gene structure, the genetic code, and mutation. Thus we begin by considering bacterial genetics in this chapter and the next. Because of the distinctive characteristics of eukaryotes and archaea, the replication and expression of their genomes are discussed in chapter 15.

Readiness Check:

Based on what you have learned previously, you should be able to:

- ✔ Define the terms genome, chromosome, plasmid, haploid, diploid, genotype, and phenotype
- ✔ Outline the flow of genetic information, identifying the processes used
- ✔ Draw a simple diagram that illustrates the chemical moieties found in nucleotides (section 12.6)
- ✔ State the base-pairing rules
- ✔ Describe the structure of bacterial genomes (section 3.6)

13.1 DNA as Genetic Material

After reading this section, you should be able to:

- Summarize Griffith's experiments on transformation
- Relate how the contributions of Avery, MacLeod, McCarty, Hershey, and Chase confirmed that DNA stores genetic material

Although it is now hard to imagine, it was once thought that DNA was too simple a molecule to store genetic information. It is composed of only four different nucleotides, and it seemed that a molecule of much greater complexity must house the genetic information of a cell. It was argued that proteins, being composed of 20 different amino acids, were the better candidate for this important cellular function.

The early work of Fred Griffith in 1928 on the transfer of virulence in the pathogen *Streptococcus pneumoniae*, commonly called pneumococcus, 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 (**figure 13.1**). 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 transformation. These investigators used enzymes to selectively destroy DNA, RNA, or protein in purified extracts of virulent pneumococci (S cells). 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 13.2**).

The publication of these studies by Avery, C. M. MacLeod, and M. J. McCarty in 1944 provided the first evidence that DNA carried genetic information.

Eight years later, Alfred Hershey and Martha Chase wanted to know if protein or DNA carried the genetic information of a bacterial virus called T2 bacteriophage. They performed experiments in which they made the virus's DNA radioactive with ^{32}P, or they labeled its protein coat with ^{35}S. They mixed radioactive virions with *Escherichia coli* and incubated the mixture for a few minutes. This allowed the virions to attach to *E. coli* and begin multiplying. The suspension was then agitated violently in a blender to shear off any adsorbed bacteriophage particles (**figure 13.3**). After centrifugation, radioactivity in the supernatant (where the phage particles remained) versus the bacterial cells in the pellet was determined. They found that most radioactive protein was released into the supernatant, whereas ^{32}P DNA remained within the bacteria. Since DNA entered the cells and T2 progeny were produced, the phage DNA must have been carrying the genetic information. Some luck was involved in their discovery, for the genetic material of many viruses is RNA and the researchers happened to select a DNA virus for their studies. Imagine the confusion if T2 had been an RNA virus! The controversy surrounding the nature of genetic information might have lasted considerably longer than it did.

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 17 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.

Retrieve, Infer, Apply

1. Briefly summarize the experiments of Griffith; Avery, MacLeod, and McCarty; and Hershey and Chase.
2. Explain how protein was ruled out as the molecule of genetic information storage in each of the experiments performed by these important microbiologists.

13.2 Nucleic Acid and Protein Structure

After reading this section, you should be able to:

- Draw schematic representations of DNA, RNA, and amino acids that show their major features
- Compare and contrast the structure of DNA and RNA
- Identify the covalent bonds that are used to link nucleotides together to form a nucleic acid and amino acids together to form a polypeptide

DNA, RNA, and proteins are often called informational molecules. The information exists as the sequence of monomers from which they are built. Here we describe the monomers and how they are linked together to form these important macromolecules.

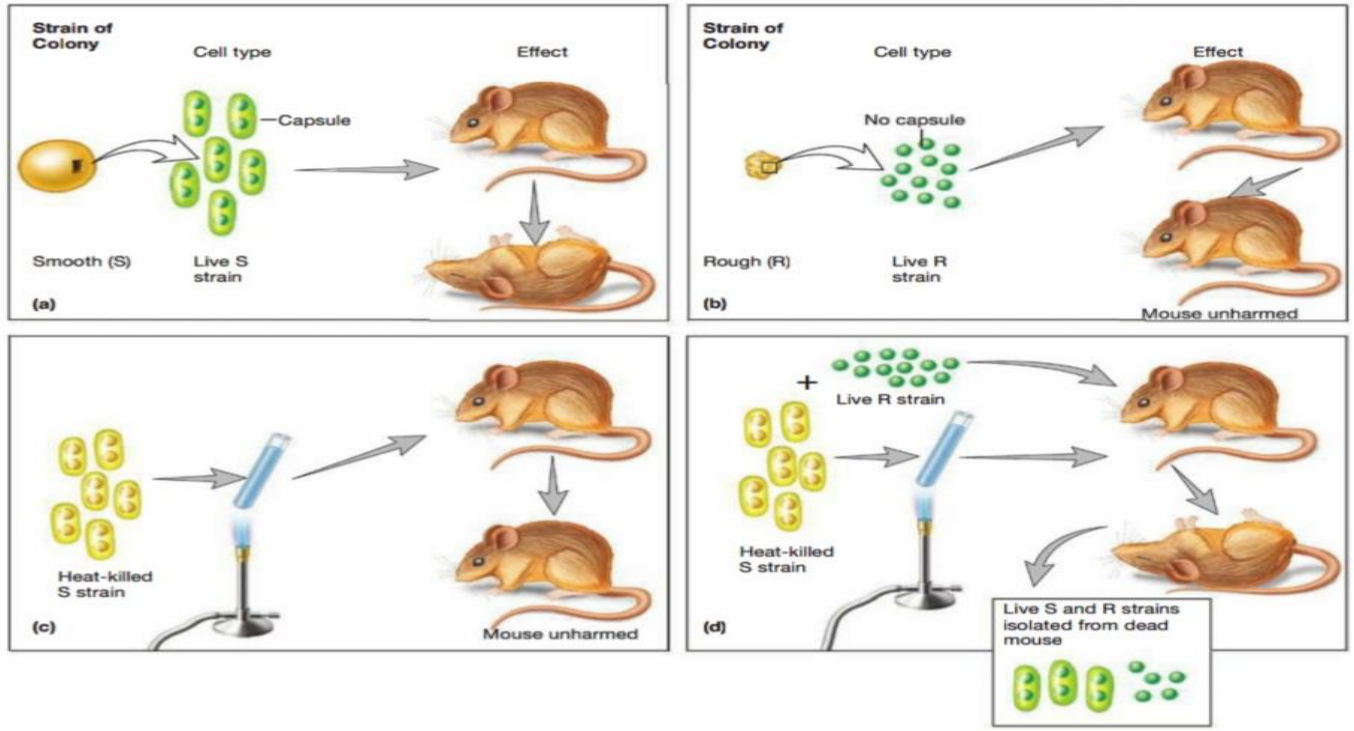


Figure 13.1 Griffith's Transformation Experiments. (a) Mice died of pneumonia when injected with pathogenic strains of pneumococci, which have a capsule and form smooth-looking colonies (S strains). (b) Mice survived when injected with nonpathogenic strains of pneumococci, which lack a capsule and form rough colonies (R strains). (c) Injection with heat-killed S strains 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.

MICRO INQUIRY *Based on what we now know about proteins, why can we conclude from this experiment that genetic information was unlikely to be carried by proteins?*

DNA Structure

Deoxyribonucleic acid (**DNA**) is a polymer of deoxyribonucleotides (**figure 13.4**) linked together by phosphodiester bonds (**figure 13.5*a***). It contains the bases adenine, guanine, cytosine, and thymine. 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 13.5). The monomers of DNA are called deoxyribonucleotides because the sugar found in them is deoxyribose (figure 13.4*b*). The bond that links the monomers together to form the polymer is called a phosphodiester bond because it consists of a phosphate that forms a bridge between the 3′-hydroxyl of one sugar and the 5′-hydroxyl of an adjacent sugar. Purine and pyrimidine bases are attached to the 1′-carbon of the deoxyribose sugars, and the bases extend toward the middle of the cylinder formed by the two chains. (The numbers designating the carbons in the sugars are given a prime to distinguish them from the numbers designating the carbons and nitrogens in the nitrogenous bases.) The bases from each strand interact with those of the other strand, forming base pairs. The base pairs are

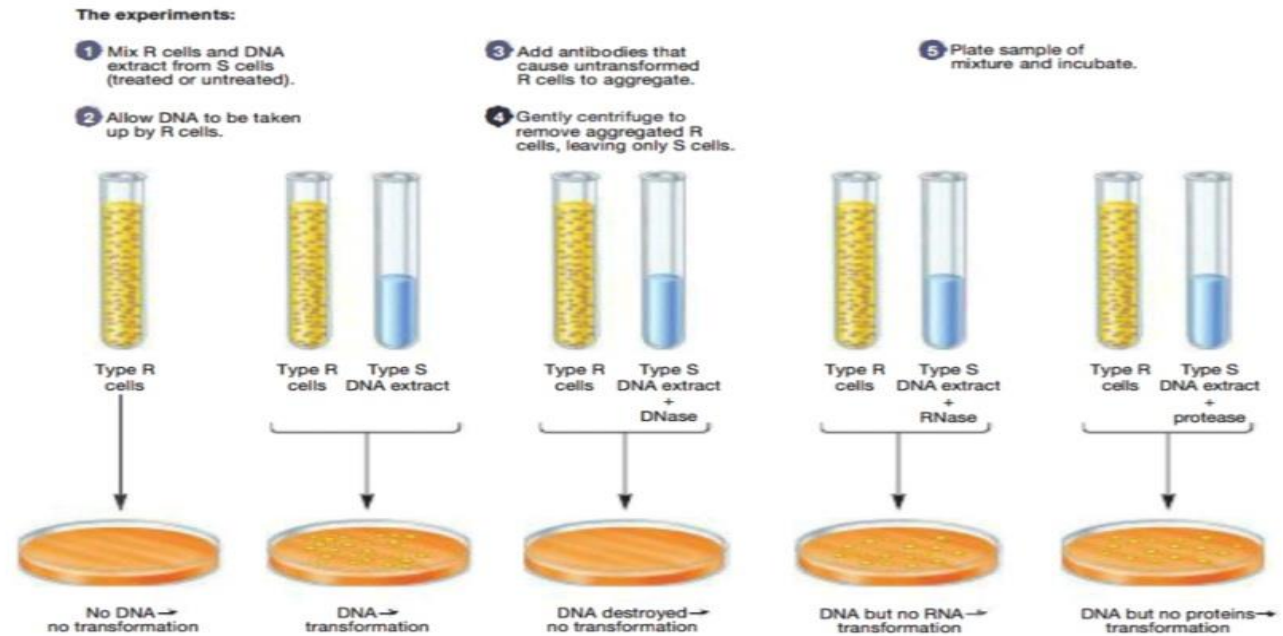


Figure 13.2 Some Experiments on the Transforming Principle. Earlier experiments done by Avery, MacLeod, and McCarty had shown that only DNA extracts from S cells caused transformation of R cells to S cells. To demonstrate that contaminating molecules in the DNA extract were not responsible for transformation, the DNA extract from S cells was treated with RNase, DNase, or protease, and then mixed with R cells. Only treatment of the DNA extract from S cells with DNase destroyed the ability of the extract to transform the R cells.

