

15 The Genetic Code and Translation



The Hutterites are a religious branch of Anabaptists who live on communal farms in the prairie states and provinces of North America. A small number of founders, coupled with a tendency to intermarry, has resulted in a high frequency of the mutation for Bowen–Conradi syndrome among Hutterites. Bowen–Conradi syndrome results from defective ribosome biosynthesis, affecting the process of translation. [Kevin Fleming/Corbis.]

HUTTERITES, RIBOSOMES, AND BOWEN–CONRADI SYNDROME

The essential nature of the ribosome—the cell’s protein factory—is poignantly illustrated by children with Bowen–Conradi syndrome. Born with a prominent nose, small head, and an unusual curvature of the small finger, these children fail to thrive and gain weight, usually dying within the first year of life.

Almost all children with Bowen–Conradi syndrome are Hutterites, a branch of Anabaptists who originated in the 1500s in the Tyrolean Alps of Austria. After years of persecution, the Hutterites immigrated to South Dakota in the 1870s and subsequently spread to neighboring prairie states and Canadian provinces. Today, the Hutterites in North America number about 40,000 persons. They live on communal farms, are strict pacifists, and rarely marry outside of the Hutterite community.

Bowen–Conradi syndrome is inherited as an autosomal recessive disorder, and the association of Bowen–Conradi syndrome with the Hutterite community is a function of the group’s unique genetic history. The gene pool of present-day Hutterites in North America can be traced to fewer than 100 persons who immigrated to South Dakota in the late 1800s. The increased incidence of Bowen–Conradi syndrome in

Hutterites today is due to the founder effect—the presence of the gene in one or more of the original founders—and its spread as Hutterites intermarried within their community. Because of the founder effect and inbreeding, many Hutterites today are as closely related as first cousins. This close genetic relationship among the Hutterites increases the probability that a child will inherit two copies of the recessive gene and have Bowen–Conradi syndrome; indeed, almost 1 in 10 Hutterites is a heterozygous carrier of the gene that causes the disease.

Although Bowen–Conradi syndrome was described in 1976, the genetic and biochemical basis of the disease long remained a mystery. Following a 7-year quest to find the causative gene, researchers at the University of Manitoba determined in 2009 that Bowen–Conradi syndrome results from the mutation of a single base pair in the *EMG1* gene, located on chromosome 12.

The discovery of the gene for Bowen–Conradi syndrome was a source of immediate insight into the biochemical nature of the disease. Although little is known about the

function of the *EMG1* gene in humans, earlier studies in yeast revealed that it encodes a protein that helps to assemble the ribosome. As discussed in Chapter 14, the ribosome comprises small and large subunits. The small subunit in humans consists of a single piece of ribosomal RNA, known as 18S rRNA, and a large number of proteins. The protein encoded by the *EMG1* gene plays an essential role in processing 18S rRNA and helps to assemble it into the small subunit of the ribosome. Because of a mutation in the *EMG1* gene, babies with Bowen–Conradi syndrome produce ribosomes that function poorly and the process of protein synthesis is universally affected.

Bowen–Conradi syndrome illustrates the extreme importance of translation, the process of protein synthesis, which is the focus of this chapter. We begin by examining the molecular relation between genotype and phenotype. Next, we study the genetic code—the instructions that specify the amino acid sequence of a protein—and then examine the mechanism of protein synthesis. Our primary focus will be on protein synthesis in bacterial cells, but we will examine some of the differences in eukaryotic cells. At the end of the chapter, we look at some additional aspects of protein synthesis.

15.1 Many Genes Encode Proteins

The first person to suggest the existence of a relation between genotype and proteins was English physician Archibald Garrod. In 1908, Garrod correctly proposed that genes encode enzymes, but, unfortunately, his theory made little impression on his contemporaries. Not until the 1940s, when George Beadle and Edward Tatum examined the genetic basis of biochemical pathways in *Neurospora*, did the relation between genes and proteins become widely accepted.

The One Gene, One Enzyme Hypothesis

Beadle and Tatum used the bread mold *Neurospora* to study the biochemical results of mutations. *Neurospora* is easy to cultivate in the laboratory. The main vegetative part of the fungus is haploid, which allows the effects of recessive mutations to be easily observed (**Figure 15.1**).

