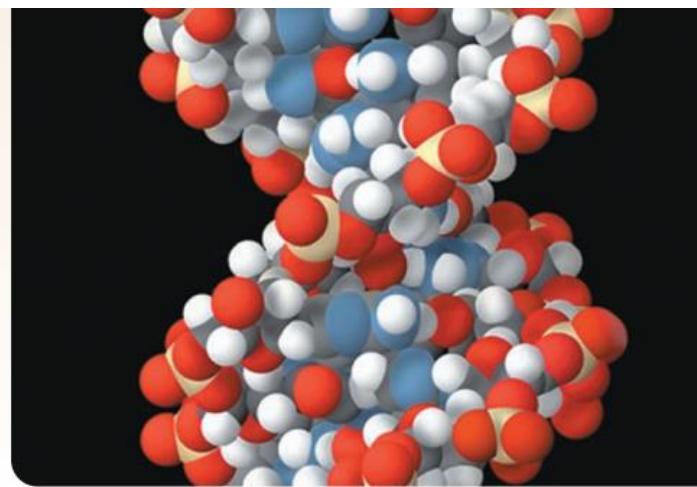


# 11

## Microbial Genetics: Gene Structure, Replication, and Expression



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

### PREVIEW

- The two kinds of nucleic acid, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), differ from one another in chemical composition and structure. In cells, DNA serves as the repository for genetic information.
- 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, which is complementary to a portion of the DNA genome.
- 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 DNA is copied.
- A gene is a nucleotide sequence that codes for a polypeptide, tRNA, or rRNA. Most prokaryotic genes have at least four major parts, each with different functions: promoters, leaders, coding regions, and trailers. When a gene directs the synthesis of a polypeptide, each amino acid is specified by a triplet codon.
- In transcription the RNA polymerase copies the appropriate sequence on the DNA template strand to produce a complementary RNA copy of the gene. Transcription differs in a number of ways among *Bacteria*, *Archaea*, and eukaryotes, even though the basic mechanism of RNA polymerase action is essentially the same.
- Translation is the process by which the nucleotide sequence of mRNA is converted into the amino acid sequence of a polypeptide through the action of ribosomes, tRNAs, aminoacyl-tRNA synthetases, ATP and GTP energy, and a variety of protein factors. As in the case of DNA replication, this complex process is designed to minimize errors.

Chapters 8 through 10 introduce the essentials of microbial metabolism. They focus on processes that provide the energy and metabolic precursors used by cells to synthesize the macromolecules needed for construction of chromosomes, ribosomes, cell walls, and other cellular components. We now turn our attention to the synthesis of three major macromolecules—DNA, RNA, and proteins—from their constituent monomers. DNA serves as the storage molecule for the genetic instructions that allow organisms to carry out metabolism and reproduction. RNA functions in the expression of genetic information so that enzymes and other proteins can be made. These proteins are used to build certain cellular structures and to do other cellular work. The study of the synthesis of DNA, RNA, and protein falls into the realms of genetics and molecular biology.

In the mid-1800s, the discipline of genetics was born from the work of Gregor Mendel, who studied the inheritance of various traits in pea plants. In the early twentieth century, Mendel's work was rediscovered and furthered by scientists working with fruit flies and plants such as corn. The use of microorganisms as models for genetic studies soon followed. Microorganisms, especially bacteria, have significant advantages as model organisms, in part because of their unique characteristics. One important feature is the nature of their genomes. The term **genome** refers to all the DNA present in a cell or virus. Prokaryotes normally have one set of genes; that is, they are haploid (1N). In addition, they often

*But 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

carry extrachromosomal genetic elements called plasmids. Eucaryotic organisms, including eucaryotic microorganisms, usually have two sets of genes, or are diploid ( $2N$ ), and they rarely have plasmids. Viral genomes differ significantly from those of cellular organisms, and their genetics and molecular biology are discussed in chapters 16–18. [Plasmids \(section 3.5\)](#)

In this chapter we review 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, gene structure, and how genes function (i.e., gene expression). Based on the foundation provided in chapter 11, chapter 12 considers the regulation of gene expression. The regulation of gene expression is important because it links the **genotype** of an organism—the specific set of genes it possesses—to the **phenotype** of an organism—the collection of characteristics that are observable. All genes are not expressed at the same time or in the same place,

and the environment profoundly influences which genes are expressed at any given time. Finally, chapter 13 contains information on the nature of mutation, DNA repair, and genetic recombination. These three chapters provide the background needed for understanding the material on recombinant DNA technology (chapter 14) and microbial genomics (chapter 15). Much of the information presented in chapters 11 through 13 will be familiar to those who have taken an introductory genetics course. Because of the importance of bacteria as model organisms, primary emphasis is placed on their genetics. The genetics of the *Archaea* is discussed in chapter 20.

Although modern genetic analysis began with studies of fruit flies and corn, the nature of genetic information, gene structure, the genetic code, and mutation were elucidated by elegant experiments involving bacteria and bacterial viruses. 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”—which have guided much of modern research.



## Historical Highlights

### 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 are large molecules with complex amino acid sequences that 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.

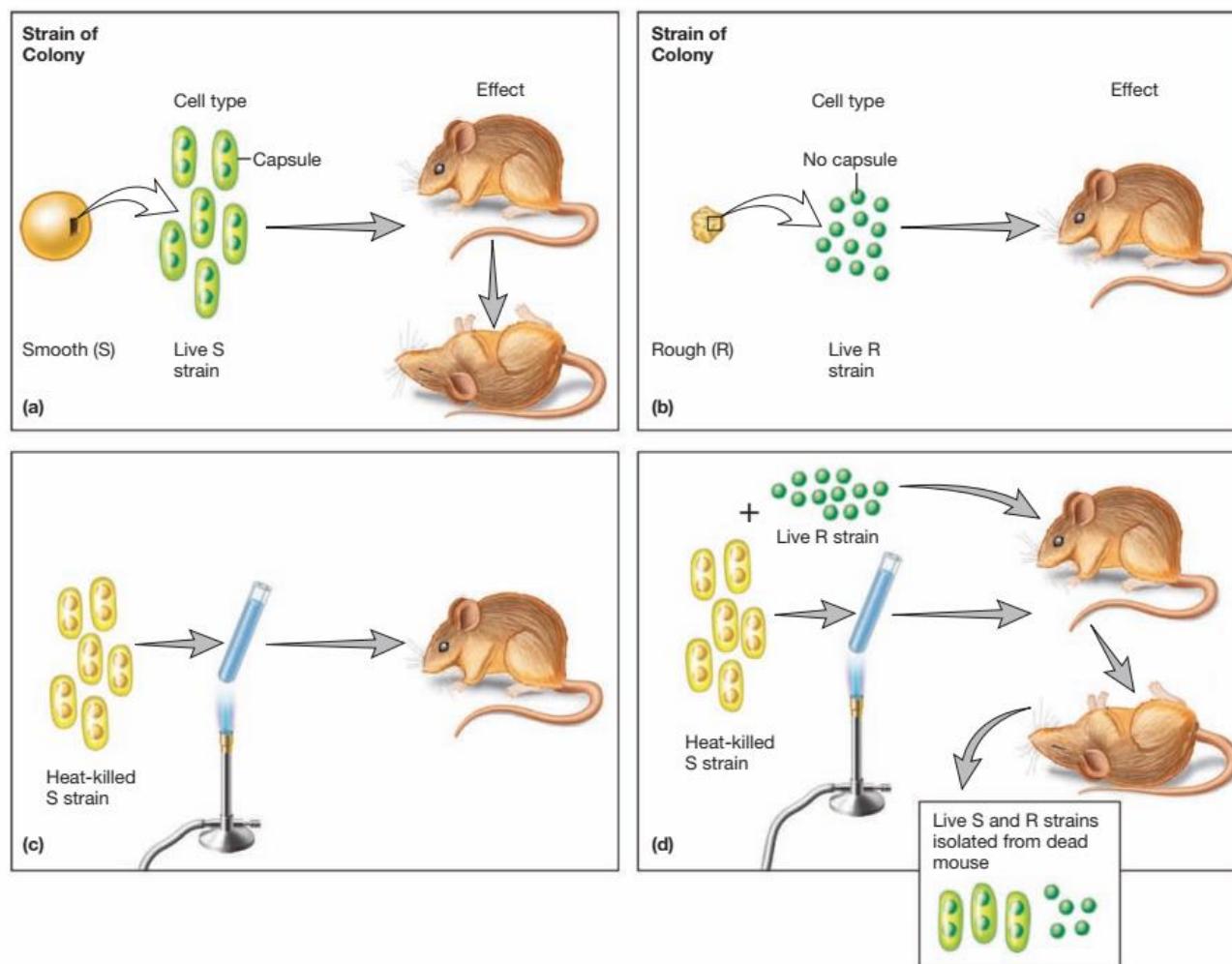
Without actually doing any experiments themselves, Watson and Crick constructed their model by combining Chargaff's rules on base composition with Franklin's X-ray data and their predictions about how genetic material should behave. By building models, they found that a smooth, two-stranded helix of 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. Unfortunately, Franklin could not be considered for the prize because she had died of cancer in 1958 at the age of thirty-seven.

### 11.1 DNA AS GENETIC MATERIAL

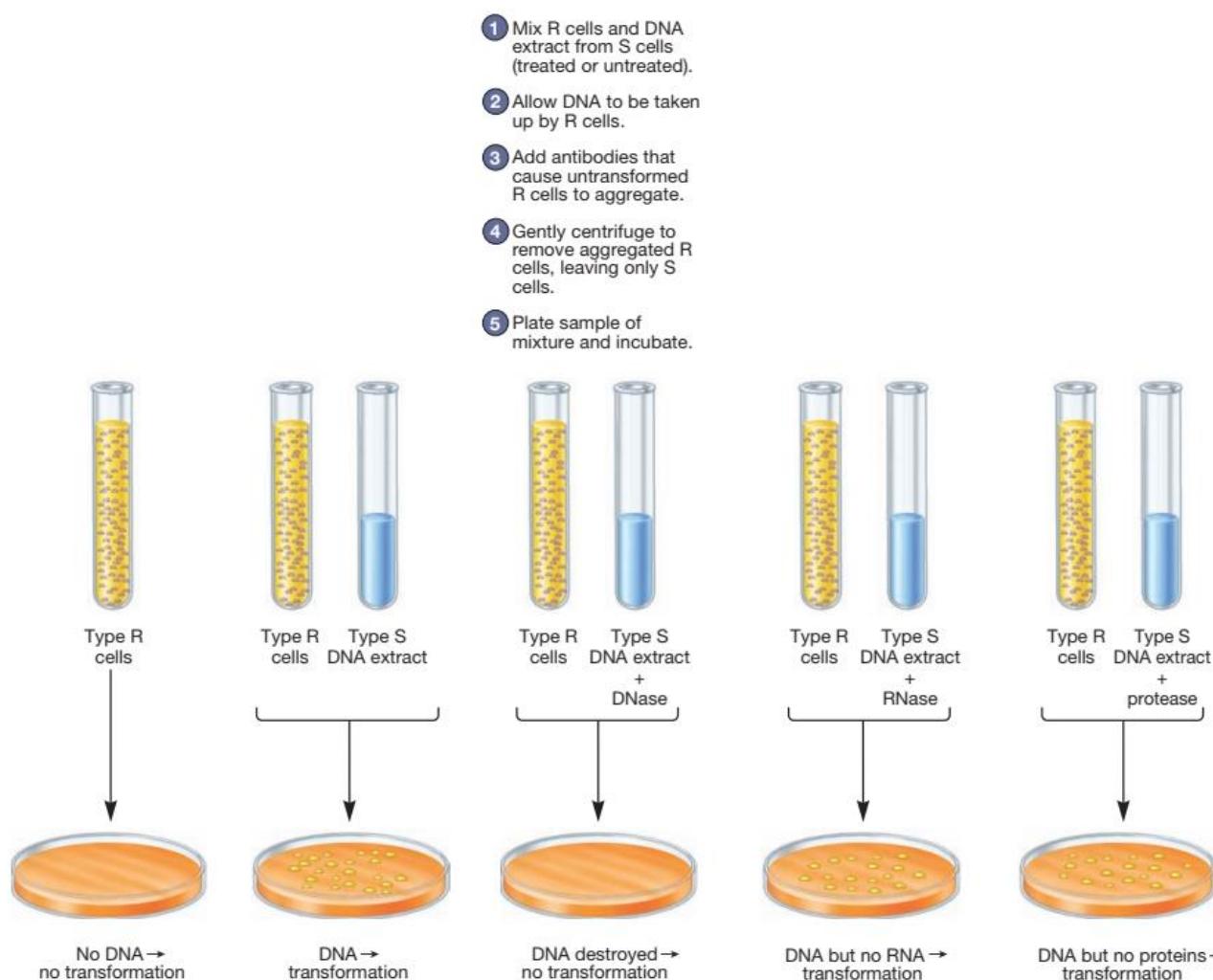
Although it is now hard to imagine, it was once thought that DNA was too simple a molecule to store genetic information (**Historical Highlights 11.1**). The early work of **Fred Griffith** in 1928 on the transfer of virulence in the pathogen *Streptococcus pneumoniae*, commonly called pneumococcus (**figure 11.1**), set the stage for research showing that DNA was indeed 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 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 (S cells), using enzymes that would hydrolyze DNA, RNA, or protein. They then exposed nonvirulent pneumococcal strains (R 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 Avery, C. M. MacLeod, and M. J. McCarty in 1944



**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.

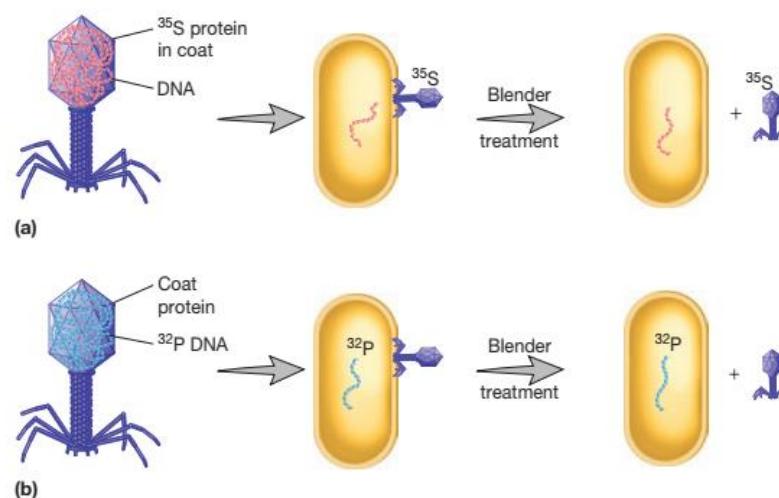


**Figure 11.2 Some Experiments on the Transforming Principle.** Earlier experiments done by Avery, MacLeod, and McCarty had shown that only DNA extracts from S cells caused transformation of R cells to S cells. To demonstrate that contaminating molecules in the DNA extract were not responsible for transformation, the DNA extract from S cells was treated with RNase, DNase, and protease and then mixed with R cells. Time was allowed for the DNA from S cells to be taken up by the R cells and expressed, transforming R cells into S cells. Then, antibodies (immune system proteins that recognize specific structures) that recognized R cells, but not S cells, were added to the mixture. The addition of antibodies caused the R cells (i.e., those R cells that had not been transformed) to aggregate. These aggregated R cells were removed from the mixture by gentle centrifugation. Thus, the only cells remaining in the mixture were cells that had been transformed and were now S cells. Only treatment of the DNA extract from S cells with DNase destroyed the ability of the extract to transform the R cells.

provided the first evidence that Griffith's transforming principle was DNA and therefore that DNA carried genetic information.

Some years later (1952), **Alfred Hershey** and **Martha Chase** performed several experiments indicating that DNA was the genetic material in a bacterial virus called 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 informa-

tion might have lasted considerably longer than it did. Hershey and Chase made the virus's DNA radioactive with  $^{32}\text{P}$  or they labeled its protein coat with  $^{35}\text{S}$ . They mixed radioactive bacteriophage with *Escherichia coli* and incubated the mixture for a few minutes. The suspension was then agitated violently in a blender to shear off any adsorbed bacteriophage particles (**figure 11.3**). After centrifugation, radioactivity in the supernatant (where the virus remained) versus the bacterial cells in the pellet was determined. They found that most radioactive protein was released into the supernatant,



**Figure 11.3 The Hershey-Chase Experiment.** (a) When *E. coli* was infected with a T2 phage containing  $^{35}\text{S}$  protein, most of the radioactivity remained outside the host cell. (b) When a T2 phage containing  $^{32}\text{P}$  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.

whereas  $^{32}\text{P}$  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. [Virulent double-stranded DNA phages \(section 17.2\)](#)

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

1. Define 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?

## 11.2 THE FLOW OF GENETIC INFORMATION

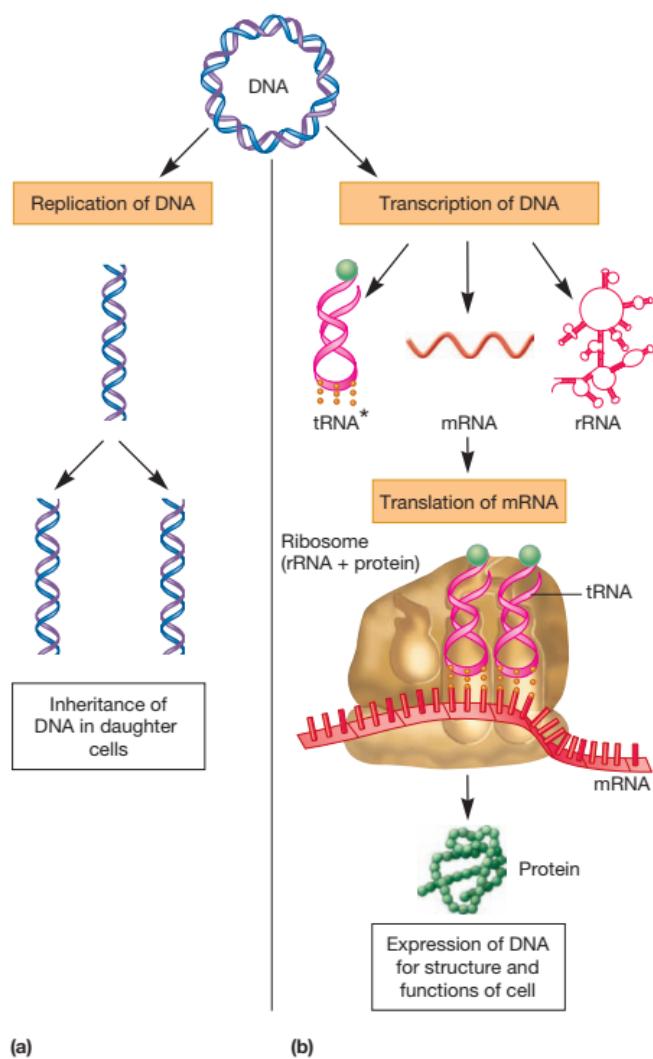
Biologists have long recognized a relationship among DNA, RNA, and protein, and this recognition has guided a vast amount of research over the past decades. The pathway from DNA to RNA and RNA to protein is conserved in all forms of life and is often called the **central dogma**. **Figure 11.4** illustrates two essential concepts: the flow of genetic information from one generation to the next (replication); and the flow of information within a single cell, a process also called gene expression.

The transmission of genetic information from one generation to the next is shown in figure 11.4a. DNA functions as a storage molecule, holding genetic information for the lifetime of a cellular organism, and allowing that information to be duplicated and

passed on to its progeny. Synthesis of the duplicate DNA is directed by the parental molecule and is called **replication**. The process is catalyzed by DNA polymerase enzymes.

The genetic information stored in DNA is divided into units called **genes**. In order for an organism to function properly and reproduce, its genes must be expressed at the appropriate time and place. Gene expression begins with the synthesis of an RNA copy of the gene. This process of DNA-directed RNA synthesis is called **transcription** because the DNA base sequence is being written into an RNA base sequence. RNA polymerase enzymes catalyze transcription. Although DNA has two complementary strands, only one strand, the template strand, of a particular gene is transcribed. If both strands of a single gene were transcribed, two different RNA molecules would result and cause genetic confusion. However, *different* genes may be encoded on opposite strands, thus both strands of DNA can serve as templates for RNA synthesis depending on the orientation of the gene on the DNA. Transcription yields three different types of RNA depending on the gene being transcribed. These are messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA) (figure 11.4b).

During the last phase of gene expression, **translation**, genetic information in the form of an RNA base sequence in a **messenger RNA (mRNA)** is decoded and used to govern the synthesis of a polypeptide. 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 complementary to a portion of the DNA genome. In addition to mRNA, translation also requires the activities of transfer RNA and ribosomal RNA. Thus all three types of RNA are involved in the production of protein, based on the code present in the DNA.



\*The sizes of RNA are not to scale—tRNA and mRNA are enlarged to show details.

**Figure 11.4 Summary of the Flow of Genetic Information in Cells.** DNA serves as the storehouse for genetic information. **(a)** During cellular reproduction, DNA is replicated and passed to progeny cells. **(b)** In order to function properly, a cell must express the genetic information stored in DNA. This is accomplished when the genetic code is transcribed into mRNA molecules. The information in the mRNA is then translated into protein.

1. Describe the general relationship between DNA, RNA, and protein.
2. What are the products of replication, transcription, and translation?
3. Until relatively recently, the “one gene-one protein” hypothesis was used to define the role of genes in organisms. Refer to figure 11.4 and explain why this description of a gene no longer applies.

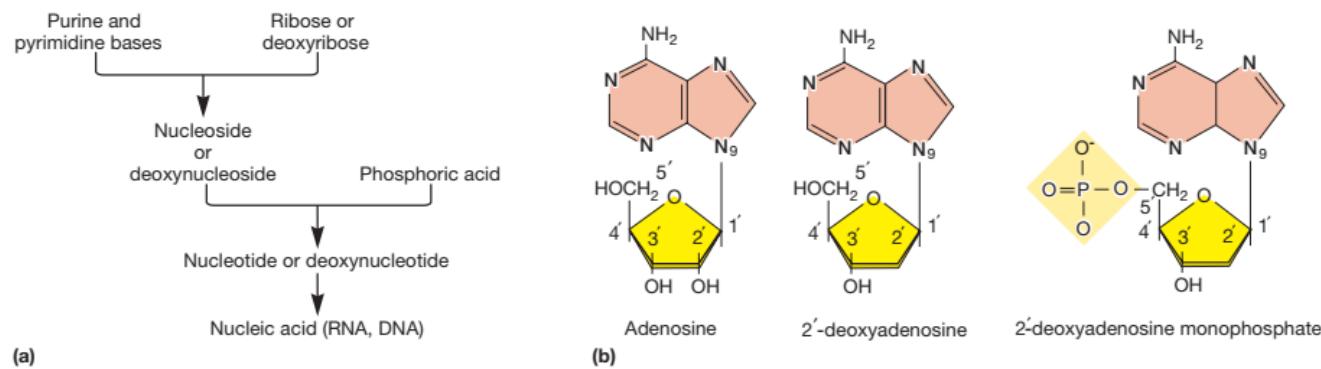
### 11.3 NUCLEIC ACID STRUCTURE

The nucleic acids, DNA and RNA, are polymers of nucleotides (figure 11.5) linked together by phosphodiester bonds (figure 11.6a). However, DNA and RNA differ in terms of the nitrogenous bases they contain, the sugar component of their nucleotides, and whether they are double or single stranded. **Deoxyribonucleic acid (DNA)** contains the bases adenine, guanine, cytosine, and thymine. The sugar found in the nucleotides is deoxyribose, and DNA molecules are usually double stranded. **Ribonucleic acid (RNA)**, on the other hand, contains the bases adenine, guanine, cytosine, and uracil (instead of thymine, although tRNA contains a modified form of thymine). Its sugar is ribose, and most RNA molecules are single stranded. The structure of DNA and RNA is described in more detail next.

#### DNA Structure

The discovery that DNA is the genetic material set into motion a fierce competition to determine the precise structure of DNA (Historical Highlights 11.1). DNA molecules are very large and are usually composed of two polynucleotide chains coiled together to form a double helix 2.0 nm in diameter (figure 11.6 and figure 11.7). Each chain contains purine and pyrimidine deoxyribonucleosides joined by **phosphodiester linkages** (figure 11.6a). That is, a phosphoric acid molecule forms a bridge between a 3'-hydroxyl of one sugar and a 5'-hydroxyl of an adjacent sugar. 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) of one strand is always paired with the pyrimidine thymine (T) of the opposite strand by two hydrogen bonds. The purine guanine (G) pairs with cytosine (C) by three hydrogen bonds. This AT and GC base pairing means that the two strands in a DNA double helix are **complementary**. In other words, the bases in one strand match up with those of the other according to specific 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 organisms, including a variety of microbes. [Microbial genomics \(chapter 15\)](#)

The two polynucleotide strands fit together much like the pieces in a jigsaw puzzle because of complementary base pairing. Inspection of figure 11.6b,c, 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**, which means they run in opposite directions with respect to the orientation of their sugars. One end of each strand has



**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—adenosine and 2'-deoxyadenosine—and the nucleotide 2'-deoxyadenosine monophosphate. The carbons of nucleoside and nucleotide sugars are indicated by numbers with primes.

an exposed 5'-hydroxyl group, often with phosphates attached, whereas the other end has a free 3'-hydroxyl group (figure 11.6a). If one 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).

### RNA Structure

RNA differs chemically from DNA, and is usually single stranded rather than double stranded. However, an RNA strand can coil back on itself to form a hairpin-shaped structure with complementary base pairing and helical organization. The three different types of RNA—messenger RNA, ribosomal RNA, and transfer RNA—differ from one another in function, site of synthesis in eucaryotic cells, and structure.

### The Organization of DNA in Cells

Although DNA exists as a double helix in all cells, its organization differs among cells in the three domains of life. DNA is organized in the form of a closed circle in all *Archaea* and most bacteria. This circular double helix is further twisted into supercoiled DNA (figure 11.8). In *Bacteria*, DNA is associated with basic proteins that appear to help organize it into a coiled, chromatinlike structure.