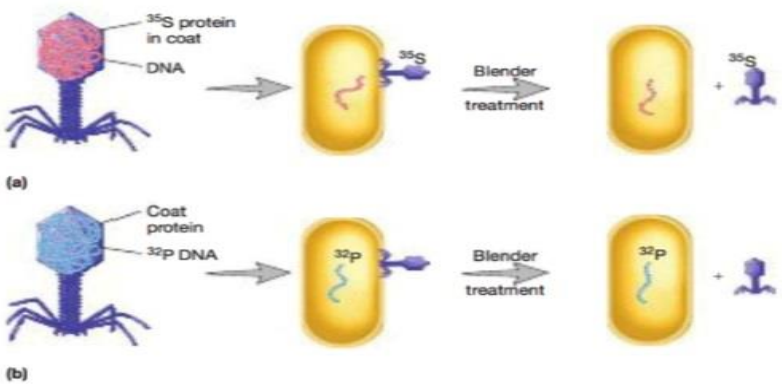


Figure 13.3 The Hershey-Chase Experiment. (a) When *E. coli* was infected with a T2 phage containing ^{35}S protein, most of the radioactivity remained outside the host cell. (b) When a T2 phage containing ^{32}P DNA was mixed with the host bacterium, the radioactive DNA was injected into the cell and phages were produced. Thus DNA was carrying the virus's genetic information.

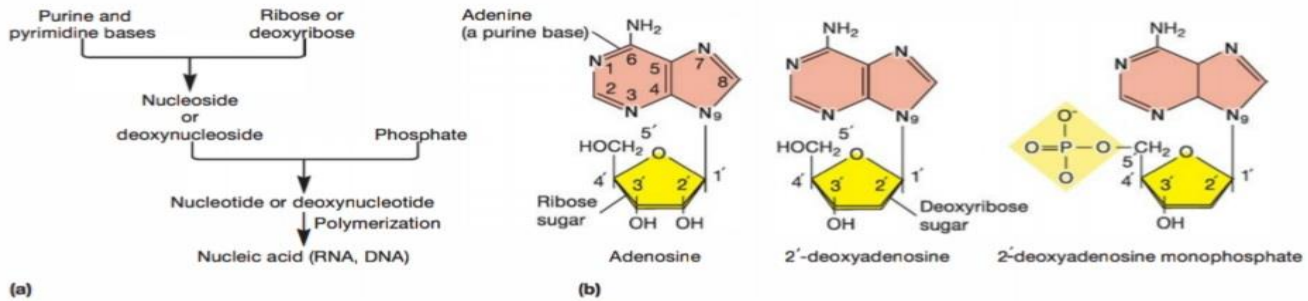


Figure 13.4 The Composition of Nucleic Acids. (a) A diagram showing the relationships of various nucleic acid components. Combination of a purine or pyrimidine base with ribose or deoxyribose gives a nucleoside (a ribonucleoside or deoxyribonucleoside). A nucleotide contains a nucleoside and one or more phosphates. 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 to distinguish them from the carbons in the bases.

MICRO INQUIRY *To which carbon of ribose (deoxyribose) is each of the following bonded: adenine and the hydroxyls of adenosine and deoxyadenosine?*

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 hundreds of microbes. ▶▶| *Microbial genomics (chapter 18)*

The two polynucleotide strands of DNA fit together much like the pieces in a jigsaw puzzle. Inspection of figure 13.5*b*,*c* shows that the two strands are not positioned directly opposite one another. Therefore when the strands twist about one another, a wide major groove and narrower minor groove are formed by the backbone. There are 10.5 base pairs per turn of the helix, and each turn of the helix has a vertical length of 3.4 Å. 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 an exposed 5′-hydroxyl group, often with phosphates attached, whereas the other end has a free 3′-hydroxyl group (figure 13.5*a*). In a given direction, one strand is oriented 5′ to 3′ and the other, 3′ to 5′ (figure 13.5*b*).

The structure of DNA just described is that of the B form, the most common form in cells. Two other forms of DNA have been identified. The A form primarily differs from the B form in that it has 11 base pairs per helical turn, rather than 10.5, and a vertical length of 2.6 Å, rather than 3.4. Thus it is wider than the B form. The Z form is dramatically different, having a left-handed helical structure, rather than right-handed as seen in the B and A forms. The Z form has 12 base pairs per helical turn and a vertical rise of 3.7 Å. Thus it is more slender than the B form. At this time, it is unclear whether the A form is found in cells. However, evidence exists that small portions of chromosomes can be in the Z form. The role, if any, for these stretches of Z DNA is unknown. ✿ *DNA Structure*

There is another property of DNA that needs to be addressed: supercoiling. DNA is helical; that is, it is a coil. Whenever the rotation of a coil is restrained in some way, it causes the coil to coil on itself. The coiling of a coil is supercoiling. Recall that most bacterial chromosomes are closed, circular double-stranded DNA molecules. In this state, the two strands are unable to rotate freely relative to each other, and the molecule is said to be strained. The strain is relieved by supercoiling. There are two types of supercoiling: positive and negative. For DNA, these are defined by the change in number of base pairs per turn in the double helix. As just discussed, the B form of DNA has 10.5 base pairs per turn of the helix. Supercoiling that decreases the number of base pairs per turn is said to be negative supercoiling. Likewise, supercoiling that increases the number of base pairs per turn is called positive supercoiling. Bacterial chromosomes are generally negatively supercoiled.

What is the importance of supercoiling? Supercoiling helps compact DNA so that it fits into the cell. Importantly for this chapter, supercoiling also "loosens" up the DNA, making it easier to separate the two strands from each other. Separation of the two strands is an important early step in both DNA replication and transcription, as we discuss in sections 13.3 and 13.5, respectively. Furthermore, positive supercoiling is often introduced into DNA during DNA replication. This can interfere with DNA replication and must be removed, as we discuss in section 13.3.

FORMS OF DNA

DNA can exist in the A, B, C and D forms. Sugar puckering is the most important characteristic for distinguishing the DNA forms. The A form has 3′-endo puckering, the B and C form 3′-exo puckering, and the C form 2′-endo puckering.

The *B form* (B-DNA) is the structure proposed by Watson and Crick, and is the native conformation of DNA in solution. The B-form X-ray diffraction pattern is obtained at high humidity (>66%) and in the presence of excess salt. It consists of a right-handed antiparallel double helix of sugar-phosphate backbone, with purine-pyrimidine base pairs roughly perpendicular to the axis of the helix. The tilt of a base pair

(γ) to the normal of the helix is 6.3°. One turn of the helix consists of 10 base pairs (10-fold helix). The rise of the helix per base pair (h) is 3.37A.

The *A form* (A-DNA) has 11 base pairs (11-fold helix). The base pairs are considerably tilted from the axis of the helix (γ = 20.2°). Because of this displacement the depth of the deep groove is increased and that of the shallow groove decreased. The axial rise (h) is only 2.56A. The B-form is observed at lesser hydration and salt content than the A form.

Reduction of hydration of the B form below 66%, with excess of salt still present, results in the *C-form* (C-DNA). The helix symmetry of the C form is 28/3, or $9\frac{1}{3}$, i.e. there are fewer residues per turn than in the B form. The axis of the helix is located in or near the minor groove, and there is a pronounced negative tilt of the base pair plane (γ=7.8°). The helix rise per base pair is 3.32A.

The *D form* poly(dA–dT) and poly(dG–dC) has 8 base pairs per turn of the helix (8-fold symmetry). As in the C form the base pairs are displaced backwards relative to the axis of the helix (γ=16.7°). The helix rise per base pair is 3.03A.

The B form is the metabolically stable configuration. It undergoes transition to the A, C or D conformations at relative humidity lower than 92%, depending upon the nature of the counterion, the nucleotide sequence and the concentration of excess salt. D→A transition has been observed at low relative humidities. The A form is, however, metastable and rapidly returns to the D form. RNA always occurs in the A form only. RNA-DNA hybrids are also in the A form leading to the belief that DNA must undergo a B→A transition during transcription.

Table 5.4. Characteristics of the different forms of DNA.

	A form	B form	C form	D form
Abbreviation	A–DNA	B–DNA	C–DNA	D–DNA
Base pairs per turn of the helix	11	10	$9\frac{1}{3}$	8
Axial rise (h)	2.5 A	3.37A	3.32A	3.03A
Tilt of base pairs (γ)	20.2°	6.3°	—7.8°	—16.7°
Pitch of the helix	28.15A	34A	31A	.....
Sugar puckering	3′–endo	3′–exo	2′–endo	3′–exo

RL helix and Z-DNA

According to the Watson-Crick model DNA exists in the form of a right handed double helix. Recently G. A. Rodley's group working in

99. REPLICATION IS A SEMI-CONSERVATIVE PROCESS

Watson and Crick were aware that any model of DNA structure should be able to explain replication. Delbruck suggested that the Watson-Crick model of DNA could theoretically replicate by three modes, *conservative*, *semi-conservative* and *dispersive*. (Fig. 8.4).

(1) According to the *conservative* mode, of the two double helices formed one would be entirely of old material and the other entirely of new material. Thus the old parent double helix would be unchanged.

(2) According to the *semi-conservative* mode proposed by Watson and Crick, each strand of the two double helices formed would have one old and one new strand.

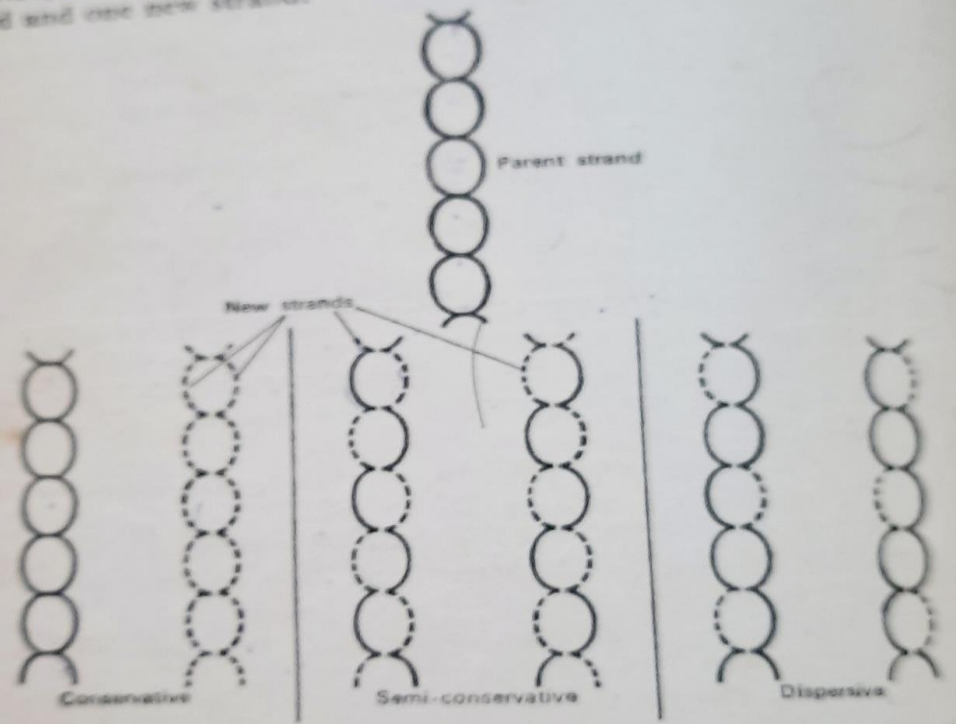


Fig. 8.4. The three theoretically possible modes of replication, conservative, semi-conservative and dispersive.

(3) According to the *dispersive* method of replication the DNA double helix would break at several points forming many pieces. Each piece would replicate, and then the pieces would reconnect at random. Thus the two double helices formed would have a patchwork of old and new pieces.

Taylor *et al* (1957) demonstrated by autoradiography that both chromatids during prophase have half old and half new material. This indicated that DNA replication was semi-conservative. The work of Meselson and Stahl (1958) has conclusively demonstrated the semi-conservative nature of replication (Fig. 8.5). *Escherichia coli* bacteria were grown for several generations in a medium containing "heavy" nitrogen (N_{15}), an istope of nitrogen. All the nitrogen of the bacteria, including that of DNA, became N_{15}. This DNA is heavier than ordinary DNA, from which it can be distinguished by an ultracentrifuge. The labelled N_{15} cells were now grown in ordinary "light" N_{14} media, an allowed to divide several times. After the first division the DNA was extracted and all of it was found to be a hybrid ($N_{14}+N_{15}$). This hybrid was not as heavy as N_{15} nor as light as N_{14}, but had an intermediate density. After the second division two kinds of DNA were found, normal N_{14} DNA (half) and hybrid $N_{14}+N_{15}$ DNA (half). After the third divison 3/4 of the DNA was normal N_{14} and 1/4 was hybrid $N_{14}+N_{15}$.