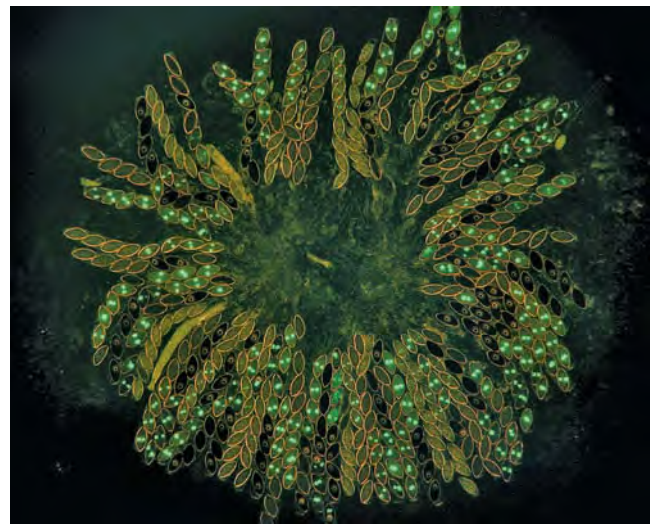
Wild-type *Neurospora* grows on minimal medium, which contains only inorganic salts, nitrogen, a carbon source such as sucrose, and the vitamin biotin. The fungus can synthesize all the biological molecules that it needs from these basic compounds. However, mutations may arise that disrupt fungal growth by destroying the fungus's ability to synthesize one or more essential biological molecules. These nutritionally deficient mutants, termed auxotrophs, will not grow on minimal medium, but they can grow on medium that contains the substance that they are no longer able to synthesize.

Beadle and Tatum first irradiated spores of *Neurospora* to induce mutations (**Figure 15.2**). After irradiation, they placed individual spores into different culture tubes containing complete medium (medium having all the biological substances needed for growth). Next, they transferred spores from each

culture to tubes containing minimal medium. Fungi containing auxotrophic mutations grew on complete medium but would not grow on minimal medium, which allowed Beadle and Tatum to identify cultures that contained mutations.

After they had determined that a particular culture had an auxotrophic mutation, Beadle and Tatum set out to determine the specific *effect* of the mutation. They transferred spores of each mutant strain from complete medium to a series of tubes (see **Figure 15.2**), each of which possessed minimal medium plus one of a variety of essential biological molecules, such as an amino acid. If the spores in a tube grew, Beadle and Tatum were able to identify the added substance as the biological molecule whose synthesis had been affected by the mutation. For example, an auxotrophic mutant that would grow only on minimal medium to which arginine had been added must have possessed a mutation that disrupts the synthesis of arginine.

Adrian Srb and Norman H. Horowitz patiently applied this procedure to genetically dissect the multistep biochemical pathway of arginine synthesis (**Figure 15.3**). They first isolated a series of auxotrophic mutants whose growth required arginine. They then tested these mutants for their ability to grow on minimal medium supplemented with

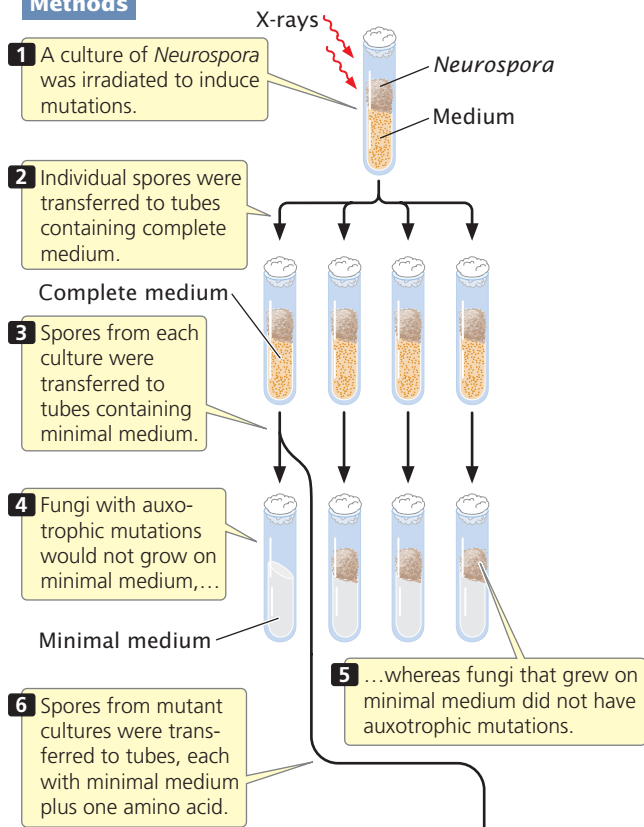


15.1 Beadle and Tatum used the fungus *Neurospora*, which has a complex life cycle, to work out the relation of genes to proteins. [Namboori B. Raju.]