DNA is much more highly organized in eucaryotic chromatin 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 11.9a). DNA coils around the surface of the ellipsoid approximately 1 3/4 turns or 166 base pairs before proceeding on to the next. This combination of histones plus DNA, or nucleo-

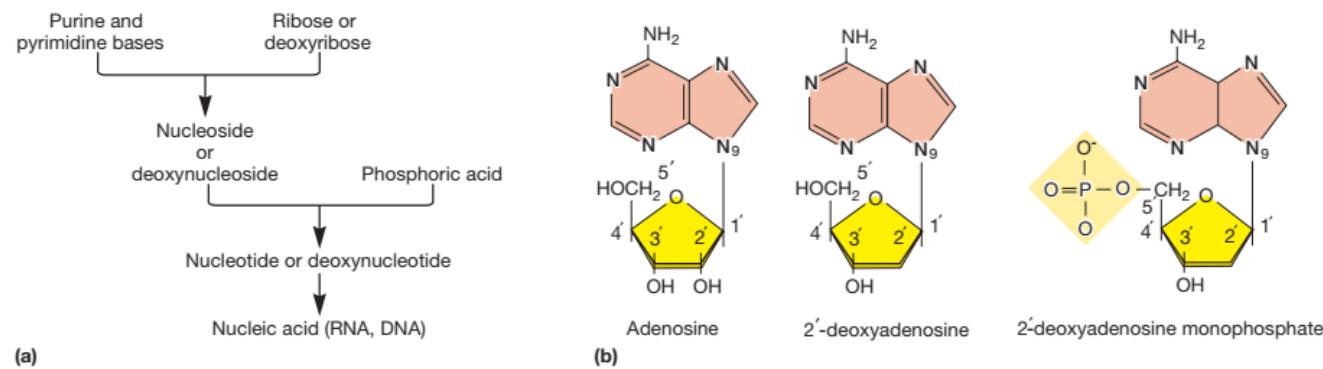
protein complex, is called a **nucleosome**. DNA gently isolated from chromatin looks like a string of beads. The stretch of DNA between the beads or nucleosomes, the linker region, varies in length from 14 to over 100 base pairs. Histone H1 associates with the linker regions to aid the folding of DNA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

Although the *Archaea* share the prokaryotic style of cellular organization with the *Bacteria*, there are some important differences. Thus far, all archaeal genomes examined are circular. In many archaea, the DNA is complexed with histone proteins. Like eucaryotic histones, archaeal histones form nucleoprotein complexes called **archaeal nucleosomes**. In archaea, tetramers of histones (that is, four histone proteins) interact with about 60 base pairs each, protecting the DNA from DNA-digesting nucleases. The structure of archaeal histones and their interaction with DNA strongly suggests that archaeal nucleosomes are evolutionarily similar to the eucaryotic nucleosomes formed by DNA and histones H3 and H4. [Microbial evolution \(section 19.1\)](#)

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 *Bacteria*, *Archaea*, and eucaryotes.

### 11.4 DNA REPLICATION

DNA replication is an extraordinarily important and complex process upon which all life depends. At least 30 proteins are required to replicate the *E. coli* chromosome (table 11.1). Presumably, much of the complexity is necessary for accuracy in copying



**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—adenosine and 2'-deoxyadenosine—and the nucleotide 2'-deoxyadenosine monophosphate. The carbons of nucleoside and nucleotide sugars are indicated by numbers with primes.

an exposed 5'-hydroxyl group, often with phosphates attached, whereas the other end has a free 3'-hydroxyl group (figure 11.6a). If one 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).

### RNA Structure

RNA differs chemically from DNA, and is usually single stranded rather than double stranded. However, an RNA strand can coil back on itself to form a hairpin-shaped structure with complementary base pairing and helical organization. The three different types of RNA—messenger RNA, ribosomal RNA, and transfer RNA—differ from one another in function, site of synthesis in eucaryotic cells, and structure.

### The Organization of DNA in Cells

Although DNA exists as a double helix in all cells, its organization differs among cells in the three domains of life. DNA is organized in the form of a closed circle in all *Archaea* and most bacteria. This circular double helix is further twisted into supercoiled DNA (figure 11.8). In *Bacteria*, DNA is associated with basic proteins that appear to help organize it into a coiled, chromatinlike structure.

DNA is much more highly organized in eucaryotic chromatin 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 11.9a). DNA coils around the surface of the ellipsoid approximately 1 3/4 turns or 166 base pairs before proceeding on to the next. This combination of histones plus DNA, or nucleo-

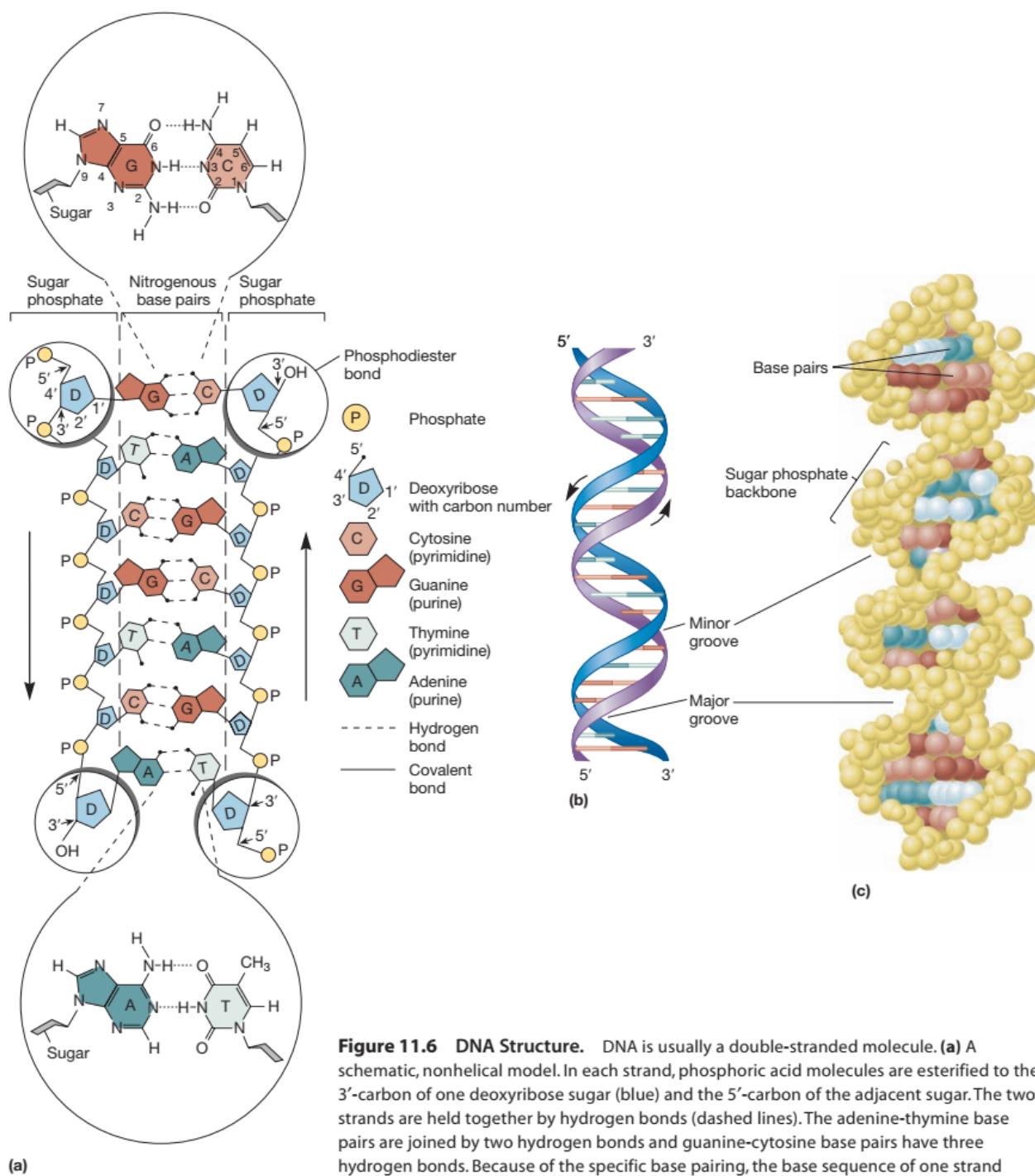
protein complex, is called a **nucleosome**. DNA gently isolated from chromatin looks like a string of beads. The stretch of DNA between the beads or nucleosomes, the linker region, varies in length from 14 to over 100 base pairs. Histone H1 associates with the linker regions to aid the folding of DNA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

Although the *Archaea* share the prokaryotic style of cellular organization with the *Bacteria*, there are some important differences. Thus far, all archaeal genomes examined are circular. In many archaea, the DNA is complexed with histone proteins. Like eucaryotic histones, archaeal histones form nucleoprotein complexes called **archaeal nucleosomes**. In archaea, tetramers of histones (that is, four histone proteins) interact with about 60 base pairs each, protecting the DNA from DNA-digesting nucleases. The structure of archaeal histones and their interaction with DNA strongly suggests that archaeal nucleosomes are evolutionarily similar to the eucaryotic nucleosomes formed by DNA and histones H3 and H4. [Microbial evolution \(section 19.1\)](#)

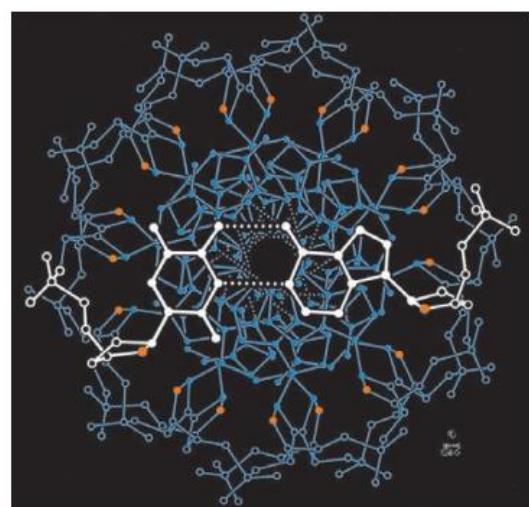
1. What are nucleic acids? How do DNA and RNA differ in structure?
2. Describe in some detail the structure of the DNA double helix. What does it mean to say that the two strands are complementary and antiparallel?
3. What are histones and nucleosomes? Describe the way in which DNA is organized in the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

### 11.4 DNA REPLICATION

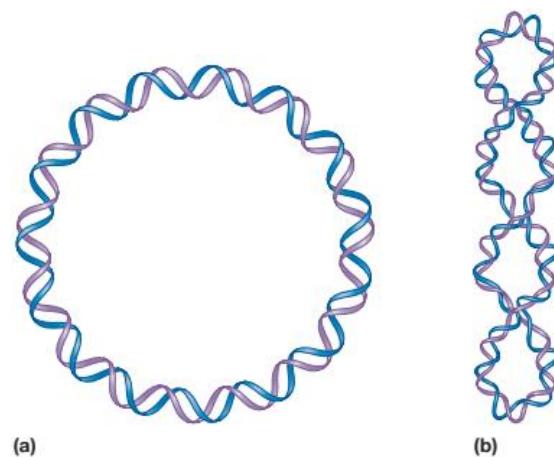
DNA replication is an extraordinarily important and complex process upon which all life depends. At least 30 proteins are required to replicate the *E. coli* chromosome (table 11.1). Presumably, much of the complexity is necessary for accuracy in copying



**Figure 11.6 DNA Structure.** DNA is usually a double-stranded molecule. **(a)** A schematic, nonhelical model. In each strand, phosphoric acid molecules are esterified to the 3'-carbon of one deoxyribose sugar (blue) and the 5'-carbon of the adjacent sugar. The two strands are held together by hydrogen bonds (dashed lines). The adenine-thymine base pairs are joined by two hydrogen bonds and guanine-cytosine base pairs have three hydrogen bonds. Because of the specific base pairing, the base sequence of one strand determines the sequence of the other. The two strands are antiparallel—that is, the backbones run in opposite directions as indicated by the two arrows, which point in the 5' to 3' direction. **(b)** Simplified model that highlights the antiparallel arrangement and the major and minor grooves. **(c)** Space-filling model of the B form of DNA. Note that the sugar-phosphate backbone spirals around the outside of the helix and the base pairs are embedded inside.



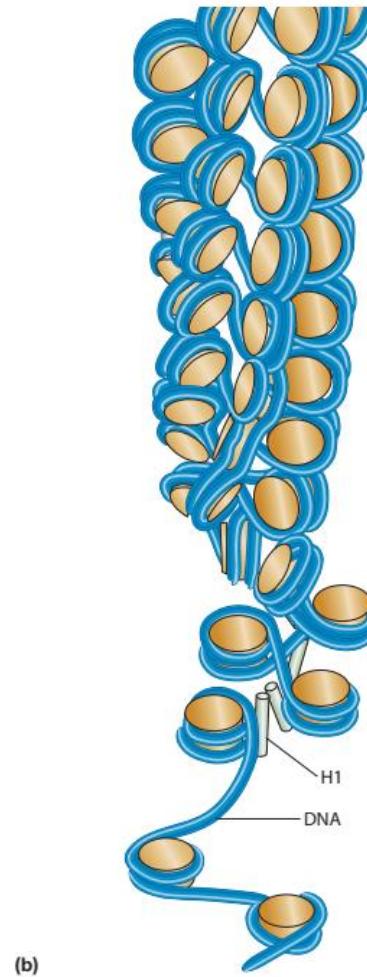
**Figure 11.7 Structure of the DNA Double Helix.** End view of a double helix showing the outer backbone and the bases stacked in the center of the cylinder. The ribose ring oxygens are red. The nearest base pair, an AT base pair, is highlighted in white.



**Figure 11.8 DNA Forms.** (a) The DNA double helix of most prokaryotes 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.



**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.



**Table 11.1 Components of the *E. coli* Replication Machinery**

Protein	Function
DnaA protein	Initiation of replication; binds origin of replication ( <i>oriC</i> )
DnaB protein	Helicase (5'→3'); breaks hydrogen bonds holding two strands of double helix together; promotes DNA primase activity; involved in primosome assembly
DNA gyrase	Relieves supercoiling of DNA produced as DNA strands are separated by helicases; separates daughter molecules in final stages of replication
SSB proteins	Bind single-stranded DNA after strands are separated by helicases
DnaC protein	Helicase loader; helps direct DnaB protein (helicase) to DNA template
n' protein	Component of primosome; helicase (3'→5')
n protein	Primosome assembly; component of primosome
n'' protein	Primosome assembly
I protein	Primosome assembly
DNA primase	Synthesis of RNA primer; component of primosome
DNA polymerase III holoenzyme	Complex of about 20 polypeptides; catalyzes most of the DNA synthesis that occurs during DNA replication; has 3'→5' exonuclease (proofreading) activity
DNA polymerase I	Removes RNA primers; fills gaps in DNA formed by removal of RNA primer
Ribonuclease H	Removes RNA primers
DNA ligase	Seals nicked DNA, joining DNA fragments together
DNA replication terminus site-binding protein	Termination of replication
Topoisomerase IV	Segregation of chromosomes upon completion of DNA replication

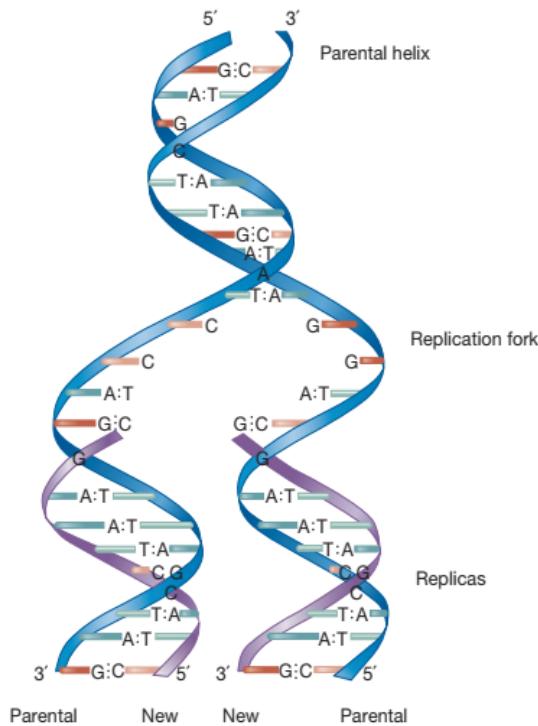
DNA. It would be dangerous for an organism to make many errors during replication because that would certainly be lethal. In fact, *E. coli* makes errors with a frequency of only  $10^{-9}$  or  $10^{-10}$  per base pair replicated (or about one in a million [ $10^{-6}$ ] per gene per generation). Despite its complexity and accuracy, replication is very rapid. In *Bacteria*, replication rates approach 750 to 1,000 base pairs per second. Eucaryotic replication is slower, about 50 to 100 base pairs per second.

During DNA replication, the two strands of the double helix are separated; each then serves as a template for the synthesis of a complementary strand according to the base pairing rules. Each of the two progeny DNA molecules consists of one new strand and one old strand. Thus DNA replication is semiconservative (**figure 11.10**). Watson and Crick suggested **semiconservative replication** of DNA just one month after they published their paper on DNA structure in April 1953; subsequent research confirmed their hypothesis and elucidated the details of replication observed in prokaryotes and eucaryotes.

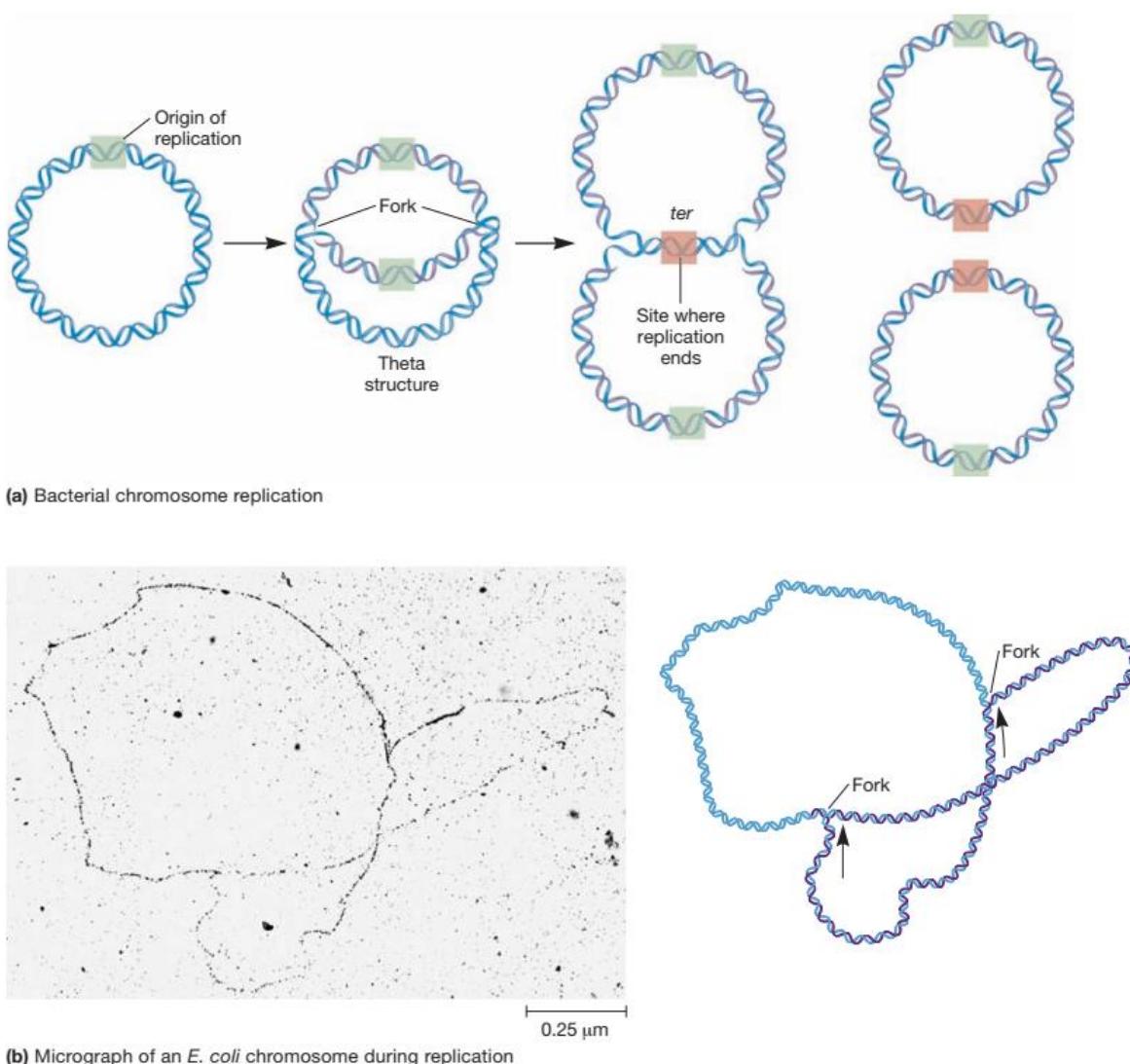
In this section, we first discuss the various patterns of DNA replication observed in cells and viruses. We will then consider the mechanism of DNA replication in *E. coli*, beginning with an examination of the replication machinery and then events at the replication fork.

### Patterns of DNA Synthesis

Replication patterns are somewhat different in *Bacteria*, *Archaea*, 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



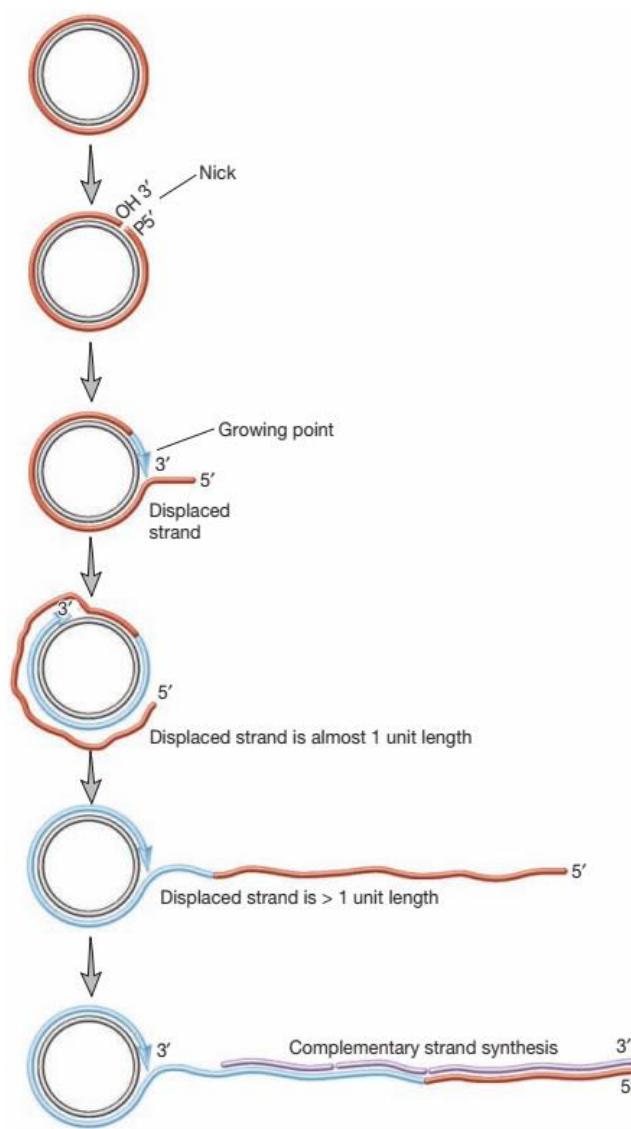
**Figure 11.10 Semiconservative DNA Replication.** The replication fork of DNA showing the synthesis of two progeny strands. Newly synthesized strands are purple. Each copy contains one new and one old strand. This process is called semiconservative replication.



**Figure 11.11 Bidirectional Replication of the *E. coli* Chromosome.** (a) Replication begins at one site on the chromosome, called the origin of replication. Two replication forks proceed in opposite directions from the origin until they meet at a special site called the replication termination site (*ter*). The theta structure is a commonly observed intermediate of the process. (b) An autoradiograph of a replicating *E. coli* chromosome; about one-third of the chromosome has been replicated. To the right is a schematic representation of the chromosome. Parental DNA is blue; new DNA strands are purple, arrow represents direction of fork movement.

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 ( $\theta$ ) 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. Until recently, it was thought that all prokaryotes have a single origin of replication. However, two members of the archaeal genus *Solfolobus* have more than one origin.

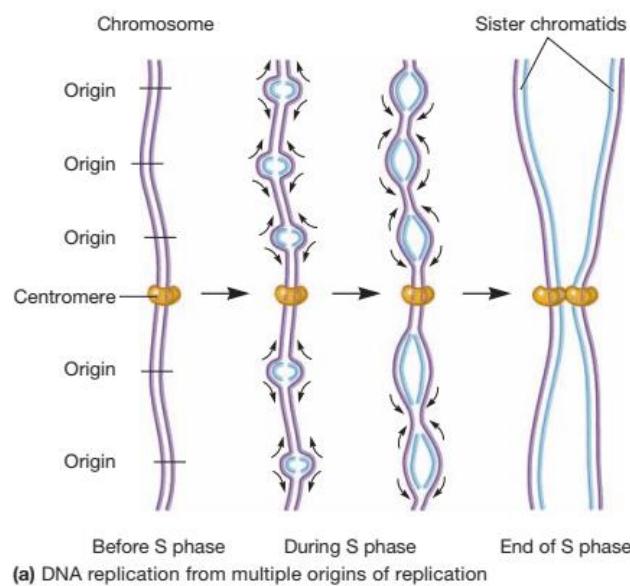
A different pattern of DNA replication occurs during *E. coli* conjugation, a type of genetic exchange mechanism observed in many bacteria. The pattern is called **rolling-circle replication**, and it is also observed during plasmid replication and the reproduction of some viruses (e.g., phage lambda). During rolling-circle replication (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, much like the peel of an apple is displaced by a knife as an apple is pared. The single-stranded tail may be converted to the double-stranded



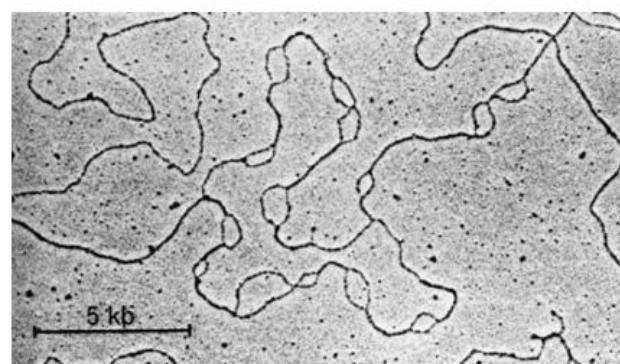
**Figure 11.12 Rolling-Circle Replication.** A single-stranded tail, often composed of more than one genome copy, is generated and can be converted to the double-stranded form by synthesis of a complementary strand. The “free end” of the rolling-circle strand is probably bound to the primosome. OH 3' is the 3'-hydroxyl and P 5' is the 5'-phosphate group created when the DNA strand is nicked.

form by complementary strand synthesis. This mechanism is particularly useful to viruses because it allows the rapid, continuous production of many genome copies from a single initiation event.

