, there is more than one codon for most amino acids.
18. A gene may be defined as the nucleic acid sequence that codes for a polypeptide, tRNA, or rRNA.
19. The template strand of DNA carries genetic information and directs the synthesis of the RNA transcript.
20. RNA polymerase binds to the promoter region, which contains RNA polymerase recognition and RNA polymerase binding sites (**figure 11.22**).
21. The gene also contains a coding region and a terminator; it may have a leader and a trailer (**figure 11.21**). Regulatory segments such as operators may be present.
22. The genes for tRNA and rRNA often code for a precursor that is subsequently processed to yield several products.
23. A mutation is a stable, heritable change in the nucleotide sequence of the genetic material, usually DNA.
24. Mutations can be divided into many categories based on their effects on the phenotype, some major types are morphological, lethal, conditional, biochemical, and resistance mutations.
25. Spontaneous mutations can arise from replication errors (transitions, transversions, and frameshifts), from DNA lesions (apurinic sites, apyrimidinic sites, oxidations), and from insertions.
26. Induced mutations are caused by mutagens. Mutations may result from the incorporation of base analogs, specific mispairing due to alteration of a base, the presence of intercalating agents, and a bypass of replication because of severe damage. Starvation and environmental stresses may stimulate mutator genes and lead to hypermutation.
27. The mutant phenotype can be restored to wild type by either a true reverse mutation or a suppressor mutation (**table 11.2**).
28. There are four important types of point mutations: silent mutations, missense mutations, nonsense mutations, and frameshift mutations (**table 11.2**).
29. It is essential to have a sensitive and specific detection technique to isolate mutants; an example is replica plating (**figure 11.30**) for the detection of auxotrophs (a direct detection system).
30. One of the most effective isolation techniques is to adjust environmental conditions so that the mutant will grow while the wild-type organism does not.
31. Because many carcinogens are also mutagenic, one can test for mutagenicity with the Ames test (**figure 11.32**) and use the results as an indirect indication of carcinogenicity.
32. Mutations and DNA damage are repaired in several ways; for example: proofreading by replication enzymes, excision repair, removal of lesions (e.g., photoreactivation), postreplication repair (mismatch repair), and recombination repair.

Key Terms

Ames test *253*
apurinic site *246*
apyrimidinic site *246*
auxotroph *245*
back mutation *248*
base analog *246*
cistron *241*
clone *228*
code degeneracy *240*
coding region *244*
codon *240*
complementary *231*
conditional mutation *245*
directed or adaptive mutation *246*
deoxyribonucleic acid (DNA) *230*
DNA gyrase *237*
DNA ligase *239*
DNA methylation *254*
DNA polymerase *236*
excision repair *254*
forward mutation *248*
frameshift *246*
frameshift mutation *251*
gene *241*
genome *228*
helicase *236*
histone *234*
hypermutation *246*
intercalating agent *248*
leader sequence *244*
major groove *231*
messenger RNA (mRNA) *230*
minor groove *231*
mismatch repair system *254*
missense mutation *250*
mutagen *246*
mutation *244*
nonsense mutation *251*
nucleosome *235*
Okazaki fragment *239*
photoreactivation *254*
point mutation *249*
primase *239*
primosome *239*
promoter *242*
proofreading *254*
prototroph *245*
reading frame *241*
recA protein *255*
recombination repair *255*
replica plating *252*
replication *230*
replication fork *235*
replicon *235*
reversion mutation *248*
ribonucleic acid (RNA) *230*
RNA polymerase binding site or Pribnow box *243*
rolling-circle mechanism *236*
sense codons *240*
Shine-Dalgarno sequence *244*
silent mutation *249*
single-stranded DNA binding proteins (SSBs) *237*
SOS repair *255*
specific mispairing *246*
stop or nonsense codons *241*
suppressor mutation *248*
template strand *242*
terminator sequence *244*
topoisomerase *237*
trailer sequence *244*
transcription *230*
transformation *228*
transition mutation *246*
translation *230*
transversion mutation *246*
wild type *248*
wobble *241*