Experiment

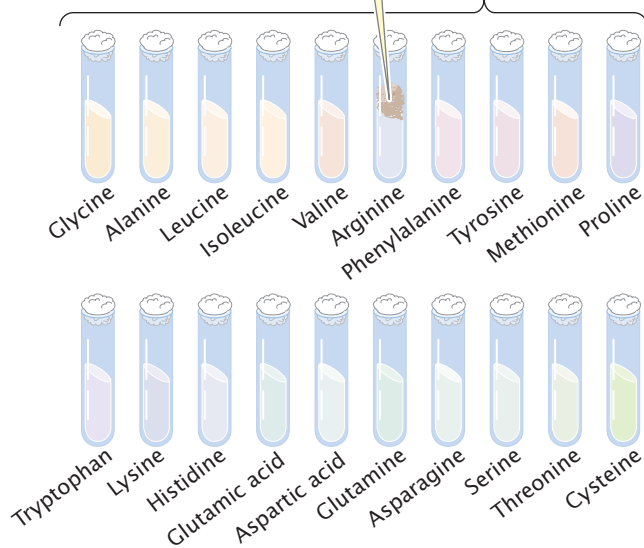
Question: How can auxotrophic mutations be isolated and identified?

Methods



Results

7 The mutant *Neurospora* grew only when supplemented with arginine, indicating that the mutant was defective in the synthesis of arginine.

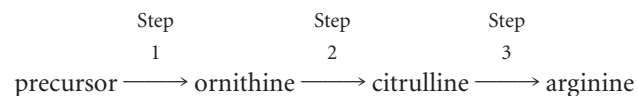


Conclusion: The mutation could be identified by the spores' ability to grow in minimal medium supplemented by the substance that the spores could not synthesize.

15.2 Beadle and Tatum developed a method for isolating auxotrophic mutants in *Neurospora*.

three compounds: ornithine, citrulline, and arginine. From the results, they were able to place the mutants into three groups (Table 15.1) on the basis of which of the substances allowed growth. Group I mutants grew on minimal medium supplemented with ornithine, citrulline, or arginine. Group II mutants grew on minimal medium supplemented with either arginine or citrulline but did not grow on medium supplemented only with ornithine. Finally, group III mutants grew only on medium supplemented with arginine.

Srb and Horowitz therefore proposed that the biochemical pathway leading to the amino acid arginine has at least three steps:



They concluded that the mutations in group I affect step 1 of this pathway, mutations in group II affect step 2, and mutations in group III affect step 3. But how did they know that the order of the compounds in the biochemical pathway was correct?

Notice that, if step 1 is blocked by a mutation, then the addition of either ornithine or citrulline allows growth, because these compounds can still be converted into arginine (see Figure 15.3). Similarly, if step 2 is blocked, the addition of citrulline allows growth, but the addition of ornithine has no effect. If step 3 is blocked, the spores will grow only if arginine is added to the medium. The underlying principle is that an auxotrophic mutant cannot synthesize any compound that comes after the step blocked by a mutation.

Using this reasoning with the information in Table 15.1, we can see that the addition of arginine to the medium allows all three groups of mutants to grow. Therefore, biochemical steps affected by all the mutants precede the step that results

Table 15.1 Growth of arginine auxotrophic mutants on minimal medium with various supplements

Mutant Strain Number	Ornithine	Citrulline	Arginine
Group I	+	+	+
Group II	-	+	+
Group III	-	-	+

Note: A plus sign (+) indicates growth; a minus sign (-) indicates no growth.

Experiment

Question: What do the effects of genetic mutation on a biochemical pathway tell us about the gene–protein relation?

Methods

Spores of auxotrophic mutants whose growth requires arginine are placed on minimal medium and on minimal medium containing a supplement.

Results

Group I mutants can grow on minimal medium supplemented with ornithine, citrulline, or arginine. The mutation blocks a step prior to the synthesis of ornithine, citrulline, and arginine.