The pattern of chromosome replication in eucaryotes differs from that in prokaryotes in part because eucaryotic DNA is much longer than prokaryotic DNA. For instance, *E. coli* DNA is about 1,300  $\mu\text{m}$  in length, whereas the 46 chromosomes in the human nucleus have a total length of 1.8 m (almost 1,400 times longer).



(a) DNA replication from multiple origins of replication

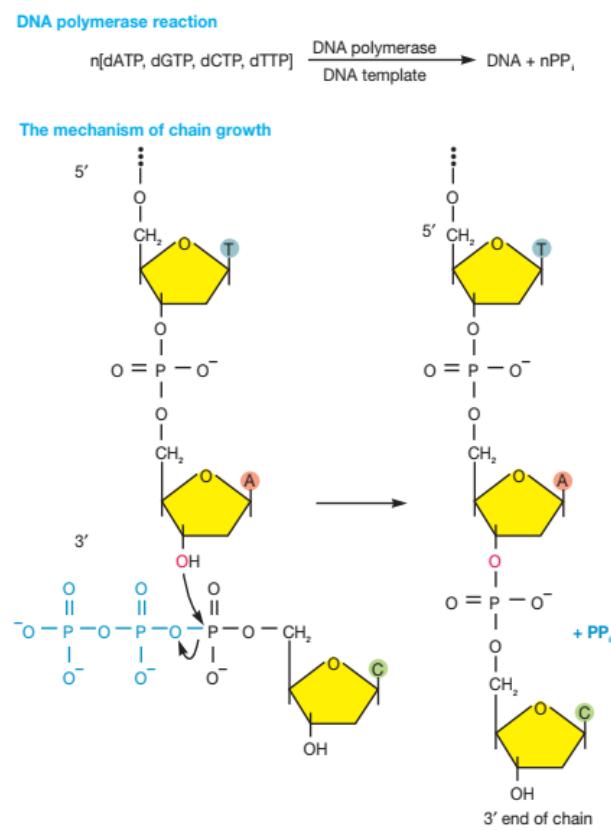


(b) A micrograph of a replicating, eucaryotic chromosome

**Figure 11.13 The Replication of Eucaryotic DNA.** Replication is initiated every 10 to 100  $\mu\text{m}$  and the replication forks travel away from the origin. Newly copied DNA is in blue.

Clearly many replication forks must copy eucaryotic DNA simultaneously so that the molecule can be duplicated in a relatively short period. Therefore many replicons are spaced such that there is an origin about every 10 to 100  $\mu\text{m}$  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.

Another reason for the different pattern of replication in eucaryotes is that their chromosomes are linear. Linear chromosomes present cells with a dilemma: how to replicate the ends of the chromosomes. However, the reason for this dilemma and the mechanisms by which it is resolved can only be grasped by first understanding the mechanisms of DNA replication. Therefore we will consider that first, and then return to the problem of replicating the ends of linear chromosomes.

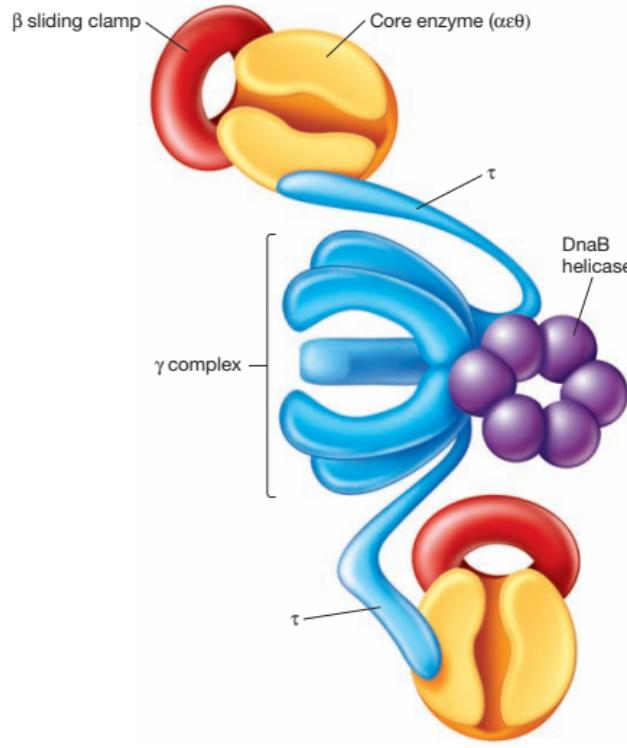


**Figure 11.14 The DNA Polymerase Reaction and Its Mechanism.** The mechanism involves a nucleophilic attack by the hydroxyl of the 3' terminal deoxyribose on the alpha phosphate group of the nucleotide substrate (in this example, adenine attacks cytidine triphosphate).

### The Replication Machinery

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 discussion. The overall process in other bacteria, *Archaea*, and eucaryotes is thought to be similar.

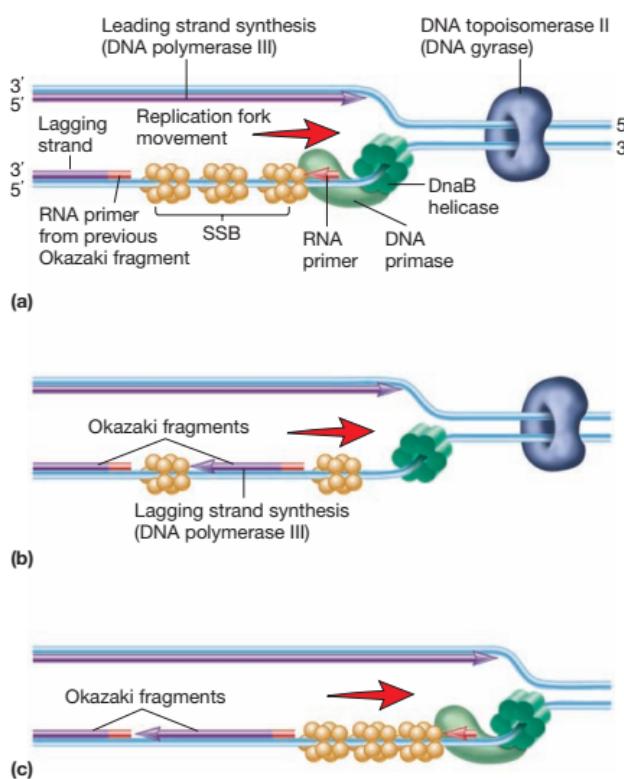
Enzymes called **DNA polymerase** catalyze DNA synthesis. All known DNA polymerase enzymes catalyze the synthesis of DNA in the 5' to 3' direction. This is because the 3'-hydroxyl group of the deoxyribose of the nucleotide at the end of the growing DNA strand attacks the alpha phosphate (the phosphate closest to the 5' carbon) of the deoxynucleoside triphosphate to be incorporated (figure 11.14). This results in the formation of a **phosphodiester bond**; the energy needed to form this covalent bond is provided by the release of the terminal diphosphate (the beta and gamma phosphates) from the nucleotide that is added to the growing chain. Thus deoxynucleoside *triphosphates* (dNTPs: dATP, dTTP, dCTP, dGTP) serve as DNA polymerase substrates while deoxynucleoside *monophosphates* (dNMPs: dAMP, dTMP, dCMP, dGMP) are incorporated into the growing chain.



**Figure 11.15 DNA Polymerase III Holoenzyme.** The holoenzyme consists of two core enzymes (three subunits each;  $\alpha$ ,  $\epsilon$ ,  $\theta$ , not shown) and several other subunits. The two tau ( $\tau$ ) subunits connect the two core enzymes to a large complex called the gamma ( $\gamma$ ) complex. Each core enzyme is associated with a  $\beta$  sliding clamp, which tethers a DNA template to each core.

In order for DNA polymerases to catalyze the synthesis of a complementary strand of DNA, three things are needed: (1) a template, read in the 3' to 5' direction, that directs the synthesis of a complementary DNA strand; (2) a primer (e.g., an RNA strand or a DNA strand) to provide a free 3'-hydroxyl group to which nucleotides can be added (figure 11.14); and (3) dNTPs. *E. coli* has five different DNA polymerase enzymes (DNA Pol I-V). DNA polymerase III plays the major role in replication, although it is assisted by DNA polymerase I.

**DNA polymerase III holoenzyme** is a complex of 10 proteins including two core enzymes, each composed of three protein subunits (figure 11.15). As we shall see, the core enzymes are responsible for catalyzing DNA synthesis and proofreading the product to ensure fidelity of replication. A dimer of another subunit (tau) connects the two core enzymes. Associated with each core enzyme is a subunit called the  $\beta$  sliding clamp. This protein tethers the core enzyme to one strand of the DNA molecule. Another complex of proteins, called the  $\gamma$  complex, is responsible for loading the  $\beta$  sliding clamp onto the DNA. Because there are two core enzymes, both strands of DNA are bound by a single DNA polymerase III holoenzyme.



**Figure 11.16 Bacterial DNA Replication.** A general diagram of DNA replication in *E. coli*. A single replication fork showing both leading strand and lagging strand synthesis is illustrated. The lagging strand is synthesized in short fragments called Okazaki fragments. A new primer is required for the synthesis of each Okazaki fragment.

In *E. coli*, replication begins when a collection of **DnaA proteins** binds to specific nucleotide sequences (DnaA boxes) within the origin of replication. The DnaA proteins hydrolyze ATP to break or “melt” the hydrogen bonds between the DNA strands, thus making this localized region single stranded. Although this provides the initial template for replication, DNA polymerase III cannot by itself unwind and maintain the single-stranded DNA. These activities are provided by the action of other proteins, many of which are found in the **replisome**, a huge complex of proteins that includes DNA polymerase III holoenzyme. These other proteins include helicases, single-stranded DNA binding proteins, and topoisomerases (**figure 11.16**). **Helicases** are responsible for separating (unwinding) the DNA strands. These enzymes also use energy from ATP to unwind short stretches of helix just ahead of the replication fork. **Single-stranded DNA binding proteins (SSBs)** keep the strands apart once they have been separated, and **topoisomerases** relieve the tension generated by the rapid unwinding of the double helix (the replication fork may rotate as rapidly as 75 to 100 revolutions per second). This is important because rapid unwinding can lead to the 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), and these can impede replication if not removed. Topoisomerases change the structure of DNA by transiently breaking one or two strands in such a way that the nucleotide sequence of the DNA 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 important topoisomerase in *E. coli*.

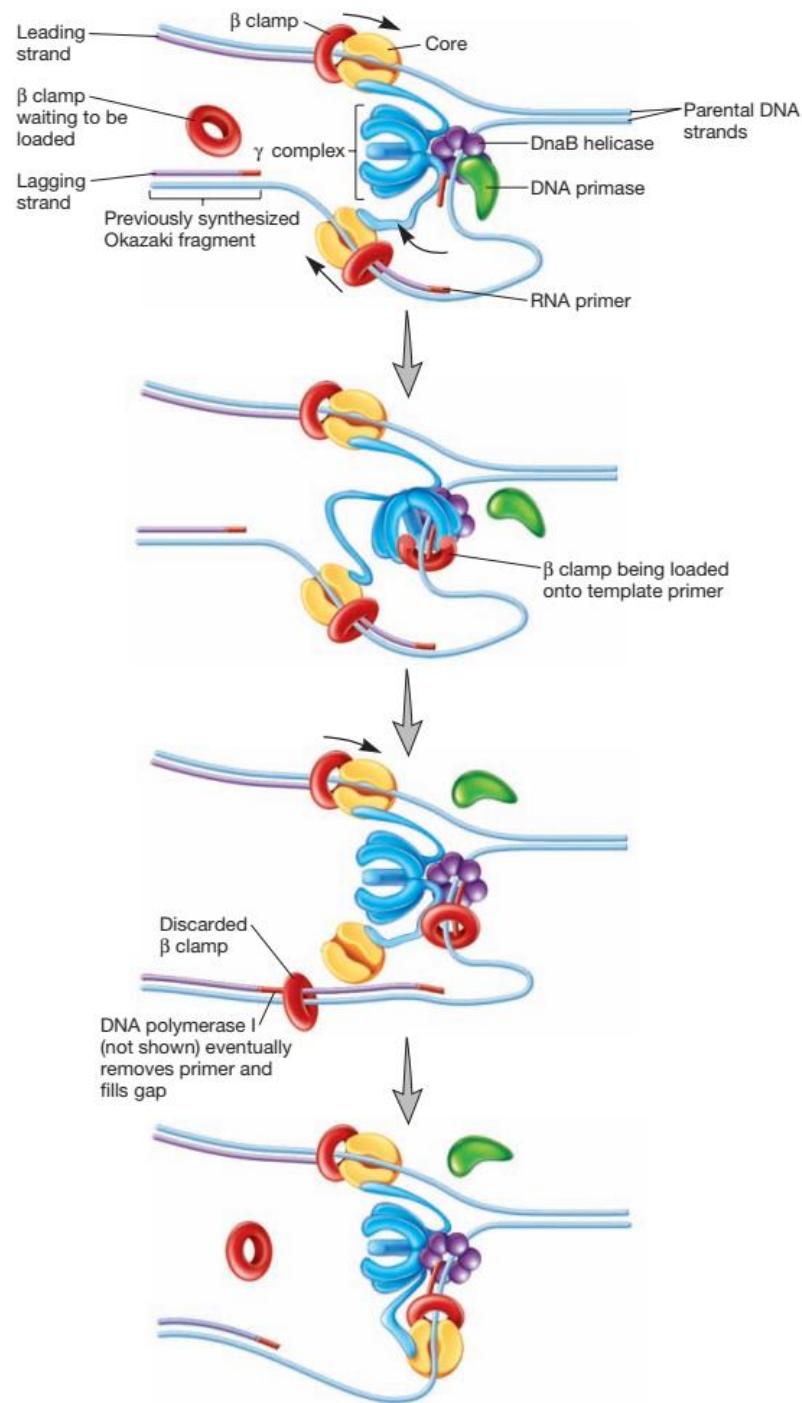
Once the template is prepared, the primer needed by DNA polymerase III can be synthesized. A special polymerase called **primase** synthesizes a short RNA strand, usually around 10 nucleotides long and complementary to the DNA; this serves as the primer (figure 11.16). RNA is used as the primer because unlike DNA polymerase, RNA polymerases (such as primase) can initiate RNA synthesis without adding a nucleotide to an existing 3'-OH. It appears that the primase requires the assistance of several other proteins, and the complex of the primase and its accessory proteins is called the **primosome** (table 11.1). The primosome is another important component of the replisome.

Because DNA polymerase enzymes must synthesize DNA in the 5' to 3' direction, only one of the strands, called the **leading strand**, can be synthesized continuously at its 3' end as the DNA unwinds (figure 11.16). The other strand, called the **lagging strand**, cannot be extended in the same direction because there is no free 3'-OH to which a nucleotide can be added. As a result, the lagging strand is synthesized discontinuously in the 5' to 3' direction as a series of fragments, called **Okazaki fragments** after their discoverer, Reiji Okazaki. Discontinuous synthesis occurs as primase adds many RNA primers along the single-stranded lagging strand. DNA polymerase III then extends these primers with DNA to form short fragments. These fragments are finally joined to form a complete strand; the steps of this process are detailed next. Thus while the leading strand requires only one RNA primer (and only one primosome) to initiate synthesis, the lagging strand has many RNA primers (and primosomes) that must eventually be removed. Okazaki fragments are about 1,000 to 2,000 nucleotides long in *Bacteria* and approximately 100 nucleotides long in eukaryotic cells.

### Events at the Replication Fork

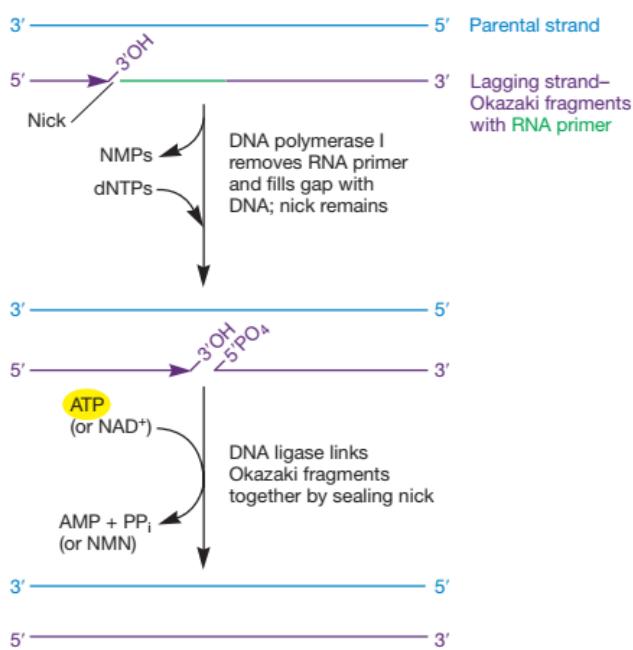
The details of DNA replication are outlined in a diagram of the replication fork (**figure 11.17**). In *E. coli*, DNA replication is initiated at specific nucleotides called the *oriC* locus (for *origin of chromosomal replication*). Here we present replication as a series of discrete steps, but it should be remembered that synthesis is extremely rapid and occurs simultaneously on both the leading and lagging strands.

1. To initiate replication, as many as 40 DnaA proteins bind *oriC* while hydrolyzing ATP. Binding of the DnaA proteins causes the DNA to bend around the protein complex, resulting in separation of the double-stranded DNA at regions within the origin that have many A-T base pairs. Recall that adenines pair with thymines using only two hydrogen bonds, so A-T rich segments of DNA become single stranded more readily than do G-C rich regions. Once the

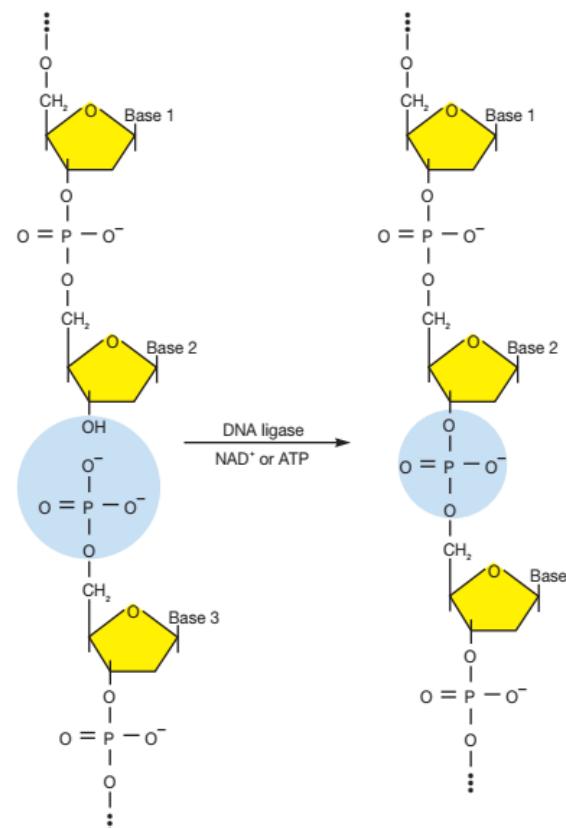


**Figure 11.17 A Model for Activity at the Replication Fork.** DNA polymerase III holoenzyme and other components of the replisome are responsible for the synthesis of both leading and lagging strands. The arrows show the movement of each DNA core polymerase. After completion of each new Okazaki fragment, the old  $\beta$  sliding clamp is discarded and a new one loaded onto the template DNA (step 3). This is achieved by the activity of the  $\gamma$  complex (see figure 11.15), which is also known as the clamp loader. The Okazaki fragments are eventually joined together (see figure 11.18) after removal of the RNA primer and synthesis of DNA to fill the gap, both catalyzed by DNA polymerase I; DNA ligase then seals the nick and joins the two fragments (see figure 11.19).

- strands have separated, the replication process then proceeds through four stages.
2. Helicases unwind the helix with the aid of topoisomerases like DNA gyrase (figure 11.17, step 1). It appears that the DnaB protein is the helicase most actively involved in replication, but the n' protein also may participate in unwinding. The single strands are kept separate by SSBs.
  3. Primase synthesizes RNA primers as needed (figure 11.17, step 1) and a single DNA polymerase III holoenzyme catalyzes both leading strand and lagging strand synthesis from the RNA primers. Lagging strand synthesis is particularly amazing because of the “gymnastic” feats performed by the replisome. It must discard old  $\beta$  sliding clamps (figure 11.17, step 3), load new  $\beta$  sliding clamps (figure 11.17, step 2), and tether the template to the core enzyme with each new round of Okazaki fragment synthesis. All of this occurs as DNA polymerase III is synthesizing DNA. Thus DNA polymerase III is a multifunctional enzyme.
  4. After most of the single-stranded region of the lagging strand has been replicated by the formation of Okazaki fragments, **DNA polymerase I** or (more rarely) RNaseH removes the RNA primer. DNA polymerase I can do this because, unlike other DNA polymerases, it has 5' to 3' exonuclease activity—that is, it can snip off nucleotides one at a time starting at the 5' end. Thus DNA polymerase I begins its exonuclease activity at the free end of the RNA primer. With the removal of each ribonucleotide, the adjacent 3'-OH from the deoxynucleotide is used by DNA polymerase I to fill the gap between Okazaki fragments (**figure 11.18**).



**Figure 11.18** Completion of Lagging Strand Synthesis.



**Figure 11.19** The DNA Ligase Reaction. The groups being altered are shaded in blue. Bacterial ligases use the pyrophosphate bond of  $\text{NAD}^+$  as an energy source; many other ligases employ ATP.

5. Finally, the Okazaki 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 11.19**).

As we have seen, DNA polymerase III is an amazing multi-protein complex, with multiple enzymatic activities. In *E. coli*, the polymerase component is encoded by the *dnaE* gene. Genome sequencing of other bacteria has revealed that some have a second *dnaE* gene. In *Bacillus subtilis*, a gram-positive bacterium that is another important experimental model, this second polymerase gene is called *dnaE<sub>Bs</sub>*, and its protein product appears to be responsible for replicating the lagging strand. Thus while the overall mechanism by which DNA is replicated is highly conserved, there can be variations in replisome components.

Amazingly, DNA polymerase III, like all DNA polymerases, has an additional function that is critically important: **proofreading**. Proofreading is the removal of a mismatched base immediately after it has been added; its removal must occur before the next base is incorporated. Recall that the polymerase III core is

composed of three subunits:  $\alpha$ ,  $\epsilon$ , and  $\theta$ . While we have discussed the  $\alpha$  subunit polymerase activity, the  $\epsilon$  subunit has 3' to 5' exonuclease activity. This enables it to check each newly incorporated base to see that it forms stable hydrogen bonds. In this way mismatched bases can be detected. If the wrong base has been mistakenly added, this subunit is able to remove it. Because it has exonuclease activity (exo meaning outside or in this case, from the end), it can remove a mismatched base, as long as it is still at the 3' end of the growing strand. Once removed, holoenzyme backs up and adds the proper nucleotide in its place. DNA proofreading is not 100% efficient and, as discussed in chapter 12, the mismatch repair system is the cell's second line of defense.

### Termination of Replication

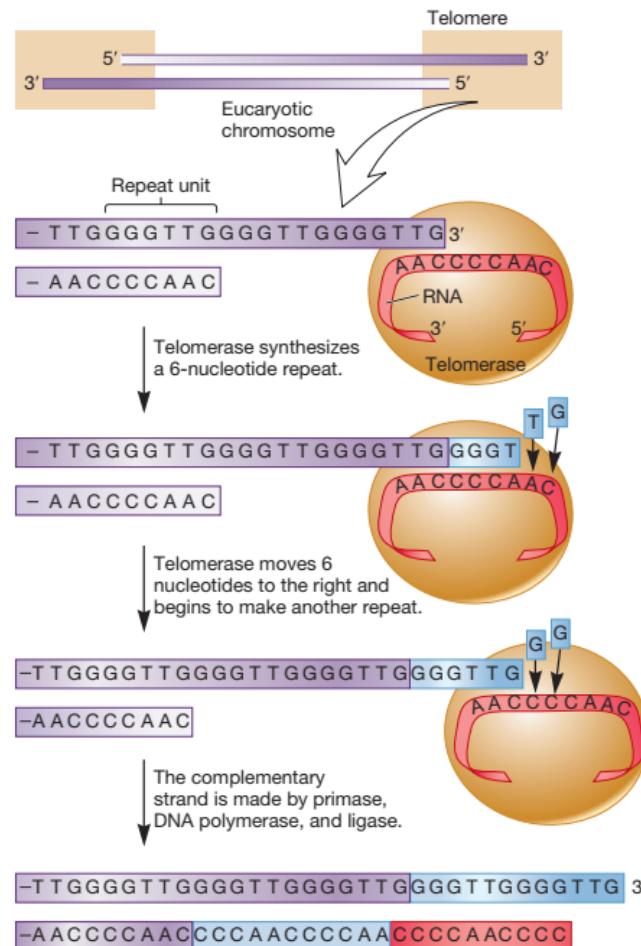
In *E. coli*, DNA replication stops when the replisome reaches a termination site (*ter*) on the DNA. The **Tus protein** binds to the *ter* sites and halts progression of the forks. In many other bacteria, replication stops randomly when the forks meet. Regardless of how fork movement is stopped, there is often a problem to be solved by the replisome: separation of daughter molecules. When replication of a circular chromosome is complete, the two circular daughter chromosomes may remain intertwined. Such interlocked chromosomes are called **catenanes**. This is obviously a problem if each daughter cell is to inherit a single chromosome. Fortunately, topoisomerases solve the problem by temporarily breaking DNA molecules, so that the strands can be separated.

### Replication of Linear Chromosomes

The fact that eucaryotic chromosomes are linear poses a problem during replication because of DNA polymerase's need for a primer, providing a free 3'-OH. At the ends (telomeres) of eucaryotic chromosomes, space is not available for synthesis of a primer on the lagging strand, and therefore it should be impossible to replicate the end of that strand. Over numerous rounds of DNA replication and cell division, this would lead to a progressively shortened chromosome. Ultimately the chromosome would lose critical genetic information, which would be lethal to the cell.

