The Blueprint of Life, from DNA to Protein

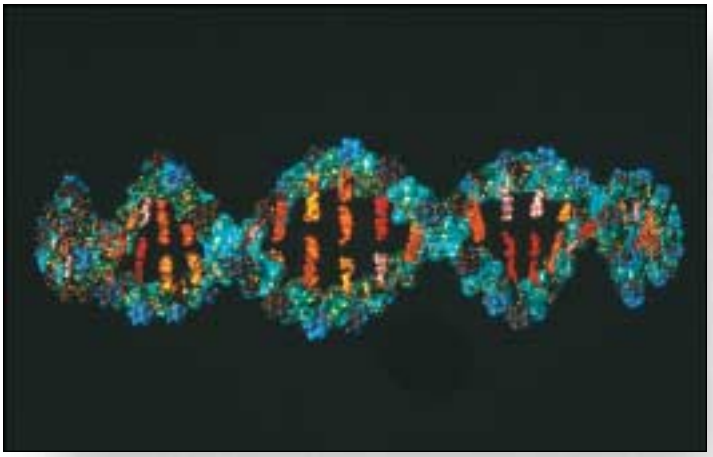
In 1866, the Czech monk Gregor Mendel showed that traits are inherited by means of physical units, which we now call genes. It was not until 1941, however, that the precise function of genes was revealed when George Beadle, a geneticist, and Edward Tatum, a chemist, published a scientific paper reporting that genes determine the structure of enzymes. Biochemists had already shown that enzymes catalyze the conversion of one compound into another in a biochemical pathway.

Beadle and Tatum studied Neurospora crassa, a common bread mold that grows on a very simple medium containing sugar and simple inorganic salts. Beadle and Tatum created N. crassa strains with altered properties, ***mutants,*** *by treating cells with X rays, which were known to alter genes. Some of these mutants could no longer grow on the glucose-salts medium unless growth factors such as vitamins were added to the medium. To isolate these Beadle and Tatum had to laboriously screen thousands of progeny to find the relatively few that required the growth factors. Each mutant presumably contained a defective gene.*

The next task for Beadle and Tatum was to identify the specific biochemical defect of each mutant. To do this, they added different growth factors, one at a time, to each mutant culture. The one that allowed a particular mutant to grow had presumably bypassed the function of a defective enzyme. In this manner, they were able to pinpoint in each mutant the specific step in the biochemical pathway that was defective. Then, using these same mutants, Beadle and Tatum showed that the requirement for each growth factor was inherited as a single gene, ultimately leading to their conclusion that a single gene determines the production of one enzyme. Their conclusion has been modified somewhat, because we now know that some enzymes are made up of more than one protein. A single gene determines the production of one protein. In 1958, Beadle and Tatum shared the Nobel Prize in Medicine, largely for these pioneering studies that ushered in the era of modern biology.

As so often occurs in science, the answer to one question raised many more questions. How do genes specify the synthesis of enzymes? What are genes made of? How do genes replicate? Numerous other investigators won more Nobel Prizes for answering these questions, many of which are covered in this chapter.

—A Glimpse of History



DNA double helix

CONSIDER FOR A MOMENT THE VAST DIVERSITY OF cellular life forms that exist. Our world contains a remarkable variety of microorganisms and specialized cells that make up plants and animals. Every characteristic of each of these cells, from its shape to its function, is dictated by information contained in its deoxyribonucleic acid (DNA). DNA encodes the master plan, the blueprint, for all cell structures and processes. Yet for all the complexity this would seem to require, DNA is a string composed of only four different nucleotides, each containing a particular nitrogenous base: adenine (A), thymine (T), cytosine (C), or guanine (G). ■ **nucleotides, p. 31**

While it might seem improbable that the vast array of life forms can be encoded by a molecule consisting of only four different units, think about how much information can be transmitted by binary code, the language of all computers, which has a base of only two. A simple series of ones and zeros can code for each letter of the alphabet. String enough of these series together in the right sequence and the letters become words, and the words can become complete sentences, chapters, books, or even whole libraries.

The four nucleotides of a DNA molecule create information in a similar fashion. A set of three nucleotides encodes a specific amino acid. In turn, a string of amino acids makes up a protein, the function of which is dictated by the order of the amino acid subunits. Some proteins serve as structural components of a cell. Others, such as enzymes, mediate cellular activities including biosynthesis and energy conversion.

Together, proteins synthesized by a cell are responsible for every aspect of that cell. Thus, the sequential order of nucleotide bases in a cell's DNA ultimately dictates the characteristics of that cell. ■ **amino acids, p. 25** ■ **protein structure, p. 27** ■ **flagella, p. 63** ■ **enzymes, pp. 131, 138**

This chapter will focus on the cellular process of converting the information encoded within DNA into proteins, concentrating primarily on the mechanisms used by prokaryotic cells. The eukaryotic processes have many similarities, but they are considerably more complicated and will only be discussed briefly.

7.1 Overview

The complete set of genetic information for a cell is referred to as its **genome.** Technically, this includes plasmids as well as the chromosome; however, the term genome is often used interchangeably with chromosome. The genome of all cells is composed of DNA, but some viruses have an RNA genome. The functional unit of the genome is a **gene.** A gene encodes a product, the **gene product,** most commonly a protein. The study of the function and transfer of genes is called **genetics,** whereas the study and analysis of the nucleotide sequence of DNA is called **genomics.** ■ **chromosome, p. 66** ■ **plasmid, p. 66**

All living cells must accomplish two general tasks in order to multiply. The double-stranded DNA must be duplicated before cell division so that its encoded information can be passed on to future generations. This is the process of **DNA replication.** In addition, the information encoded by the DNA must be deciphered, or **expressed,** so that the cell can synthesize the necessary gene products at the appropriate time. Gene expression involves two interrelated processes, transcription and translation. **Transcription** copies the information encoded in DNA into a slightly different molecule, RNA. The RNA serves as a transitional, temporary form of the genetic information and is the one that is actually deciphered. **Translation** interprets information carried by RNA to synthesize the encoded protein. The chemistry and structure of DNA and RNA ensure that each of these processes can occur with great accuracy.

The flow of information from DNA to RNA to protein is often referred to as the **central dogma of molecular biology (figure 7.1).** It was once believed that information flow proceeded only in this direction. Although this direction is by far the most common, certain viruses, such as the one that causes AIDS, have an RNA genome but copy that information into the form of DNA.

Characteristics of DNA

A single strand of DNA is composed of a series of deoxyribonucleotide subunits, more commonly called nucleotides. These are joined in a chain by a covalent bond between the $5'PO_4$ (5 prime phosphate) group of one nucleotide and the 3′OH (3 prime hydroxyl) group of the next. Note that the designations 5′ and 3′ refer to the numbered carbon atoms of the pentose sugar of the nucleotide (see figure 2.22). Joining of the nucleotides in this manner creates a series of alternating sugar and phosphate moieties, called the **sugar-phosphate backbone.** Connected to each sugar is one of the nitrogenous bases, an adenine (A), thymine (T), guanine (G), or cytosine (C). Because of the chemical structure of the nucleotides and how they are joined, a single strand of DNA will always have a $5'PO_4$ group at one end and a 3′OH group at the other. These ends are often referred to as the **5′ end** and the **3′ end** and have important implications in DNA and RNA synthesis that will be discussed later. ■ **deoxyribonucleic acid (DNA), p. 31** ■ **nucleotides, p. 31**

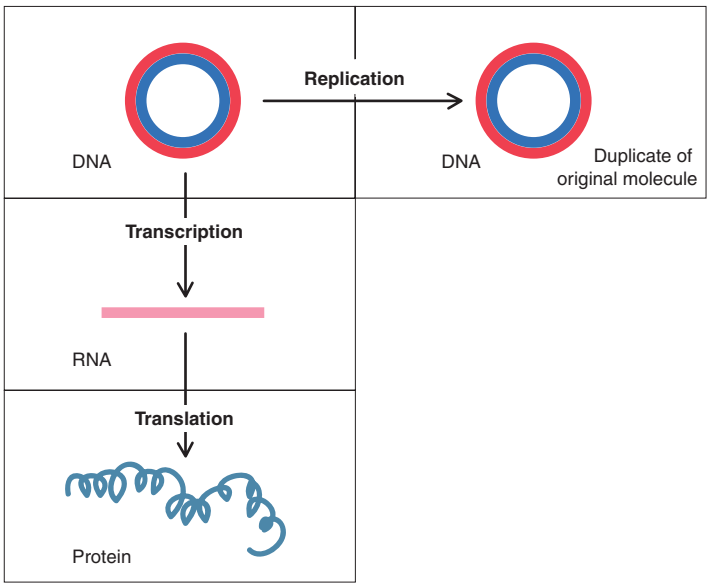


Figure 7.1 Overview of Replication, Transcription, and Translation DNA replication is the process that duplicates DNA so that its encoded information can be passed on to future generations. Transcription is the process that copies the genetic information into a transitional form, RNA. Translation is the process that deciphers the encoded information to synthesize a specific protein.

The DNA in a cell usually occurs as a double-stranded, helical structure **(figure 7.2).** The two strands are held together by weak hydrogen bonds between the nitrogenous bases of the opposing strands. While individual hydrogen bonds are readily broken, the duplex structure of double-stranded DNA is quite stable because of the sheer number of bonds that occurs along its length. Because short fragments of DNA have correspondingly fewer hydrogen bonds, they are readily separated into single-stranded pieces. Separating the two strands is called **denaturing.** ■ **hydrogen bonds, p. 21**

The two strands of double-stranded DNA are complementary **(figure 7.3).** Wherever an adenine is in one strand, a thymine is in the other; these two opposing nucleotides are held together by two hydrogen bonds between them. Similarly, wherever a cytosine is in one strand, a guanine is in the other. These are held together by the formation of three hydrogen bonds, a slightly stronger attraction than that of an A:T pair. The characteristic bonding of A to T and G to C is called **base-pairing** and is fundamental to the remarkable functionality of DNA. Because of the rules of base-pairing, one strand can always be used as a **template** for the synthesis of the complementary opposing strand. ■ **complementary, p. 33**

While the two strands of DNA in the double helix are complementary, they are also **antiparallel.** That is, they are oriented in opposite directions. One strand is oriented in the 5′ to

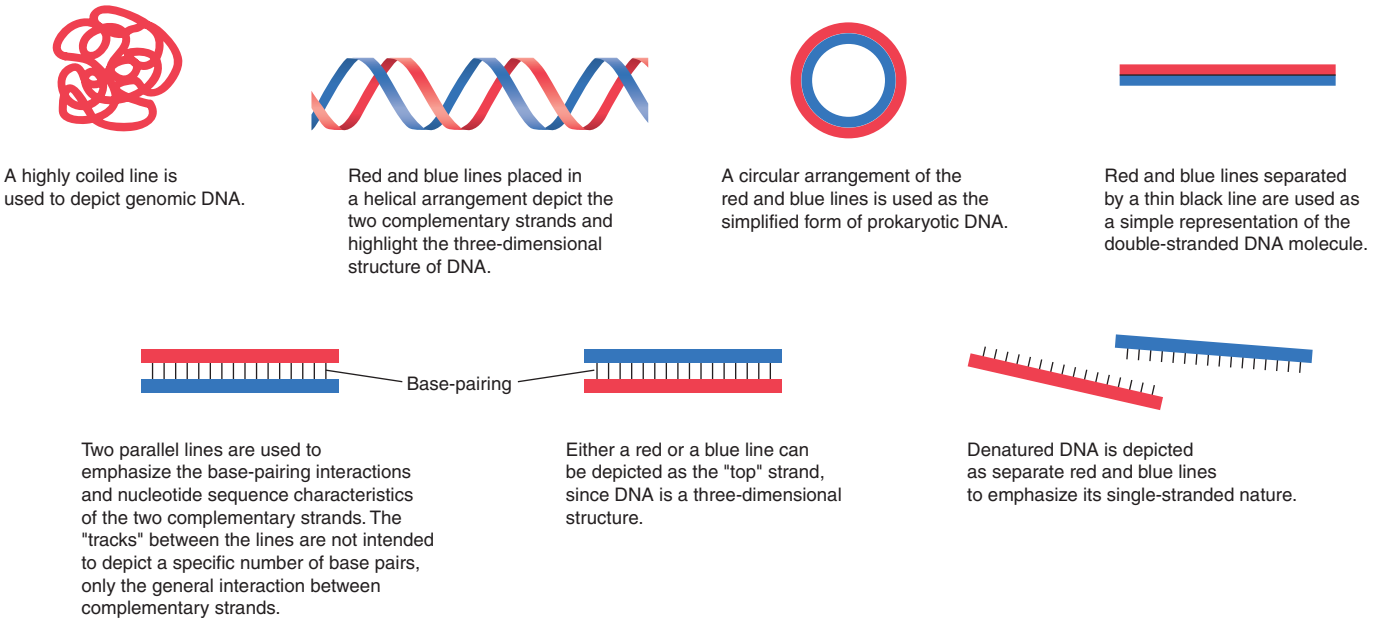


Figure 7.2 Diagrammatic Representations of the Structure of DNA Although DNA is a double-stranded helical structure, explanatory diagrams may depict it in a number of different ways. In this chapter we will use these color-coded representations.

3′ direction and its complement is oriented in the 3′ to 5′ direction. This also has important implications in the function and synthesis of nucleic acids.

Characteristics of RNA

RNA is in many ways comparable to DNA, but with some important exceptions. One difference is that RNA is made up of ribonucleotides rather than deoxynucleotides, although in both cases these are usually referred to simply as nucleotides. Another distinction is that RNA contains the nitrogenous base uracil in place of the thymine found in DNA. Like DNA, RNA consists of a sequence of nucleotides, but RNA usually exists as a single-stranded linear molecule that is much shorter than DNA. ■ **ribonucleic acid (RNA), p. 33** ■ **nucleotide, p. 31**

A fragment of RNA, a **transcript,** is synthesized using a region of one of the two strands of DNA as a template. In making the RNA transcript, the same base-pairing rules of DNA apply except uracil, rather than thymine, base-pairs with adenine. This base-pairing is only transient, however, and the molecule quickly leaves the DNA template. Numerous different RNA transcripts can be generated from a single chromosome using specific regions as templates. Either strand may serve as the template. In a region the size of a single gene, however, only one of the two strands is generally transcribed. As a result, two complementary strands of RNA are not normally generated.