Group II mutants grow on medium supplemented with either arginine or citrulline but not ornithine. The mutation blocks a step prior to the synthesis of citrulline and arginine.

Group III mutants grow only on medium supplemented with arginine. The mutation blocks a step prior to the synthesis of arginine.

Supplements to minimal medium

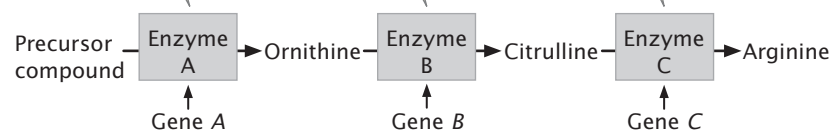
		None	Ornithine	Citrulline	Arginine
Group					
Wild type					
I					
II					
III					

Group I is blocked at this step.

Group II is blocked at this step.

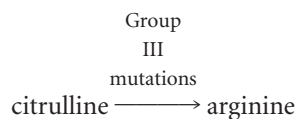
Group III is blocked at this step.

15.3 Method used to determine the relation between genes and enzymes in *Neurospora*. This biochemical pathway leads to the synthesis of arginine in *Neurospora*. Steps in the pathway are catalyzed by enzymes affected by mutations.

Interpretation of data

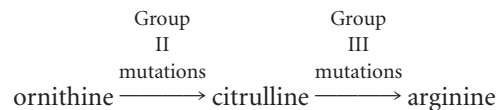
Conclusion: Each gene encodes a separate protein—in this case, an enzyme.

in arginine. The addition of citrulline allows group I and group II mutants to grow but not group III mutants; therefore, group III mutations must affect a biochemical step that takes place after the production of citrulline but before the production of arginine.

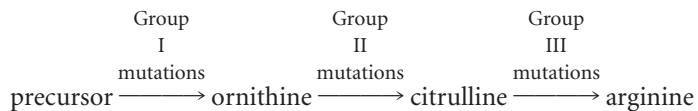


The addition of ornithine allows the growth of group I mutants but not group II or group III mutants; thus, mutations in groups II and III affect steps that come after the

production of ornithine. We've already established that group II mutations affect a step before the production of citrulline; so group II mutations must block the conversion of ornithine into citrulline.



Because group I mutations affect some step before the production of ornithine, we can conclude that they must affect the conversion of some precursor into ornithine. We can now outline the biochemical pathway yielding ornithine, citrulline, and arginine.



Importantly, this procedure does not necessarily detect all steps in a pathway; rather, it detects only the steps producing the compounds tested.

Using mutations and this type of reasoning, Beadle, Tatum, and others were able to identify genes that control several biosynthetic pathways in *Neurospora*. They established that each step in a pathway is controlled by a different enzyme, as shown in Figure 15.3 for the arginine pathway. The results of genetic crosses and mapping studies demonstrated that mutations affecting any one step in a pathway always map to the same chromosomal location. Beadle and Tatum reasoned that mutations affecting a particular biochemical step occurred at a single locus that encoded a particular enzyme. This idea became known as the **one gene, one enzyme hypothesis**: genes function by encoding enzymes, and each gene encodes a separate enzyme. When research findings showed that some proteins are composed of more than one polypeptide chain and that different polypeptide chains are encoded by separate genes, this model was modified to become the **one gene, one polypeptide hypothesis**. [TRY PROBLEM 16 →](#)

CONCEPTS

Beadle and Tatum's studies of biochemical pathways in the fungus *Neurospora* helped define the relation between genotype and phenotype by establishing the one gene, one enzyme hypothesis, the idea that each gene encodes a separate enzyme. This hypothesis was later modified to the one gene, one polypeptide hypothesis.

✓ CONCEPT CHECK 1

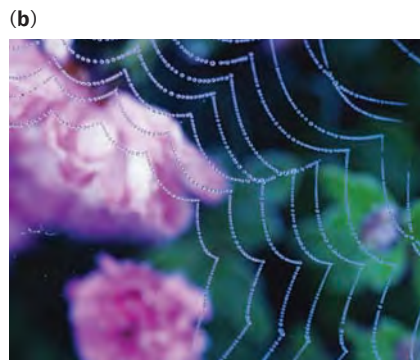
Auxotrophic mutation 103 grows on minimal medium supplemented with A, B, or C; mutation 106 grows on medium supplemented with A and C but not B; and mutation 102 grows only on medium supplemented with C. What is the order of A, B, and C in a biochemical pathway?