Questions for Thought and Review

1. How do replication patterns differ between procaryotes and eucaryotes? Describe the operation of replication forks in the generation of theta-shaped intermediates and in the rolling-circle mechanism.
2. Outline the steps involved in DNA synthesis at the replication fork. How do DNA polymerases correct their mistakes?
3. Currently a gene is described in several ways. Which definition do you prefer and why?
4. How could one use small deletion mutations to show that codons are triplet (i.e., that the nucleotide sequence is read three bases at a time rather than two or four)?
5. Sometimes a point mutation does not change the phenotype. List all the reasons you can why this is so.
6. Why might a mutation leading to an amino acid change at a protein's surface not result in a phenotypic change while the substitution of an internal amino acid will?
7. Describe how you would isolate a mutant that required histidine for growth and was resistant to penicillin.
8. How would the following DNA alterations and replication errors be corrected (there may be *more* than one way): base addition errors by DNA polymerase III during replication, thymine dimers, AP sites, methylated guanines, and gaps produced during replication?

Critical Thinking Questions

1. Mutations are often considered harmful. Give an example of a mutation that would be beneficial to a microorganism. What gene would bear the mutation? How would the mutation alter the gene's role in the cell, and what conditions would select for this mutant allele?
2. Mistakes made during transcription affect the cell, but are not considered "mutations." Why not?
3. Given what you know about the difference between procaryotic and eucaryotic cells, give two reasons why the Ames test detects only about half of potential carcinogens, even when liver extracts are used.
4. Suppose that you have isolated a microorganism from a soil sample. Describe how you would go about determining the nature of its genetic material.

Additional Reading

General

Dale, J. W. 1998. *Molecular genetics of bacteria,* 3rd ed. New York: John Wiley and Sons.

Griffiths, A. J. F.; Miller, J. H.; Suzuki, D. T.; Lewontin, R. C.; and Gelbart, W. M. 2000. *An introduction to genetic analysis,* 7th ed. New York: W. H. Freeman.

Hartwell, L. H.; Hood, L.; Goldberg, M. L.; Reynolds, A. E.; Silver, L. M.; and Veres, R. C. 2000. *Genetics: From genes to genomes.* New York: McGraw-Hill.

Holloway, B. W. 1993. Genetics for all bacteria. *Ann. Rev. Microbiol.* 47:659–84.

Joset, F., and Guespin-Michel, J. 1993. *Prokaryotic genetics: Genome organization, transfer and plasticity.* Boston: Blackwell.

Kendrew, J., editor. 1994. *The encyclopedia of molecular biology.* Boston: Blackwell Scientific Publications.

Klug, W. S., and Cummings, M. R. 1997. *Concepts of Genetics,* 5th ed. Upper Saddle River, N.J.: Prentice-Hall.

Lewin, B. 2000. *Genes,* 7th ed. New York: Oxford University Press.

Maloy, S. R.; Cronan, J. E., Jr.; and Freifelder, D. 1994. *Microbial genetics,* 2d ed. Boston: Jones and Bartlett.

Russell, P. J. 1998. *Genetics,* 5th ed. New York: Harper Collins.

Scaife, J.; Leach, D.; and Galizzi, A., editors. 1985. *Genetics of bacteria.* New York: Academic Press.

Smith-Keary, P. 1989. *Molecular genetics of Escherichia coli.* New York: Guilford Press.

Snyder, L., and Champness, W. 1997. *Molecular Genetics of Bacteria.* Washington, D.C.: ASM Press

Weaver, R. F. 1999. *Molecular biology.* Dubuque, Iowa: WCB McGraw-Hill.

Weaver, R. F., and Hedrick, P. W. 1997. *Genetics,* 3d ed. Dubuque, Iowa: Wm. C. Brown.

11.2 Nucleic Acid Structure

Bauer, W. R.; Crick, F. H. C.; and White, J. H. 1980. Supercoiled DNA. *Sci. Am.* 243(1):118–33.