Clearly, eucaryotic cells must have evolved a mechanism for replicating their telomeres. The solution to the "end replication problem" is the enzyme **telomerase**. Telomerase has two components: a protein that can synthesize DNA using an RNA template (telomerase reverse transcriptase) and an internal RNA template. The internal RNA is complementary to the single strand of DNA jutting out from the end of the chromosome (figure 11.20) and acts as the template for DNA synthesis to elongate that strand (i.e., the 3'-OH of the telomere DNA strand serves as the primer for DNA synthesis). After being lengthened sufficiently, the single strand of telomere DNA can serve as the template for synthesis of the complementary strand by DNA polymerase III. Thus the length of the chromosome is maintained.

Telomerase has solved the problem of end replication for eucaryotes, but recall that some bacteria also have linear chromosomes. How do they replicate the ends of their chromosomes?



**Figure 11.20** Replication of the Telomeres of Eucaryotic Chromosomes by Telomerase. Telomerase contains an RNA molecule that can base pair with a small portion of the 3' overhang. The RNA serves as a template for DNA synthesis catalyzed by the reverse transcriptase activity of the enzyme. The 3'-OH of the telomere DNA serves as the primer and is lengthened. The process shown is repeated many times until the 3' overhang is long enough to serve as the template for the complementary telomere DNA strand.

Unfortunately, little is known about the replication of linear bacterial chromosomes. However, a recent discovery in *Streptomyces*, an important group of soil bacteria, has led to speculation that a telomerase-like process may function in these bacteria. The ends of the linear chromosome of *Streptomyces coelicolor* are associated with a complex of proteins, including one with in vitro reverse transcriptase activity. No RNA has been found in the *Streptomyces* complex, so it is unclear if the protein functions as a reverse transcriptase in cells and what it might use as a template, if it does.

- Define the following terms: origin of replication, replicon, replication fork, primosome, and replisome.
- Describe 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, lagging strand, primase, and telomerase.
- How do replication patterns differ between prokaryotes and eukaryotes? Describe the operation of replication forks in the generation of theta-shaped intermediates.
- How does rolling-circle replication differ from the usual type of replication observed for cellular chromosomes?
- Outline the steps involved in DNA synthesis at the replication fork. How do DNA polymerases correct their mistakes?

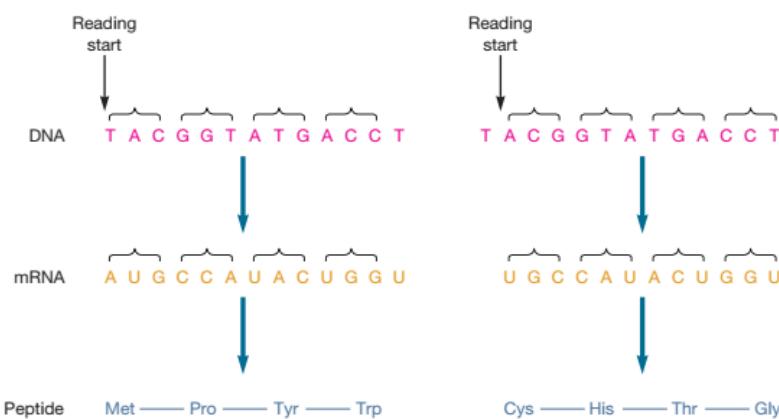
## 11.5 GENE STRUCTURE

DNA replication allows genetic information to be passed from one generation to the next. But how is the genetic information used? To answer that question, we must first look at how genetic information is organized. The basic unit of genetic information is the gene. 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 (e.g., virus, bacterium) with that from another and is responsible for generating much of the genetic variability found in viruses and living organisms. With the discovery and characterization of DNA, the gene was defined more precisely as a linear sequence of nucleotides with fixed start and end points.

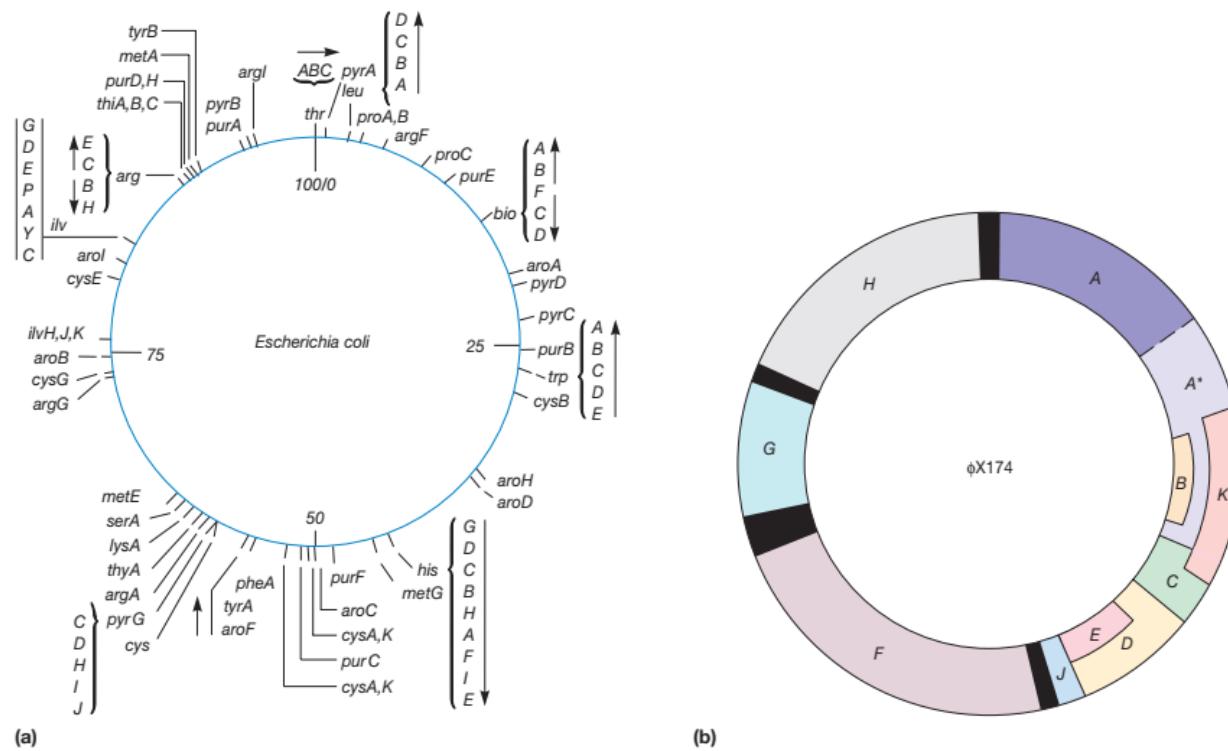
[Creating genetic variability: Recombination at the molecular level \(section 13.4\)](#)

At first, it was thought that a gene contained information for the synthesis of one enzyme, the one gene-one enzyme hypothesis. This was next 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. Historically, a segment of DNA that encodes a single polypeptide was termed a **cistron**; this term is still sometimes used. However, not all genes encode proteins; some code instead for rRNA and tRNA (see figure 11.4). In addition, it is now known that some eukaryotic genes encode more than one protein. Thus a gene might be defined as a polynucleotide sequence that codes for a functional product (i.e., a polypeptide, tRNA, or rRNA). The nucleotide sequences of protein-coding genes are distinct from RNA-coding genes and noncoding regions because when transcribed, the resulting mRNA can be “read” in discrete sequences of sets of three nucleotides, each set being a **codon**. Each codon codes for a single amino acid. The sequence of codons is “read” in 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 11.21). Each strand of DNA therefore usually consists of gene sequences that do not overlap one another (figure 11.22a). However, there are exceptions to the rule. Some viruses such as the phage  $\phi$ X174 have overlapping genes (figure 11.22b), and parts of genes overlap in some bacterial genomes.

Prokaryotic and viral gene structure differs greatly from that of eukaryotes. In prokaryotic and viral systems, the coding information within a gene normally is continuous. However, in eukaryotic organisms, many genes contain coding information (exons) interrupted periodically by noncoding sequences (in-



**Figure 11.21 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 11.22 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  $\phi$ X174 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.

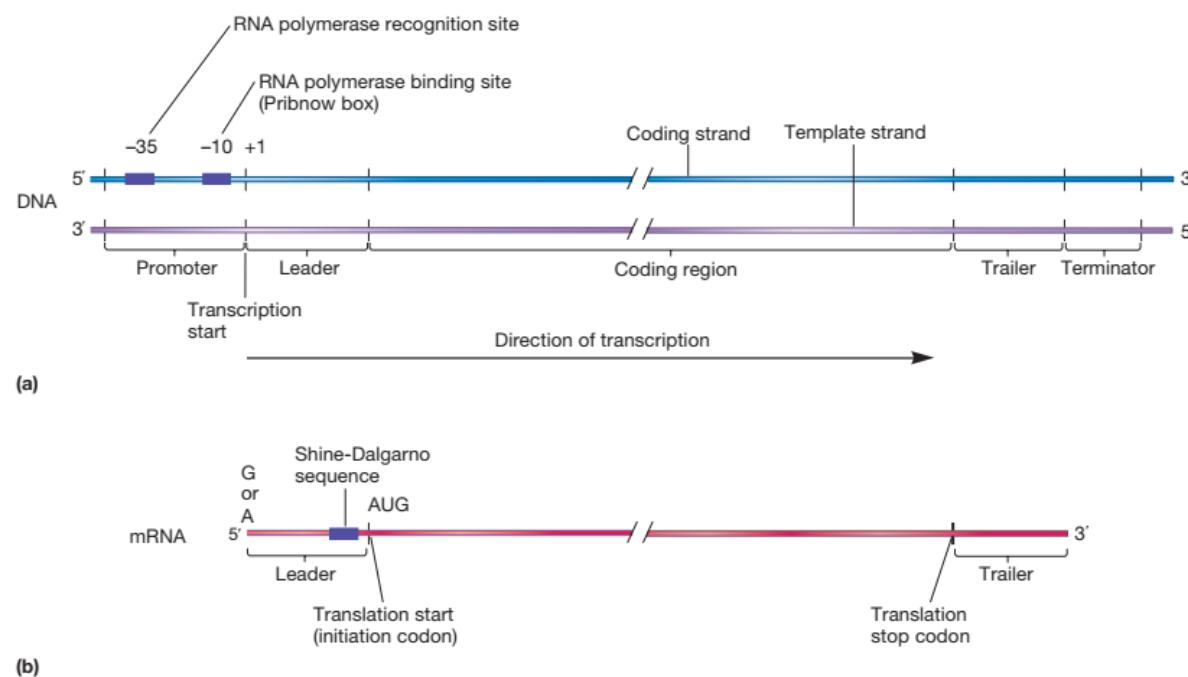
trons). The introns must be cut, or spliced, out of the mRNA before the protein is made. As we will see, this affords eucaryotes the ability to cut and paste mRNA molecules so that they can encode more than one polypeptide, a process known as **alternative splicing**. An interesting exception to this 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.

#### Genes That Code for Proteins

In order for genetic information in the DNA to be used, it must first be transcribed to form an RNA molecule. The RNA product of a gene that codes for a protein is messenger RNA (mRNA). Recall from the discussion of information flow that although DNA is double stranded, only one strand of a gene contains coded information and directs RNA synthesis. This strand is called the **template strand**, and the complementing strand is known as the coding strand because it is the same nucleotide sequence as the mRNA, except in DNA bases (**figure 11.23**). Because the mRNA

is made from the 5' to the 3' end, the polarity of the DNA template strand is 3' to 5'. Therefore the beginning of the gene is at the 3' end of the template strand. An important site, the **promoter**, is located at the start of the gene. The promoter is a recognition/binding site for RNA polymerase, the enzyme that synthesizes RNA. The promoter is neither transcribed nor translated; it functions strictly to orient RNA polymerase a specific distance from the first DNA nucleotide that will serve as a template. As we will see in chapter 12, the promoter is also very important in regulating when and where a gene will be transcribed or expressed.

The transcription start site (labeled +1 in figure 11.23) represents the first nucleotide in the mRNA synthesized from the gene. However, the initially transcribed portion of the gene does not necessarily code for amino acids. Instead it is a **leader sequence** that is transcribed into mRNA, but is not translated into amino acids. The leader sequence includes a region called the **Shine-Dalgarno sequence** that is important in the initiation of translation. The leader sometimes is also involved in regulation of transcription and translation. [Regulation of transcription elongation \(section 12.3\)](#); [Regulation at the level of translation \(section 12.4\)](#)



**Figure 11.23 A Bacterial Structural Gene and Its mRNA Product.** (a) 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 proceeds to the right as shown. The template is read in the 3' to 5' direction. (b) Messenger RNA product of the gene shown in part a. The first nucleotide incorporated into mRNA is usually GMP or AMP. Translation of the mRNA begins with the AUG initiation codon. Regulatory sites are not shown.

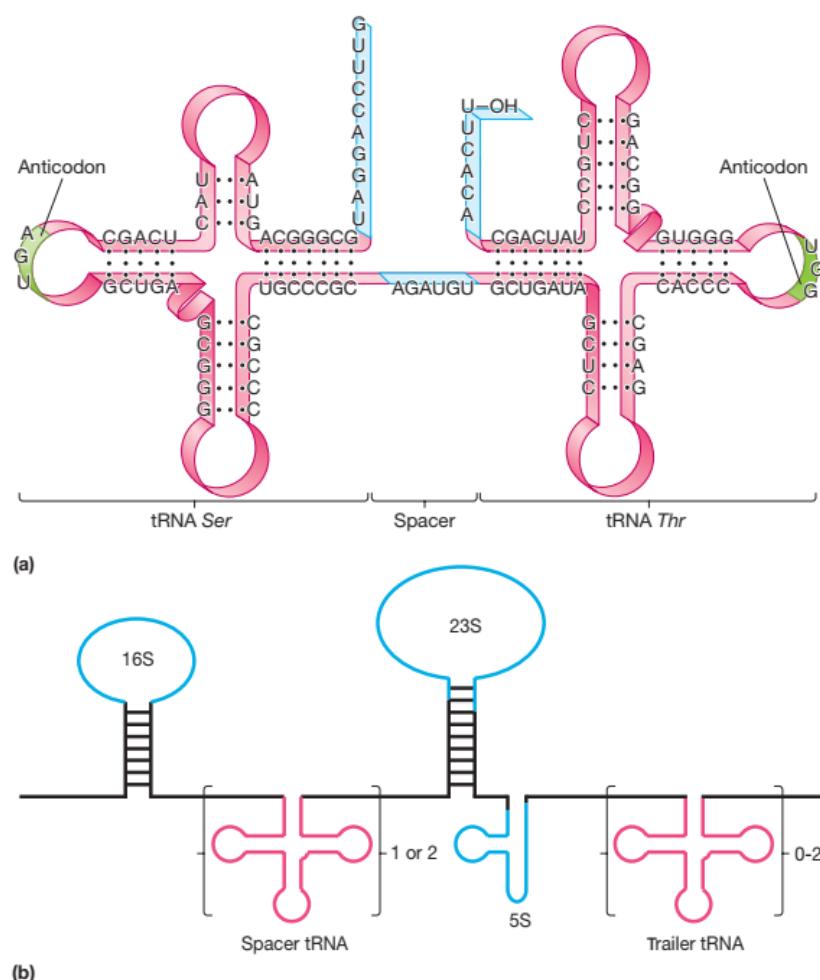
Immediately next to (and downstream of) the leader is the most important part of the gene, the **coding region** (figure 11.23). In genes that direct the synthesis of proteins, the coding region typically begins with the template DNA sequence 3'-TAC-5'. This produces the codon 5'-AUG-3', which in bacteria codes for *N*-formylmethionine, a specially modified amino acid used to initiate protein synthesis. The remainder of the coding region consists of a sequence of codons that specifies the sequence of amino acids for that particular protein. The coding region ends with a special codon called the **stop codon**, which signals the end of the protein and stops the ribosome during translation. The stop codon is immediately followed by the **trailer sequence** (figure 11.23), which is needed for proper expression of the coding region of the gene. The stop codon is not recognized by RNA polymerase during transcription. Instead, a **terminator sequence** is used to stop transcription by dislodging the RNA polymerase from the template DNA.

Besides these basic components—the promoter, leader, coding region, trailer, and terminator—many bacterial 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, operator and activator binding sites, are discussed in sections 12.2 and 12.5. Certainly everything is not known about genes and their structure. With the ready availability of cloned genes and DNA sequencing technology, major discoveries continue to be made in this area.

#### 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 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 process of tRNA maturation (figure 11.24a). The precise function of the leader is not clear; however, the trailer is required for transcription termination. Genes coding for tRNA may code for more than a single tRNA molecule or type of tRNA (figure 11.24a). The segments coding for tRNAs are separated by short spacer sequences that are removed after transcription by special ribonu-



**Figure 11.24 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.

cleases, at least one of which contains catalytic RNA. RNA molecules with catalytic activity are called **ribozymes** (**Microbial Tidbits 11.2**).

The genes for rRNA also are similar in organization to genes coding for proteins because they have promoters, trailers, and terminators (figure 11.24b). 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 prokaryotes.

1. Define or describe the following: gene, template and coding strands, promoter, leader, coding region, reading frame, trailer, and terminator.
2. How do the genes of prokaryotes 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?



## Microbial Tidbits

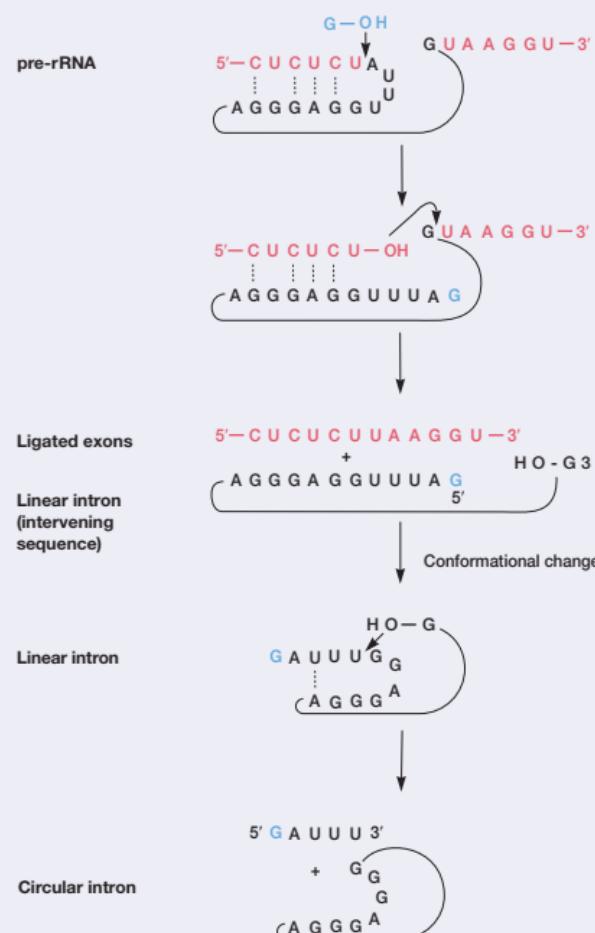
### 11.2 Catalytic RNA (Ribozymes)

Biologists once thought that all cellular reactions were catalyzed by proteins called enzymes (see section 8.7). The discovery during 1981–1984 by Thomas Cech and Sidney Altman that RNA also can sometimes catalyze reactions has transformed our way of thinking about topics as diverse as catalysis and the origin of life. It is now clear that some RNA molecules, called ribozymes, catalyze reactions that alter either their own structure or that of other RNAs.

This discovery has stimulated scientists to hypothesize that the early Earth was an “RNA world” in which RNA acted as both the genetic material and a reaction catalyst. Experiments showing that introns from *Tetrahymena thermophila* can catalyze the formation of polycytidylic acid under certain circumstances have further encouraged such speculations. Some have suggested that RNA viruses are “living fossils” of the original RNA world. **The first self-replicating entity: The RNA world (section 19.1)**

The best-studied ribozyme activity is the self-splicing of RNA. This process is widespread and occurs in *Tetrahymena* pre-rRNA; the mitochondrial rRNA and mRNA of yeast and other fungi; chloroplast tRNA, rRNA, and mRNA; in mRNA from some bacteriophages (e.g., the T4 phage of *E. coli*); and in the hepatitis delta virusoid. The 413-nucleotide rRNA intron of *T. thermophila* provides a good example of the self-splicing reaction. The reaction occurs in three steps and requires the presence of guanosine (see Box figure). First, the 3'-OH group of guanosine attacks the intron's 5'-phosphate group and cleaves the phosphodiester bond. Second, the new 3'-hydroxyl on the left exon attacks the 5'-phosphate of the right exon. This joins the two exons and releases the intron. Finally, the intron's 3'-hydroxyl attacks the phosphate bond of the nucleotide 15 residues from its end. This releases a terminal fragment and cyclizes the intron. Self-splicing of this rRNA occurs about 10 billion times faster than spontaneous RNA hydrolysis. Just as with enzyme proteins, the RNA's shape is essential to catalytic efficiency. The ribozyme even has Michaelis-Menten kinetics (see figure 8.18). The ribozyme from the hepatitis delta virusoid catalyzes RNA cleavage that is involved in its replication. It is unusual in that the same RNA can fold into two shapes with quite different catalytic activities: the regular RNA cleavage activity and an RNA ligation reaction.

The discovery of ribozymes has many potentially important practical consequences. Ribozymes act as “molecular scissors” and will enable researchers to manipulate RNA easily in laboratory experiments. It also might be possible to protect hosts by specifically removing RNA from pathogenic viruses, bacteria, and fungi. For example, ribozymes are being tested against the AIDS, herpes, and tobacco mosaic viruses.



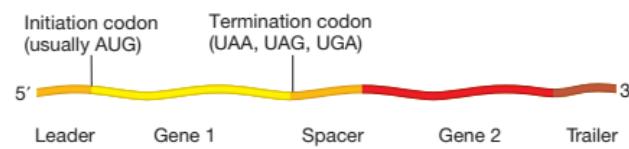
**Ribozyme Action.** The mechanism of *Tetrahymena thermophila* pre-rRNA self-splicing. See text for details.

### 11.6 TRANSCRIPTION

As mentioned earlier, synthesis of RNA under the direction of DNA is called transcription. The RNA product has a sequence complementary to the DNA template directing its synthesis (table 11.2). Thymine is not normally found in mRNA and rRNA. Although adenine directs the incorporation of thymine

during DNA replication, it usually codes for uracil during RNA synthesis. Transcription generates three kinds of RNA. Messenger RNA (mRNA) bears the message for protein synthesis. In *Bacteria* and *Archaea*, the mRNA often bears coding information transcribed from adjacent genes. Therefore it is said to be polygenic or polycistronic (figure 11.25). Eucaryotic mRNAs, on the other hand, are usually monocistronic, containing infor-

Table 11.2 RNA Bases Coded for by DNA	
DNA Base	Purine or Pyrimidine Incorporated into RNA
Adenine	Uracil
Guanine	Cytosine
Cytosine	Guanine
Thymine	Adenine

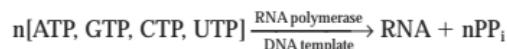


**Figure 11.25 A Polycistronic Bacterial Messenger RNA.**

mation for a single polypeptide. **Transfer RNA (tRNA)** carries amino acids during protein synthesis, and **ribosomal RNA (rRNA)** molecules are components of ribosomes. The synthesis of bacterial mRNA is described first.

### Transcription in Bacteria

RNA is synthesized under the direction of DNA by the enzyme **RNA polymerase**. The reaction is quite similar to that catalyzed by DNA polymerase (figure 11.14). ATP, GTP, CTP, and UTP are used to produce an RNA complementary to the DNA template. As mentioned earlier, these nucleotides contain ribose rather than deoxyribose (figure 11.15).



RNA synthesis, like DNA synthesis, proceeds in a 5' to 3' direction with new nucleotides being added to the 3' end of the growing chain at a rate of about 40 nucleotides per second at 37°C (figure 11.26). In both DNA and RNA polymerase reactions, pyrophosphate ( $\text{PP}_i$ ) is produced. It is then hydrolyzed to orthophosphate in a reaction catalyzed by the pyrophosphatase enzyme. Hydrolysis of the pyrophosphate product makes DNA and RNA synthesis irreversible. If the pyrophosphate level were too high, DNA and RNA would be degraded by a reversal of the polymerase reactions.

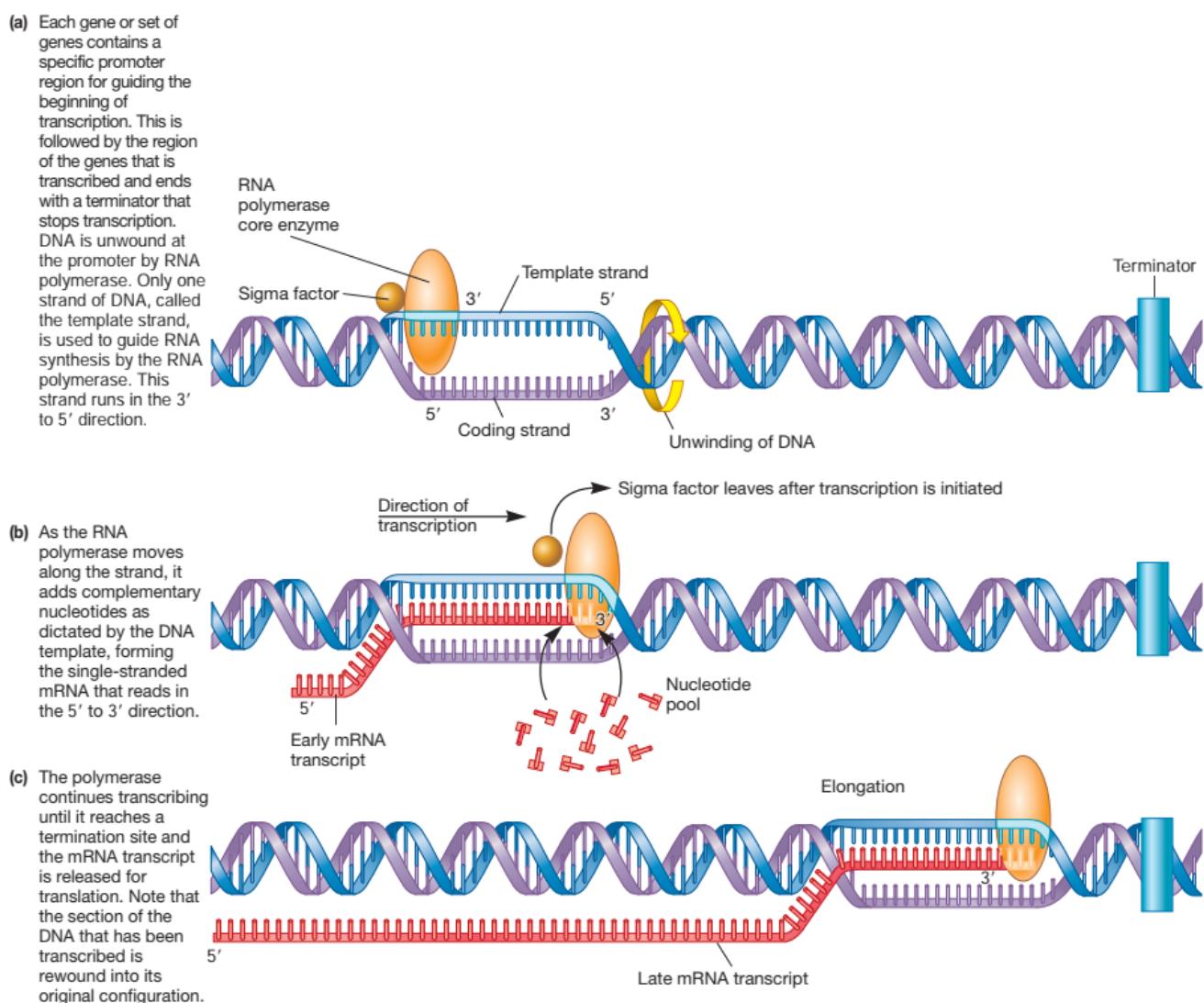
Most bacterial RNA polymerases contain five types of polypeptide chains:  $\alpha$ ,  $\beta$ ,  $\beta'$ ,  $\omega$ , and  $\sigma$  (figure 11.27). The **core enzyme** is composed of five chains ( $\alpha_2$ ,  $\beta$ ,  $\beta'$ , and  $\omega$ ) and catalyzes RNA synthesis. The **sigma factor** ( $\sigma$ ) has no catalytic activity but helps the core enzyme recognize the start of genes. When sigma is bound to core enzyme, the six-subunit complex is termed **RNA polymerase holoenzyme**. Only holoenzyme can begin transcription, but as we will see, core enzyme completes RNA synthesis once it has been initiated. The precise functions of the  $\alpha$ ,  $\beta$ ,  $\beta'$ , and  $\omega$  polypeptides are not yet clear. The  $\alpha$  subunits

seem to be involved in the assembly of the core enzyme, recognition of promoters, and interaction with some regulatory factors. The binding site for DNA is on  $\beta'$ , and the  $\omega$  subunit seems to be involved in stabilizing the conformation of the  $\beta'$  subunit. The  $\beta$  subunit binds ribonucleotide substrates. Rifampin, an RNA polymerase inhibitor, binds to the  $\beta$  subunit.