There are three different functional groups of RNA molecules, each one transcribed from different genes. Most genes encode proteins and are transcribed into **messenger RNA (mRNA).** These molecules are translated during protein synthesis. Encrypted information in mRNA is deciphered according to the **genetic code,** which correlates each set of three nucleotides, called a **codon,** to a particular amino acid. Some genes are never translated into proteins; instead the RNAs themselves are the ultimate products. These

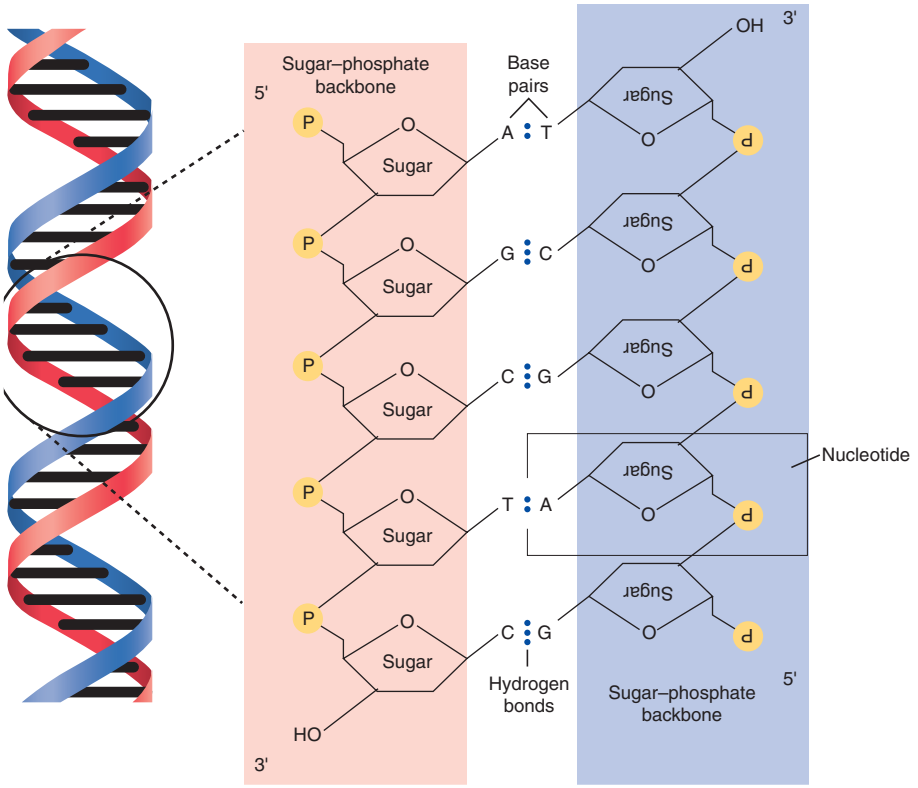


Figure 7.3 The Double Helix of DNA The two strands of DNA are antiparallel; one strand is oriented in the 5′ to 3′ direction, and its complement is oriented in the 3′ to 5′ direction. Hydrogen bonding occurs between the complementary base pairs; three bonds form between a G-C base pair, and two bonds form between an A-T base pair.

genes encode either **ribosomal RNA (rRNA)** or **transfer RNA (tRNA),** each of which plays a different but critical role in protein synthesis.

Regulating the Expression of Genes

While the basic structure of DNA and RNA is relatively simple, the information the molecules encode is extensive and complex. The nucleotide sequence contains genes that encode the amino acid sequence of proteins, and it also encodes mechanisms to regulate expression of those genes. Not all proteins are required by a cell in the same quantity and at all times; therefore, mechanisms that determine the extent and duration of their synthesis are needed.

One of the key mechanisms a cell uses to control protein synthesis is to regulate the synthesis of mRNA molecules. Unless a gene is transcribed into mRNA, the encoded protein cannot be synthesized. The number of mRNA copies of the gene also influences the level of expression. If transcription of a gene ceases, the level of gene expression rapidly declines. This is because mRNA is short-lived, often only a few minutes, due to the activity of enzymes called **RNases** that rapidly degrade it.

MICROCHECK 7.1

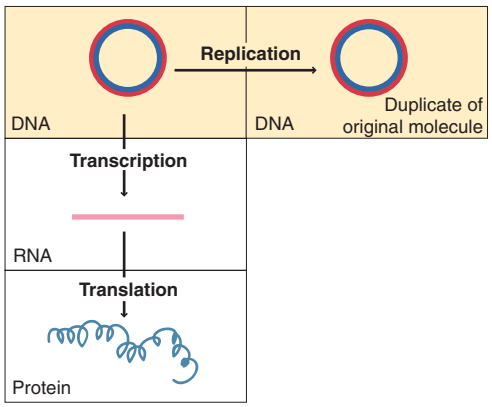
Replication is the process of duplicating double-stranded DNA. Transcription is the process of copying the information encoded in DNA into RNA. Translation is the process of interpreting the information carried by messenger RNA in order to synthesize the encoded protein.

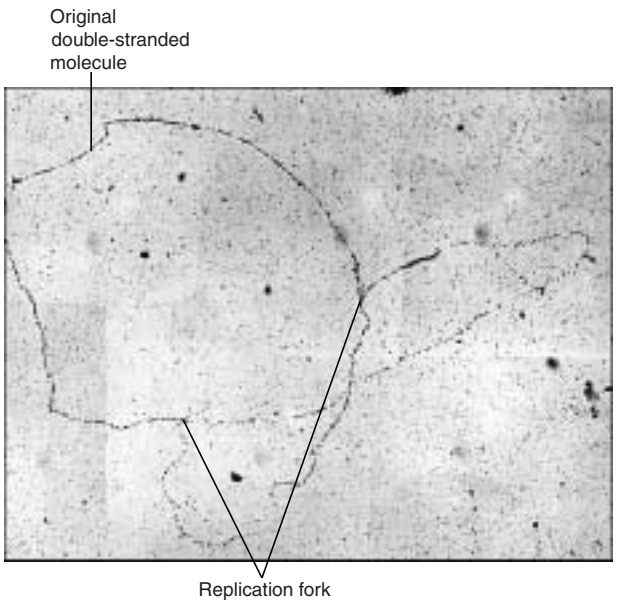
- How does the 5′ end of DNA differ from the 3′ end?
- If the nucleotide sequence of one strand of DNA is 5′ ACGTTGCA 3′, what is the sequence of the complementary strand?
- Why is a short-lived RNA important in cell control mechanisms?

7.2 DNA Replication

DNA is replicated in order to create a second DNA molecule, identical to the original. Each of the two cells generated during binary fission then receives one complete copy. ■ **binary fission, p. 84**

DNA replication is generally **bidirectional.** From a distinct starting point in circular DNA, replication proceeds in opposite directions, creating an ever-expanding "bubble" of two identical replicated portions of the chromosome **(figure 7.4).** Bidirectional replication allows an entire chromosome to





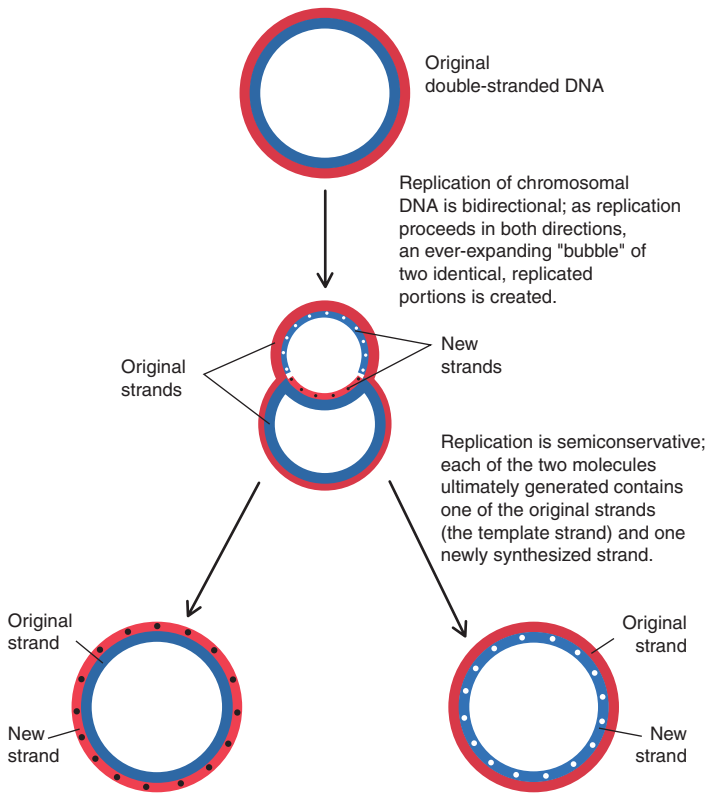


Figure 7.4 Replication of Chromosomal DNA of Prokaryotes

be replicated in half the time it would take if replication were unidirectional.

Replication of double-stranded DNA is **semiconservative.** Each of the two molecules generated contains one of the original strands (the template strand) and one newly synthesized strand. Thus, the two cells produced as a result of division each have one of the original strands of DNA paired with a new complementary strand.

The process of DNA replication requires the coordinated action of many different enzymes and other proteins **(table 7.1).** The most critical of these exist together as a complex called a **replisome.** The replisome appears to act as a fixed DNA-synthesizing factory, reeling in the DNA to be replicated. **DNA polymerases** are enzymes in the replisome that synthesize DNA, using one strand as a template to generate the complementary strand. These enzymes can only add nucleotides onto a preexisting fragment of nucleic acid, either DNA or RNA. Thus, the fragment serves as a **primer** from which synthesis can continue.

DNA is synthesized one nucleotide at a time as the deoxynucleoside triphosphates (dATP, dGTP, dCTP, and dTTP) are covalently joined to the nucleotide at the 3′ end of the growing strand. Hydrolysis of a phosphate bond in the incoming molecule provides energy for the reaction. DNA polymerase always elongates the chain in the 5′ to 3′ direction. Because the two DNA strands are antiparallel, however, the enzyme must "read" the template strand in the 3′ to 5′ direction **(figure 7.5).** The base-pairing rules determine the specific nucleotides added.

The replication process is very accurate, resulting in only one mistake approximately every billion nucleotides. Part of the reason for this remarkable precision is the proofreading ability of some DNA polymerases. If an incorrect nucleotide is incorporated into the growing chain, the enzyme can edit the mistake by replacing that nucleotide before moving on.

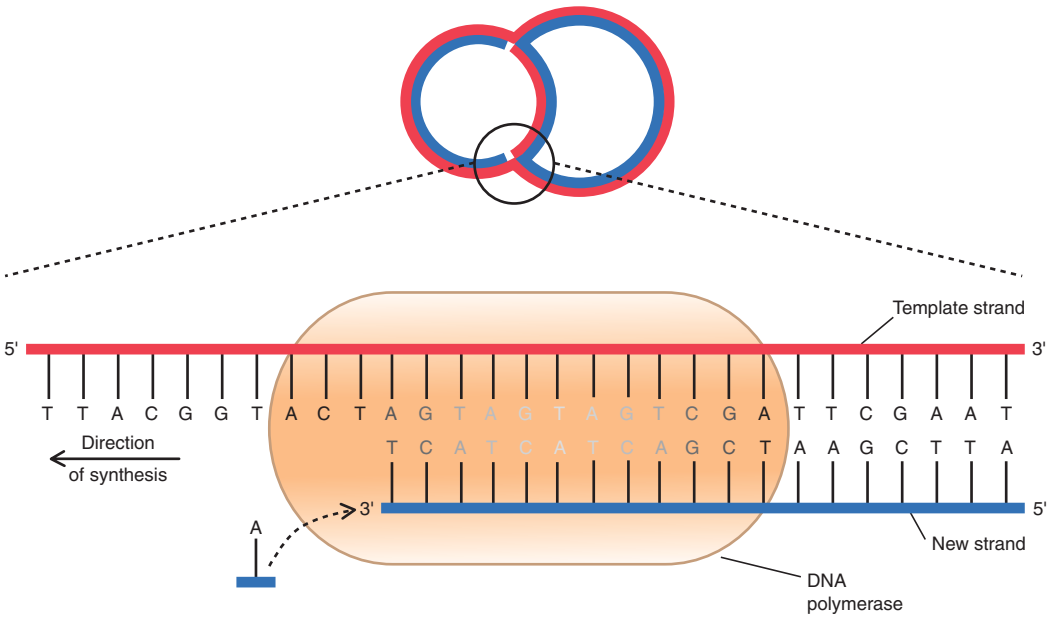


Figure 7.5 The Process of DNA Synthesis DNA polymerase synthesizes a new strand by adding one nucleotide at a time to the 3′ end of the elongating strand. Because DNA is synthesized in the 5′ to 3′ direction, the enzyme must "read" the template strand in the 3′ to 5′ direction. The base-pairing rules determine the specific nucleotides that are added.

Table 7.1 Components of DNA Replication in Prokaryotes

Component	Comments
Primer	Fragment of nucleic acid to which DNA polymerase can add nucleotides (the enzyme can only add nucleotides to a preexisting fragment).
DNA gyrase	Enzyme that temporarily breaks the strands of DNA, relieving the tension caused by unwinding the two strands of the DNA helix; it is the target of a class of antibacterial medications called fluoroquinolones.
DNA ligase	Enzyme that joins two DNA fragments by forming a covalent bond between the sugar-phosphate residues of adjacent nucleotides.
DNA polymerases	Enzymes that synthesize DNA; they use one strand of DNA as a template to generate the complementary strand. Synthesis always occurs in the 5′ to 3′ direction.
Helicases	Enzymes that unwind the DNA helix ahead of the replication fork.
Okazaki fragment	Nucleic acid fragment generated during discontinuous replication of the lagging strand of DNA.
Origin of replication	Distinct region of a DNA molecule at which replication is initiated.
Primase	Enzyme that synthesizes small fragments of RNA to serve as primers for DNA synthesis during discontinuous replication of the lagging strand.

It takes approximately 40 minutes for the chromosome of *E. coli* to be replicated, regardless of the environmental conditions. How, then, can *E. coli* sometimes multiply with a generation time of only 20 minutes? Under favorable growing conditions, a cell initiates replication before the preceding round of replication is completed. In this way, the two progeny resulting from cell division each will get one complete chromosome that has already started another round of replication. ■ **generation time, p. 85**

Initiation of DNA Replication

To begin the process of DNA replication, specific proteins must recognize and bind to a distinct region of the DNA, called an **origin of replication.** All molecules of DNA, including chromosomes and plasmids, must have this region of approximately 250 nucleotides for replication to be initiated. The binding of the proteins causes localized denaturation, or **melting,** of a specific region within the origin. Using the exposed single strands as templates, small fragments of RNA are synthesized to serve as primers for DNA synthesis. The enzymes that synthesize RNA do not require a primer.