If DNA replicated *conservatively* one would expect to find two layers, one of N_{14} and the other of N_{15}, in the first generation, and similarly

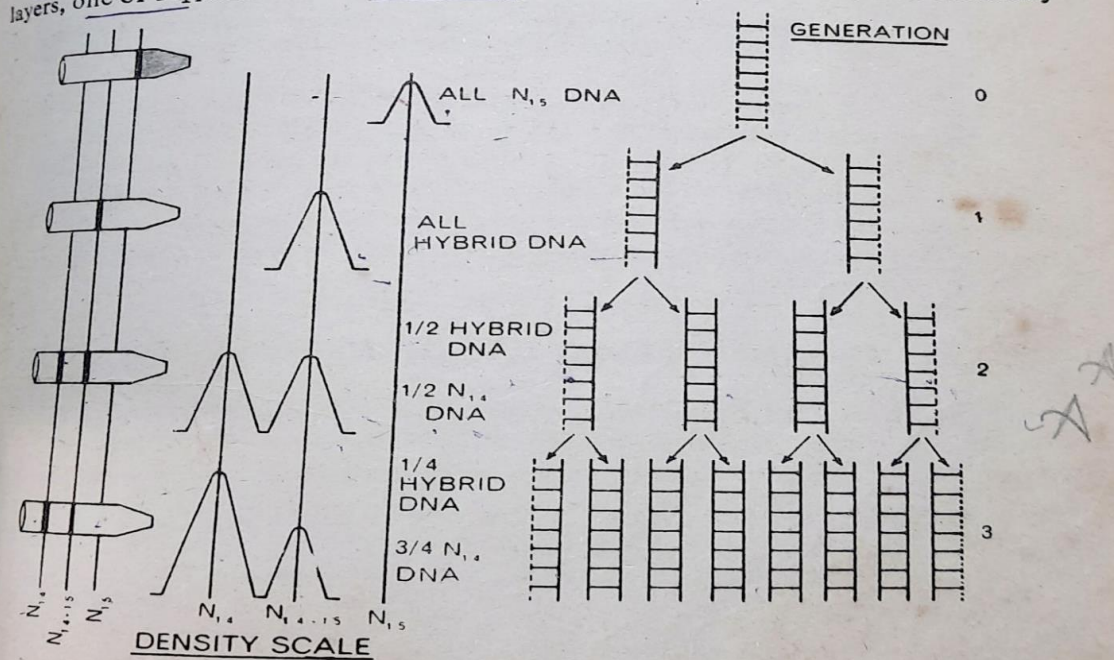


Fig. 8.5. The Meselson-Stahl experiment as evidence for the semi-conservative nature of DNA replication.

for subsequent generations (Fig. 8.6). With *dispersive* replication, tubes of all generations would be expected to show a single layer ($N_{14}+N_{15}$), since the DNA would contain both new and old material

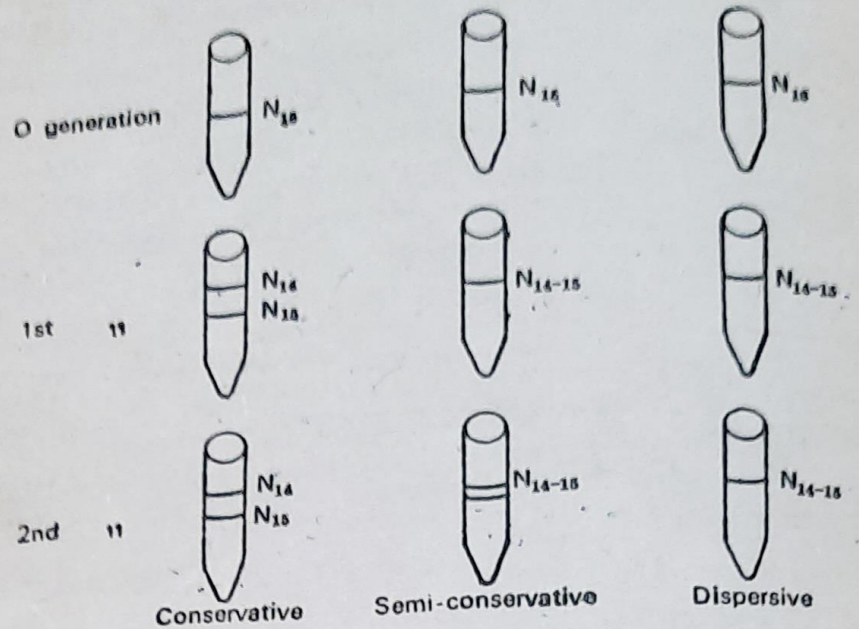


Fig. 8.6. The Meselson-Stahl experiment : expected results from conservative, semi-conservative and dispersive modes of replication.

mixed up. In *semi-conservative* replication the first generation would be expected to show a hybrid $N_{14}+N_{15}$ layer. With each generation after the second the N_{14} layer would would show a greater accumulation of material. Actual observations correspond to this expectation. This shows that *replication of DNA is of the semi-conservative manner proposed by Waston and Crick, i.e. that the double strands formed are identical to the parent strand.*

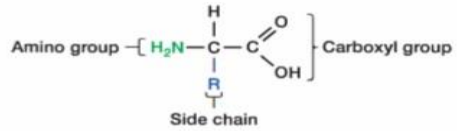


Figure 13.6 The General Structure of an Amino Acid. All amino acids have a central carbon (the α carbon) to which are attached a carboxyl group (COOH), an amino group (NH_2), and a side chain (R). Amino acids differ in terms of their side chains, which may be nonpolar, polar, negatively charged (acid), or positively charged (basic).

charged. The peptide bonds linking the amino acids together are formed by a reaction between the carboxyl group of one amino acid and the amino group of the next amino acid in the protein (**figure 13.7**). A polypeptide has polarity just as DNA and RNA do. At one end of the chain is an amino group, and at the other end is a carboxyl group. Thus a polypeptide has an amino or N terminus and a carboxyl or C terminus.

Proteins do not typically exist as extended chains of amino acids. Rather, they fold back on themselves to form three-dimensional structures, often more or less spherical in shape. The final shape is determined to a large extent by the sequence of amino acids in the polypeptide. This sequence is called the primary structure. Secondary and tertiary structures result from the folding of the chain. Finally, two or more polypeptide strands can interact to form the final, functional protein. This level of structure is called quaternary structure. These higher levels of structure are stabilized by intra- (and inter-) chain bonds. Protein structure is described in more detail in Appendix I.

Retrieve, Infer, Apply

1. What are nucleic acids? How do DNA and RNA differ in structure?
2. What does it mean to say that the two strands of the DNA double helix are complementary and antiparallel? Examine figure 13.5*b* and explain the differences between the minor and major grooves.
3. Amino acids are described as nonpolar, polar, and charged, depending on the molecular makeup of their side chains. Which type of amino acid might be found in the transmembrane portion of a polypeptide located in a cell's plasma membrane? Explain your answer.

13.3 DNA Replication in Bacteria

After reading this section, you should be able to:

- Describe a bacterial replicon
- Summarize the events that occur during the three phases of DNA replication
- Create a table or concept map that illustrates the function of the major proteins found in a bacterial replisome
- List the enzymatic and structural elements needed by DNA polymerases for DNA synthesis
- Outline the major events that occur at the replication fork

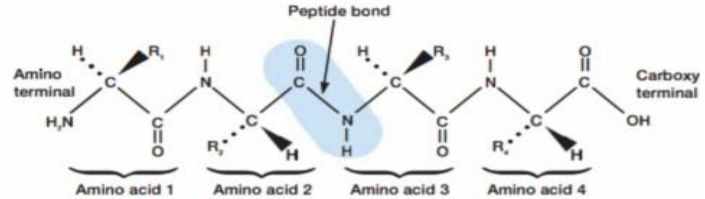


Figure 13.7 Peptide Bonds Link Amino Acids Together in Peptide Chains. A tetrapeptide chain is shown. One of the peptide bonds linking the four amino acids together is highlighted in blue. At one end of the peptide is an amino group (amino or N terminal); at the other end is a carboxyl group (carboxyl or C terminal).

MICRO INQUIRY *Identify the two other peptide bonds of the tetrapeptide chain.*

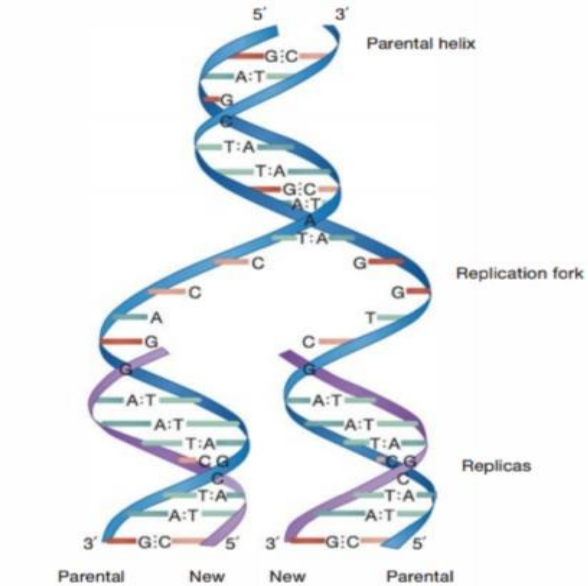


Figure 13.8 Semiconservative DNA Replication. The replication fork of DNA showing the synthesis of two progeny strands. Newly synthesized strands are purple. Each copy contains one new and one old strand.

DNA replication is an extraordinarily important and complex process upon which all life depends. 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, and DNA replication is said to be semiconservative (**figure 13.8**). DNA

replication is also extremely accurate; *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. Most of our discussion in this section is based on studies of chromosomal DNA replication in *E. coli*.

Bacterial DNA Replication Initiates from a Single Origin of Replication

The replication of chromosomal DNA begins at a single point, the **origin of replication**. Synthesis of DNA occurs at the **replication fork,** the place at which the DNA helix is unwound and individual strands are replicated. Two replication forks move outward from the origin until they have copied the whole **replicon**—the portion of the genome that contains an origin that is replicated as a unit. When the replication forks move around the circular chromosomes observed in most bacteria, a structure shaped like the Greek letter theta (θ) is formed (**figure 13.9**). Because the bacterial chromosome is a single replicon, the forks meet on the other side and two separate chromosomes are released. Less is known about the replication of linear bacterial chromosomes. One well-studied organism with linear chromosomes belongs to the spirochete genus *Borrelia*. The replication of its chromosomes is considered on p. 300. *DNA Replication Fork; Bidirectional DNA Replication*

Replication Machinery

DNA replication is essential to organisms, and a great deal of effort has been devoted to understanding its mechanism. The replication of *E. coli* DNA requires at least 30 proteins. As we describe in chapter 15, many of the archaeal and eukaryotic

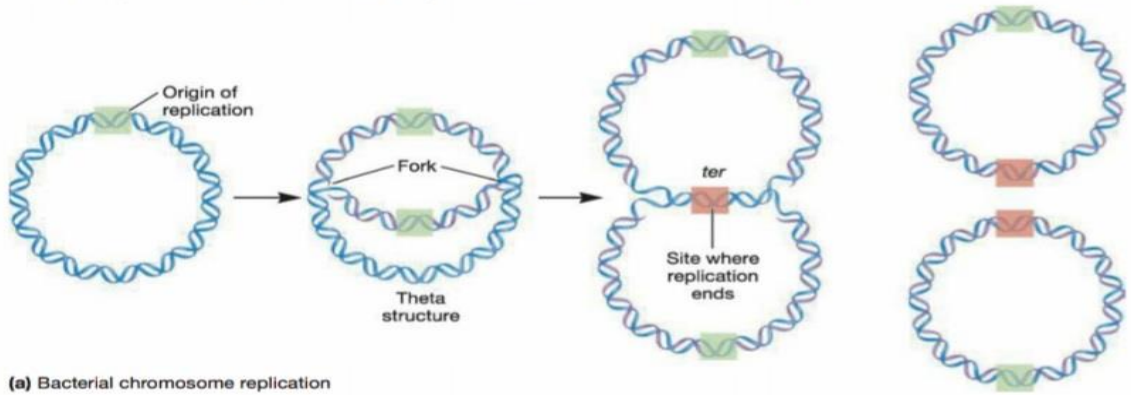


Figure 13.9 Bidirectional Replication of the *E. coli* Chromosome. (a) Replication begins at one site on the chromosome, called the origin of replication. Two replication forks proceed in opposite directions from the origin until they meet at a site called the replication termination site (*ter*). A 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.

replication enzymes and proteins differ from those used by bacterial cells. In general, the archaeal replication machinery is more similar to the eukaryotic machinery than it is to the bacterial system. Despite these differences, the overall process of DNA replication is similar in all organisms.