Recently the atomic structure of RNA polymerase from *Thermus aquaticus* has been determined (figure 11.27a,b). In this bacterium, the core enzyme is composed of four different subunits ( $\alpha_2$ ,  $\beta$ ,  $\beta'$ , and  $\omega$ ) and is complexed with the sigma factor ( $\sigma$ ). The enzyme is claw-shaped with a clamp domain that closes on an internal channel, which contains an essential magnesium and the active site. Sigma interacts extensively with the core enzyme and specifically binds to elements of the promoter. It also may widen the channel so that DNA can enter the interior of the polymerase complex.

Transcription involves three separate processes: initiation, elongation, and termination. Only a relatively short segment of DNA is transcribed (unlike replication in which the entire chromosome must be copied), and initiation begins when the RNA polymerase binds to the promoter for the gene. RNA polymerase core enzyme is not able to bind DNA tightly or specifically. This situation is drastically changed when sigma is bound to core to make the holoenzyme, which binds the promoter tightly. Recall that the promoter serves only as a target for the binding of the RNA polymerase and is not transcribed. Bacterial promoters have two characteristic features: a sequence of six bases (often TTGACA) about 35 bases pairs before the transcription starting point and a TATAAT sequence, or **Pribnow box**, usually about 10 base pairs upstream of the transcriptional start site (figure 11.28; also 11.23). These regions are called the **-35** and **-10 sites**, respectively, while the first nucleotide to be transcribed is referred to as the **+1 site**. As noted previously, RNA polymerase holoenzyme recognizes the specific sequences at the -10 and -35 sites of promoters. Because the sites must be similar in all promoters, they are called **consensus sequences**.

Once bound to the promoter site, RNA polymerase is able to unwind the DNA without the aid of helicases (figure 11.29). The -10 site is rich in adenines and thymines, making it easier to break the hydrogen bonds that keep the DNA double stranded; when the DNA is unwound at this region, it is called **open complex**. A region of unwound DNA equivalent to about two turns of the helix (about 16–20 bases pairs) becomes the “**transcription bubble**,” which moves with the RNA polymerase as it proceeds to transcribe mRNA from the template DNA strand during elongation (figure 11.30). Within the transcription bubble, a temporary RNA:DNA hybrid is formed. As the RNA polymerase progresses in the 3' to 5' direction along the DNA template, the sigma factor soon dissociates from core RNA polymerase and is available to aid another unit of core enzyme initiate transcription. The mRNA is made in the 5' to 3' direction so it is complementary and antiparallel to the template DNA. As elongation of the mRNA continues, single-stranded mRNA is released and the two strands of DNA behind the transcription bubble resume their double helical structure. As shown in figure 11.26, RNA polymerase is a remarkable

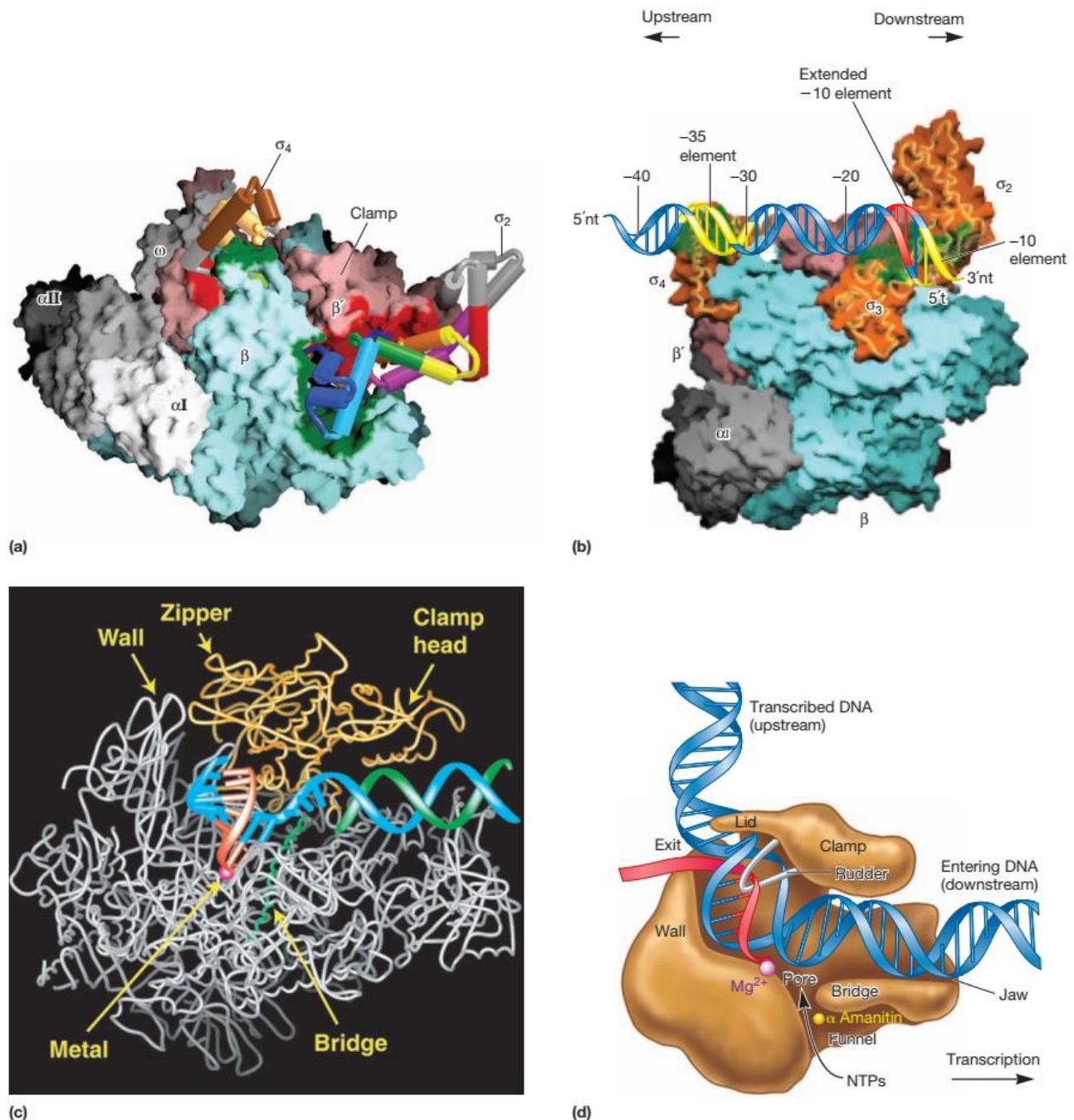


**Figure 11.26** The Major Events in Transcription.

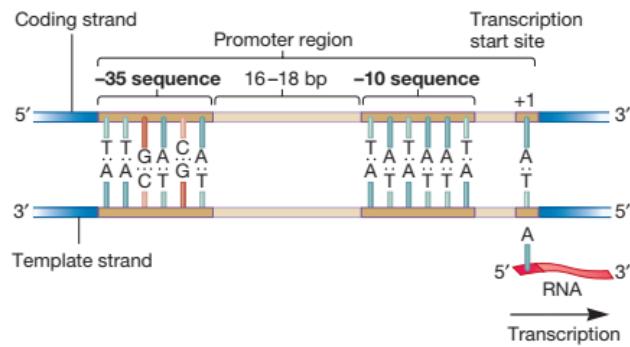
enzyme capable of several activities, including unwinding the DNA, moving along the template, and synthesizing RNA.

Termination of transcription occurs when the core RNA polymerase dissociates from the template DNA. The end of a gene or group of genes is marked by DNA sequences in the trailer (which is transcribed but not translated) and the terminator. The sequences within prokaryotic terminators often contain nucleotides that, when transcribed into RNA, form hydrogen bonds within the single-stranded RNA. This intrastrand base pairing creates a hairpin-shaped loop-and-stem structure. This structure appears to cause the RNA polymerase to pause or stop transcribing DNA. There are two kinds of terminators. The first type causes **intrinsic** or **rho-independent termination** (figure 11.31). It features the

mRNA hairpin followed by a stretch of about six uridine residues. Once the RNA polymerase has paused at the hairpin loop, the A-U base pairs in the uracil-rich region are too weak to hold the RNA:DNA duplex together and the RNA polymerase falls off. The second kind of terminator lacks a poly-U region, and often the hairpin; it requires the aid of a special protein, the **rho factor** ( $\rho$ ). This terminator causes **rho-dependent termination**. It is thought that rho binds to mRNA and moves along the molecule until it reaches the RNA polymerase that has halted at a terminator (figure 11.32). The rho factor, which has hybrid RNA:DNA helicase activity, then causes the polymerase to dissociate from the mRNA, probably by unwinding the mRNA-DNA complex.



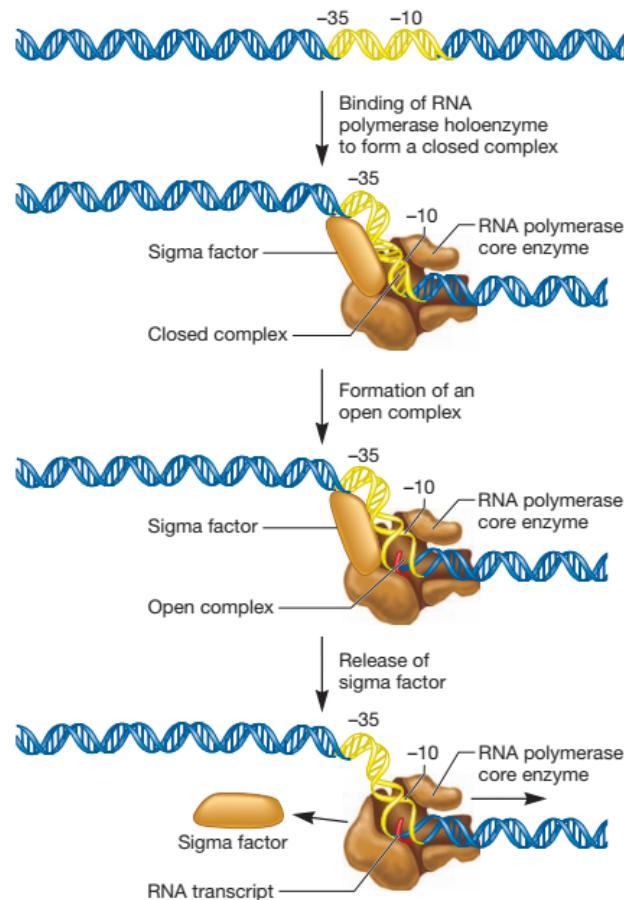
**Figure 11.27 RNA Polymerase Structure.** The atomic structures of RNA polymerase from the bacterium *Thermus aquaticus* (a and b) and yeast RNA polymerase II (c and d) are presented here. (a) The Taq RNA polymerase holoenzyme is shown with the  $\sigma$  subunit depicted as an  $\alpha$ -carbon backbone with cylinders for  $\alpha$ -helices. Two of the three  $\sigma$  factor domains are labeled. (b) The holoenzyme-DNA complex with the  $\sigma$  surface rendered slightly transparent to show the  $\alpha$ -carbon backbone inside. Protein surfaces that contact the DNA are in green and are located on the  $\sigma$  factor. The  $-10$  and  $-35$  elements in the promoter are in yellow. The internal active site is covered by the  $\beta$  subunit in this view. (c) Yeast RNA polymerase II transcribing complex with some peptide chains removed to show the DNA. The active site metal is a red sphere. A short stretch of DNA-RNA hybrid (blue and red) lies above the metal. (d) A cutaway side view of the polymerase II transcribing complex with the pathway of the nucleic acids and some of the more important parts shown. The enzyme is moving from left to right and the DNA template strand is in blue. A protein "wall" forces the DNA into a right-angle turn and aids in the attachment of nucleoside triphosphates to the growing 3' end of the RNA. The newly synthesized RNA (red) is separated from the DNA template strand and exits beneath the rudder and lid of the polymerase protein complex. The binding site of the inhibitor  $\alpha$ -amanitin also is shown.



**Figure 11.28 The conventional numbering system of promoters.** The first nucleotide that acts as a template for transcription is designated +1. The numbering of nucleotides to the left of this spot is in a negative direction, while the numbering to the right is in a positive direction. For example, the nucleotide that is immediately to the left of the +1 nucleotide is numbered -1, and the nucleotide to the right of the +1 nucleotide is numbered +2. There is no zero nucleotide in this numbering system. In many bacterial promoters, sequence elements at the -35 and -10 regions play a key role in promoting transcription.

### Transcription in Eucaryotes

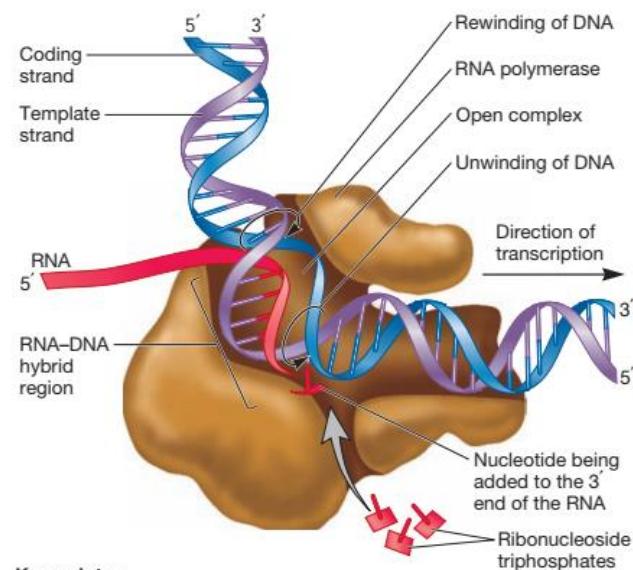
Transcriptional processes in eucaryotic microorganisms (and in other eucaryotic cells) differ in several ways from bacterial transcription. There are three major RNA polymerases, not one as in *Bacteria*. RNA polymerase II, associated with chromatin in the nuclear matrix, is responsible for mRNA synthesis. Polymerases I and III synthesize rRNA and tRNA, respectively (table 11.3). The eucaryotic RNA polymerase II is a large aggregate, at least 500,000 daltons in size, with about 10 or more subunits. The atomic structure of the 10-subunit yeast RNA polymerase II associated with DNA and RNA has been determined (figure 11.27c,d). The entering DNA is held in a clamp that closes down on it. A magnesium ion is located at the active site and the 9 base pair DNA-RNA hybrid in the transcription bubble is bound in a cleft formed by the two large polymerase subunits. The newly synthesized RNA exits the polymerase beneath the rudder and lid regions. The substrate nucleoside triphosphates probably reach the active site through a pore in the complex. Unlike bacterial polymerase, RNA polymerase II requires extra transcription factors to recognize its promoters (figure 11.33). The polymerase binds near the start point; the transcription factors bind to the rest of the promoter. Eucaryotic promoters also differ from those in *Bacteria*. They have combinations of several elements. Three of the most common are the TATA box (located about 30 base pairs before or upstream of the start point), and the GC and CAAT boxes located between 50 to 100 base pairs upstream of the start site (figure 11.34). The TFIID transcription factor (figure 11.33) plays an important role in transcription initiation in eucaryotes. This multi-



**Figure 11.29 The Initiation of Transcription in Bacteria.** The sigma factor of the RNA polymerase holoenzyme is responsible for positioning the core enzyme properly at the promoter. Sigma factor recognizes two regions in the promoter, one centered at -35 and the other centered at -10. Once positioned properly, the DNA at the -10 region unwinds to form an open complex. The sigma factor dissociates from the core enzyme as it begins transcribing the gene.

protein complex contains the TATA-binding protein (TBP). TBP has been shown to sharply bend the DNA on attachment. This makes the DNA more accessible to other initiation factors. A variety of general transcription factors, promoter specific factors, and promoter elements have been discovered in different eucaryotic cells. Each eucaryotic gene seems to be regulated differently, and more research will be required to fully understand the regulation of eucaryotic gene transcription.

Unlike bacterial mRNA, eucaryotic mRNA arises from **post-transcriptional modification** of large RNA precursors, about 5,000 to 50,000 nucleotides long, sometimes called heteroge-

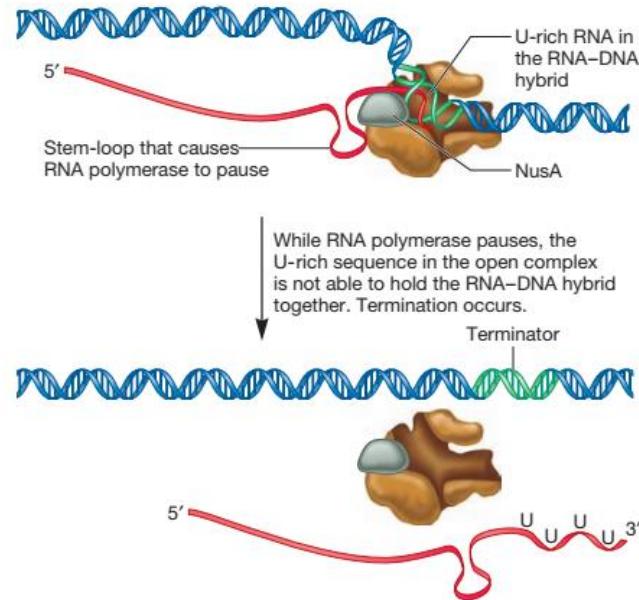
**Key points:**

- RNA polymerase slides along the DNA, creating an open complex as it moves.
- The DNA strand known as the template strand is used to make a complementary copy of RNA as an RNA-DNA hybrid.
- The RNA is synthesized in a 5' to 3' direction using ribonucleoside triphosphates as precursors. Pyrophosphate is released (not shown).
- The complementarity rule is the same as the AT/GC rule except that U is substituted for T in the RNA.

**Figure 11.30** The “Transcription Bubble.”

neous nuclear RNA (hnRNA) (figure 11.35). As hnRNA is synthesized, a **5' cap** is added. After synthesis is completed, the precursor RNA is modified by the addition of a **3' poly-A tail**. It is also processed, if necessary, to remove any introns. The 5' cap is the unusual nucleotide 7-methylguanosine. It is attached to the 5'-hydroxyl of the hnRNA by a triphosphate linkage (figure 11.36). Addition of a poly-A tail is initiated by an endonuclease that shortens the hnRNA and generates a 3'-OH group. The enzyme polyadenylate polymerase then catalyzes the addition of adenylic acid to the 3' end of hnRNA to produce a poly-A sequence about 200 nucleotides long. The functions of the 5' cap and poly-A tail are not completely clear. The 5' cap on eukaryotic mRNA may promote the initial binding of ribosomes to the mRNA. It also may protect the mRNA from enzymatic attack. Poly-A protects mRNA from rapid enzymatic degradation. The poly-A tail must be shortened to about 10 nucleotides before mRNA can be degraded. Poly-A also seems to aid in mRNA translation.

As noted earlier, many eukaryotic genes are split or interrupted, which leads to the final type of posttranscriptional pro-



**Figure 11.31** Intrinsic Termination of Transcription. This type of terminator contains a U-rich sequence downstream from a stretch of nucleotides that can form a stem-loop and stem structure. Formation of the stem loop in the newly synthesized RNA causes RNA polymerase to pause. This pausing is stabilized by the NusA protein. The U-A bonds in the uracil-rich region are not strong enough to hold the RNA and DNA together. Therefore, the RNA, DNA, and RNA polymerase dissociate and transcription stops.

cessing. **Split** or **interrupted genes** have **exons** (expressed sequences), regions coding for RNA that end up in the mRNA. Exons are separated from one another by **introns** (intervening sequences), sequences coding for RNA that is never translated into protein (figure 11.37). The initial RNA transcript contains both exon and intron sequences. Genes coding for rRNA and tRNA may also be interrupted. Except for cyanobacteria and *Archaea* (see chapters 20 and 21), interrupted genes have not been found in prokaryotes.

Introns are removed from the initial RNA transcript (also called **pre-mRNA** or **primary transcript**) by a process called **RNA splicing** (figure 11.37). The intron's borders are clearly marked for accurate removal. Exon-intron junctions have a GU sequence at the intron's 5' boundary and an AG sequence at its 3' end. These two sequences define the splice junctions and are recognized by special RNA molecules. The nucleus contains several **small nuclear RNA (snRNA)** molecules, about 60 to 300 nucleotides long. These complex with proteins to form small nuclear ribonucleoprotein particles called snRNPs or snurps. Some of the snRNPs recognize splice junctions and ensure splicing accuracy. For example, U1-snRNP recognizes the 5' splice junction, and U5-snRNP recognizes

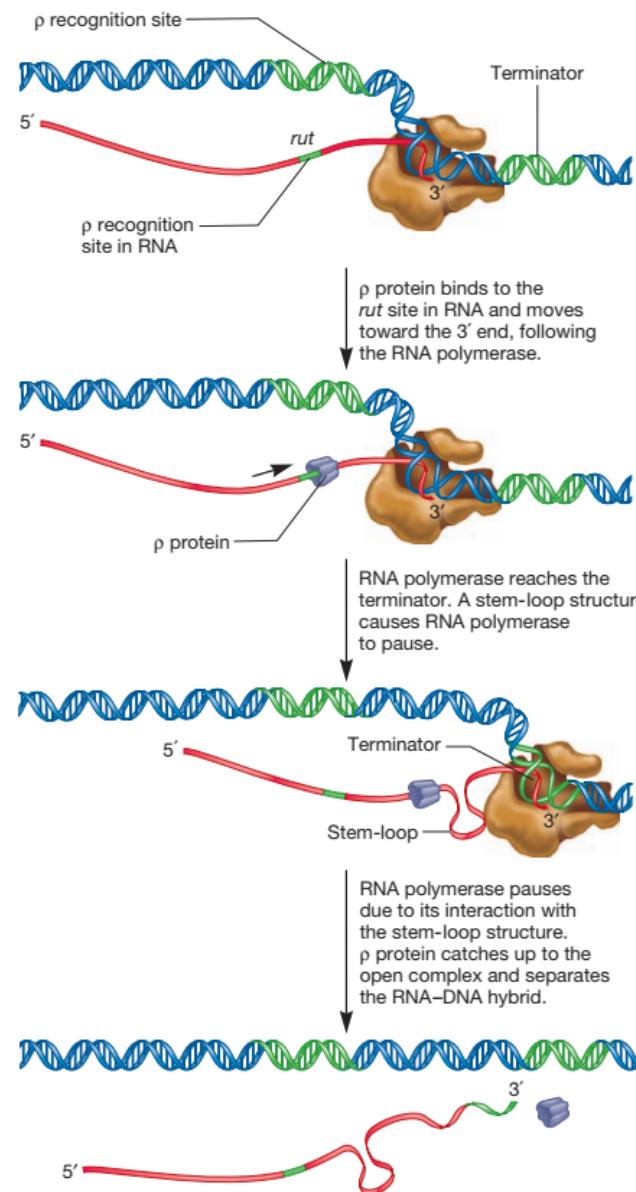
Table 11.3 Eucaryotic RNA Polymerases		
Enzyme	Location	Product
RNA polymerase I	Nucleolus	rRNA (5.8S, 18S, 28S)
RNA polymerase II	Chromatin, nuclear matrix	mRNA
RNA polymerase III	Chromatin, nuclear matrix	tRNA, 5S rRNA

the 3' junction. Splicing of pre-mRNA occurs in a large complex called a **spliceosome** that contains the pre-mRNA, at least five kinds of snRNPs, and non-snRNP splicing factors. Sometimes a pre-mRNA will be spliced so that different patterns of exons remain. This alternative splicing allows a single gene to code for more than one protein. The splice pattern determines which protein will be synthesized. Splice patterns can be cell-type specific or determined by the needs of the cell. The importance of alternative splicing in multicellular eucaryotes was emphasized when it was discovered that the human genome has only about 20,000 genes rather than the anticipated 100,000. It is thought that alternative splicing is one mechanism by which human cells produce such a vast array of proteins.