The Replication Fork

The bidirectional progression of replication around a circular DNA molecule creates two advancing Y-shaped regions where active replication is occurring. Each of these is called a **replication fork.** The template strands continue to "unzip" at each fork due to the activity of enzymes called **helicases.** Synthesis of one new strand proceeds continuously in the 5′ to 3′ direction, as fresh single-stranded template DNA is exposed **(figure 7.6).** This strand is called the **leading strand.** Synthesis of the opposing strand, the **lagging strand,** is considerably more complicated because the DNA polymerase cannot add nucleotides to the 5′ end of DNA. Instead, synthesis must be reinitiated periodically as advancement of the replication fork exposes more of the template DNA. Each initiation event must be preceded by the synthesis of an RNA primer by the enzyme **primase.** The result is the synthesis of a series of fragments, called **Okazaki fragments,** each of which begins with a short stretch of RNA. As DNA polymerase adds nucleotides to the 3′ end of an Okazaki fragment, it eventually reaches the initiating point of the previous fragment. A different type of DNA polymerase then removes those RNA primer nucleotides and simultaneously replaces them with deoxynucleotides. The enzyme **DNA ligase** seals the gaps between fragments by catalyzing the formation of a covalent bond between the adjacent nucleotides.

Several other proteins are also involved in DNA replication. Among them is **DNA gyrase,** an enzyme that temporarily breaks the strands of DNA, relieving the tension caused by the unwinding of the two strands of the DNA helix. This enzyme is one of the targets of ciprofloxacin and other members of a class of antibacterial drugs called fluoroquinolones. By inhibiting the function of gyrase, the fluoroquinolones interfere with bacterial DNA replication and prevent the growth of bacteria. ■ **fluoroquinolones, p. 517**

MICROCHECK 7.2

DNA polymerases synthesize DNA in the 5′ to 3′ direction, using one strand as a template to generate the complementary strand. Replication of DNA begins at a specific sequence, called the origin of replication, and then proceeds bidirectionally, creating two replication forks.

- Why is a primer required for DNA synthesis?
- How does synthesis of the lagging strand differ from that of the leading strand?
- If DNA replication were shown to be "conservative," what would this mean?

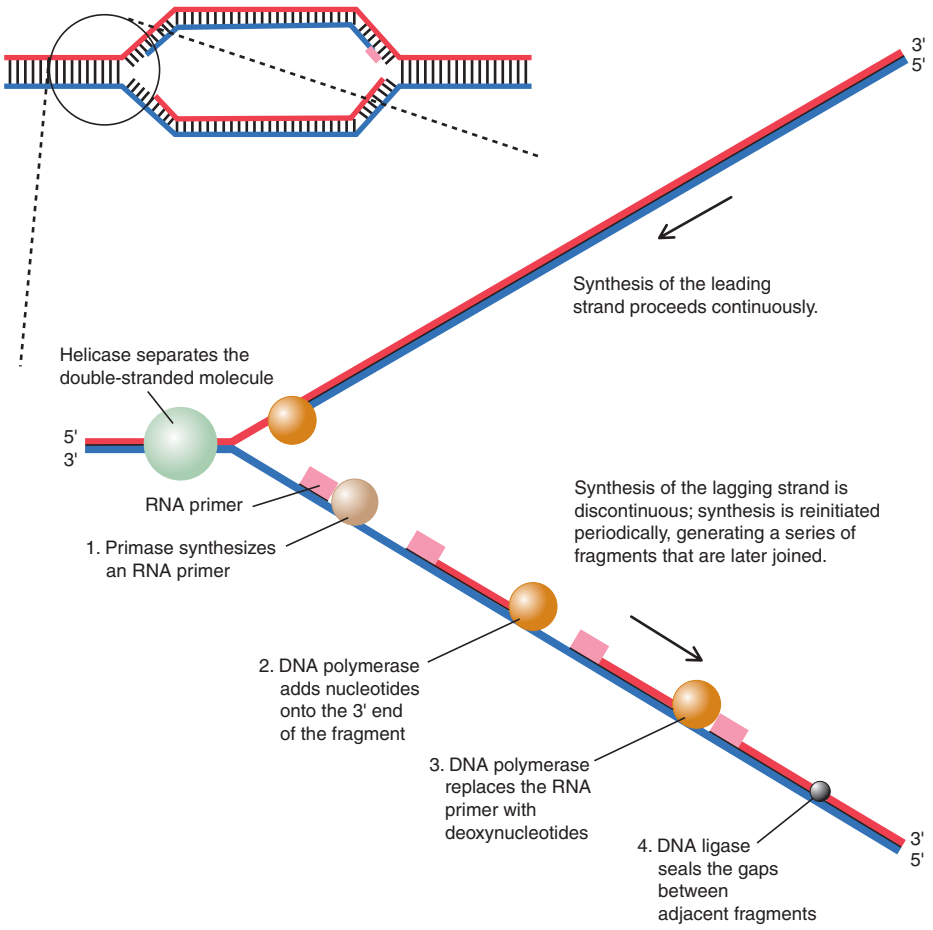


Figure 7.6 The Replication Fork This simplified diagram of the replication fork highlights the key steps in the synthesis of the lagging strand.

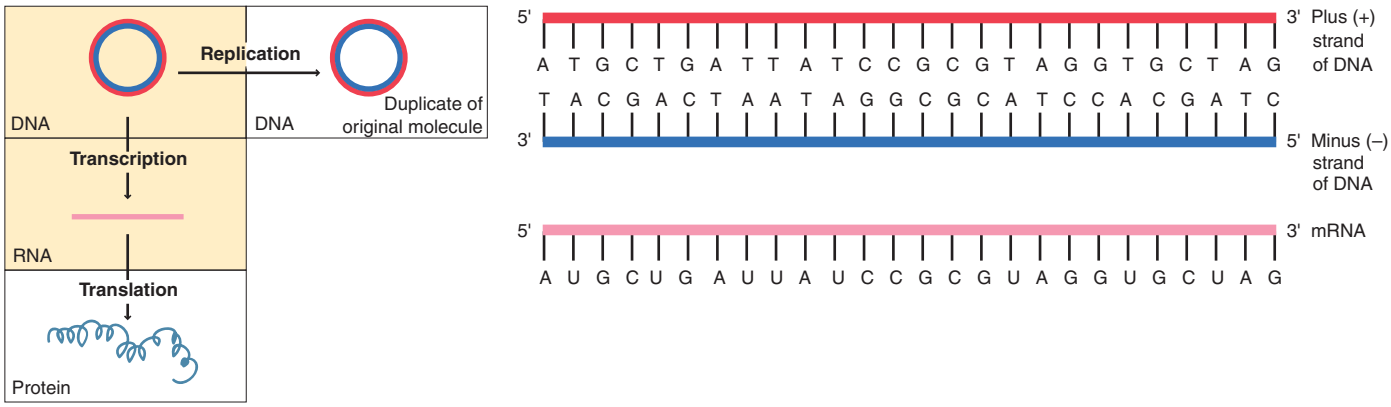


Figure 7.7 **RNA Is Transcribed from a DNA Template** The DNA strand that serves as a template for RNA synthesis is called the (–) strand of DNA. The nucleotide sequence of the transcript is analogous to that of the (+) strand, with uracil (U) occurring in place of thymine (T) in the mRNA.

7.3 Gene Expression

Gene expression involves two separate but interrelated processes, transcription and translation. Transcription is the process of synthesizing RNA from a DNA template. During translation, information encoded on an mRNA transcript is deciphered to synthesize a protein.

Transcription

The enzyme RNA polymerase catalyzes the process of transcription, producing a single-stranded RNA molecule that is complementary and antiparallel to the DNA template (**figure 7.7**). To describe the two strands of DNA in a region that is transcribed into RNA, the terms **minus (–) strand** and **plus (+) strand** are sometimes used (**table 7.2**). The strand that serves as the template for RNA synthesis is called the minus (–) strand, whereas its complement is called the plus (+) strand. Recall that the base-pairing rules of DNA and RNA are the same, except that RNA contains uracil in place of thymine. Therefore, because the RNA is complementary to the (–) strand, its nucleotide sequence is the same as the (+) strand, except it has uracil in place of thymine. Likewise, the RNA transcript has the same 5′ to 3′ direction, or **polarity,** as the (+) strand.

In prokaryotes, an mRNA molecule can carry the information for one or multiple genes. A transcript that carries one gene is called **monocistronic** (a cistron is synonymous with a gene). Those that carry multiple genes are called **polycistronic.** Generally, the proteins encoded on a polycistronic message are all involved in a single biochemical pathway. This enables the cell to express related genes in a coordinated manner.

Table 7.2 **Components of Transcription in Prokaryotes**

Component	Comments
–strand	Strand of DNA that serves as the template for RNA synthesis; the resulting RNA molecule is complementary to this strand.
+strand	Strand of DNA complementary to the one that serves as the template for RNA synthesis; the sequence of the resulting RNA molecule is analogous to this strand.
Promoter	Nucleotide sequence to which RNA polymerase binds to initiate transcription.
RNA polymerase	Enzyme that synthesizes RNA using single-stranded DNA as a template; synthesis always occurs in the 5′ to 3′ direction.
Sigma (σ) factor	Component of RNA polymerase that recognizes the promoter regions. A cell may have different types of σ factors that recognize different promoters. These may be expressed at different stages of cell growth, enabling the cell to transcribe specialized sets of genes as needed.
Terminator	Sequence at which RNA synthesis stops; the RNA polymerase falls off the DNA template and releases the newly synthesized RNA.

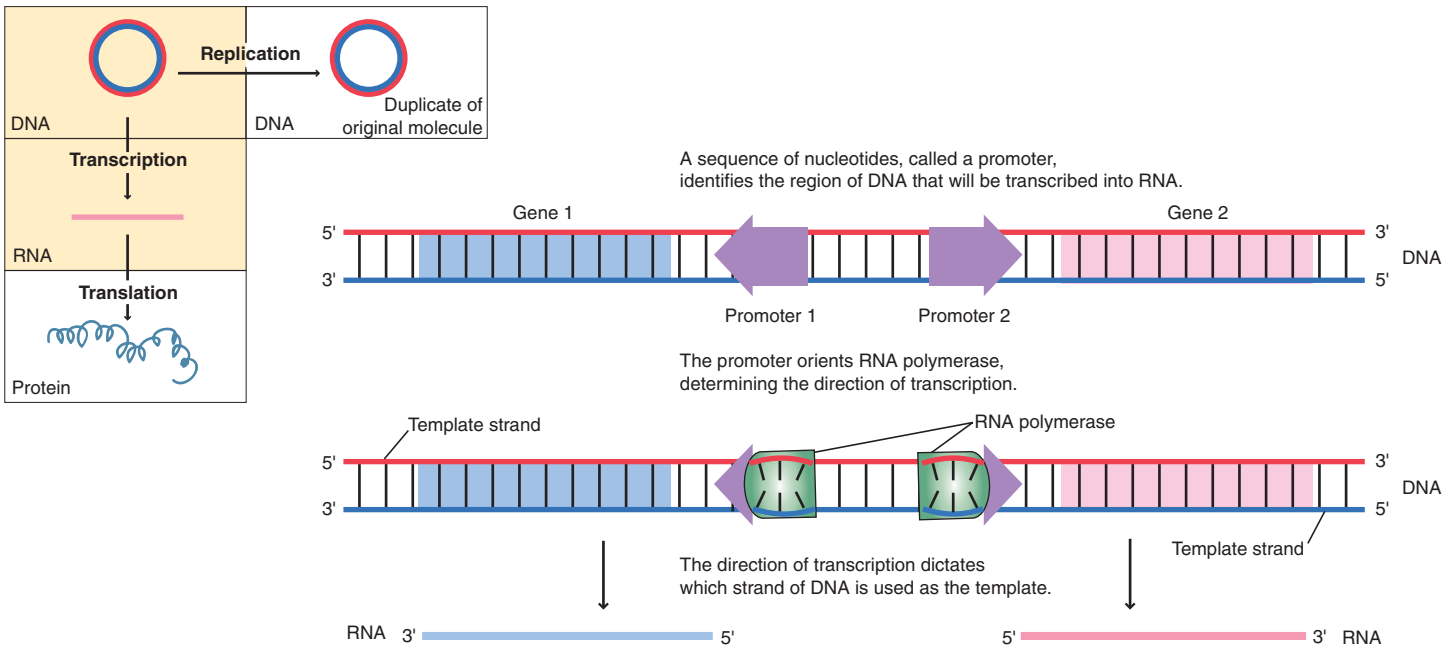


Figure 7.8 Promoters Direct Transcription A promoter not only identifies the region of DNA that will be transcribed into RNA, its orientation determines which strand will be used as the template.

Transcription begins when RNA polymerase recognizes a sequence of nucleotides on the DNA called a **promoter.** The promoter identifies the region of the DNA molecule that will be transcribed into RNA. In addition, the promoter orients the RNA polymerase in one of the two possible directions. This dictates which of the two DNA strands is used as a template **(figure 7.8).** Like DNA polymerase, RNA polymerase can only synthesize nucleic acid in the 5′ to 3′ direction and must "read" the template in the 3′ to 5′ direction. Unlike DNA polymerase, however, RNA polymerase can begin to synthesize a new chain without a primer.

The transcribed RNA molecule can be used as a reference point to describe direction on the analogous DNA. **Upstream** implies the direction toward the 5′ end of the transcribed region, whereas **downstream** implies the direction toward the 3′ end. Thus, a promoter is upstream of a gene.

Initiation of RNA Synthesis

Transcription begins when RNA polymerase recognizes and binds to a promoter region on the double-stranded DNA molecule. The binding melts a short stretch of DNA, creating a region of exposed nucleotides that serves as a template for RNA synthesis.

A particular subunit of RNA polymerase recognizes the promoter region prior to the initiation of transcription. This subunit, **sigma** (σ) **factor,** is only a loose component of the enzyme. After transcription is initiated, the σ factor dissociates from the enzyme, leaving the remaining portion of RNA polymerase, called the **core enzyme,** to complete transcription. A cell may have different types of σ factors that recognize different promoters. These may be expressed at different stages of cell growth, enabling the cell to transcribe specialized sets of genes as needed.

Elongation

In the elongation phase, the RNA polymerase moves along the template strand of the DNA, synthesizing the complementary single-stranded RNA molecule. The RNA molecule is synthesized in the 5′ to 3′ direction as the enzyme adds nucleotides to the 3′OH group at the end of the growing chain. The core RNA polymerase advances along the DNA, melting a new stretch and allowing the previous stretch to close **(figure 7.9).** This exposes a new region of the template, permitting the elongation process to continue. The rate of polymerization is about 30 nucleotides per second.

Once elongation has proceeded far enough for RNA polymerase to clear the promoter, another molecule of RNA polymerase can bind to that promoter, initiating a new round of transcription. Thus, a single gene can be transcribed multiple times in a very short time interval.