The Structure and Function of Proteins

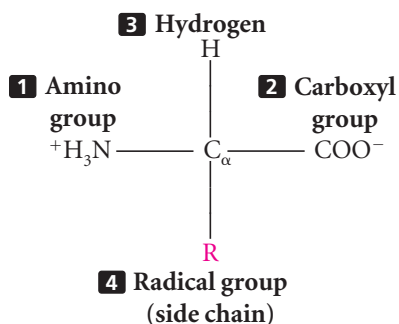
Proteins are central to all living processes (**Figure 15.4**). Many proteins are enzymes, the biological catalysts that drive the chemical reactions of the cell; others are structural components, providing scaffolding and support for membranes, filaments, bone, and hair. Some proteins help transport substances; others have a regulatory, communication, or defense function.

Amino acids All proteins are composed of **amino acids**, linked end to end. Twenty common amino acids are found in proteins; these amino acids are shown in **Figure 15.5** with both their three-letter and their one-letter abbreviations. (Other amino acids sometimes found in proteins are modified forms of the common amino acids.) The 20 common amino acids are similar in structure: each consists of a central carbon atom bonded to an amino group, a hydrogen atom, a carboxyl group, and an R (radical) group that differs for each amino acid. The amino acids in proteins are joined together by **peptide bonds** (**Figure 15.6**) to form **polypeptide** chains, and a protein consists of one or more polypeptide chains. Like nucleic acids, polypeptides have polarity, with one end having a free amino group (NH_3^+) and the other end possessing a free carboxyl group (COO^-). Some proteins consist of only a few amino acids, whereas others may have thousands.

Protein structure Like that of nucleic acids, the molecular structure of proteins has several levels of organization. The *primary structure* of a protein is its sequence of amino acids (**Figure 15.7a**). Through interactions between neighboring amino acids, a polypeptide chain folds and twists into a *secondary structure* (**Figure 15.7b**); two common secondary structures found in proteins are the beta (β) pleated sheet and the alpha (α) helix. Secondary structures interact and fold further to form a *tertiary structure* (**Figure 15.7c**), which is the overall, three-dimensional shape of the protein. The secondary and tertiary structures of a protein

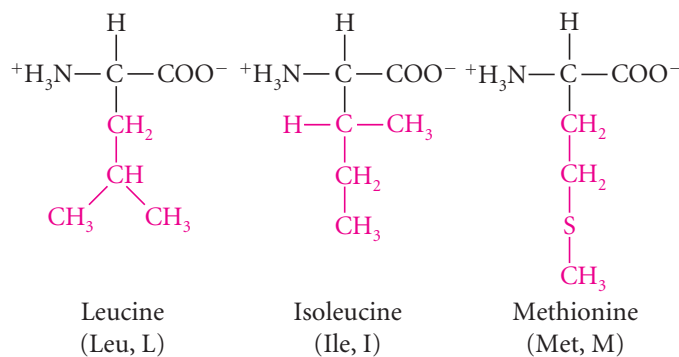
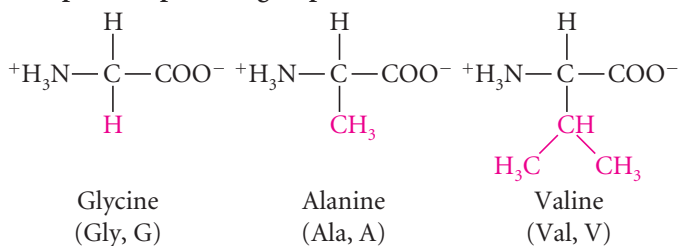


15.4 Proteins serve a number of biological functions. (a) The light produced by fireflies is the result of a light-producing reaction between luciferin and ATP catalyzed by the enzyme luciferase. (b) The protein fibroin is the major structural component of spider webs. (c) Castor beans contain a highly toxic protein called ricin. [Part a: Phil Degginger/Alamy. Part b: Rosemary Calvert/Imagestate. Part c: Gerald & Buff Corsi/Visuals Unlimited.]