Darnell, J. E., Jr. 1985. RNA. *Sci. Am.* 253(4):68–78.

Drlica, K. 2000. Chromosome, Bacterial. In *Encyclopedia of microbiology,* 2d ed., vol. 1, J. Lederberg, editor-in-chief, 808–21. San Diego: Academic Press.

Drlica, K., and Rouviere-Yaniv, J. 1987. Histonelike proteins of bacteria. *Microbiol. Rev.* 51(3):301–19.

Felsenfeld, G. 1985. DNA. *Sci. Am.* 253(4):59–67.

Kornberg, R. D., and Klug, A. 1981. The nucleosome. *Sci. Am.* 244(2):52–64.

Rich, A., and Kim, S. H. 1978. The three-dimensional structure of transfer RNA. *Sci. Am.* 238(1):52–62.

11.3 DNA Replication

Baker, T. A., and Bell, S. P. 1998. Polymerases and the replisome: Machines within machines. *Cell* 92:295–305.

Cook, P. R. 1999. The organization of replication and transcription. *Science* 284: 1790–95.

Dickerson, R. E. 1983. The DNA helix and how it is read. *Sci. Am.* 249(6):94–111.

Hejna, J. A., and Moses, R. E. 2000. DNA replication. In *Encyclopedia of microbiology,* 2d ed., vol. 2, J. Lederberg, editor-in-chief, 82–90. San Diego: Academic Press.

Johnson, K. A. 1993. Conformational coupling in DNA polymerase fidelity. *Ann. Rev. Biochem.* 62:685–713.

Joyce, C. M., and Steitz, T. A. 1994. Function and structure relationships in DNA polymerases. *Ann. Rev. Biochem.* 63:777–822.

Kornberg, A. 1992. *DNA replication,* 2d ed. San Francisco: W. H. Freeman.

Lohman, T. M.; Thorn, K.; and Vale, R. D. 1998. Staying on track: Common features of DNA helicases and microtubule motors. *Cell* 93:9–12.

Losick, R., and Shapiro, L. 1998. Bringing the mountain to Mohammed. *Science* 282:1430–31.

Marians, K. J. 1992. Prokaryotic DNA replication. *Annu. Rev. Biochem.* 61:673–719.

Matson, S. W., and Kaiser-Rogers, K. A. 1990. DNA helicases. *Annu. Rev. Biochem.* 59:289–329.

McHenry, C. S. 1988. DNA polymerase III holoenzyme of *Escherichia coli. Annu. Rev. Biochem.* 57:519–50.

Meyer, R. R., and Laine, P. S. 1990. The single-stranded DNA-binding protein of *Escherichia coli. Microbiol. Rev.* 54(4):342–80.

Radman, M., and Walker, R. 1988. The high fidelity of DNA duplication. *Sci. Am.* 259(2):40–46.

Roca, J. 1995. The mechanisms of DNA topoisomerases. *Trends Biochem. Sci.* 20(4):156–60.

Stillman, B. 1994. Smart machines at the DNA replication fork. *Cell* 78:725–28.

Waga, S., and Stillman, B. 1998. The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* 67:721–51.

Wang, J. C. 1982. DNA topoisomerases. *Sci. Am.* 247(1):94–109.

11.4 The Genetic Code

Andersson, S. G. E., and Kurland, C. G. 1990. Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54(2):198–210.

Osawa, S.; Jukes, T. H.; Watanabe, K.; and Muto, A. 1992. Recent evidence for evolution of the genetic code. *Microbiol. Rev.* 56(1):229–64.

11.5 Gene Structure

Berlyn, M. K. B.; Low, K. B.; and Rudd, K. E. 1996. Linkage map of *Escherichia coli* K–12, edition 9. In *Escherichia coli and Salmonella: Cellular and molecular biology,* 2d ed., vol. 2, F. C. Neidhardt, editor-in-chief, 1715–1902. Washington, D.C.: ASM Press.

Breathnach, R., and Chambon, P. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.* 50:349–83.