Termination

Just as an initiation of transcription occurs at a distinct site on the DNA, so does termination. When RNA polymerase encounters a **terminator,** it falls off the DNA template and releases the newly synthesized RNA. The terminator is a sequence of nucleotides in the DNA that, when transcribed, permits two complementary regions of the resulting RNA to base-pair, forming a hairpin loop structure. For reasons that are not yet understood, this causes the RNA polymerase to stall, resulting in its dissociation from the DNA template and release of the

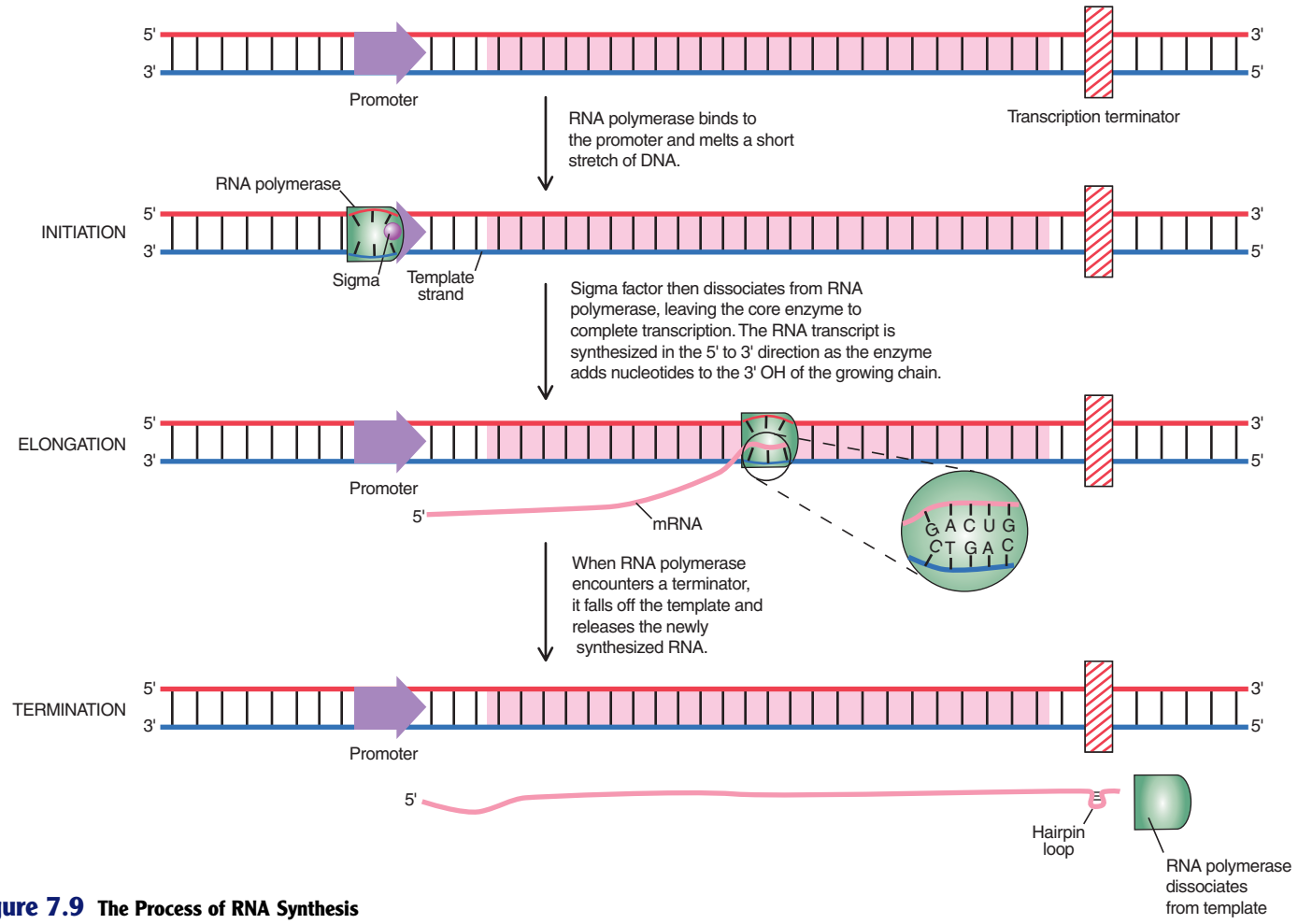


Figure 7.9 **The Process of RNA Synthesis**

RNA. The termination of transcription should not be confused with the termination of translation, which occurs by a totally different mechanism and will be discussed shortly.

Translation

Translation is the process of decoding the information carried on the mRNA to synthesize the specified protein. Proteins are synthesized by adding amino acid subunits sequentially to the carboxyl group at the end of an elongating polypeptide chain. Each amino acid added is specified by one codon of the mRNA, as directed by the genetic code. The process of translation requires three major components—mRNA, ribosomes, and tRNAs—in addition to various accessory proteins **(table 7.3).**

■ **carboxyl group, p. 26**

The Role of mRNA

The mRNA is a temporary copy of genetic information; it carries encoded instructions for synthesis of a specific polypeptide, or in the case of a polycistronic message, a specific group of polypeptides. That information is deciphered using the **genetic code (figure 7.10).** This is a universal code, used by all living things, and correlates each series of three nucleotides, a codon, with one amino acid. Because a codon is a sequence of any combination of the four nucleotides, there are 64 different codons (4^3). Three of these are stop codons; these will be discussed later. The remaining 61 translate to the 20 different amino acids. This means that more than one codon can encode a specific amino acid. For example, both ACA and ACG encode the amino acid threonine. Because of this redundancy, the genetic code is said to be **degenerate.** Note, however, that two different amino acids are never coded for by the same codon.

An equally important aspect of mRNA is that it carries the information that indicates where the coding region actually begins. This is critical because the genetic code is read as groups of three nucleotides. Thus, any given sequence has three possible **reading frames,** or ways in which triplets can be grouped **(figure 7.11).** If translation occurs in the wrong reading frame, a very different, and generally non-functional, polypeptide would be synthesized.

The Role of Ribosomes

Ribosomes serve as the sites of translation, and their structure facilitates the joining of one amino acid to another. Ribosomes bring each amino acid into a favorable position so that an

PERSPECTIVE 7.1 Making Sense of Antisense RNA

The knowledge gained through basic science research is fundamental to commercially valuable applications. For example, by understanding how genes are transcribed and translated, scientists can develop methods to suppress expression of certain genes. We know that only one strand of DNA is transcribed into a single strand of mRNA. This mRNA, the **sense strand** or **plus (+) strand,** is translated into a sequence of amino acids. An RNA molecule that is the complement of the sense strand is called an **antisense strand** or a **minus (–) strand.** Antisense RNA, which is not typically made by a cell, can base-pair with the sense strand to form a double-stranded RNA molecule, which cannot be translated.

Short fragments of antisense RNA can be chemically synthesized and used to interfere with gene expression. Recently, the first therapeutic drug based on antisense technology, fomivirsen, was approved for treating eye infections by cytomegalovirus (CMV) in AIDS patients. Fomivirsen is antisense RNA that is complementary to the mRNA of CMV; it prevents expression of two proteins required for viral replication. ■ **cytomegalovirus, p. 757**

Cells can also be genetically engineered to produce antisense RNA by introducing a copy of the gene with the promoter upstream from the antisense strand rather than from the sense strand **(figure 1).** Exploiting this principle, a plant biotechnology company genetically engineered tomato plants to synthesize antisense RNA of the gene that codes for the enzyme polygalacturonase. This plant enzyme breaks down plant cell walls and is responsible for the mushiness of ripe tomatoes a few days after they are picked. As a result of the genetic engineering, the tomatoes with antisense RNA to polygalacturonase do not get mushy for several weeks after they are picked, since the antisense RNA prevents polygalacturonase from being synthesized. Such technological achievement, however, does not guarantee economic success; commercially, the bioengineered tomatoes were a failure. ■ **genetic engineering, pp. 220, 230**

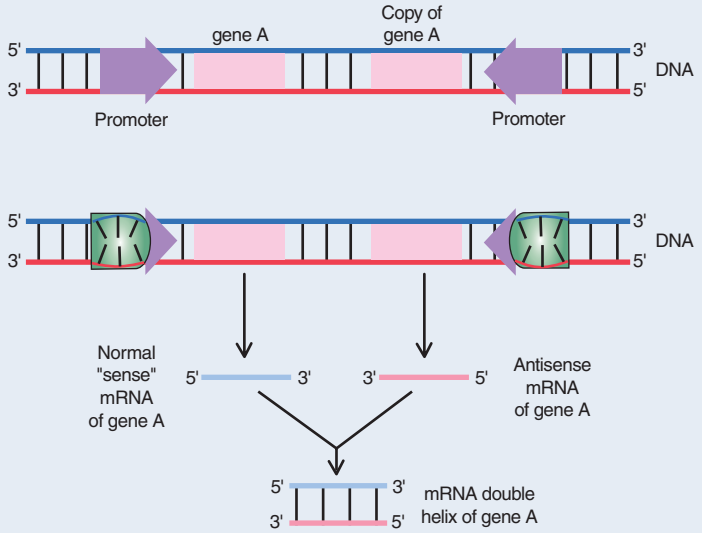


Figure 1 Formation of a Double-Stranded RNA Molecule The two copies of Gene A are oriented in opposite directions with respect to their promoters. When these genes are both transcribed, two complementary copies of mRNA (sense and antisense) are generated. These two molecules base-pair, forming a double-stranded molecule that cannot be translated into protein.

Middle Letter

First Letter	Reading frame 5′ 3′	U	C	A	G	Last Letter
U	UUU	Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine	U
	UUC	Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine	C
	UUA	Leucine	UCA Serine	UAA (Stop)	UGA (Stop)	A
	UUG	Leucine	UCG Serine	UAG (Stop)	UGG Tryptophan	G
C	CUU	Leucine	CCU Proline	CAU Histidine	CGU Arginine	U
	CUC	Leucine	CCC Proline	CAC Histidine	CGC Arginine	C
	CUA	Leucine	CCA Proline	CAA Glutamine	CGA Arginine	A
	CUG	Leucine	CCG Proline	CAG Glutamine	CGG Arginine	G
A	AUU	Isoleucine	ACU Threonine	AAU Asparagine	AGU Serine	U
	AUC	Isoleucine	ACC Threonine	AAC Asparagine	AGC Serine	C
	AUA	Isoleucine	ACA Threonine	AAA Lysine	AGA Arginine	A
	AUG	Methionine (Start)	ACG Threonine	AAG Lysine	AGG Arginine	G
G	GUU	Valine	GCU Alanine	GAU Aspartate	GGU Glycine	U
	GUC	Valine	GCC Alanine	GAC Aspartate	GGC Glycine	C
	GUA	Valine	GCA Alanine	GAA Glutamate	GGA Glycine	A
	GUG	Valine	GCG Alanine	GAG Glutamate	GGG Glycine	G

Figure 7.10 The Genetic Code The genetic code correlates each series of three nucleotides, a codon, with one amino acid. Three of the codons do not code for an amino acid and instead serve as a stop codon, terminating translation. AUG functions as a start codon.

enzyme can catalyze the formation of a peptide bond between them. The ribosome also helps to identify key punctuation sequences on the mRNA molecule, such as the point at which protein synthesis should be initiated. The ribosome moves along the mRNA in the 5′ to 3′ direction, "presenting" each codon in a sequential order for deciphering, while maintaining the correct reading frame.

A prokaryotic ribosome is composed of a 30S subunit and a 50S subunit, each of which is made up of protein and rRNA; the "S" stands for Svedberg unit, which is a unit of size **(figure 7.12).** Some of the ribosomal components are important in other aspects of microbiology as well. For example, comparison of the nucleotide sequences of rRNA molecules is playing an increasingly prominent role in the establishment of the genetic relatedness of various organisms. Medically, ribosomal proteins and rRNA are significant because they are the targets of several groups of antimicrobial drugs. ■ **ribosomal subunits, p. 67** ■ **rRNA sequencing, p. 255** ■ **antimicrobial drugs, p. 511**

Table 7.3 Components of Translation in Prokaryotes

Component	Comments
Anticodon	Sequence of three nucleotides in a tRNA molecule that is complementary to a particular codon in mRNA. The anticodon allows the tRNA to recognize and bind to the appropriate codon.
mRNA	Type of RNA molecule that contains the genetic information deciphered during translation.
Polyribosome (polysome)	Multiple ribosomes attached to a single mRNA molecule.
Reading frame	Grouping of a stretch of nucleotides into sequential triplets; an mRNA molecule has three reading frames, but only one is typically used in translation.
Ribosome	Structure that facilitates the joining of amino acids during the process of translation; composed of protein and ribosomal RNA. The prokaryotic ribosome (70S) consists of a 30S and 50S subunit; it is the target of several groups of antibacterial drugs.
Ribosome-binding site	Sequence of nucleotides in mRNA to which a ribosome binds; the first time the codon for methionine (AUG) appears after that site, translation generally begins.
rRNA	Type of RNA molecule present in ribosomes.
Start codon	Codon at which translation is initiated; it is typically the first AUG after a ribosome-binding site.
Stop codon	Codon that terminates translation, signaling the end of the protein; there are three stop codons.
tRNA	Type of RNA molecule that act as keys that interpret the genetic code; each tRNA molecule carries a specific amino acid.

The Role of Transfer RNA

The tRNAs are segments of RNA able to carry specific amino acids, thus acting as keys that interpret the genetic code. They each recognize and base-pair with a specific codon and in the process deliver the appropriate amino acid to that site. This recognition is made possible because each tRNA has an **anticodon,** three nucleotides complementary to a particular codon in the mRNA. The amino acid each tRNA carries is dictated by its anticodon and the genetic code **(figure 7.13).**

Although there are 64 different codons, there are fewer different tRNA molecules. For example, there are three **stop codons,** which signal the end of the protein; they have no corresponding tRNA molecules that recognize them. Also, the anticodon of some tRNA molecules can recognize more than one codon. It appears that a certain amount of "wobbling" is tolerated in the base-pairing so that recognition of the third nucleotide of the codon is not always precise. Due to the degeneracy of the genetic code, however, the correct amino acid is still incorporated into the polypeptide chain.

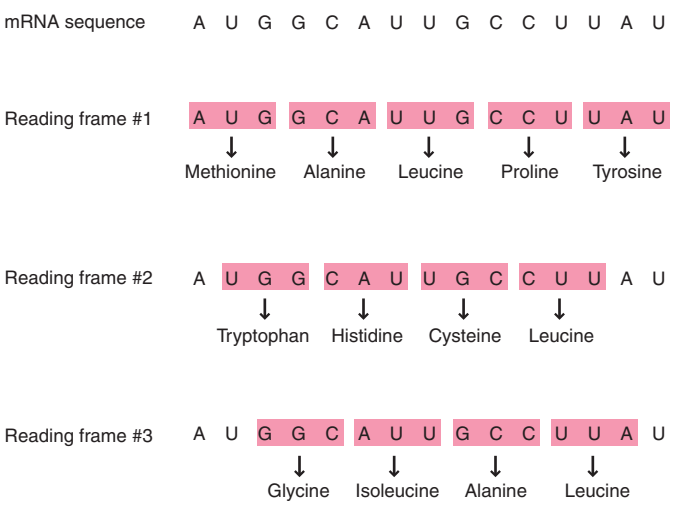


Figure 7.11 Reading Frames A nucleotide sequence has three potential reading frames. Because each reading frame encodes a very different order of amino acids, translation of the correct reading frame is important.