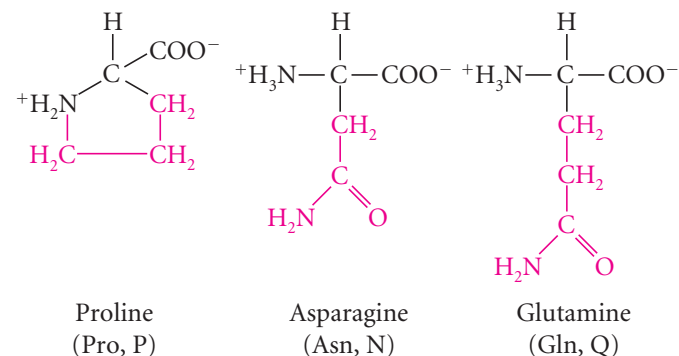
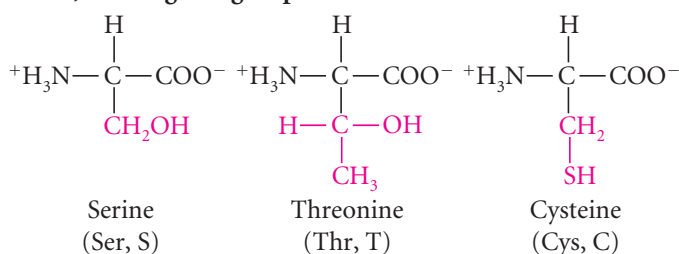


15.5 The common amino acids have similar structures. Each amino acid consists of a central carbon atom (C_α) attached to: (1) an amino group (NH_3^+); (2) a carboxyl group (COO^-); (3) a hydrogen atom (H); and (4) a radical group, designated R. In the structures of the 20 common amino acids, the parts in black are common to all amino acids and the parts in red are the R groups.

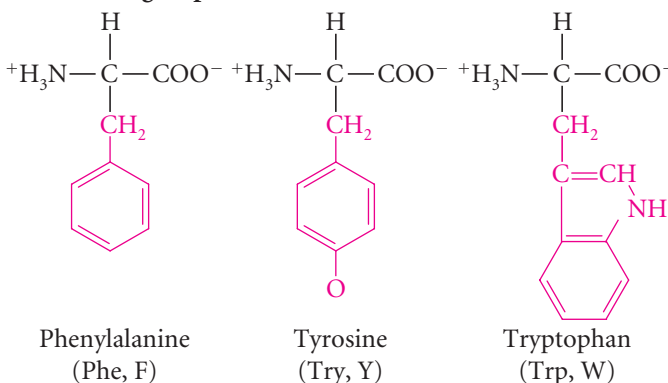
Nonpolar, aliphatic R groups



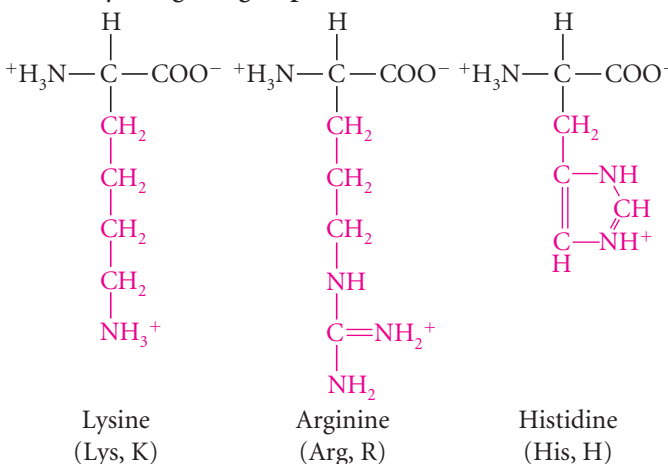
Polar, uncharged R groups



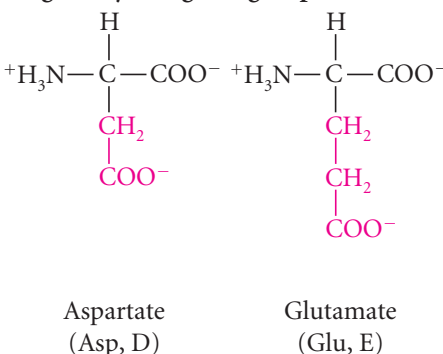
Aromatic R groups

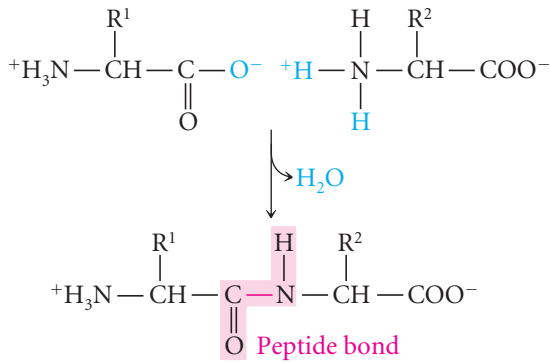


Positively charged R groups



Negatively charged R groups





15.6 Amino acids are joined together by peptide bonds. In a peptide bond (red), the carboxyl group of one amino acid is covalently attached to the amino group of another amino acid.