As just mentioned, a few rRNA genes also have introns. Some of these pre-rRNA molecules are self-splicing, that is, the pre-rRNA is a ribozyme (Microbial Tidbits 11.2). Thomas Cech first discovered that pre-rRNA from the ciliate protozoan *Tetrahymena thermophila* is self-splicing. Sidney Altman then showed that ribonuclease P, which cleaves a fragment from one end of pre-tRNA, contains a piece of RNA that catalyzes the reaction. Several other self-splicing rRNA introns have since been discovered. Cech and Altman received the 1989 Nobel Prize in chemistry for these discoveries. [Microbial evolution \(section 19.1\)](#)

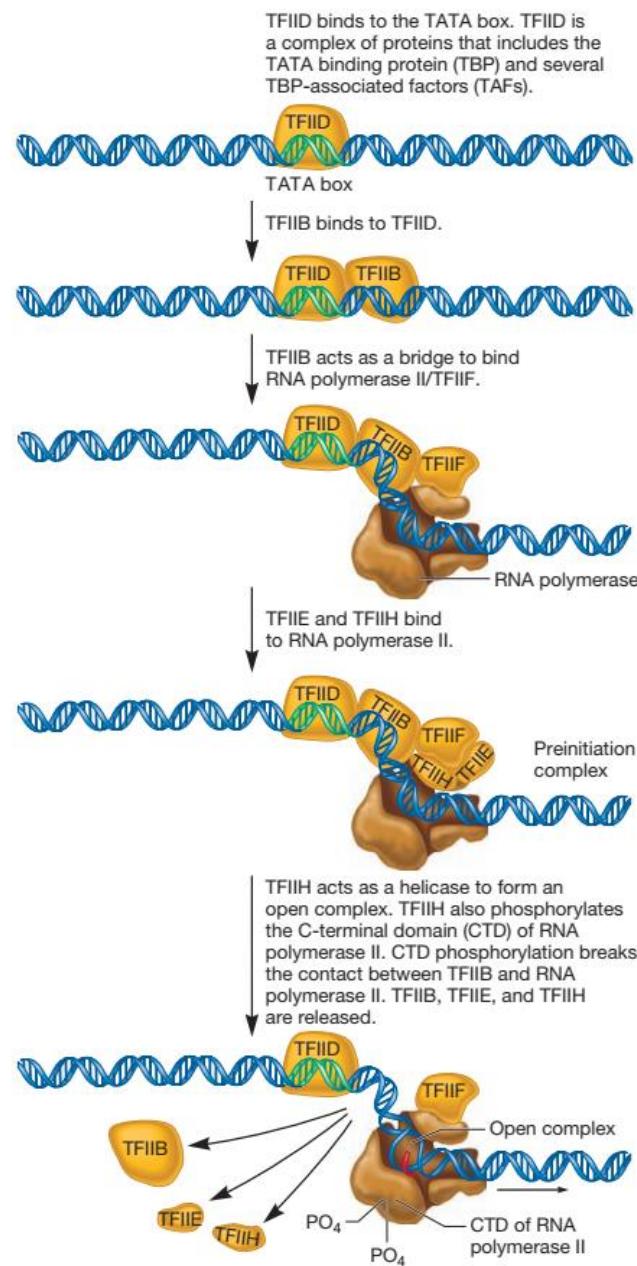
### Transcription in the Archaea

Transcription in the *Archaea* is similar to and distinct from what is observed in *Bacteria* and eucaryotes. Each archaeon has a single RNA polymerase responsible for transcribing all genes in the cell (as in *Bacteria*). However, the RNA polymerase is larger and contains more subunits, many of which are similar to subunits in RNA polymerase II of eucaryotes. The promoters of archaeal genes are similar to those of eucaryotes in having a TATA box; binding of the archaeal RNA polymerase to its promoter requires a TATA-binding protein, just as in eucaryotes. Like the eucaryotic counterpart, the archaeal RNA polymerase also needs several additional transcription factors to function properly. Furthermore, some archaeal genes have introns, which must be removed by posttranscriptional processing. Finally, the mRNA molecules produced by transcription in *Archaea* are usually polycistronic, as in *Bacteria*. This intriguing mixture of bacterial and eucaryotic features in the *Archaea* has fueled a great deal of speculation about the evolution of all three domains of life. [Introduction to the Archaea: Genetics and molecular biology \(section 20.1\)](#)



**Figure 11.32 Rho-Factor ( $\rho$ )-Dependent Termination of Transcription.** The *rut* site stands for *rho utilization site*.

1. Define the following terms: polygenic mRNA, RNA polymerase core enzyme, sigma factor, RNA polymerase holoenzyme, and rho factor.
2. Define or describe posttranscriptional modification, heterogeneous nuclear RNA, 3' poly-A sequence, 5' capping, split or interrupted genes, exon, intron, RNA splicing, snRNA, spliceosome, and ribozyme.
3. Describe how RNA polymerase transcribes bacterial DNA. How does the polymerase know when to begin and end transcription?
4. How do bacterial RNA polymerases and promoters differ from those of *Archaea* and eucaryotes?



**Figure 11.33 Initiation of Transcription in Eucaryotes.** The TATA box is a major component of eucaryotic promoters. TATA-binding protein (TBP), which is a component of a complex of proteins called TFIID (transcription factor IID), binds the TATA box. Note that numerous other transcription factors are required for initiation of transcription, unlike *Bacteria* where only the sigma factor is needed. Initiation of transcription in *Archaea* is similar to that seen in eucaryotes.

## 11.7 THE GENETIC CODE

The final step in the expression of genes that encode proteins is translation. The mRNA nucleotide sequence is translated into the amino acid sequence of a polypeptide chain. Protein synthesis is called translation because it is a decoding process. The information encoded in the language of nucleic acids must be rewritten in the language of proteins. Therefore, before we discuss protein synthesis, we will examine the nature of the genetic code.

### Establishment of the Genetic Code

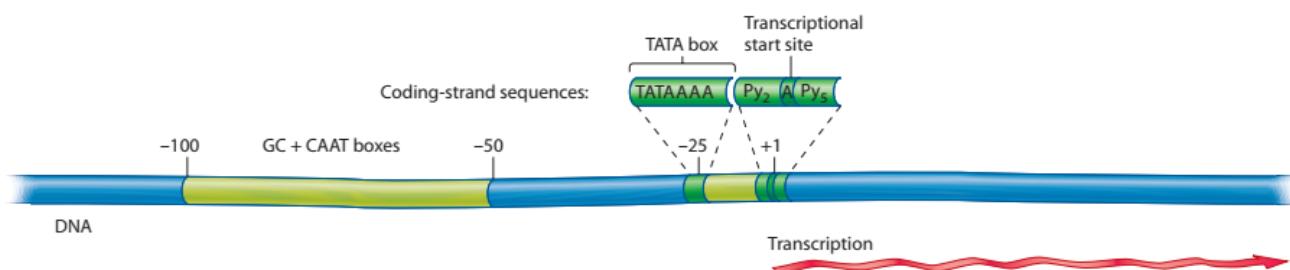
The realization that DNA is the genetic material triggered efforts to understand how genetic instructions are stored and organized in the DNA molecule. Early studies on the nature of the genetic code showed that the DNA base sequence corresponds to the amino acid sequence of the polypeptide specified 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.

Theoretical considerations directed much of the early work on deciphering the code. Scientists reasoned that because only 20 amino acids normally are present in proteins, there must be at least 20 different code words in DNA. Therefore the code must be contained in some sequence of the four nucleotides commonly found in DNA.

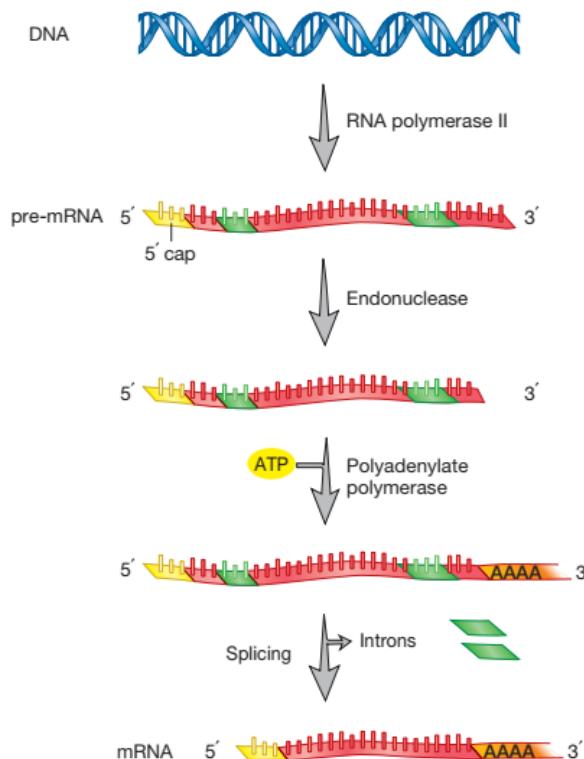
If the code words were two nucleotides in length, there would be only 16 possible combinations ( $4^2$ ) of the four nucleotides and this would not be enough to code for all 20 amino acids. Therefore a code word, or codon, had to consist of at least nucleotide triplets even though this would give 64 possible combinations ( $4^3$ ), many more than the minimum of 20 needed to specify the common amino acids. Research eventually confirmed this and the code was deciphered in the early 1960s by Marshall Nirenberg, Heinrich Matthaei, Philip Leder, and Har Gobind Khorana. In 1968 Nirenberg and Khorana shared the Nobel Prize with Robert 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 11.4**. 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 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



**Figure 11.34** The TATA Box and Other Elements of Eucaryotic Promoters.



**Figure 11.35** Eucaryotic mRNA Synthesis. The production of eucaryotic messenger RNA. The 5' cap is added shortly after synthesis of the mRNA begins.

with variation in the third position (table 11.4). This somewhat loose base pairing is known as **wobble** and relieves cells of the need to synthesize so many tRNAs (figure 11.38). Wobble also decreases the effects of DNA mutations.

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

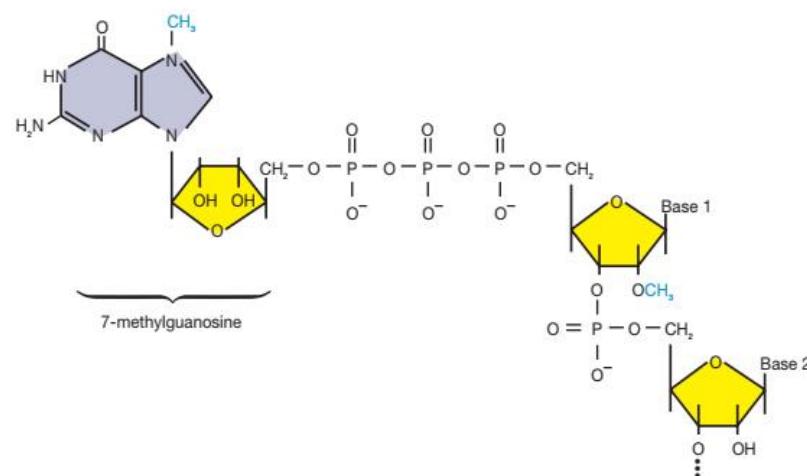
## 11.8 TRANSLATION

Translation involves decoding mRNA and covalently linking amino acids together to form a polypeptide. Just as DNA and RNA synthesis proceeds in one direction (5' to 3'), so too does protein synthesis. Polypeptides are synthesized by the addition of amino acids to the end of the chain with the free  $\alpha$ -carboxyl group (the C-terminal end). That is, the synthesis of polypeptides begins with the amino acid at the end of the chain with a free amino group (the N-terminal) and moves in the C-terminal direction. The ribosome is the site of protein synthesis. Protein synthesis is not only quite accurate but also very rapid. In *E. coli* synthesis occurs at a rate of at least 900 residues per minute; eucaryotic translation is slower, about 100 residues per minute. **Proteins (appendix I)**

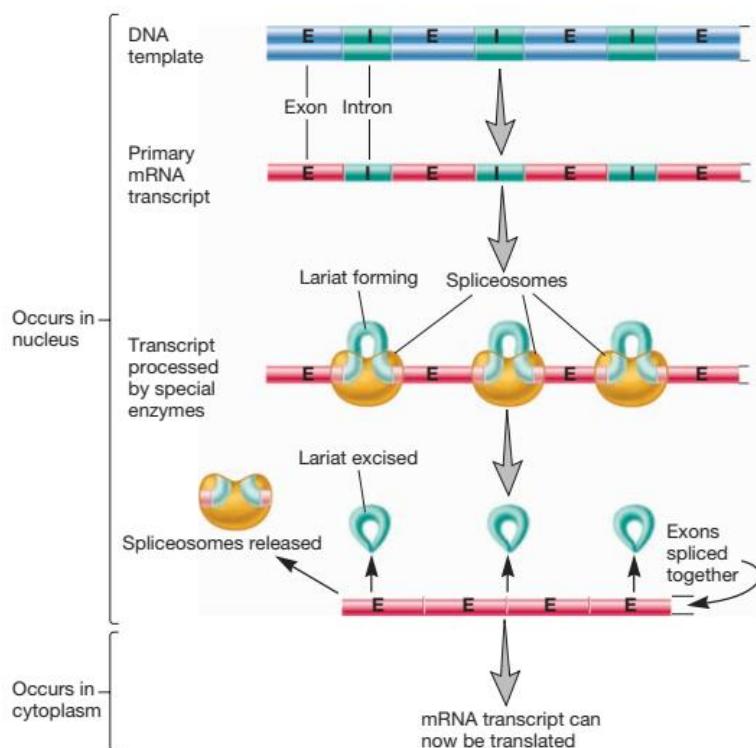
Cells that grow quickly must use each mRNA with great efficiency to synthesize proteins at a sufficiently rapid rate. The two subunits of the ribosome (the 50S subunit and the 30S subunit in *Bacteria* and *Archaea*; 60S and 40S in eucaryotes) are free in the cytoplasm if protein is not being synthesized. They come together to form the complete ribosome only when translation occurs. Frequently mRNAs are simultaneously complexed with several ribosomes, each ribosome reading the mRNA message and synthesizing a polypeptide. At maximal rates of mRNA use, there may be a ribosome every 80 nucleotides along the messenger or as many as 20 ribosomes simultaneously reading an mRNA that codes for a 50,000 dalton polypeptide. A complex of mRNA with several ribosomes is called a **polyribosome** or polysome (figure 11.39). Polysomes are present in both prokaryotes and eucaryotes. *Bacteria* and *Archaea* can further increase the efficiency of gene expression through coupled transcription and translation (figure 11.39b). While RNA polymerase is synthesizing an mRNA, ribosomes can already be attached to the messenger so that transcription and translation occur simultaneously. Coupled transcription and translation is possible in prokaryotes because a nuclear envelope does not separate the translation machinery from DNA as it does in eucaryotes (see figure 3.16).

### Transfer RNA and Amino Acid Activation

The first stage of protein synthesis is **amino acid activation**, a process in which amino acids are attached to transfer RNA molecules. These RNA molecules are normally between 73 and 93 nu-



**Figure 11.36** The 5' Cap of Eucaryotic mRNA. Methyl groups are in blue.



**Figure 11.37** Splicing of Eucaryotic mRNA Molecules.

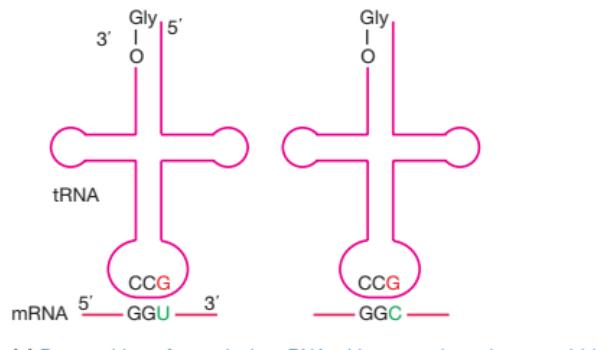
cleotides in length and possess several characteristic structural features. The structure of tRNA becomes clearer when its chain is folded in such a way to maximize the number of normal base pairs, which results in a cloverleaf conformation of five stems and loops (figure 11.40). The acceptor or amino acid stem holds the

activated amino acid on the 3' end of the tRNA. The 3' end of all tRNAs has the same —C—C—A sequence; the amino acid is attached to the terminal adenylic acid. At the other end of the cloverleaf is the anticodon arm, which contains the **anticodon** triplet complementary to the mRNA codon triplet. There are two

**Table 11.4** The Genetic Code

	Second Position					
	U	C	A	G		
First Position (5' End) <sup>a</sup>	UUU	UCU	UAU	UGU		U
	UUC	UCC	UAC	UGC		C
	UUA	UCA	UAA	UGA	STOP	A
	UUG	UCG	UAG	UGG	Trp	G
	CUU	CCU	CAU	CGU		U
	CUC	CCC	CAC	CGC		C
	CUA	CCA	CAA	CGA		A
	CUG	CCG	CAG	CGG		G
	AUU	ACU	AAU	AGU		U
	AUC	ACC	AAC	AGC		C
	AUA	ACA	AAA	AGA		A
	AUG	ACG	AAG	AGG		G
G	GUU	GCU	GAU	GGU		U
	GUC	GCC	GAC	GGC		C
	GUA	GCA	GAA	GGA		A
	GUG	GCG	GAG	GGG		G

<sup>a</sup>The code is presented in the RNA form. Codons run in the 5' to 3' direction. See text for details.



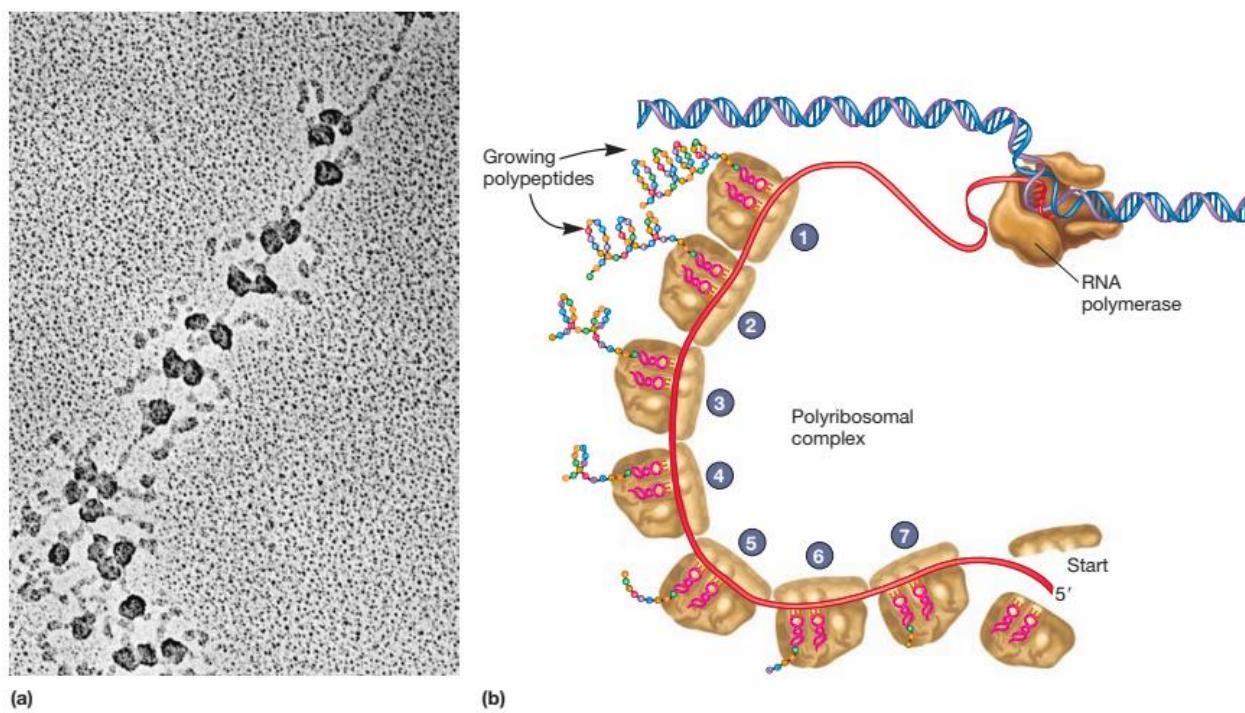
(a) Base pairing of one glycine tRNA with two codons due to wobble

Glycine mRNA codons: GGU, GGC, GGA, GGG (5' → 3')

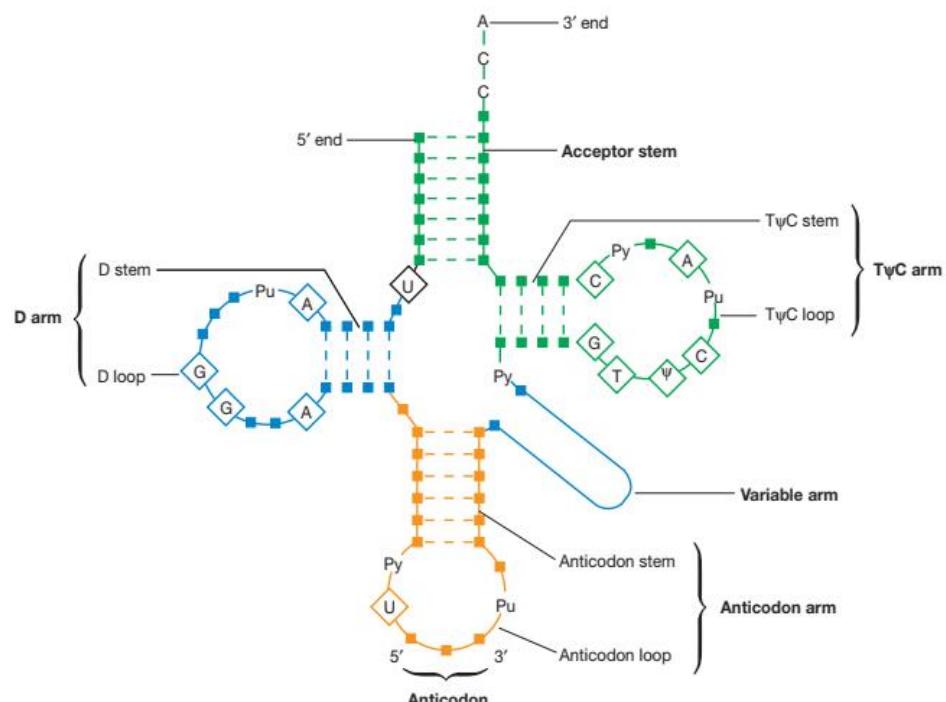
Glycine tRNA anticodons: CCG, CCU, CCC (3' → 5')

(b) Glycine codons and anticodons

**Figure 11.38 Wobble and Coding.** The use of wobble in coding for the amino acid glycine. (a) Because of wobble, G in the 5' position of the anticodon can pair with either C or U in the 3' position of the codon. Thus two codons can be recognized by the same tRNA. (b) Because of wobble, only three tRNA anticodons are needed to translate the four glycine (Gly) codons.



**Figure 11.39 Coupled Transcription and Translation in Prokaryotes.** (a) A transmission electron micrograph showing coupled transcription and translation. (b) A schematic representation of coupled transcription and translation. As the DNA is transcribed, ribosomes bind the free 5' end of the mRNA. Thus translation is started before transcription is completed. Note that there are multiple ribosomes bound to the mRNA, forming a polyribosomal complex.

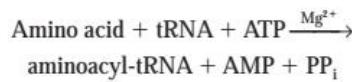


**Figure 11.40 tRNA Structure.**  
The cloverleaf structure for tRNA in prokaryotes and eucaryotes. Bases found in all tRNAs are in diamonds; purine and pyrimidine positions in all tRNAs are labeled Pu and Py respectively.

other large arms: the D or DHU arm has the unusual pyrimidine nucleoside dihydrouridine; and the T or T $\Psi$ C arm has ribothymidine (T) and pseudouridine ( $\Psi$ ), both of which are unique to tRNA. Finally, the cloverleaf has a variable arm whose length changes with the overall length of the tRNA; the other arms are fairly constant in size.

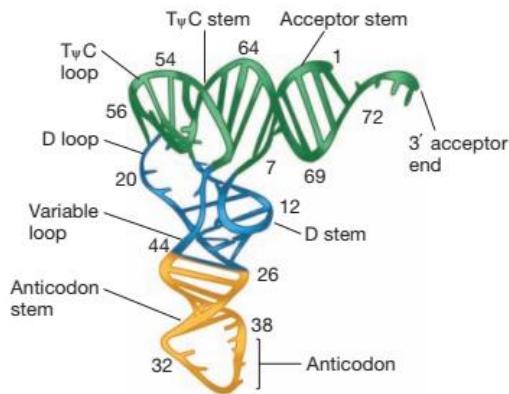
Transfer RNA molecules are folded into an L-shaped structure (figure 11.41). The amino acid is held on one end of the L, the anticodon is positioned on the opposite end, and the corner of the L is formed by the D and T loops. Because there must be at least one tRNA for each of the 20 amino acids incorporated into proteins, at least 20 different tRNA molecules are needed. Actually more tRNA species exist.

Amino acids are activated for protein synthesis through a reaction catalyzed by **aminoacyl-tRNA synthetases** (figure 11.42).

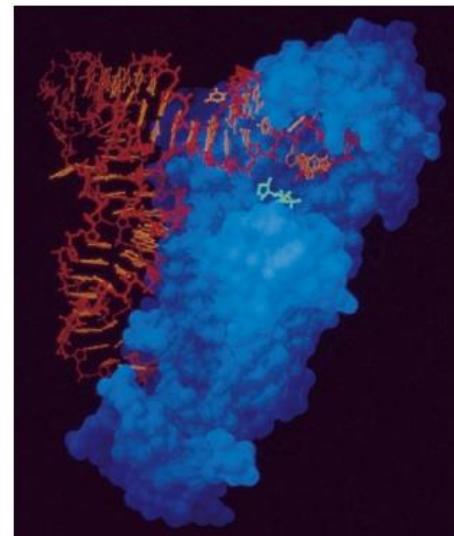


Just as is true of DNA and RNA synthesis, the reaction is driven to completion when the pyrophosphate product is hydrolyzed to two orthophosphates. The amino acid is attached to the 3'-hydroxyl of the terminal adenylic acid on the tRNA by a high-energy bond (figure 11.43), and is readily transferred to the end of a growing peptide chain. This is why the amino acid is said to be activated.