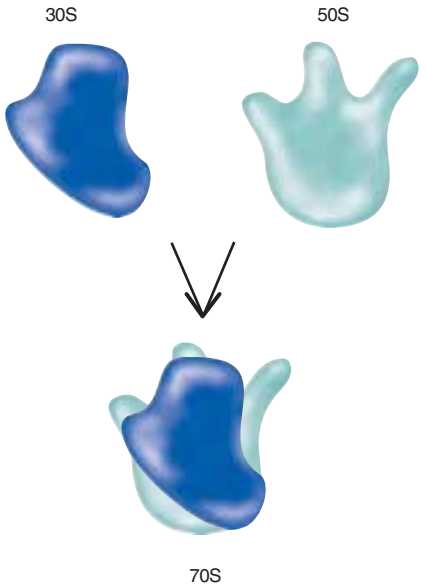


Figure 7.12 The Structure of the 70S Ribosome The 70S ribosome is composed of a 30S subunit and a 50S subunit.

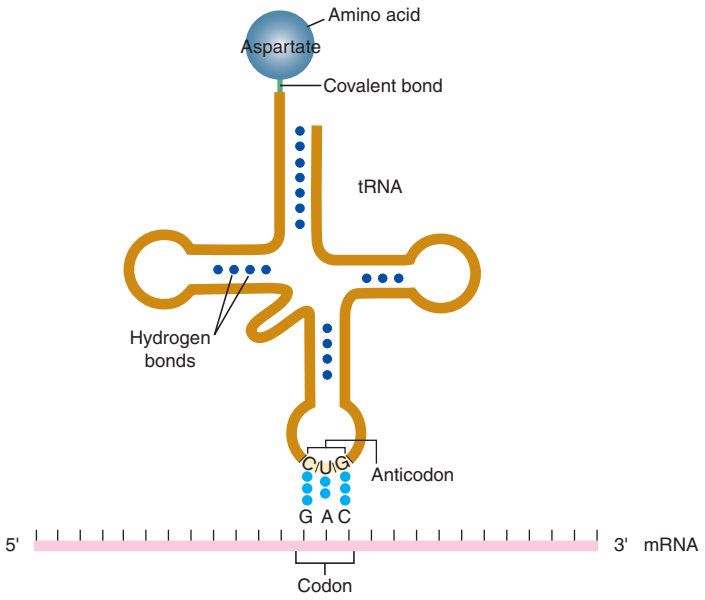


Figure 7.13 The Structure of Transfer RNA (tRNA) The anticodon of the tRNA base-pairs with a specific codon in the mRNA; by doing so, the appropriate amino acid is delivered to the site. The amino acid that the tRNA carries is dictated by the genetic code. The tRNA that recognizes the codon GAC carries the amino acid aspartate.

Initiation of Translation

In prokaryotes, translation begins as the mRNA is still being synthesized **(figure 7.14).** The 30S subunit of the ribosome binds to a sequence in mRNA called the **ribosome-binding site.** The first time the codon for methionine (AUG) appears after that site, translation generally starts. That first AUG is typically 7 nucleotides downstream of the ribosome-binding site. Note that AUG functions as a **start codon** only when preceded by a ribosome-binding site; at other sites, it simply encodes methionine. The position of the first AUG is critical, as it determines the reading frame used for translation of the remainder of that protein.

At that first AUG, the ribosome begins to assemble. An **initiation complex,** consisting of the 30S ribosomal subunit, a tRNA that carries a chemically altered form of the amino acid methionine, *N*-formylmethionine or **f-Met,** and proteins called **initiation factors,** forms. Shortly thereafter, the 50S subunit of the ribosome joins that complex and the initiation factors leave, forming the 70S ribosome. The elongation phase then begins.

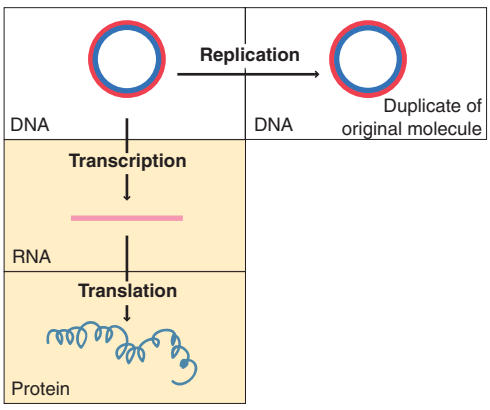
Elongation

The 70S ribosome has two sites to which tRNA-carrying amino acids can bind **(figure 7.15).** One is called the **P-site** (peptidyl site), and the other is called the **A-site** (aminoacyl site, commonly referred to as the acceptor site). The initiating tRNA, carrying the f-Met, binds to the P-site. A tRNA that recognizes the next codon on the mRNA then fills the unoccupied A-site. An enzyme joins the f-Met carried by the tRNA in the P-site to the amino acid carried by the tRNA that just entered the A-site. This transfers the amino acid from the initiating tRNA to the amino acid carried by the incoming tRNA. The ribosome advances, or **translocates,** a distance of one codon, and the tRNA that carried the f-Met is released through an adjacent site called the **E-site** (exit site). Translocation requires several different proteins, called **elongation factors.** As a result of translocation the remaining tRNA, which now carries the two-amino-acid chain, occupies the P-site; the A-site is transiently vacant. A tRNA that recognizes the next codon then quickly fills the empty A-site, and the process repeats.

Once translation has progressed far enough for the ribosome to clear the ribosome-binding site and the first AUG, another ribosome can bind to begin another round of synthesis of the encoded polypeptide. Thus, at any one time, multiple ribosomes can be translating a single mRNA molecule. This allows the maximal expression of protein from a single mRNA template. The assembly of multiple ribosomes attached to a single mRNA molecule is called a **polyribosome** or a **polysome.**

Termination

Elongation of the polypeptide terminates when the ribosome reaches a **stop codon,** a codon that does not code for an amino acid and is not recognized by a tRNA. At this point, enzymes called **release factors** free the newly synthesized polypeptide by breaking the covalent bond that joins it to the tRNA. The ribosome falls off the mRNA and dissociates into its two component subunits, 30S and 50S. These can then be reused to initiate translation at other sites.



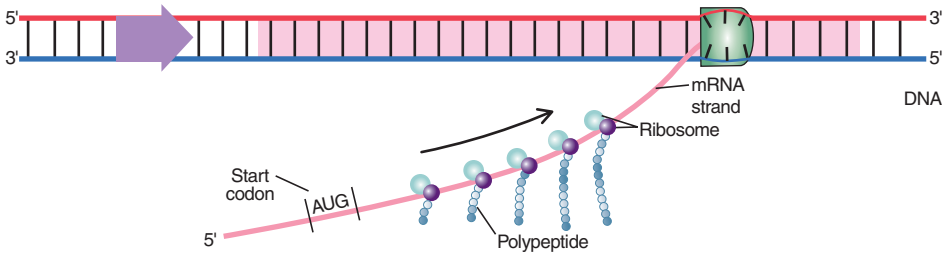


Figure 7.14 In Prokaryotes, Translation Begins As the mRNA Molecule Is Still Being Synthesized Ribosomes begin translating the 5′ end of the transcript even as the 3′ end is still being synthesized. More than one ribosome can be translating the same mRNA molecule.

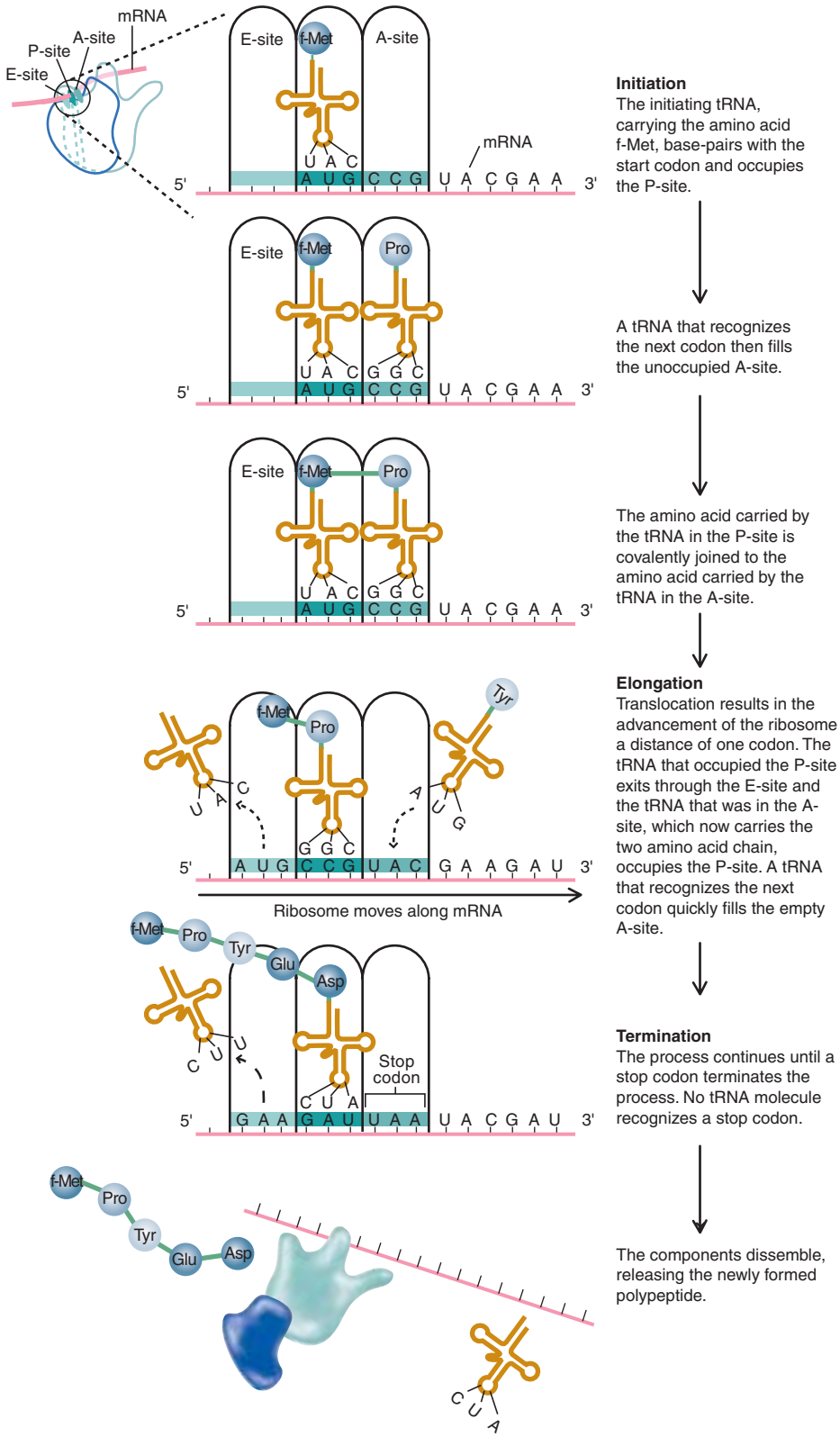


Figure 7.15 The Process of Translation

Post-Translational Modification

Proteins must often be modified after they are synthesized. For example, some proteins require the assistance of another protein, a **chaperone,** to fold into the final functional shape. Those proteins destined for transport outside of the cytoplasmic membrane also must be modified. Such proteins have a characteristic series of hydrophobic amino acids, a **signal sequence,** at their amino terminal end, which "tags" them for transport through the membrane. The signal sequence is removed when the protein leaves the cytoplasm. ■ chaperones, p. 29 ■ hydrophobic amino acids, p. 26

MICROCHECK 7.3

RNA polymerase recognizes sequences in DNA called promoters and at those sites initiates synthesis of RNA using one strand of DNA as a template. The RNA is synthesized in the 5′ to 3′ direction and synthesis stops when RNA polymerase encounters a terminator. Translation occurs as ribosomes move along mRNA in the 5′ to 3′ direction, with the ribosomes serving as the structure that facilitates the joining of one amino acid to another. tRNAs carry specific amino acids, thus acting to decode the genetic code.

- How does the orientation of the promoter dictate which strand is used as a template for RNA synthesis?
- Explain why it is important for the translation machinery to recognize the correct reading frame.
- Could two mRNAs have different nucleotide sequences and yet code for the same protein?

7.4 Differences Between Eukaryotic and Prokaryotic Gene Expression

Eukaryotes differ significantly from prokaryotes in several aspects of transcription and translation **(table 7.4).** For example, in eukaryotic cells, most mRNA molecules are extensively modified, or **processed,** in the nucleus during and after transcription. Shortly after transcription begins, the 5′ end of the transcript is modified, or **capped,** by the addition of a methylated guanine derivative, creating what is called a **cap.** The cap likely stabilizes the transcript and enhances translation. The 3′ end of the molecule is also modified, even before transcription

Table 7.4 Major Differences Between Prokaryotic and Eukaryotic Transcription and Translation

Prokaryotes	Eukaryotes
mRNA is not processed.	A cap is added to the 5′ end of mRNA, and a poly A tail is added to the 3′ end.
mRNA does not contain introns.	mRNA contains introns, which are removed by splicing.
Translation of mRNA begins as it is being transcribed.	The mRNA transcript is transported out of the nucleus so that it can be translated in the cytoplasm.
mRNA is often polycistronic; translation usually begins at the first AUG that follows a ribosome-binding site.	mRNA is monocistronic; translation begins at the first AUG.

has been terminated. This process, called **polyadenylation,** involves cleaving the transcript at a specific sequence of nucleotides and then adding approximately 200 adenine derivatives to the newly exposed 3′ end. This creates what is called a **poly A tail,** which is thought to stabilize the transcript as well as enhance translation. Another important modification is **splicing,** a process that removes specific segments of the transcript **(figure 7.16).** Splicing is necessary because eukaryotic genes are not always contiguous; they are often interrupted by noncoding nucleotide sequences. These intervening sequences, or **introns,** are transcribed along with the expressed regions, or **exons,** generating what is called **precursor mRNA.** The introns must be removed from precursor mRNA to form the mature mRNA that is then translated.