Enzymes called **DNA polymerases** catalyze DNA synthesis. All known DNA polymerases catalyze the synthesis of DNA in the 5′ to 3′ direction, and the nucleotide to be added is a deoxynucleoside triphosphate (dNTP). Deoxynucleotides are linked by phosphodiester bonds formed by a reaction between the hydroxyl group at the 3′ end of the growing DNA strand and the phosphate closest to the 5′ carbon (the α-phosphate) of the incoming deoxynucleotide (**figure 13.10**). The energy needed to form the phosphodiester bond is provided by release of the terminal two phosphates as pyrophosphate (PP_i) from the nucleotide that is added. The PP_i is subsequently hydrolyzed to two separate phosphates (P_i). Thus the deoxynucleoside triphosphates dATP, dTTP, dCTP, and dGTP serve as DNA polymerase substrates while deoxynucleoside monophosphates (dNMPs: dAMP, dTMP, dCMP, dGMP) are incorporated into the growing chain. ✿ *How Nucleotides Are Added in DNA Replication*

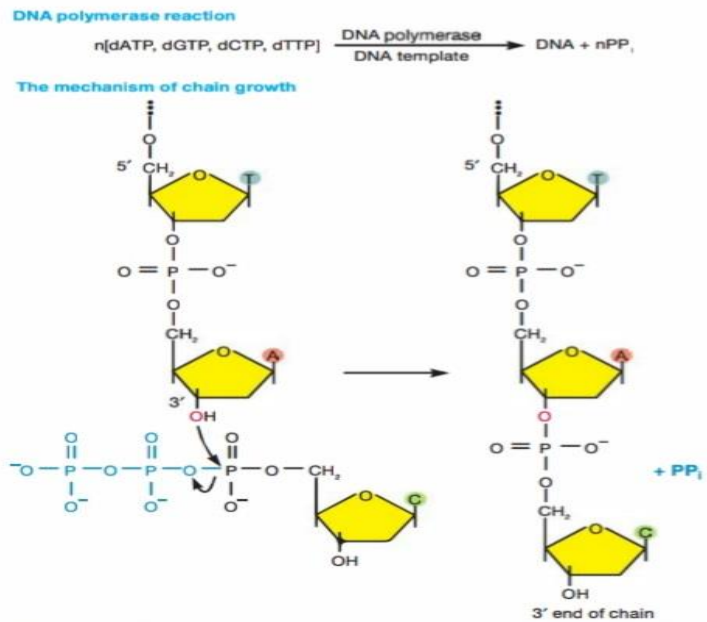


Figure 13.10 The DNA Polymerase Reaction. The hydroxyl of the 3′ terminal deoxyribose makes a nucleophilic attack on the α-phosphate of the nucleotide substrate (in this example, adenosine attacks cytidine triphosphate).

MICRO INQUIRY *What provides the energy to fuel this reaction?*

For DNA polymerase to catalyze the synthesis of DNA, it needs three things. The first is a template, which is read in the 3′ to 5′ direction and is used to direct the synthesis of a complementary DNA strand. The second is 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 13.10). The third is a set of dNTPs. *E. coli* has five different DNA polymerases (DNA polymerase 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 multifunctional enzyme composed of 10 different proteins. Most evidence suggests that within the complex are found two core enzymes (**figure 13.11**), although some evidence for three core polymerases exists. Each core enzyme binds one strand of DNA and is responsible for catalyzing DNA synthesis and proofreading the product to ensure fidelity of replication. Associated with each core enzyme is a subunit called the β clamp. The β clamp tethers a core enzyme to the DNA. At the center of the holoenzyme, and represented by an octopus-like structure in figure 13.11, is a complex of proteins called the clamp loader, which is responsible for loading the β clamp onto DNA. A dimer of another protein (tau) holds the holoenzyme together. Because there are two core enzymes, both strands of DNA are bound by a single DNA polymerase III holoenzyme.

DNA polymerase III holoenzyme is only one component of a huge complex of proteins called the **replisome** (**table 13.1**).

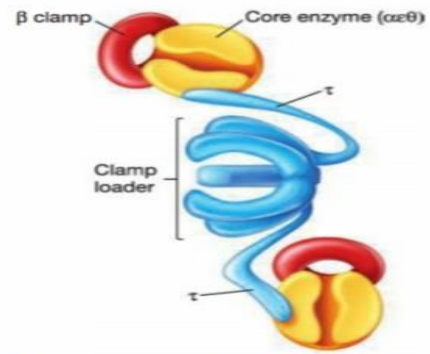


Figure 13.11 DNA Polymerase III Holoenzyme. The holoenzyme consists of two core enzymes (each composed of three different proteins; α, ε, θ, not shown) and several other proteins. Each core enzyme is associated with a β clamp, which tethers a DNA template to the core enzyme. The two tau (τ) proteins connect the two core enzymes to a large complex called the clamp loader, so named because it loads β clamps onto DNA.

Table 13.1 Components of the *E. coli* Replication Machinery

Protein	Function
DnaA	Initiation of replication; binds origin of replication (*oriC*)
DnaB	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	Helicase loader; helps direct DnaB protein (helicase) to DNA template
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
Tus	Termination of replication
Topoisomerase IV	Separation of chromosomes upon completion of DNA replication

Other proteins found in the replisome include helicases, single-stranded DNA binding proteins, and topoisomerases (figure 13.12). **Helicases** are responsible for separating (unwinding) the DNA strands just ahead of the replication fork, using energy from ATP hydrolysis. **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 positive supercoils in the helix ahead of the replication fork, and these can impede replication if not removed. Topoisomerases change the structure of DNA by transiently breaking one or two strands without altering the nucleotide sequence of the DNA (e.g., a topoisomerase might tie or untie a knot in a DNA strand). **DNA gyrase** is an important topoisomerase in *E. coli*. It is not only important during DNA replication but also for introducing negative supercoiling in the bacterial chromosome that helps compact it. ◀◀ *Nucleoid (section 3.6)*

Once the template is prepared, the primer needed by DNA polymerase III can be synthesized. An enzyme called **primase** synthesizes a short RNA strand, usually around 10 nucleotides long and complementary to the DNA; this serves as the primer (figure 13.12). RNA is used as the primer because unlike DNA polymerase, RNA polymerases (such as primase) can initiate RNA synthesis without an existing 3′-OH. It appears that primase requires the assistance of several other proteins, and the complex of primase and its accessory proteins is called the **primosome** (table 13.1). The primosome is another important component of the replisome.

As noted, DNA polymerase enzymes synthesize DNA in the 5′ to 3′ direction. Therefore only one of the strands, called the **leading strand,** can be synthesized continuously at its 3′ end as the DNA unwinds (figure 13.12). The other strand, called the **lagging strand,** cannot be extended in the same direction as the movement of the replication fork 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 (i.e., in the direction opposite of the movement of the replication fork) and produces a series of fragments called **Okazaki fragments,** after their discoverer, Reiji Okazaki. Discontinuous synthesis occurs as primase makes many RNA primers along the template strand. DNA polymerase III then extends these primers with DNA, and eventually the Okazaki fragments are joined to form a complete strand. Thus while the leading strand requires only one RNA primer to initiate synthesis, the lagging strand has many RNA primers 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. ● *DNA Replication; Structural Basis of DNA Replication*

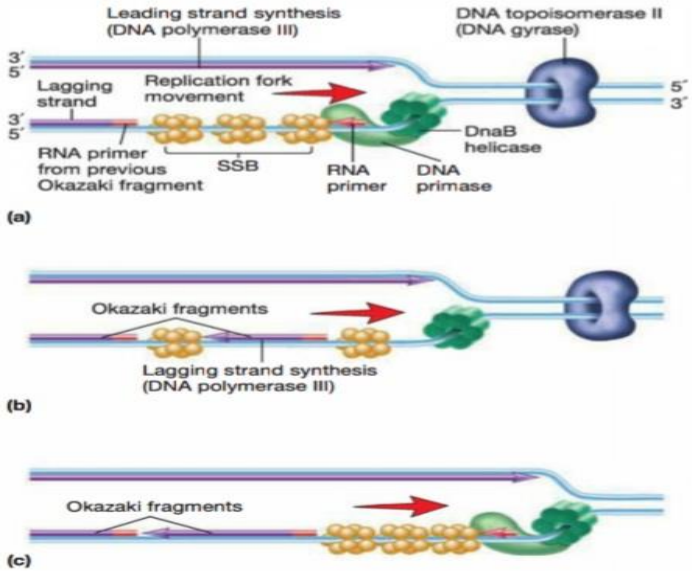


Figure 13.12 Other Replisome Proteins. A single replication fork showing the activity of replisome proteins other than DNA polymerase III holoenzyme (not shown) is illustrated. DnaB helicase is responsible for separating the two strands of parental DNA. The strands are kept apart by single-stranded DNA binding proteins (SSB), which allows for synthesis of an RNA primer by DNA primase. DNA gyrase eases the strain introduced into the DNA double helix by helicase activity. Both leading strand and lagging strand synthesis are illustrated. The lagging strand is synthesized in short fragments called Okazaki fragments. A new RNA primer is required for the synthesis of each Okazaki fragment.