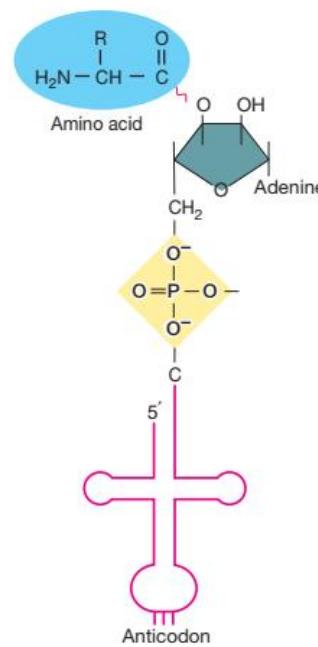
There are at least 20 aminoacyl-tRNA synthetases, each specific for a single amino acid and its tRNAs (cognate tRNAs). It is critical that each tRNA attach the corresponding amino acid because if an incorrect amino acid is attached to a tRNA, it will be incorporated into a polypeptide in place of the correct amino acid. The protein synthetic machinery recognizes only the anticodon of the aminoacyl-tRNA and cannot tell whether the correct amino acid is attached. Some aminoacyl-tRNA synthetases proofread just like DNA polymerases do. If the wrong amino acid is attached to tRNA, the enzyme hydrolyzes the amino acid from the tRNA rather than release the incorrect product.



**Figure 11.41 Transfer RNA Conformation.** The three-dimensional structure of tRNA. The various regions are distinguished with different colors.



**Figure 11.42 An Aminoacyl-tRNA Synthetase.** A model of *E. coli* glutamyl-tRNA synthetase complexed with its tRNA and ATP. The enzyme is in blue, the tRNA in red and yellow, and ATP in green.



**Figure 11.43 Aminoacyl-tRNA.** The amino acid is attached to the 3'-hydroxyl of adenylic acid by a high-energy bond (red).

### The Ribosome

The actual process of protein synthesis takes place on ribosomes that serve as workbenches, with mRNA acting as the blueprint. Prokaryotic ribosomes have a sedimentation value of 70S and a mass of 2.8 million daltons. A rapidly growing *E. coli* cell may have as many as 15,000 to 20,000 ribosomes, about 15% of the cell mass.

Each of the two subunits of the prokaryotic ribosome is an extraordinarily complex structure constructed from one or two rRNA molecules and many polypeptides (figure 11.44). The shape of the subunits and their association to form the 70S ribosome are depicted in figure 11.45. The region of the ribosome directly responsible for translation is called the translational domain (figure 11.45d). Both subunits contribute to this domain. The growing peptide chain emerges from the large subunit at the exit domain. This is located on the side of the subunit opposite the central protuberance (figure 11.45b). X-ray diffraction studies have now confirmed this general picture of ribosome structure (figure 11.45e–g).

Ribosomal RNA is thought to have three roles. It obviously contributes to ribosome structure. The 16S rRNA of the 30S subunit is needed for the initiation of protein synthesis in *Bacteria*. There is evidence that the 3' end of the 16S rRNA complexes with a site on the mRNA called the Shine-Dalgarno sequence, which is located in the **ribosome-binding site (RBS)**. This helps position the mRNA on the ribosome. The 16S rRNA also binds a protein needed to initiate translation, initiation factor 3 and the 3' CCA end of aminoacyl-tRNA. Finally, it appears that the 23S rRNA has a catalytic role in protein synthesis.

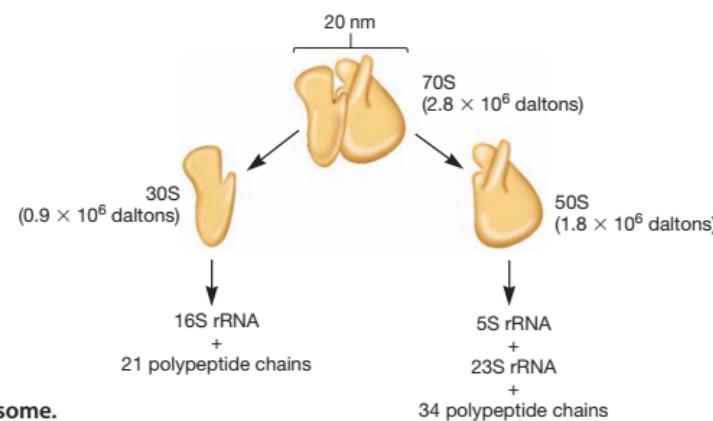
### Initiation of Protein Synthesis

Like transcription, protein synthesis may be divided into three stages: initiation, elongation, and termination. In the initiation stage, *E. coli* and most bacteria begin protein synthesis with a specially modified aminoacyl-tRNA, *N*-formylmethionyl-tRNA<sup>fMet</sup> (figure 11.46). Because the  $\alpha$ -amino is blocked by a formyl group, this aminoacyl-tRNA can be used only for initiation. When methionine is to be added to a growing polypeptide chain, a normal methionyl-tRNA<sup>Met</sup> is employed. Eucaryotic protein synthesis

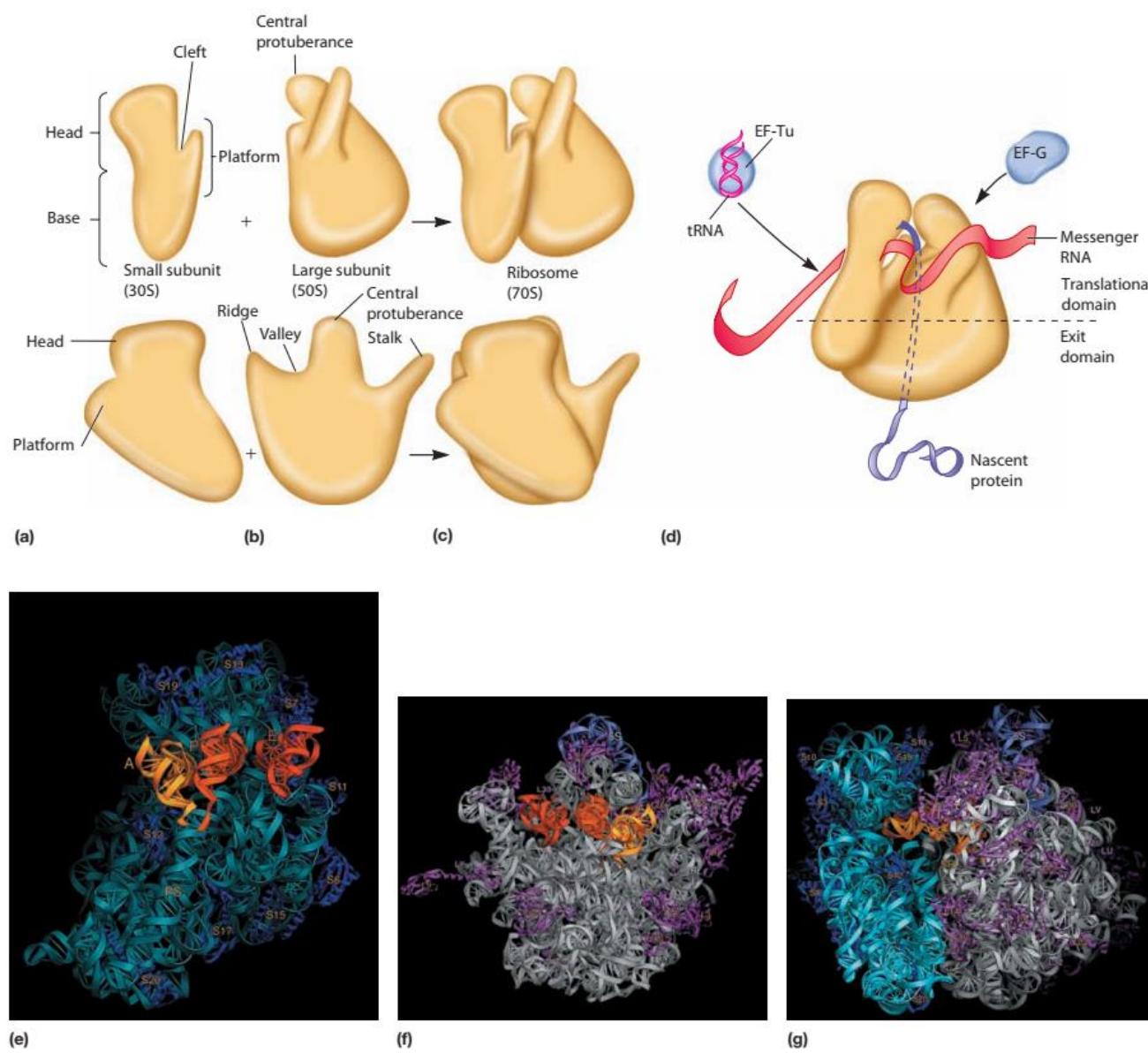
(except in the mitochondrion and chloroplast) and archaeal protein synthesis begin with a special initiator methionyl-tRNA<sup>Met</sup>. Although most bacteria start protein synthesis with formylmethionine, the formyl group does not remain but is hydrolytically removed. In fact, one to three amino acids may be removed from the amino terminal end of the polypeptide after synthesis.

The initiation stage is crucial for the translation of the mRNA into the correct polypeptide (figure 11.47). In *Bacteria*, it begins when initiator *N*-formylmethionyl-tRNA<sup>fMet</sup> (fMet-tRNA) binds to a free 30S ribosomal subunit. As noted earlier, the 30S subunit possesses a molecule of 16S rRNA with nucleotide sequences that are complementary to the Shine-Dalgarno sequence in the leader sequence of the mRNA. Recall that the leader is transcribed but not translated. This is because the role of the leader sequence is to align the mRNA with complementary bases on the 16S rRNA of the 30S ribosomal subunit such that the codon for the initiator fMet-tRNA is translated first. Messenger RNAs have a special **initiator codon** (AUG or sometimes GUG) that specifically binds with the fMet-tRNA anticodon. Finally, the 50S subunit binds to the 30S subunit-mRNA forming an active ribosome-mRNA complex. The fMet-tRNA must be positioned at the peptidyl or P site (see description of the elongation cycle). There is some uncertainty about the exact initiation sequence, and mRNA may bind before fMet-tRNA in *Bacteria*. Eucaryotic and archaeal initiation appears to begin with the binding of the initiator Met-tRNA to the small subunit, followed by attachment of the mRNA.

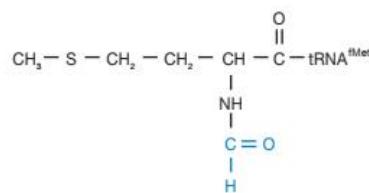
In *Bacteria*, three protein **initiation factors** are required (figure 11.47). Initiation factor 3 (IF-3) prevents 30S subunit binding to the 50S subunit and promotes the proper mRNA binding to the 30S subunit. IF-2, the second initiation factor, binds GTP and fMet-tRNA and directs the attachment of fMet-tRNA to the P site of the 30S subunit. GTP is hydrolyzed during association of the 50S and 30S subunits. The third initiation factor, IF-1, appears to be needed for release of IF-2 and GDP from the completed 70S ribosome. IF-1 may aid in the binding of the 50S subunit to the 30S subunit. It also blocks tRNA binding to the A site. Eucaryotes require more initiation factors; otherwise the process is quite similar to that of *Bacteria*.



**Figure 11.44** The 70S Ribosome.



**Figure 11.45 Prokaryotic Ribosome Structure.** Parts (a)–(d) illustrate *E. coli* ribosome organization; parts (e)–(g) show the molecular structure of the *Thermus thermophilus* ribosome. (a) The 30S subunit. (b) The 50S subunit. (c) The complete 70S ribosome. (d) A diagram of ribosomal structure showing the translational and exit domains. The locations of elongation factor and mRNA binding are indicated. The growing peptide chain probably remains unfolded and extended until it leaves the large subunit. (e) Interior interface view of the 30S subunit of the *T. thermophilus* 70S ribosome showing the positions of the A, P, and E site tRNAs. (f) Interior interface view of the *T. thermophilus* 50S subunit and portions of its three tRNAs. (g) The complete *T. thermophilus* 70S ribosome viewed from the right-hand side with the 30S subunit on the left and the 50S subunit on the right. The anticodon arm of the A site tRNA is visible in the interface cavity. The components in figures (e)–(g) are colored as follows: 16S rRNA, cyan; 23S rRNA, gray; 5S rRNA, light blue; 30S proteins, dark blue; 50S proteins, magenta; and A, P, and E site tRNAs (gold, orange, and red, respectively).

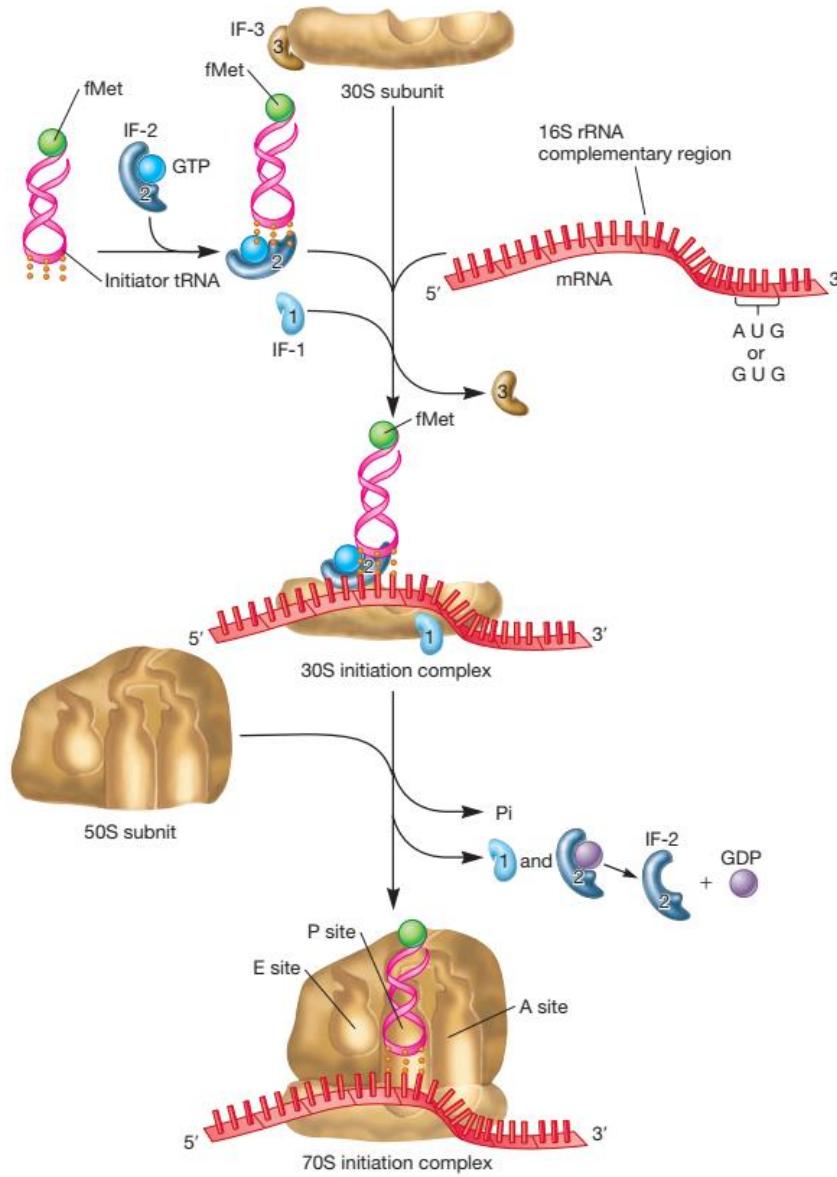


**Figure 11.46 Bacterial Initiator tRNA.** The initiator aminoacyl-tRNA, *N*-formylmethionyl-tRNA<sup>fMet</sup>, is used by *Bacteria*. The formyl group is in color. *Archaea* and eucaryotes use methionyl-tRNA for initiation.

The initiation of protein synthesis is very elaborate. Apparently the complexity is necessary to ensure that the ribosome does not start synthesizing a polypeptide chain in the middle of a gene—a disastrous error.

#### Elongation of the Polypeptide Chain

Every amino acid addition to a growing polypeptide chain is the result of an **elongation cycle** composed of three phases: aminoacyl-tRNA binding, the transpeptidation reaction, and translocation. The process is aided by special protein **elongation factors** (just as with the initiation of protein synthesis). In each turn of the cycle, an



**Figure 11.47 Initiation of Protein Synthesis.** The initiation of protein synthesis in *Bacteria*. The following abbreviations are employed: IF-1, IF-2, and IF-3 stand for initiation factors 1, 2, and 3; initiator tRNA is *N*-formylmethionyl-tRNA<sup>fMet</sup>. The ribosomal locations of initiation factors are depicted for illustration purposes only. They do not represent the actual initiation factor binding sites. See text for further discussion.

amino acid corresponding to the proper mRNA codon is added to the C-terminal end of the polypeptide chain. The bacterial elongation cycle is described next.

The ribosome has three sites for binding tRNAs: (1) the **peptidyl or donor site** (the **P site**), (2) the **aminoacyl or acceptor site** (the **A site**), and (3) the **exit site** (the **E site**). At the beginning of an elongation cycle, the peptidyl site is filled with either *N*-formylmethionyl-tRNA<sup>fMet</sup> or peptidyl-tRNA, and the aminoacyl and exit sites are empty (figure 11.48). Messenger RNA is bound to the ribosome in such a way that the proper codon interacts with the P site tRNA (e.g., an AUG codon for fMet-tRNA). The next codon is located within the A site and is ready to accept an aminoacyl-tRNA.

The first phase of the elongation cycle is the aminoacyl-tRNA binding phase. The aminoacyl-tRNA corresponding to the codon in the A site is inserted so its anticodon is aligned with the codon on the mRNA. GTP and the elongation factor EF-Tu, which carries the aminoacyl-tRNA to the ribosome, are required for this insertion. When GTP is bound to EF-Tu, the protein is in its active state and delivers aminoacyl-tRNA to the A site. This is followed by GTP hydrolysis, and the EF-Tu · GDP complex leaves the ribosome. EF-Tu · GDP is converted to EF-Tu · GTP with the aid of a second elongation factor, EF-Ts. Subsequently another aminoacyl-tRNA binds to EF-Tu · GTP (figure 11.48).

Aminoacyl-tRNA binding to the A site initiates the second phase of the elongation cycle, the transpeptidation reaction (figure 11.48 and figure 11.49). This is catalyzed by the 23S rRNA ribozyme activity called **peptidyl transferase**, located on the 50S subunit. The  $\alpha$ -amino group of the A site amino acid nucleophilically attacks the  $\alpha$ -carboxyl group of the C-terminal amino acid on the P site tRNA (figure 11.49). The peptide chain attached to the tRNA in the P site is transferred to the A site as a peptide bond is formed between the chain and the incoming amino acid. No extra energy source is required for peptide bond formation because the bond linking an amino acid to tRNA is high in energy (figure 11.43). Evidence strongly suggests that 23S rRNA contains the peptidyl transferase function, and is therefore a ribozyme. Almost all protein can be removed from the 50S subunit, leaving the 23S rRNA and protein fragments and the remaining complex still has peptidyl transferase activity. The high-resolution structure of the large subunit has been obtained by X-ray crystallography. There is no protein in the active site region. A specific adenine base seems to participate in catalyzing peptide bond formation. Thus the 23S rRNA appears to be the major component of the peptidyl transferase and contributes to both A and P site functions.

The final phase in the elongation cycle is **translocation**. Three things happen simultaneously: (1) the peptidyl-tRNA moves about 20 Å from the A site to the P site; (2) the ribosome moves one codon along mRNA so that a new codon is positioned in the A site; and (3) the empty tRNA leaves the P site. Instead of immediately being ejected from the ribosome when the ribosome moves along the mRNA, the empty tRNA is moved from the P site to the E site and then leaves the ribosome. Ribosomal proteins are involved in these tRNA movements. The intricate process also requires the participation of the EF-G or translocase

protein and GTP hydrolysis. The ribosome changes shape as it moves down the mRNA in the 5' to 3' direction.

### Termination of Protein Synthesis

Protein synthesis stops when the ribosome reaches one of three nonsense codons—UAA, UAG, and UGA (figure 11.50). The nonsense (stop) codon is found on the mRNA immediately before the trailer region. Three **release factors** (RF-1, RF-2, and RF-3) aid the ribosome in recognizing these codons. Because there is no cognate tRNA for a nonsense codon, the ribosome stops. The peptidyl transferase hydrolyzes the peptide free from the tRNA in the P site, and the empty tRNA is released. GTP hydrolysis seems to be required during this sequence, although it may not be needed for termination in *Bacteria*. Next the ribosome dissociates from its mRNA and separates into 30S and 50S subunits. IF-3 binds to the 30S subunit to prevent it from reassociating with the 50S subunit until the proper stage in initiation is reached. Thus ribosomal subunits associate during protein synthesis and separate afterward. The termination of eucaryotic protein synthesis is similar except that only one release factor appears to be active.

Protein synthesis is a very expensive process. Three GTP molecules probably are used during each elongation cycle, and two ATP high-energy bonds are required for amino acid activation (ATP is converted to AMP rather than to ADP). Therefore five high-energy bonds are required to add one amino acid to a growing polypeptide chain. GTP also is used in initiation and termination of protein synthesis (figures 11.47 and 11.50). Presumably this large energy expenditure is required to ensure the fidelity of protein synthesis. Very few mistakes can be tolerated.

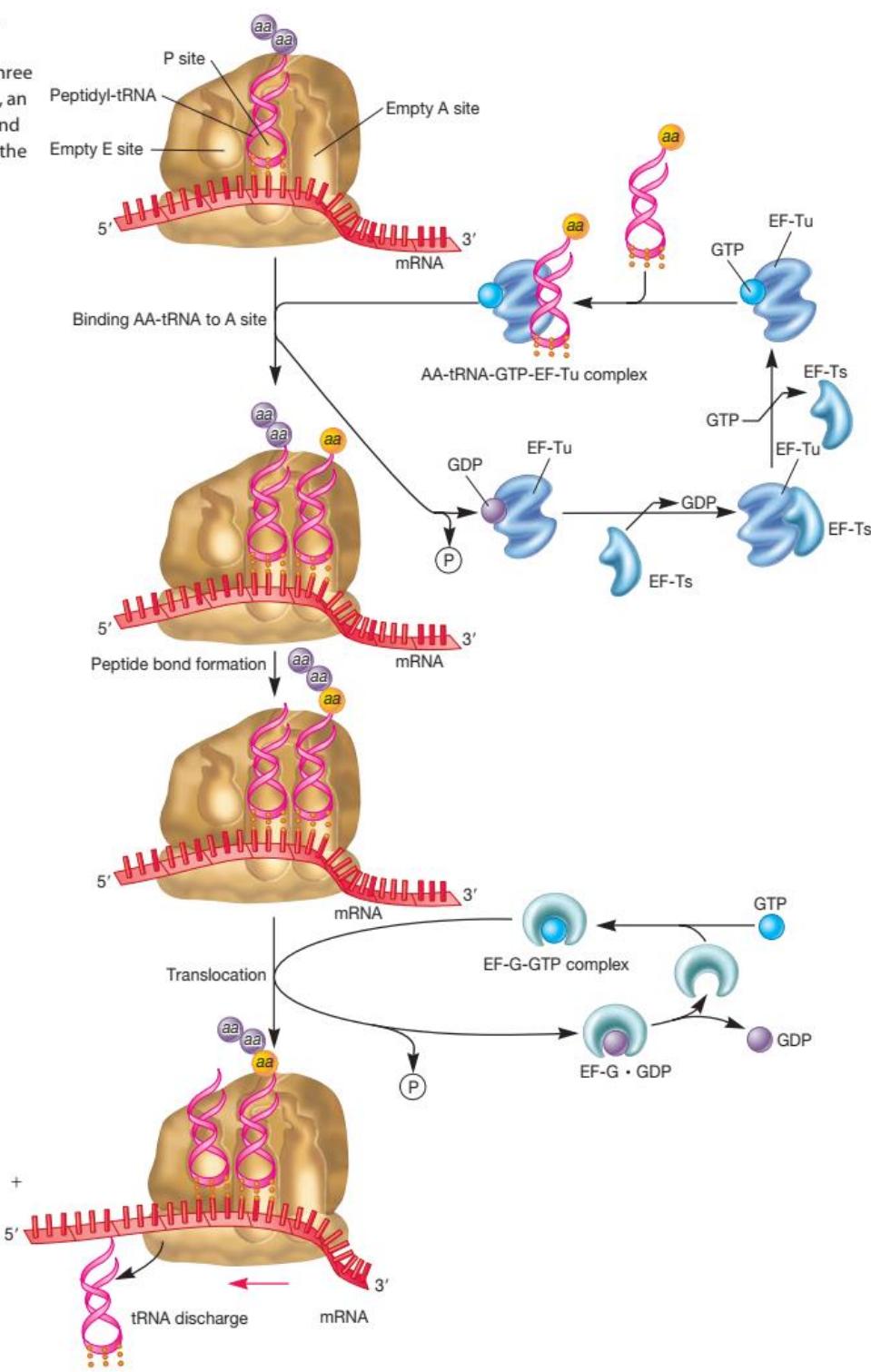
Although the mechanism of protein synthesis is similar in *Bacteria* and eucaryotes, bacterial ribosomes differ substantially from those in eucaryotes. This explains the effectiveness of many important antibacterial agents. Either the 30S or the 50S subunit may be affected. For example, streptomycin binding to the 30S ribosomal subunit inhibits protein synthesis and causes mRNA misreading. Erythromycin binds to the 50S subunit and inhibits peptide chain elongation. [Antibacterial drugs \(section 34.4\)](#)

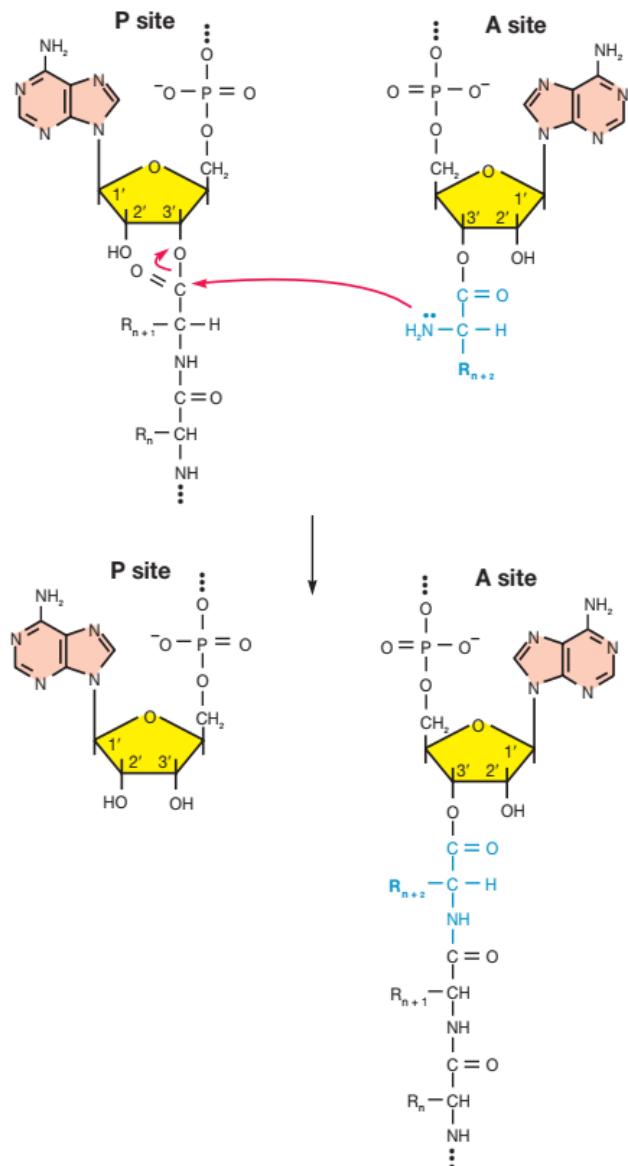
### Protein Folding and Molecular Chaperones

For many years it was believed that polypeptides spontaneously folded into their final native shape, either as they were synthesized by ribosomes or shortly after completion of protein synthesis. Although the amino acid sequence of a polypeptide does determine its final conformation, it is now clear that special helper proteins aid the newly formed or nascent polypeptide in folding to its proper functional shape. These proteins, called **molecular chaperones** or chaperones, recognize only unfolded polypeptides or partly denatured proteins and do not bind to normal, functional proteins. Their role is essential because the cytoplasmic matrix is filled with nascent polypeptide chains and proteins. Under such conditions it is quite likely that new polypeptide chains often will fold improperly and aggregate to form nonfunctional complexes. Molecular chaperones suppress incorrect folding and may reverse

**Figure 11.48 Elongation Cycle.**

The elongation cycle of protein synthesis. The ribosome possesses three sites, a peptidyl or donor site (P site), an aminoacyl or acceptor site (A site), and an exit site (E site). The arrow below the ribosome in the translocation step shows the direction of mRNA movement. See text for details.