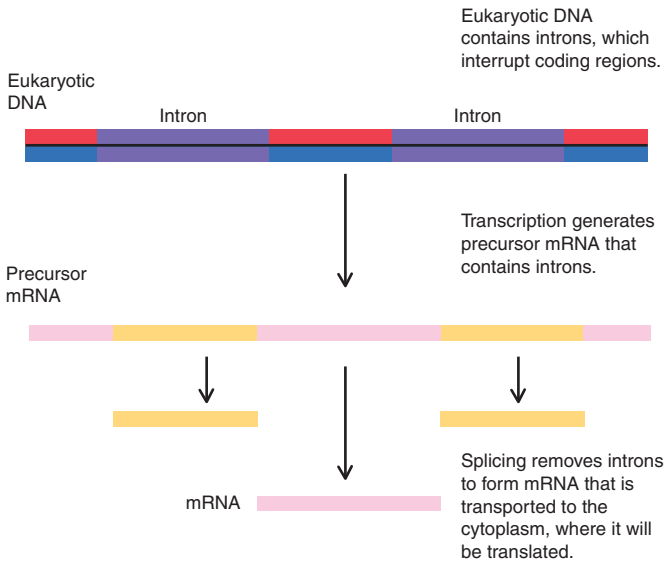


Figure 7.16 Splicing of Eukaryotic RNA

The mRNA in eukaryotic cells must be transported out of the nucleus before it can be translated in the cytoplasm. Thus, the same mRNA molecule cannot be transcribed and translated at the same time or even in the same cellular location. Unlike in prokaryotes, the mRNA of eukaryotes is generally monocistronic. Translation of the message generally begins at the first occurrence of AUG in the molecule.

The ribosomes of eukaryotes are different from those of prokaryotes. Whereas the prokaryotic ribosome is 70S, made up of 30S and 50S subunits, the eukaryotic ribosome is 80S, made up of 40S and 60S subunits. The differences in ribosome structure account for the ability of certain types of antibiotics to kill bacteria without causing significant harm to mammalian cells.

Some of the proteins that play essential roles in translation differ between eukaryotic and prokaryotic cells. Diphtheria toxin, which selectively kills eukaryotic but not prokaryotic cells, illustrates this difference. This toxin is produced by *Corynebacterium diphtheriae;* it binds to and inactivates one of the elongation factors of eukaryotes. Since this protein is required for translocation of the ribosome, translation ceases and the eukaryotic cell dies, resulting in the typical symptoms of diphtheria. ■ **diphtheria toxin, p. 570**

MICROCHECK 7.4

Eukaryotic mRNA must be processed, which involves capping, polyadenylation, and splicing. In eukaryotic cells, the mRNA must be transported out of the nucleus before it can be translated in the cytoplasm. Eukaryotic mRNA is monocistronic.

- What is an intron?
- Explain the mechanism of action of diphtheria toxin.
- Would a deletion of two base pairs have a greater consequence if it occurred in an intron or in an exon?

7.5 Genomics

Increasingly rapid methods of determining the nucleotide sequence of DNA have led to exciting advancements in genomics. Fueled by the commitment to sequence the entire human genome, scientists honed the methodologies by first sequencing the genomes of select microorganisms. In 1995, the sequence of the chromosome of *Haemophilus influenzae* was published, marking the first complete genomic sequence ever determined. The genome sequences of more than 75 other organisms have now been determined **(table 7.5).** A draft of the human genome is also complete.

Although sequencing methodologies are becoming more rapid, analyzing the resulting data and extracting the pertinent information is far more complex than it might initially seem. One of the most difficult steps is to locate and characterize the potential protein-encoding regions. Imagine trying to determine the amino acid sequence of a protein encoded by a 1,000-base-pair (bp) stretch of DNA without

PERSPECTIVE 7.2 RNA: The First Macromolecule?

The 1989 Nobel Prize in Chemistry was awarded to two Americans, Sidney Altman of Yale University and Thomas Cech of the University of Colorado, who independently made the surprising and completely unexpected observation that RNA molecules can act as enzymes. Before their studies, it was believed that only proteins had enzymatic activity. The key observation was made by Cech in 1982 when he was trying to understand how introns were removed from mRNA that coded for ribosomal RNA in the eukaryotic protozoan *Tetrahymena*. Since he was convinced that proteins were responsible for cutting out these introns, he added all of the protein in the cells' nuclei to the mRNA that still contained the introns.

As expected, the introns were cut out. As a control, Cech looked at the ribosomal RNA to which no nuclear proteins had been added, fully expecting that nothing would happen. Much to his surprise, the introns were also removed. It did not make any difference whether the protein was present—the introns were removed regardless. Thus, Cech could only conclude that the RNA acted on itself to cut out pieces of RNA.

The question remained of how widespread this phenomenon was. Did RNA have catalytic properties other than that of cutting out introns from rRNA? The studies of Altman and his colleagues, carried out simultaneously to and independently of Cech's, provided answers to these further questions. Altman's group found that RNA could convert a tRNA molecule from a precursor form to its final functional state. Additional studies have shown that enzymatic reactions in which catalytic RNAs, termed **ribozymes,** play a role are very widespread. Ribozymes have been shown to occur in the mitochondria of eukaryotic cells and to catalyze other reactions that resemble the polymerization of RNA. Whether catalytic RNA cuts out introns from mRNA in the nucleus is not known.

These observations have profound implications for evolution: Which came first, proteins or nucleic acids? The answer seems to be that nucleic acids came first, specifically RNA, which acted both as a carrier of genetic information as well as an enzyme. Billions of years ago, before the present universe in which DNA, RNA, and protein are found, probably the only macromolecule that existed was RNA. Once tRNA became available, these adapters could carry amino acids present in the environment to specific nucleotide sequences on a strand of RNA. In this scenario, the RNA functions as the genes as well as the mRNA.

knowing anything about the orientation of the promoter or the reading frame of the transcribed mRNA. Since either strand of the double-stranded DNA molecule could be the template strand, two entirely different mRNA molecules could potentially code for the protein. In turn, each of those two molecules has three reading frames, for a total of six reading frames. Yet only one of these actually codes for the protein. Understandably, computers are an invaluable aid and are used extensively in deciphering the meaning of the raw sequence data. In turn, this has resulted in the emergence of a new field,

Table 7.5 Representative Microorganisms Whose Genome Sequences Have Been Determined

Name of Organism	Genome Size (10^6 base pairs)	Important Characteristics
Agrobacterium tumefaciens	5.67	Plant pathogen; causes crown gall. Scientists use its plasmid to introduce desired genes into plant cells.
Bacillus subtilis	4.20	Endospore-former; has served as a model for studies of Gram-positive bacteria.
Borrelia burgdorferi	1.44	Important human pathogen; causative agent of Lyme disease.
Deinococcus radiodurans	3.28	Radiation-resistant bacterium. Genome consists of two chromosomes, a large plasmid and a small plasmid.
Escherichia coli K12	4.64	Has served as model for studies of Gram-negative bacteria. Common inhabitant of the intestinal tract.
Escherichia coli O157:H7	5.53	Important human pathogen; causes hemorrhagic colitis (bloody diarrhea).
Haemophilus influenzae	1.83	First bacterial genome sequenced; important human pathogen; causes ear and respiratory infections and meningitis, mostly in children.
Helicobacter pylori	1.66	Important human pathogen; causes gastric diseases, including stomach ulcers.
Lactococcus (Streptococcus) lactis	2.35	Important bacterium to the dairy industry; used to make cheeses and other fermented milk products.
Methanococcus jannaschii	1.75	First archaeal genome sequenced; also the first autotrophic organism sequenced; hyperthermophile isolated from a hydrothermal vent; strict anaerobe; methane producer.
Mycobacterium tuberculosis	4.40	Important human pathogen; causes tuberculosis.
Mycoplasma genitalium	0.58	Smallest known bacterial genome; represents what might be the minimal genome; human pathogen.
Pseudomonas aeruginosa	6.3	Important cause of the infection in burn victims and people who have cystic fibrosis.
Saccharomyces cerevisiae	13	Yeast; first eukaryotic genome completed.
Sinorhizobium (Rhizobium) meliloti	6.7	Fixes nitrogen; forms a symbiotic relationship with legumes.
Synechocystis species	3.57	Cyanobacterium; has served as a model for studies of photosynthesis.
Treponema pallidum	1.14	Important human pathogen; causes syphilis; has not been cultured *in vitro*.

bioinformatics, which has created the computer technology to store, retrieve, and analyze nucleotide sequence data.

Analyzing a Prokaryotic DNA Sequence

When analyzing a DNA sequence, the nucleotide sequence of the (+) strand is used to infer information contained in the corresponding RNA transcript. Because of this, terms like start codon, which actually refers to a sequence in mRNA, are used to describe sequences in DNA. For example, to locate the start codon AUG, which would be found in mRNA, one would look for the analogous sequence, ATG, in the (+) strand of the DNA molecule. In most cases it is not initially known which of the two strands is actually used as a template for RNA synthesis, so that both strands are potentially a (+) strand. Only after a promoter is located is it known which strand in a given region is actually the (+) strand.

To locate protein-encoding regions, computers are used to search for **open reading frames (ORFs),** stretches of DNA, generally longer than 300 bp, that begin with a start codon and end with a stop codon. An ORF potentially encodes a protein. Other characteristics, such as the presence of an upstream sequence that can serve as a ribosome-binding site, also indicate that an ORF encodes a protein.

The nucleotide sequence of the ORF or deduced amino acid sequence of the encoded protein can be compared with other known sequences by searching computerized databases of published sequences. Not surprisingly, as genomes of more organisms are being sequenced, information contained in these databases is growing at a remarkable rate. If the encoded protein shows certain amino acid similarities, or **homology,** to characterized proteins, a putative function can sometimes be assigned. For example, proteins that bind DNA share amino acid sequences in certain regions. Likewise, regulatory regions in DNA such as promoters can sometimes be identified based on the nucleotide homologies to known sequences. ■ **Regulatory proteins that bind DNA, p. 183**

The 580,070-base-pair genome of *Mycoplasma genitalium* is the smallest bacterial genome known. The predicted coding regions and the functional role of the genes are shown in **figure 7.17.**

Figure 7.17 Map of the *Mycoplasma genitalium* genome The wide arrows indicate the predicted protein-encoding regions. The orientation of the arrows indicates the direction of transcription; the color indicates the functional role.

MICROCHECK 7.5

The first genomic sequence of a microorganism was completed in 1995. The sequencing methodologies are quickly becoming more rapid, but analyzing the data and extracting the pertinent information is difficult.

- What is an open reading frame?
- Describe two things that can be learned by searching a computerized database for sequences that have homologies to a newly sequenced gene.
- There are some characteristic differences in the nucleotide sequences of the leading and lagging strands. Why might this be so?

7.6 Regulating Gene Expression

To cope with changing conditions in their environment, microorganisms have evolved elaborate control mechanisms to synthesize the maximum amount of cell material from a limited supply of energy. This is critical, because generally a microorganism must reproduce more rapidly than its competitors in order to be successful.

Consider the situation of *Escherichia coli*. For over 100 million years, it has successfully inhabited the gut of mammals, where it reaches concentrations of 10^6 cells per milliliter. In this habitat, it must cope with alternating periods of feast and famine. For a limited time after a mammal eats, *E. coli* in the large intestine prosper, wallowing in the milieu of amino acids, vitamins, and other nutrients. The cells actively take up these compounds they would otherwise synthesize, expending minimal energy. Simultaneously, the cells shut down their biosynthetic pathways, channeling the conserved energy into the rapid synthesis of macromolecules, including DNA, RNA, and protein. Under these conditions, the cells divide at their most rapid rate. Famine, however, follows the feast. Between meals, which may be many days in the case of some mammals, the rich source of nutrients is depleted. Now the cells' biosynthetic pathways must be activated, using energy and markedly slowing cell division. Cells dividing several times an hour in a nutrient-rich environment may divide only once every 24 hours in a famished mammalian gut.

A cell controls its metabolic pathways by two general mechanisms. The most immediate of these is the allosteric inhibition of enzymes. The most energy-efficient strategy, however, is to control the actual synthesis of the enzymes, making only what is required. To do this, cells have the ability to control expression of certain genes. ■ allosteric regulation, p. 140

Principles of Regulation

Not all genes are subjected to the same type of regulation. Many are routinely expressed, whereas others are either turned on or off by certain conditions. Enzymes are often described according to characteristics of the regulation that governs their synthesis:

- **Constitutive enzymes** are constantly synthesized; the genes that encode these enzymes are always active. Constitutive enzymes usually play indispensable roles in the central metabolic pathways. For example, the enzymes of glycolysis are constitutive. ■ central metabolic pathways, p. 142
- **Inducible enzymes** are not regularly produced; instead, their synthesis is turned on by certain conditions. Inducible enzymes are often involved in the utilization of specific energy sources. A cell would waste precious resources if it synthesized the enzyme when the energy source is not present. An example of an inducible enzyme is **β-galactosidase,** whose sole function is to break down the disaccharide lactose into its two component monosaccharides, glucose and galactose. The mechanisms by which the cell controls β-galactosidase synthesis serve as an important model for regulation and will be described shortly.
- **Repressible enzymes** are routinely synthesized, but they can be turned off by certain conditions. Repressible enzymes are generally involved in biosynthetic pathways, such as those that produce amino acids. Cells require a sufficient amount of a given amino acid to multiply; thus, the amino acid must be either synthesized or available as a component of the growth medium. If a certain amino acid is not present in the medium, then the cell must synthesize the enzymes involved in its manufacture. When the amino acid is supplied, however, synthesis of the enzymes would waste energy.

Mechanisms to Control Transcription

The mechanisms that a cell uses to prevent or facilitate transcription must be readily reversible, allowing cells to effectively control the relative number of transcripts made. In some cases, the control mechanisms may affect the transcription of only a limited number of genes; in other cases, a wide array of genes is coordinately controlled. For example, in *E. coli* the expression of more than 300 different genes is affected by the availability of glucose as an energy source. The simultaneous regulation of numerous genes unrelated in function is called **global control.**

Transcription of genes is often controlled by means of a regulatory region near the promoter to which a specific protein can bind, acting as a sophisticated on/off switch. When a regulatory protein binds to DNA, it can either act as a repressor, which blocks transcription, or an activator, which facilitates transcription. A set of adjacent genes coordinately controlled by a regulatory protein and transcribed as a single polycistronic message is called an **operon.**

Repressors

A **repressor** is a regulatory protein that blocks transcription. It does this by binding to DNA at a region, called the **operator,** located immediately downstream of a promoter. This effectively prevents RNA polymerase from progressing past that region to initiate transcription. Regulation involving a repressor is called **negative control.**

Specific molecules may bind to the repressor and, by doing so, alter the ability of the repressor to bind to DNA. This can occur because repressors are allosteric proteins, having a distinct site to which another molecule can bind. When that molecule binds, the shape of the repressor is altered. In turn, this affects the ability of the repressor to bind to DNA. As shown in **figure 7.18,** there are two general mechanisms by which different repressors can function:

1. The repressor is synthesized as a form that alone cannot bind to the operator. When a molecule, termed a **corepressor** binds to it, however, its shape is altered so that it can bind to the operator, blocking transcription initiation.
2. The repressor is synthesized as a form that effectively binds to the operator, blocking transcription initiation. The binding to it of a molecule that functions as an **inducer,** alters the shape of the repressor so that it no longer binds to the operator. Consequently, the gene can be transcribed.