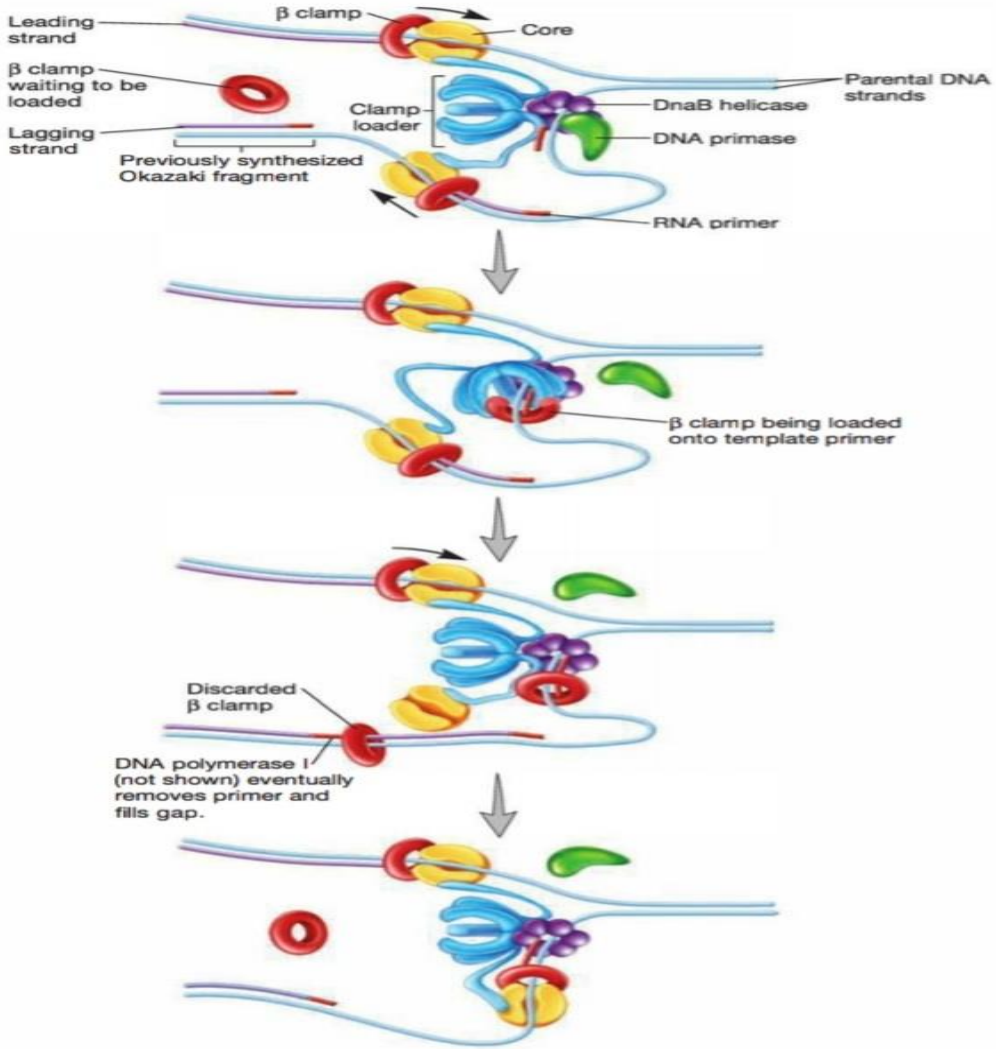
MICRO INQUIRY *What is the difference between helicase and gyrase? Which is a topoisomerase?*

Events at the Replication Fork

The details of DNA replication are outlined in **figure 13.13**. We present replication as a series of discrete steps, but in the cell these events occur quickly and simultaneously on both the leading and lagging strands. In *E. coli,* DNA replication is initiated at specific nucleotides called the *oriC* locus (for *ori*gin of *c*hromosomal replication). This site is AT rich. Recall that adenines pair with thymines using only two hydrogen bonds, so AT-rich segments of DNA become single stranded more readily than do GC-rich regions. This is important for initiation of replication.

1. The bacterial initiator protein DnaA is responsible for triggering DNA replication. DnaA proteins bind regions in *oriC* throughout the cell cycle, but to initiate replication, DnaA proteins must bind a few particular *oriC* sequences. The presence of DnaA at these sites recruits a helicase (usually DnaB helicase) to the origin.
2. The helicase unwinds the helix with the aid of topoisomerases such as DNA gyrase (figure 13.13, step 1). The single strands are kept separate by SSBs.
3. Primase synthesizes RNA primers as needed (figure 13.13, step 1). 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 holoenzyme. It must discard old β clamps (figure 13.13, step 3), load new β clamps (figure 13.13, 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.
4. After most of the lagging strand has been synthesized by the formation of Okazaki fragments, DNA polymerase I removes the RNA primers. DNA polymerase I does this because, unlike other DNA polymerases, it has the ability to snip off nucleotides one at a time starting at the 5′ end while moving toward the 3′ end of the RNA primer. This ability is referred to as 5′ to 3′ exonuclease activity. DNA polymerase I begins its exonuclease activity at the free 5′ end of each 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 13.14**).
5. Finally, the Okazaki fragments are joined by the enzyme **DNA ligase,** which forms a phosphodiester bond between the 3′-OH of the growing strand and the 5′-phosphate of an Okazaki fragment (**figure 13.15**).

Amazingly, DNA polymerase III, like all DNA polymerases (except DNA polymerases called reverse transcriptases), 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. One of the protein subunits of the DNA polymerase III core enzyme (the ε subunit) has 3′ to 5′ exonuclease activity. This activity enables the polymerase core enzyme 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, the exonuclease activity is used to remove it (**figure 13.16**). A mismatched base can be removed only 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 16, the mismatch repair system is the cell's second line of defense against the potential harm caused by the incorporation of the incorrect nucleotide. *Proofreading Function of DNA Polymerase*



1. The leading-strand core DNA polymerase synthesizes DNA as the parental DNA strands are unwound by DnaB helicase. The lagging strand core DNA polymerase is nearing completion of an Okazaki fragment. DNA primase begins synthesis of the RNA primer for the next Okazaki fragment to be synthesized.

2. Upon completion of the new RNA primer, DNA primase dissociates and the clamp loader loads a β clamp onto the template primer.

3. The lagging-strand core polymerase reaches the previously synthesized Okazaki fragment and dissociates from the DNA.

4. The lagging-strand core polymerase associates with the newly loaded β clamp, and synthesis of a new Okazaki fragment begins.

Figure 13.13 Events 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 β clamp is discarded and a new one loaded onto the template DNA (steps 2 and 3). This is achieved by the activity of the clamp loader (figure 13.11). Okazaki fragments are eventually joined together after removal of the RNA primer and synthesis of DNA to fill the gap, both catalyzed by DNA polymerase I (figure 13.14); DNA ligase then seals the nick and joins the two fragments (figure 13.15).

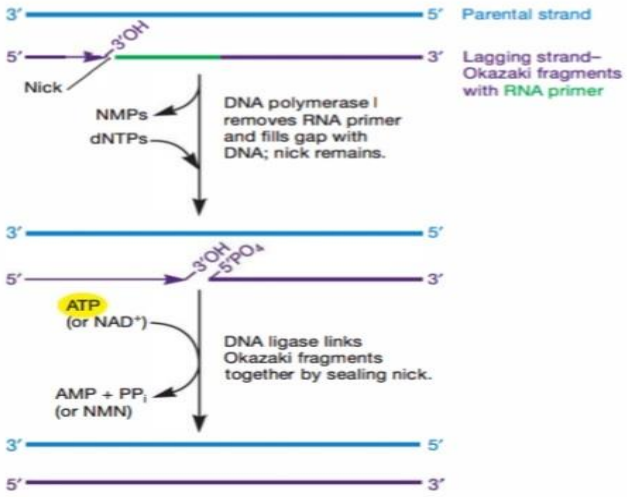


Figure 13.14 Completion of Lagging Strand Synthesis. NMPs, nucleoside monophosphates; dNTPs, deoxynucleoside triphosphates; NAD^+, nicotinamide adenine dinucleotide; NMN, nicotinamide mononucleotide.

As we have seen, DNA polymerase III is a remarkable multiprotein complex, with several enzymatic activities. In *E. coli*, the polymerase component is encoded by the *dnaE* gene. *Bacillus subtilis*, a Gram-positive bacterium that is another important experimental model, has a second polymerase gene called $dnaE_{Bs}$. Its protein product appears to be responsible for synthesizing the lagging strand. Thus while the overall mechanism by which DNA is replicated is highly conserved, there can be variations in replisome components.

Termination of Replication

In *E. coli*, DNA replication stops when the replisome reaches a termination site (*ter*) on the DNA. A protein called Tus binds to the *ter* sites and halts progression of the forks. In many other bacteria, replication stops spontaneously when the forks meet. Regardless of how fork movement is stopped, there are two problems that often must be solved by the replisome. One is the formation of interlocked chromosomes called **catenanes** (figure 13.17*a*). The other is a dimerized chromosome—two chromosomes joined together to form a single chromosome twice as long (figure 13.17*b*). Catenanes are produced when topoisomerases break and rejoin DNA strands to ease supercoiling ahead of the replication fork. The two daughter DNA molecules are separated by the action of other topoisomerases that break both strands of one molecule, pass the other DNA molecule

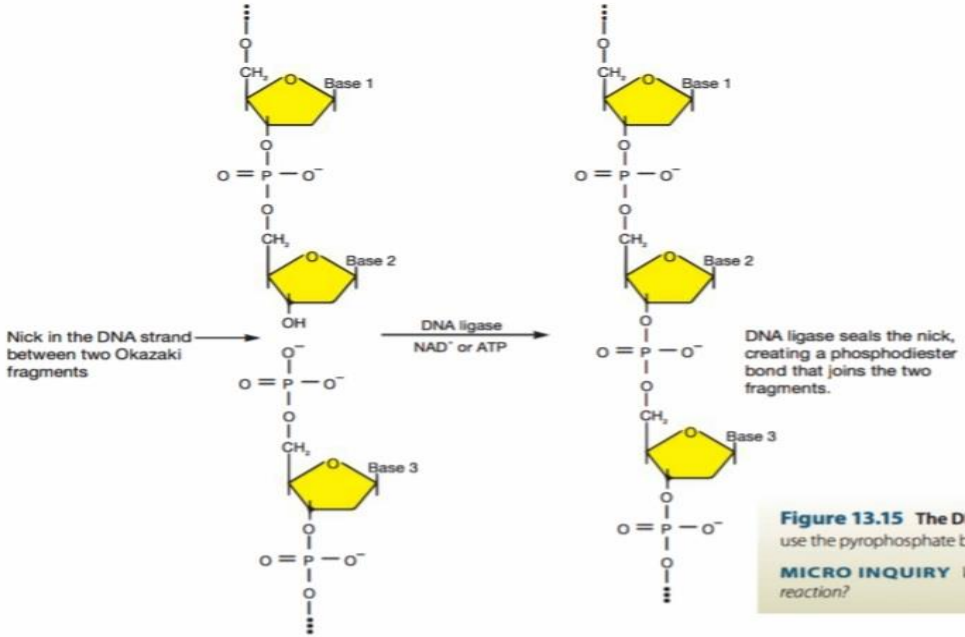


Figure 13.15 The DNA Ligase Reaction. Bacterial ligases use the pyrophosphate bond of NAD^+ or ATP as an energy source.

MICRO INQUIRY *Why can't DNA polymerase I perform this reaction?*

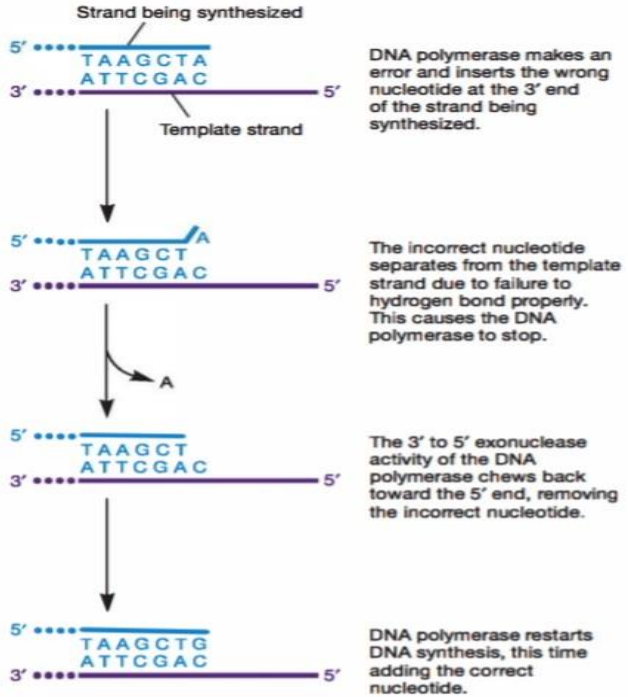


Figure 13.16 The Proofreading (Editing) Function of DNA Polymerase.

through the break, and then rejoin the strands (figure 13.17*a*). Dimerized chromosomes result from DNA recombination that sometimes occurs between two daughter molecules during DNA replication. Recombinase enzymes (e.g., XerCD in *E. coli*) catalyze an intramolecular cross-over that separates the two chromosomes (figure 13.17*b*).