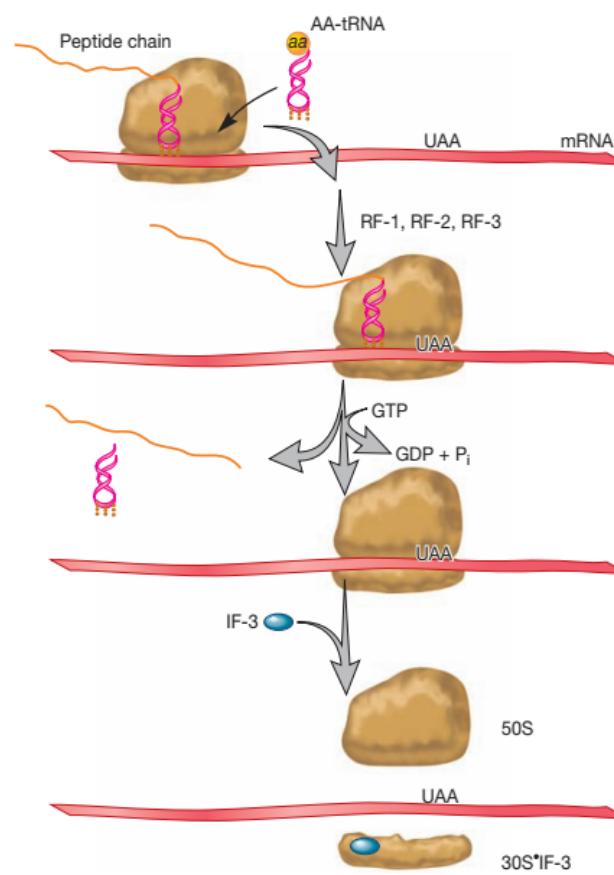




**Figure 11.49 Transpeptidation.** The peptidyl transferase reaction. The peptide grows by one amino acid and is transferred to the A site.

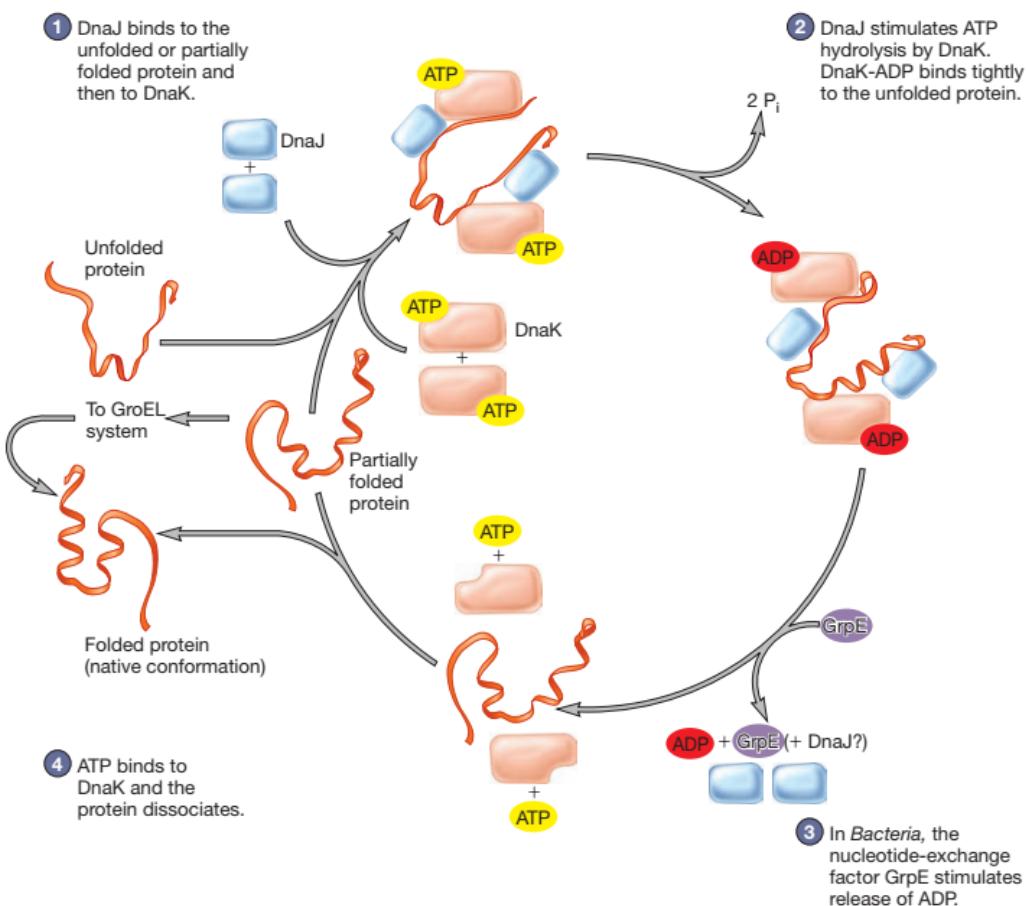
any incorrect folding that has already taken place. They are so important that chaperones are present in all cells.

Several chaperones and cooperating proteins aid proper protein folding in *Bacteria*. The process has been well studied in *E. coli* and involves at least four chaperones—DnaK, DnaJ, GroEL, and GroES—and the stress protein GrpE. After a sufficient length of nascent polypeptide extends from the ribosome, DnaJ binds to the unfolded chain (figure 11.51). DnaK, which is complexed with ATP, then attaches to the polypeptide. These



**Figure 11.50 Termination of Protein Synthesis in Bacteria.** Although three different nonsense codons can terminate chain elongation, UAA is most often used for this purpose. Three release factors (RF) assist the ribosome in recognizing nonsense codons and terminating translation. GTP hydrolysis is probably involved in termination. Transfer RNAs are in pink.

two chaperones prevent the polypeptide from folding improperly as it is synthesized. When synthesis of the polypeptide is complete, the GrpE protein binds to the chaperone-polypeptide complex and causes DnaK to release ADP. DnaJ may also be released at this step. Then ATP binds to DnaK and DnaK dissociates from the polypeptide. The polypeptide has been folding during this sequence of events and may have reached its final native conformation. If it is still only partially folded, it can bind DnaJ and DnaK again and repeat the process, or be transferred to another set of chaperones, GroEL and GroES, where the final folding takes place. As with DnaK, ATP binding to GroEL and ATP hydrolysis change the chaperone's affinity for the folding polypeptide and regulate polypeptide binding and release (polypeptide release is ATP-dependent). GroES binds to GroEL and assists in its binding and release of the refolding polypeptide.



**Figure 11.51 Chaperones and Polypeptide Folding.** The involvement of bacterial chaperones in the proper folding of a newly synthesized polypeptide chain is depicted in this diagram. Three possible outcomes of a chaperone reaction cycle are shown. A native protein may result, the partially folded polypeptide may bind again to DnaK and DnaJ, or the polypeptide may be transferred to GroEL and GroES.

Chaperones were first discovered because they dramatically increase in concentration when cells are exposed to high temperatures, metabolic poisons, and other stressful conditions. Thus many chaperones often are called **heat-shock proteins** or stress proteins. When an *E. coli* culture is switched from 30 to 42°C, the concentrations of some 20 different heat-shock proteins increase greatly within about 5 minutes. If the cells are exposed to a lethal temperature, the heat-shock proteins are still synthesized but most proteins are not. Thus chaperones protect the cell from thermal damage and other stresses as well as promote the proper folding of new polypeptides. For example, DnaK protects *E. coli* RNA polymerase from thermal inactivation in vitro. In addition, DnaK reactivates thermally inactivated RNA polymerase, especially if ATP, DnaJ, and GrpE are present. GroEL and GroES also protect intracellular proteins from aggregation. As one would expect, large quantities of chaperones are present in hyperthermophiles such as *Pyrococcus occultum*, an archaeon that will

grow at temperatures as high as 110°C. *P. occultum* has a chaperone similar to the GroEL of *E. coli*. The chaperone hydrolyzes ATP most rapidly at 100°C and makes up almost 3/4 of the cell's soluble protein when *P. occultum* grows at 108°C.

Chaperones have other functions as well. They are particularly important in the transport of proteins across membranes. For example, in *E. coli* the chaperone SecB binds to the partially unfolded forms of many proteins and keeps them in an export-competent state until they are translocated across the plasma membrane. Proteins translocated by the Sec-dependent system are synthesized with an amino-terminal signal sequence. The signal sequence is a short stretch of amino acids that helps direct the completed polypeptide to its final destination. Polypeptides associate with SecB and the chaperone then attaches to the membrane translocase. The polypeptides are transported through the membrane as ATP is hydrolyzed. When they enter the periplasm, a signal peptidase enzyme removes the signal sequence and the

protein moves to its final location. DnaK, DnaJ, and GroEL/GroES also can aid in protein translocation across membranes.

#### Protein secretion in prokaryotes (section 3.8)

Research indicates that prokaryotes and eucaryotes may differ with respect to the timing of protein folding. In terms of conformation, proteins are composed of compact, self-folding, structurally independent regions. These regions, normally around 100 to 300 amino acids in length, are called **domains**. Larger proteins such as immunoglobulins (important proteins in the immune response) may have two or more domains that are linked by less structured portions of the polypeptide chain. In eucaryotes, domains fold independently right after being synthesized by the ribosome. It appears that prokaryotic polypeptides, in contrast, do not fold until after the complete chain has been synthesized. Only then do the individual domains fold. This difference in timing may account for the observation that chaperones seem to be more important in the folding of prokaryotic proteins. Folding a whole polypeptide is more complex than folding one domain at a time and would require the aid of chaperones. [Antibodies \(section 32.7\)](#)

### Protein Splicing

A further level of complexity in the formation of proteins has been discovered. Some microbial proteins are spliced after translation. In **protein splicing**, a part of the polypeptide is removed before the polypeptide folds into its final shape. Self-splicing proteins begin as larger precursor proteins composed of one or more internal intervening sequences called **inteins** flanked by external sequences or **exteins**, the N-exteins and C-exteins ([figure 11.52a](#)). Inteins, which are between about 130 and 600 amino acids in length, are removed in an autocatalytic process involving a branched intermediate ([figure 11.52b](#)). Thus far, more than 130 inteins in 34 types of self-splicing proteins have been discovered. Over 120 inteins have been found in bacteria and archaea. Some examples are an ATPase in the yeast *Saccharomyces cerevisiae*, the RecA protein of *Mycobacterium tuberculosis*, and DNA polymerase in the archaeon *Pyrococcus*. Thus self-splicing proteins are present in all three domains of life.

1. In which direction are polypeptides synthesized? What is a polyribosome and why is it useful?

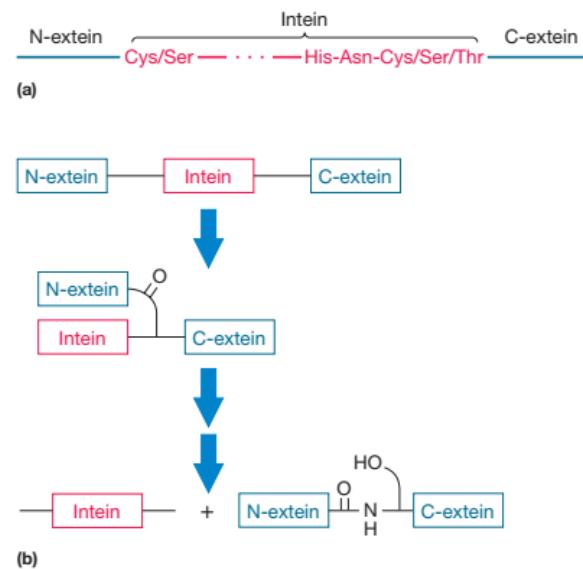
### Summary

#### 11.1 DNA as Genetic Material

- 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 ([figures 11.1–11.3](#)).

#### 11.2 The Flow of Genetic Information

- DNA serves as the storage molecule for genetic information. DNA replication is the process by which DNA is duplicated so that it can be passed on to the next generation ([figure 11.4](#)).



**Figure 11.52 Protein Splicing.** (a) A generalized illustration of intein structure. The amino acids that are commonly present at each end of the inteins are shown. Note that many are thiol or hydroxyl-containing amino acids. (b) An overview of the proposed pattern or sequence of splicing. The precise mechanism is not yet known but presumably involves the hydroxyls or thiols located at each end of the intein.

2. Briefly describe the structure of transfer RNA and relate this to its function. How are amino acids activated for protein synthesis, and why is the specificity of the aminoacyl-tRNA synthetase reaction so important?
3. What are the translational and exit domains of the ribosome? Contrast prokaryotic and eucaryotic ribosomes in terms of structure. What roles does ribosomal RNA have?
4. Describe the nature and function of the following: fMet-tRNA, initiator codon, IF-3, IF-2, IF-1, elongation cycle, peptidyl and aminoacyl sites, EF-Tu, EF-Ts, transpeptidation reaction, peptidyl transferase, translocation, EF-G or translocase, nonsense codon, and release factors.
5. What are molecular chaperones and heat-shock proteins? Describe their functions.

- During transcription, genetic information in DNA is rewritten as an RNA molecule. The three products of transcription are messenger RNA, ribosomal RNA, and transfer RNA.
- Translation converts genetic information in the form of a messenger RNA molecule into a polypeptide. Ribosomal RNA and transfer RNA participate in the decoding of genetic information during translation.

#### 11.3 Nucleic Acid Structure

- DNA differs in composition from RNA in having deoxyribose and thymine rather than ribose and uracil.

- b. 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**).
- c. RNA is normally single stranded, although it can coil upon itself and base pair to form hairpin structures.
- d. In almost all prokaryotes DNA exists as a closed circle that is twisted into supercoils. In *Bacteria*, the DNA is associated with basic proteins but not with histones.
- e. 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. The DNA of many archaea is complexed with archaeal histones (**figure 11.9**).

#### 11.4 DNA Replication

- a. Most circular prokaryotic DNAs are copied by two replication forks moving around the circle to form a theta-shaped (θ) figure (**figure 11.11**). Sometimes a rolling-circle mechanism is employed instead (**figure 11.12**).
- b. Eucaryotic DNA has many replicons and replication origins located every 10 to 100 μm along the DNA (**figure 11.13**).
- c. The replisome is a huge complex of proteins and is responsible for DNA replication.
- d. 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.
- e. The double helix is unwound by helicases with the aid of topoisomerases like DNA gyrase. DNA binding proteins keep the strands separate.
- f. DNA polymerase III holoenzyme synthesizes a complementary DNA copy beginning with a short RNA primer made by the enzyme primase.
- g. The leading strand is replicated continuously, whereas DNA synthesis on the lagging strand is discontinuous and forms Okazaki fragments (**figures 11.16** and **11.17**).
- h. DNA polymerase I excises the RNA primer and fills in the resulting gap. DNA ligase then joins the fragments together (**figures 11.18** and **11.19**).
- i. Telomerase is responsible for repeating the ends of eucaryotic chromosomes (**figure 11.20**).

#### 11.5 Gene Structure

- a. A gene may be defined as the nucleic acid sequence that codes for a polypeptide, tRNA, or rRNA.
- b. The template strand of DNA carries genetic information and directs the synthesis of the RNA transcript.
- c. The gene also contains a coding region and a terminator; it may have a leader and a trailer (**figure 11.23**).
- d. RNA polymerase binds to the promoter region, which contains RNA polymerase recognition and RNA polymerase binding sites (**figure 11.28**).
- e. The genes for tRNA and rRNA often code for a precursor that is subsequently processed to yield several products (**figure 11.24**).

#### 11.6 Transcription

- a. RNA polymerase synthesizes RNA that is complementary to the DNA template strand (**figure 11.26**).
- b. The sigma factor helps the bacterial RNA polymerase bind to the promoter region at the start of a gene (**figure 11.29**).
- c. A terminator marks the end of a gene. A rho factor is needed for RNA polymerase release from some terminators (**figures 11.31** and **11.32**).
- d. In eucaryotes, RNA polymerase II synthesizes pre-mRNA, which then undergoes posttranscriptional modification by RNA cleavage and addition of a 3' poly-A sequence and a 5' cap to generate mRNA (**figure 11.36**).
- e. Many eucaryotic genes are split or interrupted genes that have exons and introns. Exons are joined by RNA splicing. Splicing involves small nuclear RNA molecules, spliceosomes, and sometimes ribozymes (**figure 11.37**).

#### 11.7 The Genetic Code

- a. Genetic information is carried in the form of 64 nucleotide triplets called codons (**table 11.4**); sense codons direct amino acid incorporation, and stop or nonsense codons terminate translation.
- b. The code is degenerate—that is, there is more than one codon for most amino acids.

#### 11.8 Translation

- a. In translation, ribosomes attach to mRNA and synthesize a polypeptide beginning at the N-terminal end. A polysome or polyribosome is a complex of mRNA with several ribosomes (**figure 11.39**).
- b. Amino acids are activated for protein synthesis by attachment to the 3' end of transfer RNAs. Activation requires ATP, and the reaction is catalyzed by aminoacyl-tRNA synthetases (**figure 11.43**).
- c. Ribosomes are large, complex organelles composed of rRNAs and many polypeptides. Amino acids are added to a growing peptide chain at the translational domain (**figure 11.45**).
- d. Protein synthesis begins with the binding of fMet-tRNA (*Bacteria*) or an initiator methionyl-tRNA<sup>Met</sup> (eucaryotes and *Archaea*) to an initiator codon on mRNA and to the two ribosomal subunits. This involves the participation of protein initiation factors (**figure 11.47**).
- e. In the elongation cycle the proper aminoacyl-tRNA binds to the A site with the aid of EF-Tu and GTP (**figure 11.48**). Then the transpeptidation reaction is catalyzed by peptidyl transferase. Finally, during translocation, the peptidyl-tRNA moves to the P site and the ribosome travels along the mRNA one codon. Translocation requires GTP and EF-G or translocase. The empty tRNA leaves the ribosome by way of the exit site.
- f. Protein synthesis stops when a nonsense codon is reached. Bacteria require three release factors for codon recognition and ribosome dissociation from the mRNA (**figure 11.50**).
- g. Molecular chaperones help proteins fold properly, protect cells against environmental stresses, and transport proteins across membranes (**figure 11.51**).
- h. Prokaryotic proteins may not fold until completely synthesized, whereas eucaryotic protein domains fold as they leave the ribosome.
- i. Some proteins are self-splicing and excise portions of themselves before folding into their final shape.

### Key Terms

alternative splicing 265	cistron 264	deoxyribonucleic acid (DNA) 252	elongation factors 283
amino acid activation 276	code degeneracy 275	DnaA proteins 260	exit (E) site 284
aminoacyl (acceptor; A) site 284	coding region 266	DNA gyrase 260	exons 273
aminoacyl-tRNA synthetases 280	codon 264	DNA ligase 262	exteins 288
anticodon 277	complementary strand 252	DNA polymerase 259	gene 251
archaeal nucleosome 253	consensus sequence 269	domains 288	genome 247
catenanes 263	core enzyme 269	elongation cycle 283	genotype 248

heat-shock proteins 287	open complex 269	replicon 257	small nuclear RNA (snRNA) 273
helicase 260	peptidyl (donor; P) site 284	replisome 260	split (interrupted) genes 273
histone 253	peptidyl transferase 284	rho factor 270	spliceosome 274
initiation factors 281	phenotype 248	ribonucleic acid (RNA) 252	stop codon 266
initiator codon 281	polyribosome 276	ribosomal RNA (rRNA) 269	telomerase 263
inteins 288	posttranscriptional modification 272	ribosome binding site (RBS) 281	template strand 265
introns 273	Pribnow box 269	ribozyme 267	terminator sequence 266
lagging strand 260	primase 260	RNA polymerase 269	topoisomerase 260
leader sequence 265	primosome 260	RNA polymerase holoenzyme 269	trailer sequence 266
leading strand 260	promoter 265	RNA splicing 273	transcription 251
major groove 252	proofreading 262	rolling-circle replication 257	transfer RNA (tRNA) 269
messenger RNA (mRNA) 251	protein splicing 288	sense codons 275	transformation 249
minor groove 252	reading frame 264	Shine-Dalgarno sequence 265	translation 251
molecular chaperones 284	release factors 284	sigma factor 269	translocation 284
nonsense codon 275	replication 251	single-stranded DNA binding proteins (SSBs) 260	transpeptidation reaction 284
nucleosome 253	replication fork 256		wobble 276
Okazaki fragment 260			

### Critical Thinking Questions

- Many scientists say that RNA was the first of the information molecules (i.e., RNA, DNA, protein) to arise during evolution. Given the information in this chapter, what evidence is there to support this hypothesis?
  - Streptomyces coelicolor* has a linear chromosome. Interestingly, there are no genes that encode essential proteins near the ends of the chromosome in this bacterium. Why do you think this is the case?
  - You have isolated several *E. coli* mutants:  
Mutant #1 has a mutation in the -10 region of the promoter of a structural gene encoding an enzyme needed for synthesis of the amino acid serine.
- Mutant #2 has a mutation in the -35 region in the promoter of the same gene.  
 Mutant #3 is a double mutant with mutations in both the -10 and -35 region of the promoter of the same gene.  
*Only* Mutant #3 is unable to make serine. Why do you think this is so?  
 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.

### Learn More

- Bao, K., and Cohen, S. N. 2004. Reverse transcriptase activity innate to DNA polymerase I and DNA topoisomerase I proteins of *Streptomyces* telomere complex. *Proc. Natl. Acad. Sci.* 101(40):14361–66.
- Borukhov, S., and Nudler, E. 2003. RNA polymerase holoenzyme: Structure, function and biological implications. *Current Opinion Microbiol.* 6:93–100.
- Brooker, R. J. 2005. *Genetics: Analysis and principles*, 2d ed. New York: McGraw-Hill.
- Gogarten, J. P.; Senejani, A. G.; Zhaxybayeva, O.; Olendzenski, L.; and Hilaro, E. 2002. Inteins: Structure, function, and evolution. *Annu. Rev. Microbiol.* 53:263–87.
- Grabowski, B., and Kelman, Z. 2003. Archaeal DNA replication: Eukaryal proteins in a bacterial context. *Annu. Rev. Microbiol.* 57:487–516.
- Hartl, F. J., and Hayer-Hartl, M. 2002. Molecular chaperones in the cytosol: From nascent chain to folded protein. *Science* 295:1852–58.
- Johnson, A., and O'Donnell, M. 2005. Cellular DNA replicases: Components and dynamics at the replication fork. *Annu. Rev. Biochem.* 74:283–315.
- Kelman, L. M., and Kelman, Z. 2004. Multiple origins of replication in archaea. *Trends Microbiol.* 12(9):399–401.
- Laursen, B. S.; Søjrensen, H. P.; Mortensen, K. K.; and Sperling-Petersen, H. U. 2005. Initiation of protein synthesis in bacteria. *Microbiol. Mol. Biol. Rev.* 69(1):101–23.
- McKee, T., and McKee, J. R. 2003. *Biochemistry: The molecular basis of life*, 3d ed. Dubuque, Iowa: McGraw-Hill.
- Narberhaus, F. 2002. Crystallin-type heatshock proteins: Socializing minichaperones in the context of a multichaperone network. *Microbiol. Mol. Biol. Rev.* 66(1):64–93.
- Nelson, D. L., and Cox, M. M. 2005. *Lehninger: Principles of biochemistry*, 4th ed. New York: W. H. Freeman.
- Neylon, C.; Kralicek, A. V.; Hill, T. M.; and Dixon, N. E. 2005. Replication termination in *Escherichia coli*: Structure and antihelicase activity of the Tus-Ter complex. *Microbiol. Mol. Biol. Rev.* 69(3):501–26.
- Paul, B. J.; Ross, W.; Gaal, T.; and Gourse, R. L. 2004. rRNA transcription in *Escherichia coli*. *Annu. Rev. Genet.* 38:749–70.
- Reeve, J. N. 2003. Archaeal chromatin and transcription. *Mol. Microbiol.* 48(3):587–98.
- Reeve, J. N.; Bailey, K. A.; Li, W-T.; Marc, F.; Sandman, K.; and Soares, D. J. 2004. Archaeal histones: Structures, stability and DNA binding. *Biochem. Soc. Trans.* 32 (part 2):227–30.
- Smogorzewska, A., and deLange, T. 2004. Regulation of telomerase by telomeric proteins. *Annu. Rev. Biochem.* 73:177–208.
- Yarus, M.; Caporaso, J. G.; and Knight, R. 2005. Origins of the genetic code: The escaped triplet theory. *Annu. Rev. Biochem.* 74:179–98.

Please visit the Prescott website at [www.mhhe.com/prescott7](http://www.mhhe.com/prescott7) for additional references.