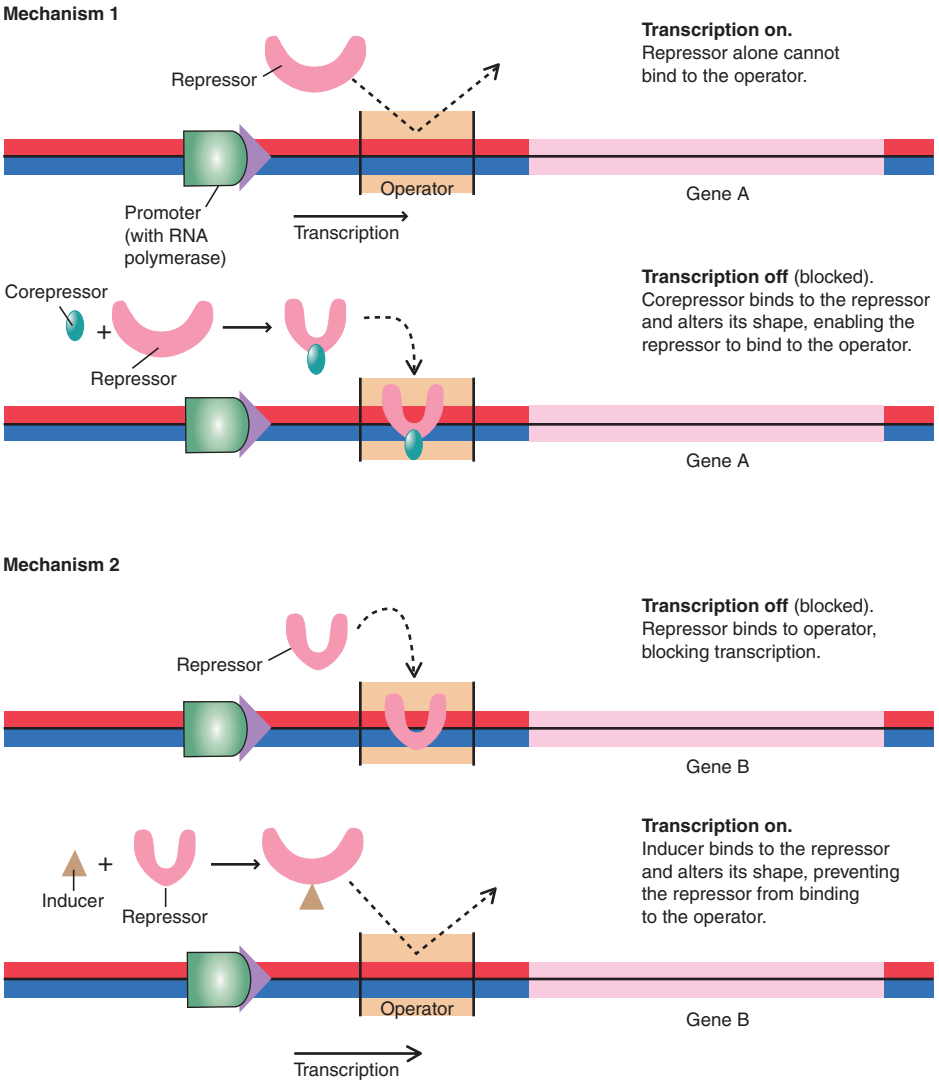


Figure 7.18 **Transcriptional Regulation by Repressors**

Activators

An **activator** is a regulatory protein that facilitates transcription. Genes that are controlled by an activator have an ineffective promoter that is preceded by an **activator-binding site.** The binding of the activator to the DNA enhances the ability of RNA polymerase to initiate transcription at that promoter. Regulation involving an activator is sometimes called **positive control.**

Like repressors, activators are allosteric proteins whose function can be modulated by the binding of other molecules. A molecule that binds to an activator and alters its shape so it can effectively bind to the activator-binding site functions as an inducer **(figure 7.19).** Thus, the term inducer applies to a molecule that turns on transcription, either by stimulating the function of an activator or interfering with the function of a repressor.

The *lac* Operon As a Model for Control of Metabolic Pathways

Originally elucidated in the early 1960s by Francois Jacob and Jacques Monod, the ***lac* operon** has served as an important model for understanding the control of gene expression in bacteria. The operon, which consists of three genes involved with lactose degradation, along with regulatory components, is subject to dual control by both a repressor and an activator **(figure 7.20).** The net effect is that the genes are expressed only when lactose is present but glucose is absent.

The Effect of Lactose on the Control of the Lactose Operon

The *lac* operon employs a repressor that prevents transcription of the genes when lactose is unavailable. When lactose is not present, the repressor binds to the operator, effectively blocking transcription. When lactose is present in the cell, however, some of the molecules are converted into a compound called allolactose. This compound binds to the repressor, altering its shape so that it can no longer bind to the operator. Thus, when lactose is present, the repressor no longer prevents RNA polymerase from transcribing the operon. Note, however, that the activator described in the next section is needed for successful transcription.

The Effect of Glucose on the Control of the Lactose Operon

Escherichia coli preferentially uses glucose over other sugars such as lactose. This can readily be demonstrated by observing growth and sugar utilization of *E. coli* in a medium containing glucose and lactose. Cells actively grow, metabolizing only glucose until its supply is exhausted **(figure 7.21).** Growth then ceases for a short period until the cells begin utilizing lactose. At this point, the cells start multiplying. This two-

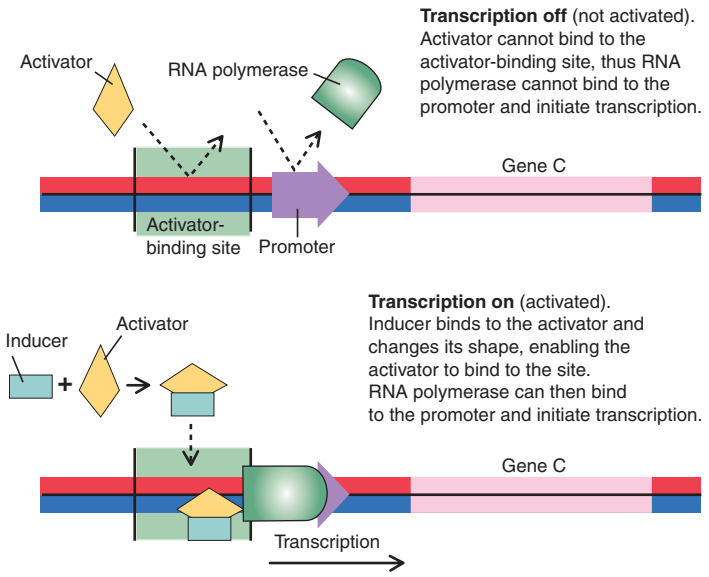


Figure 7.19 **Transcriptional Regulation by Activators**

step growth response, called **diauxic growth,** represents the ability of glucose to repress the enzymes of lactose degradation—a phenomenon called **catabolite repression.**

The regulatory mechanism of catabolite repression does not directly sense glucose in a cell. Instead, it recognizes the concentration of a nucleotide derivative, called cyclic AMP (cAMP), which is low when glucose is present and high when it is absent. cAMP is an inducer of the operon; it binds to an activator that facilitates transcription of the *lac* operon. This activator, called CAP (catabolite activator protein), is only able to bind to the *lac* promoter when cAMP is bound to it. The higher the concentration of cAMP, the more likely it is to bind to CAP. Thus, when glucose concentrations are low (and therefore cAMP levels are high), the *lac* operon can be transcribed. Note, however, that even in the presence of a functional activator, the repressor prevents transcription unless lactose is present. Catabolite repression is significant biologically because it forces the cells to first use the carbon source that is most easily metabolized. Only when the

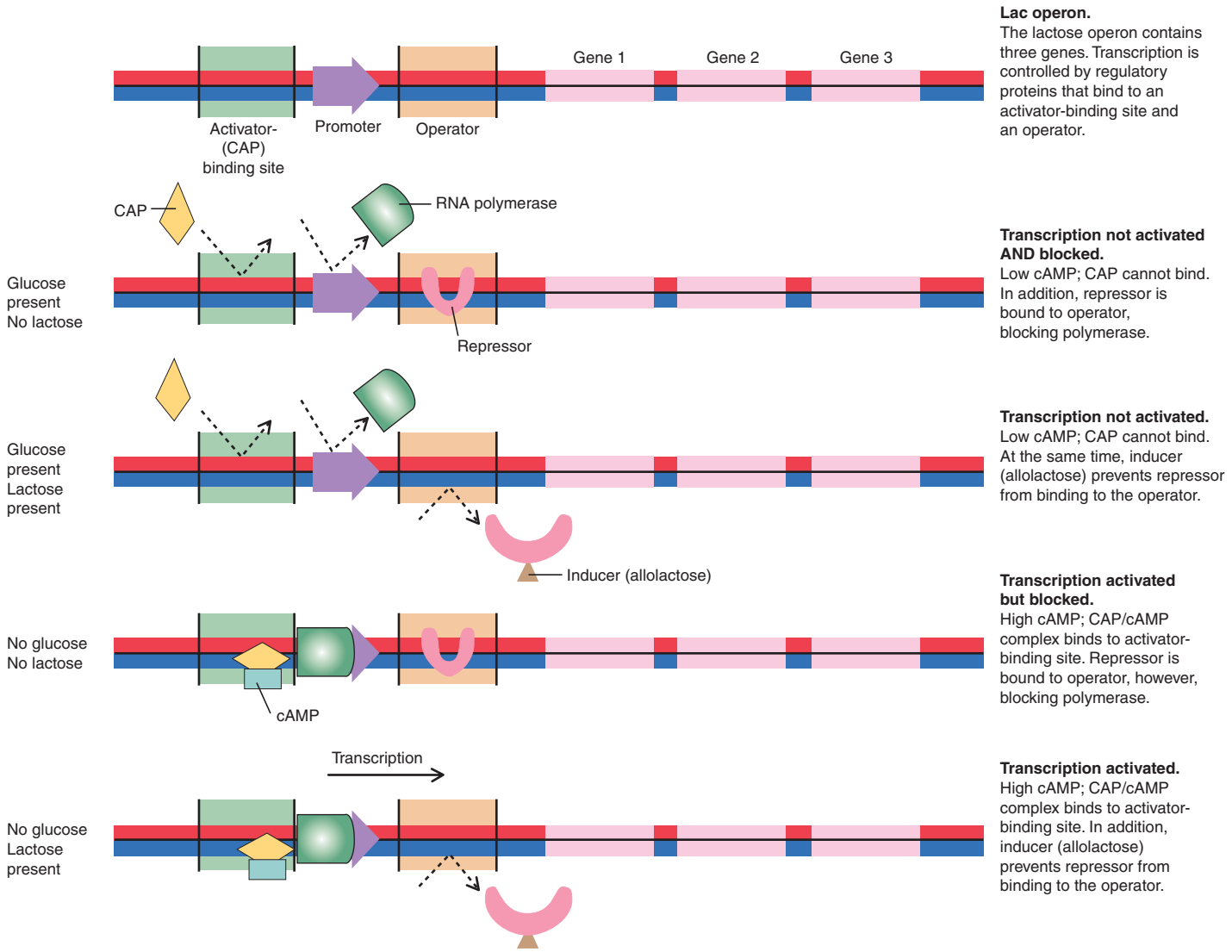


Figure 7.20 **Regulation of the *lac* Operon**

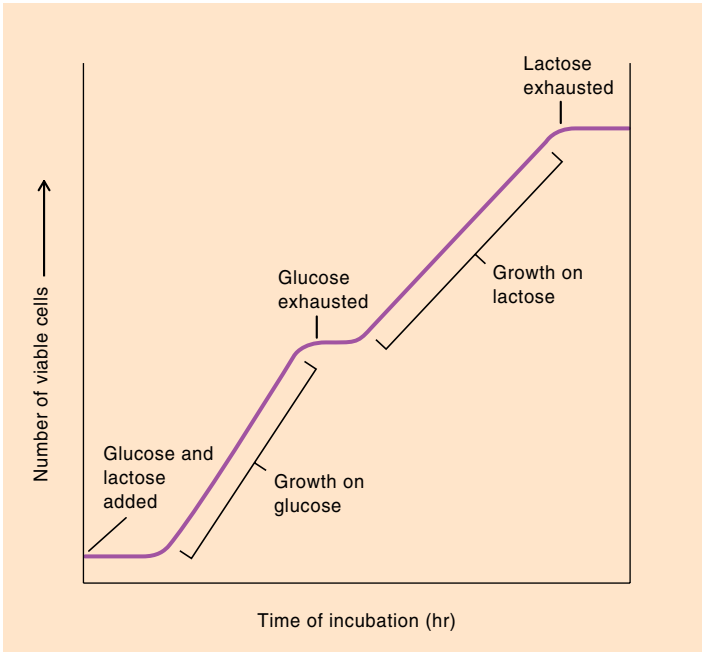


Figure 7.21 Diauxic Growth Curve of *E. coli* Growing in a Medium Containing Glucose and Lactose Cells preferentially use glucose. Only when the supply of glucose is exhausted do cells start metabolizing lactose. Note that the growth on lactose is slower than it is on glucose.

supply of glucose is exhausted do the cells begin degrading lactose, a carbon source that requires additional enzymatic steps to metabolize.

MICROCHECK 7.6

Enzymes may be constitutive, inducible, or repressible. A repressor blocks transcription when it binds to an operator. An activator enhances transcription when it binds to an activator-binding site. The functioning of specific activators and repressors may require or be blocked by other molecules.

- Explain the difference between a constitutive enzyme and an inducible enzyme.
- Explain how an inducer could affect either a repressor or an activator.
- Why would it be advantageous for a cell to control the activity of an enzyme as well as its synthesis?

7.7 Sensing and Responding to Environmental Fluctuations

Microorganisms adapt to fluctuating conditions by altering the level of expression of certain genes. For example, certain pathogenic bacteria have mechanisms to sense when they are within the tissues of an animal; in response they can activate certain genes that facilitate their survival against the impending onslaught of host defenses.

Signal Transduction

Signal transduction is a process that transmits information from outside a cell to the inside, allowing that cell to respond to changing environmental conditions. For example, cells turn on or off certain genes in response to variations in such factors as osmotic pressure, cell concentration, and nitrogen availability. Bacteria that inhabit diverse environments tend to have large numbers of regulatory genes. Approximately 10% of the genome may consist of these genes.

Two-Component Regulatory Systems

One mechanism that cells use to relay information about the external environment to the relevant genes is a **two-component regulatory system.** This system relies on the coordinated activities of two different proteins, a sensor and a response regulator. The **sensor** spans the cytoplasmic membrane so the part that recognizes changes in the environment is outside the cell. In response to specific changes in the external environment, the sensor chemically modifies a region on its internal portion, usually by phosphorylating a specific amino acid. The phosphoryl group is transferred to a **response regulator.** The modified response regulator can then act as either an activator or a repressor, turning on or off genes, depending on the system.