Fournier, M. J., and Ozeki, H. 1985. Structure and organization of the transfer ribonucleic acid genes of *Escherichia coli* K–12. *Microbiol. Rev.* 49(4):379–97.

Girons, I. S.; Old, I. G.; and Davidson, B. E. 1994. Molecular biology of *Borrelia,* bacteria with linear replicons. *Microbiology* 140:1803–16.

Lindahl, L., and Zengel, J. M. 1986. Ribosomal genes in *Escherichia coli. Annu. Rev. Genet.* 20:297–326.

Marinus, M. G. 2000. Methylation of nucleic acids and proteins. In *Encyclopedia of microbiology,* 2d ed., vol. 3, J. Lederberg, editor-in-chief, 240–44. San Diego: Academic Press.

Riley, M. 1993. Functions of the gene products of *Escherichia coli. Microbiol. Rev.* 57(4):862–952.

Weinstock, G. M. 1994. Bacterial genomes: Mapping and stability. *ASM News* 60(2):73–78.

11.6 Mutations and Their Chemical Basis

Foster, P. L. 1993. Adaptive mutation: The uses of adversity. *Annu. Rev. Microbiol.* 47:467–504.

Hall, B. G. 1991. Increased rates of advantageous mutations in response to environmental challenges. *ASM News* 57(2):82–86.

Lederberg, J. 1992. Bacterial variation since Pasteur. *ASM News* 58(5):261–65.

Lenski, R. E., and Mittler, J. E. 1993. The directed mutation controversy and neo-Darwinism. *Science* 259:188–94.

Miller, J. H. 1983. Mutational specificity in bacteria. *Annu. Rev. Genet.* 17:215–38.

Miller J. H. 1996. Spontaneous mutators in bacteria: Insights into pathways of mutagenesis and repair. *Annu. Rev. Microbiol.* 50:625–43.

Singer, B., and Kusmierek, J. T. 1982. Chemical mutagenesis. *Annu. Rev. Biochem.* 52:655–93.

11.7 Detection and Isolation of Mutants

Devoret, R. 1979. Bacterial tests for potential carcinogens. *Sci. Am.* 241(2):40–49.

11.8 DNA Repair

Claverys, J.-P., and Lacks, S. A. 1986. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol. Rev.* 50(2):133–65.

Friedberg, E. C.; Walker, G. C.; and Siede, W. 1995. *DNA repair and mutagenesis.* Herndon, Va.: ASM Press.

Grossman, L. 2000. DNA repair. In *Encyclopedia of microbiology,* 2d ed., vol. 2, J. Lederberg, editor-in-chief, 71–81. San Diego: Academic Press.

Grossman, L., and Thiagalingam, S. 1993. Nucleotide excision repair, a tracking mechanism in search of damage. *J. Biol. Chem.* 268(23):16871–74.

Howard-Flanders, P. 1981. Inducible repair of DNA. *Sci. Am.* 245(5):72–80.

Kuzminov, A. 1999. Recombinational repair of DNA damage in *Escherichia coli* and bacteriophage λ. *Microbiol. Mol. Biol. Rev.* 63(4):751–813.

McCullough, A. K.; Dodson, M. L.; and Lloyd, R. S. 1999. Initiation of base excision repair: Glycosylase mechanisms and structures. *Annu. Rev. Biochem.* 68:255–85.

Miller, R. V. 2000. recA: The gene and its protein product. In *Encyclopedia of microbiology,* 2d ed., vol. 4, J. Lederberg, editor-in-chief, 43–54. San Diego: Academic Press.

Sutherland, B. M. 1981. Photoreactivation. *BioScience* 31(6):439–44.

Van Houten, B. 1990. Nucleotide excision repair in *Escherichia coli. Microbiol. Rev.* 54(1):18–51.

Walker, G. C. 1985. Inducible DNA repair systems. *Annu. Rev. Biochem.* 54:425–57.

Winterling, K. W. 2000. SOS response. In *Encyclopedia of microbiology,* 2d ed., vol. 4, J. Lederberg, editor-in-chief, 336–43. San Diego: Academic Press.