Replication of Linear Chromosomes

All eukaryotic chromosomes and some bacterial chromosomes are linear. In both cases, this poses a problem during replication because all DNA polymerases synthesize DNA in the 5′ to 3′ direction from a primer that provides a free 3′-OH. When the RNA primer for the Okazaki fragment at the 5′ end of the daughter strand is removed, the daughter molecule is shorter than the parent molecule. Over numerous rounds of DNA replication and cell division, this leads to a progressively shortened chromosome. Ultimately the chromosome loses critical genetic information, which is lethal to the cell. This problem is called the "end replication problem," and a cell must solve it if it is to survive. Eukaryotic cells have solved the end replication problem with an

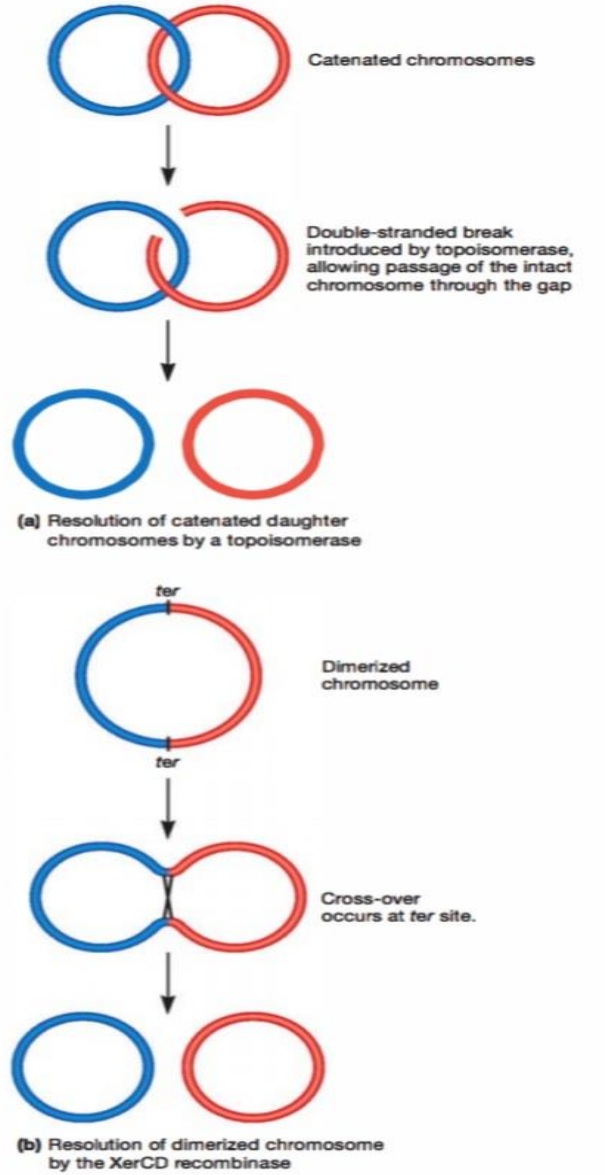


Figure 13.17 Resolving Catenated and Dimerized Chromosomes.
(a) Catenated chromosomes arise from the activity of topoisomerases during chromosome replication. Accordingly, they are resolved by topoisomerases. (b) Dimerized chromosomes arise from recombination events that can occur during chromosome replication between the daughter chromosomes. They are resolved by recombinase enzymes such as XerCD of *E. coli*. These enzymes catalyze a cross-over that separates the two chromosomes.

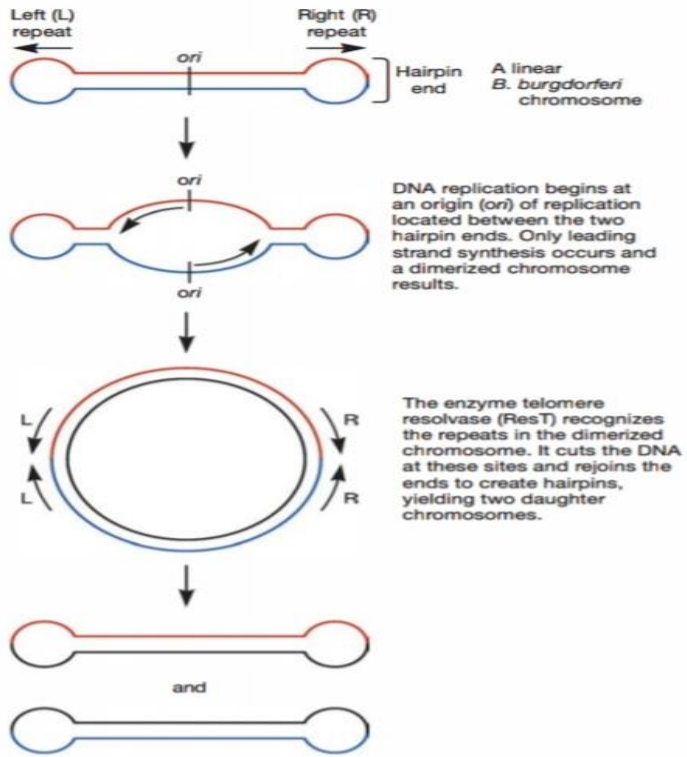


Figure 13.18 Replication of a *Borrelia burgdorferi* Linear Chromosome.

enzyme called telomerase, as we describe in chapter 15. Bacteria have taken a different approach. ▶▶ *Telomeres and telomerases (section 15.2).*

Of those bacteria having linear chromosomes, the best understood mechanism for solving the end replication problem is that used by *Borrelia burgdorferi*. The approach *B. burgdorferi* cells use is to disguise the ends so well that they aren't really ends. How are the ends disguised? Consider a typical linear, double-stranded DNA molecule. At each terminus of the double helix is a strand having a 3′-OH and the complementary strand having a 5′-phosphate. Now examine **figure 13.18**, which shows that *B. burgdorferi* chromosomes do not have these free ends. Rather, a phosphodiester bond links the two complementary strands together. This is made possible by inverted repeats at each terminus and the formation of a hairpin. The origin of replication is located between the hairpins. Interestingly, only leading strand synthesis occurs at each replication fork. When replication is complete, a circular molecule has been formed that is twice the length of the parent chromosome. Thus it is a dimerized chromosome. An enzyme called telomere resolvase (ResT) cuts the two chromosomes apart as it forms hairpin ends for each daughter molecule.

Retrieve, Infer, Apply

1. How many replicons do typical bacterial cells have (i.e., those having a single chromosome)? How many replication forks are used to replicate a circular chromosome? Is a primosome part of the replisome, or is the replisome part of the primosome? What is the function of each?
2. 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 proteins, Okazaki fragment, DNA ligase, leading strand, lagging strand, and primase.
3. Outline the steps involved in DNA synthesis at the replication fork. How do DNA polymerases correct their mistakes?
4. What is the end replication problem? How does *B. burgdorferi* solve it?

13.4 Bacterial Gene Structure

After reading this section, you should be able to:

- Draw a typical bacterial protein-coding gene, and label the important portions of the gene and the conventions for numbering base pairs in the gene
- Draw typical tRNA- and rRNA-coding genes

DNA replication allows genetic information to be passed from one generation to the next. But how is the 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. Genes have been regarded in several ways. At first, it was thought that a gene contained information for the synthesis of one enzyme—the one gene–one enzyme hypothesis. This was 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 (subunits) coded for by separate genes. A segment of DNA that encodes a single polypeptide is sometimes termed a cistron. However, not all genes encode proteins; some code instead for ribosomal RNA (rRNA) and transfer RNA (tRNA), both of which function in protein synthesis (**figure 13.19**). 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 one or more functional products (i.e., a polypeptide, tRNA, or rRNA). In this section, we consider the structure of each of these three types of genes.

ses.

VI. MODELS OF REPLICATION

The bacterial chromosome consists of a double stranded DNA ring attached to the plasma membrane. Usually there is a single replication fork during replication. This starts at a point called the *origin* and moves around the chromosome (Fig. 8.19). During rapid growth there may be one or two additional forks formed at the origin before the first one

Fig. 8.19. Diagrams showing doubling of circular DNA during replication.

completes its replication cycle. Several models have been proposed to explain the mechanism of replication in *circular* chromosomes. These include the *Cairns model*, the *Yoshikawa model* and the *rolling circle model*. Other models explain replication in *linear* chromosomes.

1. **The Cairns model** (1963) (Fig. 8.20). According to this model replication begins by denaturation of the DNA double strands at a specific site called the origin. Two growing points are established and there is bidirectional DNA synthesis. Both strands of DNA are replicated. As the growing points move apart, unwinding of the DNA double strand takes place. This unwinding creates *torque* since the parental DNA strands cannot unwind freely. The torque is transmitted to the unreplicated part of the molecule which consequently becomes *supertwisted*. Supertwisting brings about a conformational strain on the DNA molecule and prevents it from replicating further. To counterract this effect a temporary break ('nick') is brought about on one of the strands by a *swiveling protein* (ω). The break permits the parental strands to rotate freely on each other, thus relieving the strain. The swiveling protein then seals the break and replication continues.

Cairns-type replication has been demonstrated in the bacteria *E.coli* and *Bacillus subtilis*, in several viral and plasmid chromosomes and in DNA synthesis of mitochondria and chloroplasts.

3. **The rolling circle model** (Fig. 8.22) (Gilbert and Dressler, 1968; Eisen, Pareira da Silva and Jacob, 1968) is the current model for explaining replication in single stranded DNA viruses, e.g. ϕX174, and the transfer of *E.coli* sex factor (plasmid).

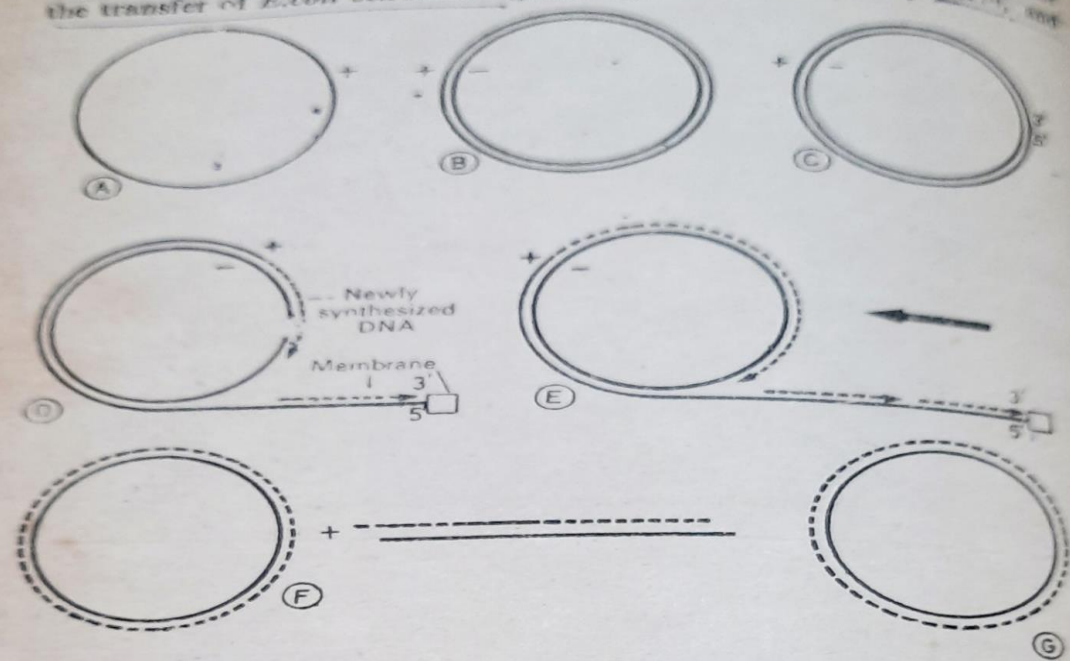


Fig. 8.22. Rolling circle model of replication of ssDNA.
- A. Single strand DNA (ssDNA) ring (+) of ϕX174.
- B. Synthesis of negative strand (—) and formation of double-stranded (dsDNA) replicative form.
- C. Nicking of one parental strand by endonuclease.
- D. Parental strand rolls and unwinds. 5′ end attaches to the host membrane. New DNA synthesized on the 3′ end and at the 'tail'.
- E. Further unwinding and synthesis. Note that the tail region synthesizes discontinuous segments.
- F. Tail is cut by a specific endonuclease into unit length progeny rod.
- G. Circularization of rods to form new circular molecules which can become new rolling circles. During circularization the gap is closed by a ligase.