Bacteria use different two-component regulatory systems to detect and respond to a wide variety of environmental cues. *E. coli,* for example, uses such systems to control the expression of genes for its alternative types of metabolism. When nitrate is present in anaerobic conditions, cells activate genes required to use nitrate as a terminal electron acceptor. Some pathogens use two-component regulatory systems to sense environmental magnesium concentrations, and then activate specific genes in response. Because the magnesium concentration within certain tissue cells is generally lower than that of extracellular sites, these pathogens are able to recognize whether or not they are within a cell. In turn, they can activate appropriate genes that help them evade the host defenses intended to protect that relative site.

Quorum Sensing

Some organisms can "sense" the density of cells within their own population—a phenomenon called **quorum sensing.** This enables them to activate genes that are only beneficial when expressed by a critical mass of cells.

The most studied example of quorum sensing is bacterial bioluminescence. The marine bacterium *Vibrio fischeri* can emit light, but only when growing in a critical mass. Light production by this organism is an intensely energy-consuming process and is presumably only beneficial when enough cells participate.

Bacteria that utilize quorum sensing synthesize one or more varieties of a **homoserine lactone (HSL)** (or AHL for **acylated homoserine lactone**). These small molecules can move freely in and out of a cell. When few cells are present, the concentration of a given HSL is very low. As the cells multiply in a confined area, however, the concentration of that HSL increases proportionally. Only when it reaches a critical level does it induce the expression of specific genes.

Natural Selection

Natural selection can also play a role in the control of gene expression. The expression of some genes changes randomly, presumably enhancing the chances of survival of at least a part of a population under certain environmental conditions. This is most readily apparent in bacteria that undergo **antigenic variation,** an alteration in the characteristics of certain surface proteins such as flagella, pili, and outer membrane proteins. Disease-causing organisms that are able to change these proteins can stay one step ahead of the body's defenses by altering the very molecules our immune systems must learn to recognize. One of the best characterized examples of bacteria that undergo antigenic variation is *Neisseria gonorrhoeae*. This organism successfully disguises itself from the immune system because it can change several of its surface proteins. One of these is **pilin,** the protein subunit that makes up pili. *N. gonorrhoeae* appears to have many different genes for this protein, yet most are not expressed. The only one that is expressed resides in a particular chromosomal location called an **expression locus.** *N. gonorrhoeae* has a mechanism to shuffle the pilin genes, randomly moving different ones in and out of the expression locus. In a population of 10^4 cells, at least one cell is expressing a different type of pilin. It appears that expression of different pilin genes is not regulated in any controlled manner but occurs randomly. Only some of the changes, however, are advantageous to a cell's survival. When the body's immune system eventually begins to respond to a specific pilin type, those cells that have already "switched" to produce a different type will survive and then multiply. Eventually, the immune system learns to recognize those, but by that time, another subpopulation will have "switched" its pilin type. Thus, natural selection serves to indirectly regulate the changes. ■ **pili, p. 65** ■ ***Neisseria gonorrhoeae*, p. 644**

Another mechanism of randomly altering gene expression is **phase variation,** the routine switching on and off of certain genes. Presumably, phase variation helps an organism adapt to selective pressures. By altering the expression of certain critical genes, at least a part of the population is poised for change and thus able to survive and multiply. For example, phase variation of genes that encode fimbriae may allow some members of a population to attach to a surface, while permitting others to detach and colonize surfaces elsewhere. ■ **fimbriae, p. 65**

MICROCHECK 7.7

End products of biosynthetic pathways typically repress expression of the associated enzymes, whereas substrates of degradative pathways typically induce the expression of the associated enzymes. Signal transduction allows a cell to respond to changing conditions outside of that cell. The expression of some genes changes randomly, presumably enhancing the chances of survival of at least a subset of a population of cells under varying environmental conditions.

- Explain the mechanism by which glucose represses the lactose operon.
- Explain the mechanism by which certain bacteria can "sense" the density of cells.
- Why would it be advantageous for a bacterium to be able to synthesize more than one type of homoserine lactone?

FUTURE CHALLENGES

Gems in the Genomes?

From a medical standpoint, one of the most exciting challenges will be to capitalize on the rapidly accruing genomic information and use that knowledge to develop new drugs and therapies. The potential gains are tremendous, particularly in the face of increasing resistance to current antimicrobial drugs. For example, by studying the genomes of pathogenic microorganisms, scientists can learn more about specific genes that enable an organism to cause disease. Already we know that many of these are encoded in large segments called ***pathogenicity islands*** *and that the genes in these segments are often coordinately regulated. By learning more about the signals and mechanisms that turn these genes on and off, scientists may be able to one day design a drug that prevents the synthesis of critical bacterial proteins. Such a drug could interfere with that pathogen's ability to survive within our body and thereby render it harmless.* ■ **resistance to antimicrobial drugs, p. 511**

Learning more about the human genome provides another means of developing drug therapies. Already, companies are searching genomic databases, a process called ***genome mining,*** *to locate ORFs that may encode proteins of medical value. What they generally look for are previously uncharacterized proteins that have certain sequence similarities to proteins of proven therapeutic value. Some of their discoveries are now in clinical trials to test their efficacy. For example, a protein involved in bone-building is being tested as a treatment for osteoporosis. Likewise, another protein discovered through genome mining may facilitate the healing of wounds. Genes encoding many other medically useful proteins are probably still hidden, waiting to be discovered.*

SUMMARY

7.1 Overview (Figure 7.1)

Characteristics of DNA (Figure 7.3)

1. A single strand of DNA has a 5′ end and a 3′ end; the two strands of DNA in the double helix are **antiparallel;** they are oriented in opposite directions.

Characteristics of RNA

1. A single-stranded RNA fragment is transcribed from one of the two strands of DNA.
2. There are three different functional groups of RNA molecules: **messenger RNA (mRNA), ribosomal RNA (rRNA),** and **transfer RNA (tRNA).**

Regulating the Expression of Genes

1. Protein synthesis is generally controlled by regulating the synthesis of mRNA molecules; mRNA is short-lived because **RNases** degrade it within minutes.

7.2 DNA Replication

1. DNA replication is generally **bidirectional** and **semiconservative.** (Figure 7.4)
2. The DNA chain always elongates in the 5′ to 3′ direction; the base-pairing rules determine the specific nucleotides that are added. (Figure 7.5)

Initiation of DNA Replication

1. DNA replication begins at the **origin of replication. DNA polymerase** synthesizes DNA in the 5′ to 3′ direction, using one strand as a **template** to generate the complementary strand.

The Replication Fork (Figure 7.6)

1. The bidirectional progression of replication around a circular DNA molecule creates two replication forks; numerous enzymes and other proteins are involved.

7.3 Gene Expression

Transcription

1. The enzyme **RNA polymerase** catalyzes the process of transcription, producing a single-stranded RNA molecule that is complementary and antiparallel to the DNA template. (Figure 7.7)
2. **Transcription** begins when RNA polymerase recognizes and binds to a **promoter.** (Figure 7.8)
3. RNA is synthesized in the 5′ to 3′ direction. (Figure 7.9)
4. When **RNA polymerase** encounters a **terminator,** it falls off the DNA template and releases the newly synthesized RNA.

Translation

1. The information encoded by mRNA is deciphered using the genetic code. (Figure 7.10)
2. A nucleotide sequence has three potential **reading frames.** (Figure 7.11)
3. **Ribosomes** function as the site of translation. (Figure 7.12)
4. **tRNAs** carry specific amino acids and act as keys that interpret the genetic code. (Figure 7.13)
5. In prokaryotes, initiation of translation begins when the ribosome binds to the **ribosome-binding site** of the mRNA molecule. Translation starts at the first AUG downstream of that site. (Figure 7.14)
6. The ribosome moves along mRNA in the 5′ to 3′ direction; translation terminates when the ribosome reaches a **stop codon.** (Figure 7.15)
7. Proteins are often modified after they are synthesized; those that contain a signal sequence are transported to the outside of the cell.

7.4 Differences Between Eukaryotic and Prokaryotic Gene Expression (Table 7.4)

1. Eukaryotic mRNA is **processed;** a **cap** and a **poly A tail** are added.
2. Eukaryotic genes often contain **introns** which are removed from **precursor mRNA** by a process called **splicing.** (Figure 7.16)
3. In eukaryotic cells, the mRNA must be transported out of the nucleus before it can be translated in the cytoplasm.

7.5 Genomics

Analyzing a Prokaryotic DNA Sequence (Figure 7.17)

1. When analyzing a DNA sequence, the nucleotide sequence of the (+) strand is used to infer information carried by the corresponding RNA transcript; computers are used to search for **open reading frames (ORFs).**

7.6 Regulating Gene Expression

Principles of Regulation

1. **Constitutive enzymes** are constantly synthesized.
2. The synthesis of **inducible enzymes** can be turned on by certain conditions.
3. The synthesis of **repressible enzymes** can be turned off by certain conditions.

Mechanisms to Control Transcription

1. A **repressor** is a regulatory protein that blocks transcription. (Figure 7.18)
2. An **activator** is a regulatory protein that enhances transcription. (Figure 7.19)

The *lac* Operon As a Model for Control of Metabolic Pathways (Figure 7.20)

1. The *lac* operon employs a repressor that prevents transcription of the genes when lactose is not available; a derivative of lactose functions as an inducer.
2. Catabolite repression prevents transcription of the *lac* operon when glucose is available.

7.7 Sensing and Responding to Environmental Fluctuations

Signal Transduction

1. **Two-component regulatory systems** utilize a **sensor** that recognizes changes outside the cell and then transmits that information to a **response regulator.**
2. Bacteria that utilize **quorum sensing** synthesize a soluble compound, a **homoserine lactone,** which can move freely in and out of a cell. Only when that compound reaches a critical concentration does it activate specific genes.

Natural Selection

1. The expression of some genes changes randomly, presumably enhancing the chances of survival of at least a subset of a population under varying environmental conditions.
2. **Antigenic variation** is a routine change in the expression of surface proteins such as flagella, pili, and outer membrane proteins.
3. **Phase variation** is the routine switching on and off of certain genes.

REVIEW QUESTIONS

Short Answer

1. Explain what the term *semiconservative* means with respect to DNA replication.
2. How can *E. coli* have a generation time of only 20 minutes when it takes 40 minutes to replicate its chromosome?
3. What is the function of primase in DNA replication? Why is this enzyme necessary?
4. What is polycistronic mRNA?
5. Explain why knowing the orientation of a promoter is critical when determining the amino acid sequence of an encoded protein.
6. What is characteristic about the nucleotide sequence of a transcription terminator?
7. What happens to a polypeptide that has a signal sequence?
8. Compare and contrast regulation by a repressor and an activator.
9. Explain how bacteria sense the density of cells in their own population.
10. Explain why it is sometimes difficult to locate genomic regions that encode a protein.

Multiple Choice Questions

1. All of the following are involved in transcription, *except*
   A. polymerase. **B.** primer. **C.** promoter.
   D. sigma factor. **E.** uracil.
2. All of the following are involved in DNA replication, *except*
   A. elongation factors. **B.** gyrase. **C.** polymerase.
   D. primase. **E.** primer.
3. All of the following are directly involved in translation, *except*
   A. promoter. **B.** ribosome. **C.** start codon.
   D. stop codon. **E.** tRNA.
4. Using the DNA strand depicted here as a template, what will be the sequence of the RNA transcript?

   5′ GCGTTAACGTAGGC 3′
   → promoter
   3′ CGCAATTGCATCCG 5′

   A. 5′ GCGUUAACGUAGGC 3′
   B. 5′ CGGAUGCAAUUGCG 3′
   C. 5′ CGCAAUUGCAUCCG 3′
   D. 5′ GCCUACGUUAACGC 3′

5. A ribosome binds to the following mRNA at the site indicated by the dark box. What are the first three amino acids that will be incorporated into the resulting polypeptide?

   5′ [■] GCCGGAAUGCUGCUGGC

A.	alanine	aspartic acid	alanine
B.	methionine	threonine	cysteine
C.	methionine	leucine	leucine
D.	alanine	glycine	methionine

6. Allolactose induces the *lac* regulon by binding to a(n)…
   A. operator. **B.** repressor.
   C. activator. **D.** CAP protein.
7. Under which of the following conditions will transcription of the *lac* operon occur?
   A. Lactose present/glucose present
   B. Lactose present/glucose absent
   C. Lactose absent/glucose present
   D. Lactose absent/glucose absent
   E. A and B
8. Which of the following statements about gene expression is *false*?
   A. More than one RNA polymerase can be transcribing a specific gene at a given time.
   B. More than one ribosome can be translating a specific transcript at a given time.
   C. Translation begins at a site called a promoter.
   D. Transcription stops at a site called a terminator.
   E. Some amino acids are coded for by more than one codon.
9. Which of the following is *not* characteristic of eukaryotic gene expression?
   A. 5′ cap is added to the mRNA.
   B. A poly A tail is added to the 3′ end of mRNA.
   C. Introns must be removed to create the mRNA that is translated.
   D. The mRNA is often polycistronic.
   E. Translation begins at the first AUG.
10. Which of the following statements is *false*?
    A. A derivative of lactose serves as an inducer of the *lac* operon.
    B. Signal transduction provides a mechanism for a cell to sense the conditions of its external environment.
    C. The function of homoserine lactone is to enable a cell to sense the density of like cells.
    D. An example of a two-component regulatory system is the lactose operon, which is controlled by a repressor and an activator.
    E. An ORF is a stretch of DNA that may encode a protein.

Applications

1. A graduate student is trying to isolate the gene coding for an enzyme found in a species of *Pseudomonas* that degrades trinitrotoluene (TNT). The student is frustrated to find that the organism does not produce the enzyme when grown in nutrient broth, making it is difficult to collect the mRNA needed to help identify the gene. What could the student do to potentially increase the amount of the desired enzyme?
2. A student wants to remove the introns from a segment of DNA coding for protein X. Devise a strategy for how this could be accomplished.

Critical Thinking

1. The study of protein synthesis often uses a cell-free system where cells are ground with an abrasive to release the cell contents and then filtered to remove the abrasive. These materials are added to the system, generating the indicated results:

Materials Added	Results
Radioactive amino acids	Radioactive protein produced
Radioactive amino acids *and* RNase (an RNA-digesting enzyme)	No radioactive protein produced

   What is the best interpretation of these observations?

2. In a variation of the experiment in the previous question, the following materials were added to three separate cell-free systems, generating the indicated results:

Materials Added	Results
Radioactive amino acids	Radioactive protein produced
Radioactive amino acids *and* DNase (a DNA-digesting enzyme)	Radioactive protein produced
Several hours after grinding: Radioactive amino acids *and* DNase	No radioactive protein produced

   What is the best interpretation of these observations?