The ϕX174 chromosome consists of a single stranded DNA ring *(positive strand)*. The chromosome first becomes double-stranded by

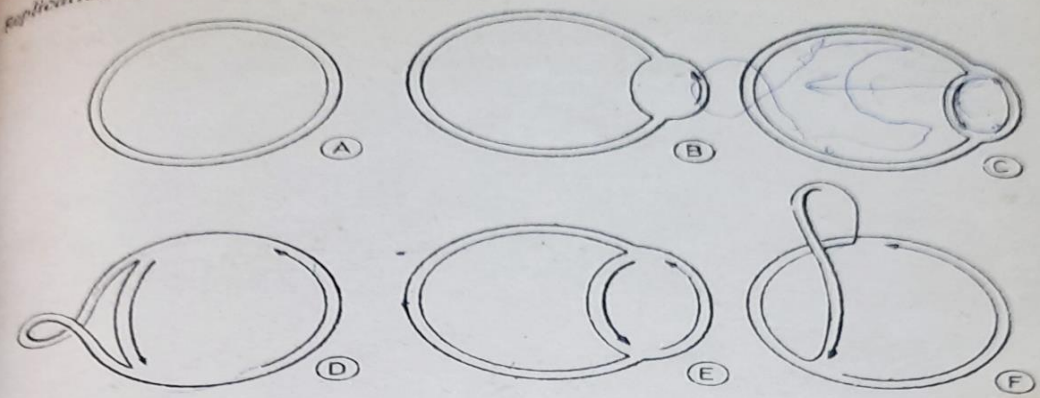


Fig. 8.20. Cairns model for circular DNA replication.

A. Double helix DNA ring.
B. Strands denatured at origin site. DNA fragment initiated *de novo*.
C. A second fragment initiated and a second growing point establiseed.
D. As a result of bidirectional replication the two growing points move forward. This causes supertwisting of unreplicated portion of DNA resulting in conformational strain on the molecule.
E. Swiveling protein (ω) relieves the strain by causing single-stranded nicks, thus allowing free rotation of the parental strands with reference to each other. The swiveling protein then seals the nicks.
F. Replication proceeds and the two growing points converge on the terminus.

2. The Yoshikawa model (Fig. 8.21). A variation of the Cairns model has been suggested by Yoshikawa. According to this model the newly formed DNA strands become covalently joined to the ends of the parental chromosomes.

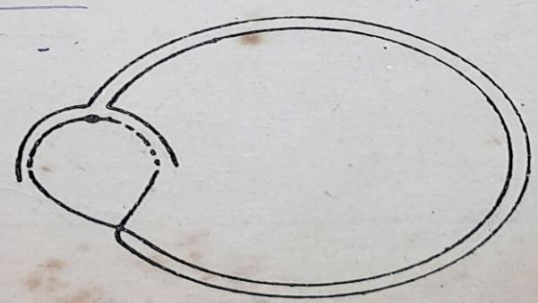


Fig. 8.21. Yoshikawa's model of replication.

Replication of DNA

the synthesis of a *negative strand*. Synthesis is presumed to begin at a specific *initiation point* on the template ring.

One strand of the parental duplex ring is now cut at a specific point by an *endonuclease*. This enzyme recognizes a particular sequence at this point. As a result of the cut ('nick') a linear strand with 3′ and 5′ ends is created.

The 3′ end serves as a primer for the synthesis of a new DNA strand under the catalytic action of *DNA polymerase*. The unbroken strand is used as the template for this purpose, and a complementary strand is synthesized. Thus the parental molecule itself is used as a primer for initiating replication.

The 5′ end of the broken strand becomes attached to the plasma membrane of the host bacterium. Such replicating phage DNA is commonly found associated with bacterial membranes.

The unbroken parental strand rolls and unwinds as synthesis proceeds, leaving a 'tail' which is attached to the membrane. New DNA is also synthesized in the tail region in discontinuous segments in the 5′→3′ direction. This synthesis is presumably preceeded by the synthesis of an *RNA primer* under the catalytic action of *RNA polymerase*.

The tail is cut off by a specific *endonuclease* into a unit length *progeny rod*. The rod may undergo *circularization* to form a new *circular molecule*. During this process the gap is closed by a *ligase*. The newly formed circular molecules can in turn become new rolling circles.

Genetic information is preserved in the single stranded template ring which remains circular and serves as an endless template. There is no swivelling problem or creation of torque in the rolling circle model. As the strands unwind the 3′ end is free to rotate on the unbroken strand. The growing point itself thus serves as a swivel.

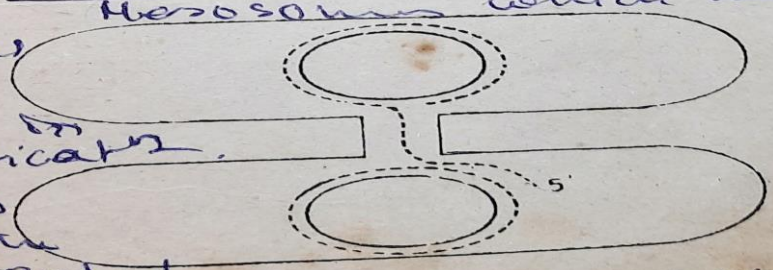


Fig. 8.23. Transfer of ssDNA during mating in bacteria through the rolling circle mechanism.

Evidence for the rolling circle model has been obtained from the replication of several viruses (M13, P2, T4, λ), replication resulting in transfer of genetic material during mating of bacteria (Fig. 8.23), and the special DNA synthesis during oogenesis in *Xenopus*.

4. **Replication in a duplex-rod prokaryote chromosome** (Fig. 8.24,A-C) In the virus T7 the chromosome is a duplex rod of about 40,000 base

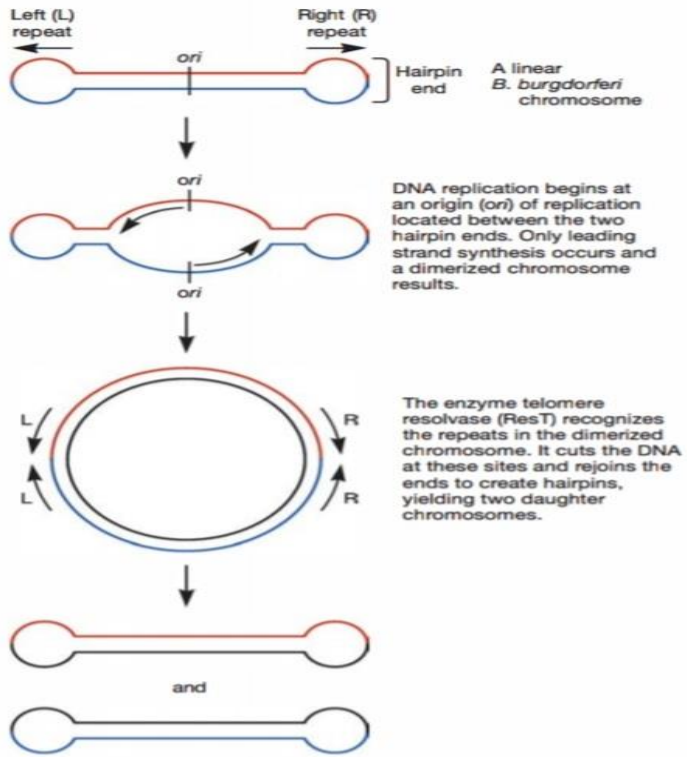


Figure 13.18 Replication of a *Borrelia burgdorferi* Linear Chromosome.

enzyme called telomerase, as we describe in chapter 15. Bacteria have taken a different approach. ▶▶ *Telomeres and telomerases (section 15.2).*

Of those bacteria having linear chromosomes, the best understood mechanism for solving the end replication problem is that used by *Borrelia burgdorferi*. The approach *B. burgdorferi* cells use is to disguise the ends so well that they aren't really ends. How are the ends disguised? Consider a typical linear, double-stranded DNA molecule. At each terminus of the double helix is a strand having a 3′-OH and the complementary strand having a 5′-phosphate. Now examine **figure 13.18**, which shows that *B. burgdorferi* chromosomes do not have these free ends. Rather, a phosphodiester bond links the two complementary strands together. This is made possible by inverted repeats at each terminus and the formation of a hairpin. The origin of replication is located between the hairpins. Interestingly, only leading strand synthesis occurs at each replication fork. When replication is complete, a circular molecule has been formed that is twice the length of the parent chromosome. Thus it is a dimerized chromosome. An enzyme called telomere resolvase (ResT) cuts the two chromosomes apart as it forms hairpin ends for each daughter molecule.

Retrieve, Infer, Apply

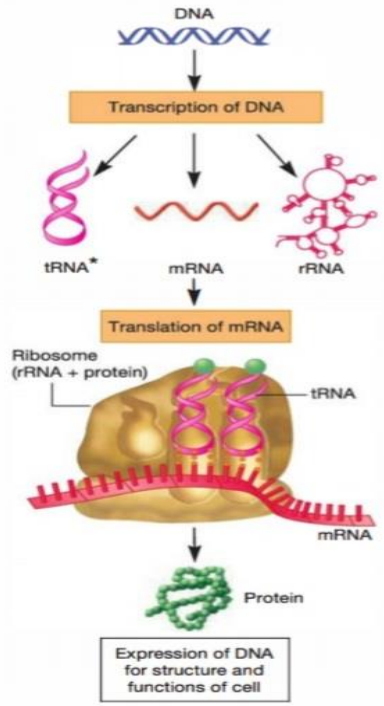
1. How many replicons do typical bacterial cells have (i.e., those having a single chromosome)? How many replication forks are used to replicate a circular chromosome? Is a primosome part of the replisome, or is the replisome part of the primosome? What is the function of each?
2. Describe the nature and functions of the following replication components and intermediates: DNA polymerases I and III, topoisomerase, DNA gyrase, helicase, single-stranded DNA binding proteins, Okazaki fragment, DNA ligase, leading strand, lagging strand, and primase.
3. Outline the steps involved in DNA synthesis at the replication fork. How do DNA polymerases correct their mistakes?
4. What is the end replication problem? How does *B. burgdorferi* solve it?

13.4 Bacterial Gene Structure

After reading this section, you should be able to:

- Draw a typical bacterial protein-coding gene, and label the important portions of the gene and the conventions for numbering base pairs in the gene
- Draw typical tRNA- and rRNA-coding genes

DNA replication allows genetic information to be passed from one generation to the next. But how is the 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. Genes have been regarded in several ways. At first, it was thought that a gene contained information for the synthesis of one enzyme—the one gene–one enzyme hypothesis. This was 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 (subunits) coded for by separate genes. A segment of DNA that encodes a single polypeptide is sometimes termed a cistron. However, not all genes encode proteins; some code instead for ribosomal RNA (rRNA) and transfer RNA (tRNA), both of which function in protein synthesis (**figure 13.19**). 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 one or more functional products (i.e., a polypeptide, tRNA, or rRNA). In this section, we consider the structure of each of these three types of genes.



*The sizes of RNA are enlarged to show details.

Figure 13.19 Transcription Yields Three Major Types of RNA Molecules. Messenger RNA (mRNA) molecules arise from transcription of protein-coding genes. They are translated into protein with the aid of the other two major types of RNA: transfer RNA (tRNA) molecules carry amino acids to the ribosome during translation; ribosomal RNA (rRNA) molecules have several functions, including catalyzing peptide bond formation.

Protein-Coding Genes

Most of the genes found in bacterial genomes encode proteins. However, DNA does not serve directly as the template for protein synthesis. Rather, the genetic information in the gene is transcribed to give rise to a messenger RNA (mRNA), which is translated (section 13.7) into a protein (figure 13.19). For this to occur, protein-coding genes must contain signals that indicate where transcription should start and stop, and signals in the resulting mRNA that indicate where translation should start and stop. As we describe in more detail in section 13.5, during transcription only one strand of a gene directs mRNA synthesis. This strand is called the **template strand,** and the complementary DNA strand is known as the coding strand because it is the same nucleotide sequence as the mRNA, except in DNA bases (**figure 13.20**). Messenger RNA is synthesized from the 5′ to the 3′ end in a manner similar to DNA synthesis. Therefore the polarity of the DNA template strand is 3′ to 5′. In other words, the beginning of the gene is at the 3′ end of the template strand.

An important site called the **promoter** is located at the start of the gene. The promoter is the binding site for RNA polymerase, the enzyme that synthesizes RNA. The promoter is neither transcribed nor translated; it functions strictly to orient RNA polymerase so it is a specific distance from the first DNA nucleotide that will serve as a template for RNA synthesis. The promoter thus specifies which strand is to be transcribed and where transcription should begin. As we discuss in chapter 14, the sequences near the promoter often are very important in regulating when and at what rate a gene is transcribed. ▶▶| *Regulation of transcription initiation (section 14.2)*

The transcription start site (labeled +1 in figure 13.20) 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** that is transcribed into mRNA but is not translated into amino acids. In bacteria, the leader includes a region called the **Shine-Dalgarno sequence,** which is important in the initiation of translation. The leader sometimes is also involved in regulation of transcription and translation. ▶▶| *Regulation of transcription elongation (section 14.3); Regulation of translation (section 14.4)*

Immediately next to (and downstream of) the leader is the most important part of the gene, the **coding region** (figure 13.20). The coding region typically begins with the template DNA sequence 3′-TAC-5′. This is transcribed into the start codon, 5′-AUG-3′, which codes for the first amino acid of the polypeptide encoded by the gene. The remainder of the coding region is transcribed into a sequence of codons that specifies the sequence of amino acids for the rest of the protein. The coding region ends with a sequence that, when transcribed, is a stop codon. It signals the end of the protein and stops the ribosome during translation. The stop codon is immediately followed by the **trailer,** which is transcribed but not translated. The trailer contains sequences that prepare the RNA polymerase for release from the template strand. Indeed, just beyond the trailer (and sometimes slightly overlapping it) is the **terminator.** The terminator is a sequence that signals the RNA polymerase to stop transcription.

In bacteria, the coding region is usually continuous, unlike the coding regions of eukaryotic genes, which are often interrupted by noncoding sequences called **introns**. Those rare bacterial genes that do contain introns are transcribed into an intron-containing mRNA. The introns are eventually removed but by a mechanism different than that used to remove introns from eukaryotic mRNAs.

tRNA and rRNA Genes

Actively growing cells need a ready supply of tRNA and rRNA molecules so that protein synthesis can occur. To ensure this, bacterial cells often have more than one gene for each of these

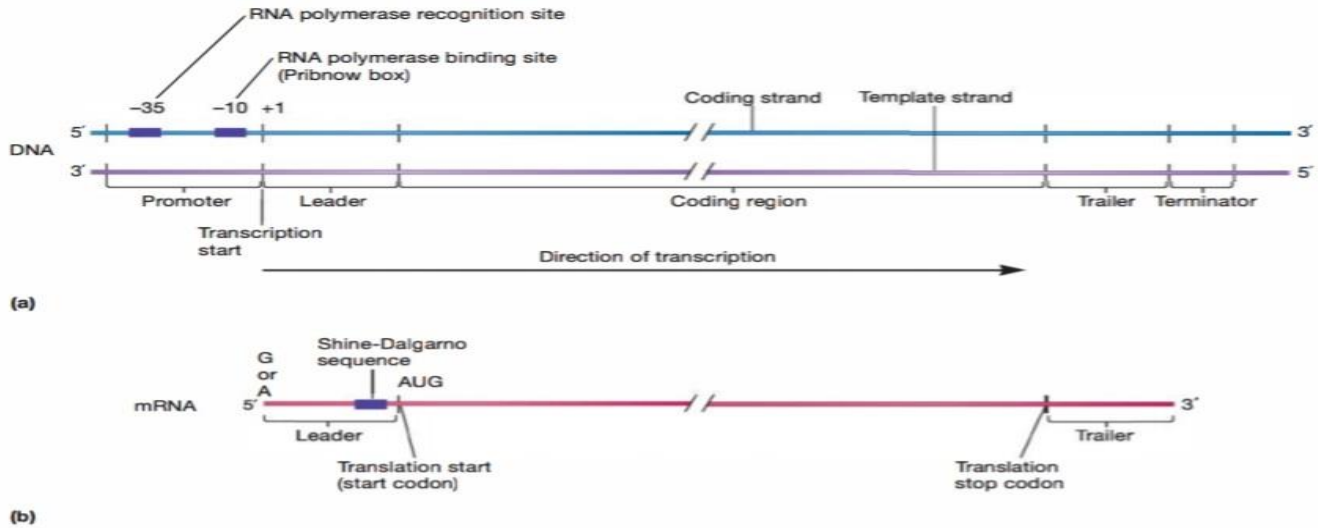


Figure 13.20 A Bacterial Structural Gene and Its mRNA Product. (a) The organization of a typical structural gene in a bacterial cell. Some genes lack leaders or trailers or both. Transcription begins at the +1 position in DNA and proceeds to the right, as shown. The numbering of nucleotides to the left 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. During transcription, 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 start codon. Regulatory sites are not shown but are usually upstream of or overlap with the promoter.

MICRO INQUIRY *Why is the nontemplate strand called the "coding strand"?*

molecules. Furthermore, it is important that the number of each tRNA or rRNA relative to other tRNAs or rRNAs be controlled. This is accomplished in part by having several tRNA or rRNA genes transcribed together, under the control of a single promoter.

In bacteria, genes for tRNA consist of a promoter, tRNA coding region, leader, and trailer. When more than one tRNA is transcribed from the promoter, the coding regions are separated by short spacer sequences (**figure 13.21*a***). Whether the gene encodes a single tRNA or multiple tRNAs, the initial transcript must be processed to remove the noncoding sequences (i.e., leader, trailer, and spacers, if present). This is called posttranscriptional modification, and it is accomplished by ribonucleases—enzymes (and in some cases ribozymes) that cut RNA. In addition, many bacterial (and archaeal) tRNA genes contain introns that must be removed during tRNA maturation.
◀◀ *Ribozymes (section 10.6)*

Bacterial cells usually contain more than one rRNA gene. Each gene has a promoter, trailer, and terminator, and encodes all three types of rRNA (figure 13.21*b*). Thus, as seen for tRNA genes, the transcript from an rRNA gene is a single, large precursor molecule that is cut up by ribonucleases to yield the final rRNA products. Interestingly, in many bacteria, the trailer regions and the spacers often contain tRNA genes. Thus the precursor rRNA encodes for both tRNA and rRNA.

Retrieve, Infer, Apply

1. The coding region of a gene is said to be "downstream" from the leader. Conversely, the leader is said to be "upstream" of the coding region. For each of the following portions of a gene, indicate whether it is downstream or upstream of the coding region: promoter, +1 nucleotide, trailer, terminator. Which portions of a gene are transcribed but not translated?
2. Which strand of a gene has sequences that correspond to the start (i.e., ATG rather than AUG) and stop codons in the mRNA product of the gene?
3. Briefly discuss the general organization of tRNA and rRNA genes. How does their expression differ from that of protein-coding genes with respect to posttranscriptional modification of the gene product?

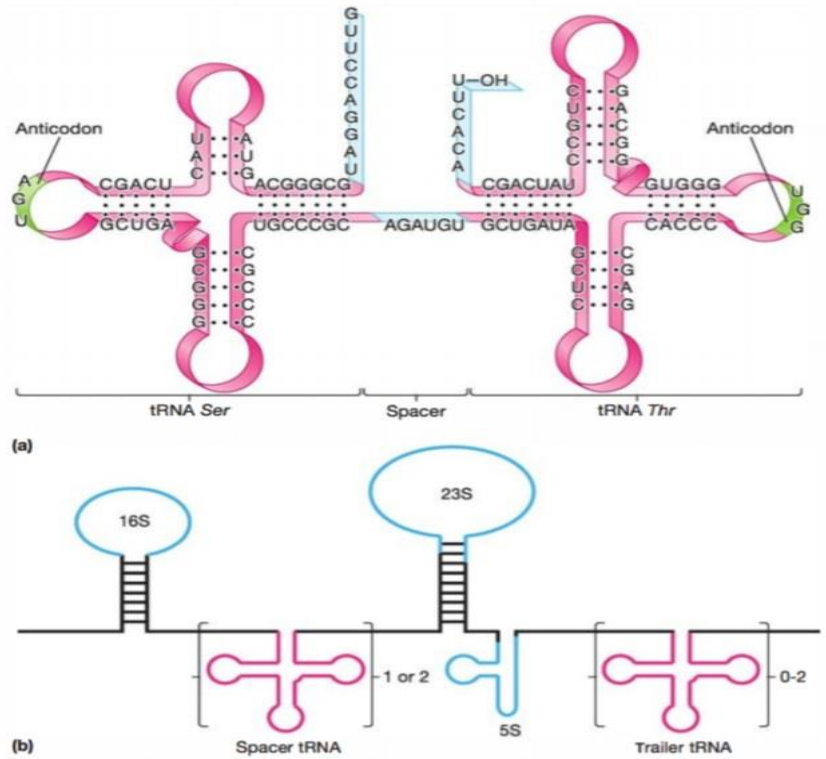


Figure 13.21 tRNA and rRNA Genes. (a) A tRNA precursor from *E. coli* that contains two tRNA molecules. The spacer and extra nucleotides at both ends are removed during processing. (b) The *E. coli* rRNA gene codes for a large transcription product that is cleaved into three rRNAs and one to three tRNAs. 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.

13.5 Transcription in Bacteria

After reading this section, you should be able to:

- Illustrate the organization of bacterial genes in a typical operon
- Describe the structure of a typical bacterial RNA polymerase holoenzyme
- Outline the events that occur during the three phases of transcription
- Discuss the role of bacterial promoters and sigma factors in transcription initiation
- Distinguish factor-independent termination of transcription from rho-dependent termination of transcription

Synthesis of RNA under the direction of DNA is called transcription, and the RNA product has a sequence complementary to the DNA template directing its synthesis. Although adenine directs the incorporation of thymine during DNA replication, it usually codes for uracil during RNA synthesis. Transcription generates three major kinds of RNA. **Transfer RNA (tRNA)** carries amino acids during protein synthesis, and **ribosomal RNA (rRNA)** molecules are components of ribosomes. **Messenger RNA (mRNA)** bears the message for protein synthesis. Bacterial genes encoding proteins involved in a related process (e.g., encoding enzymes for synthesis of an amino acid) are often located close to each other and are transcribed from a single promoter. Such a transcriptional unit is termed an **operon**. Transcription of an operon yields an mRNA consisting of a