are largely determined by the primary structure—the amino acid sequence—of the protein. Finally, some proteins consist of two or more polypeptide chains that associate to produce a *quaternary structure* (Figure 15.7d).

CONCEPTS

The products of many genes are proteins whose actions produce the traits encoded by these genes. Proteins are polymers consisting of amino acids linked by peptide bonds. The amino acid sequence of a protein is its primary structure. This structure folds to create the secondary and tertiary structures; two or more polypeptide chains may associate to create a quaternary structure.

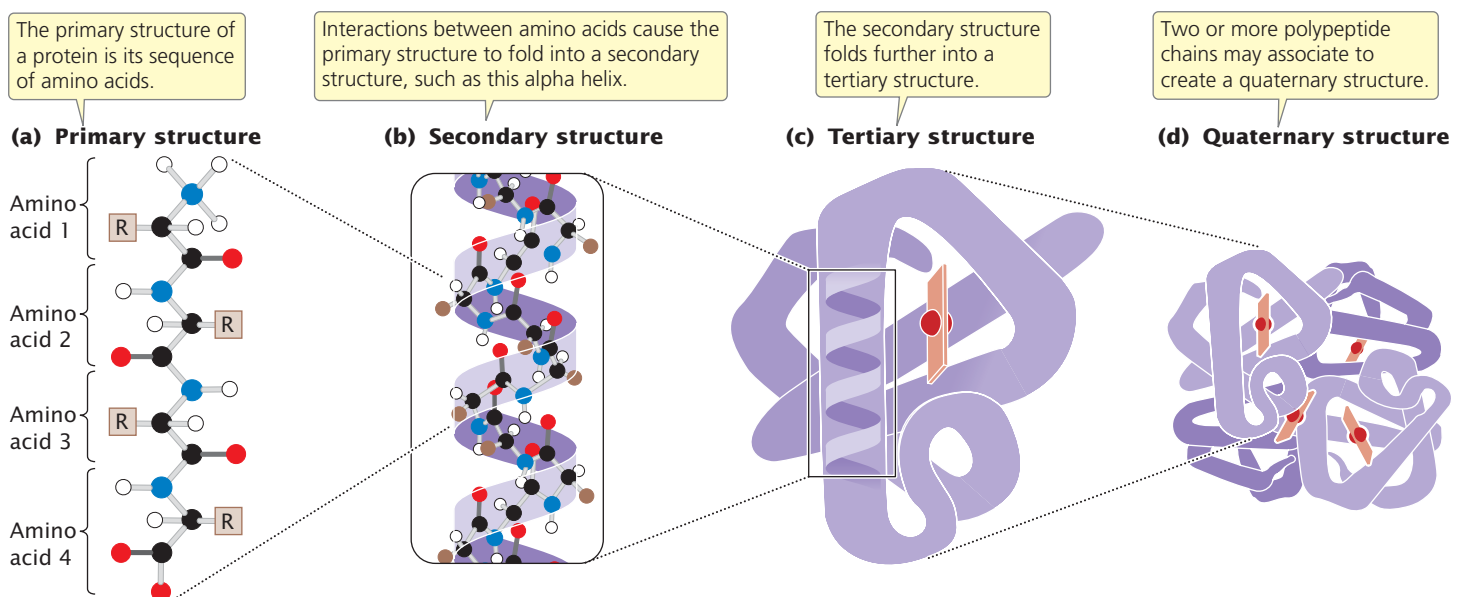
✓ CONCEPT CHECK 2

What primarily determines the secondary and tertiary structures of a protein?

15.2 The Genetic Code Determines How the Nucleotide Sequence Specifies the Amino Acid Sequence of a Protein

In 1953, James Watson and Francis Crick solved the structure of DNA and identified the base sequence as the carrier of genetic information. However, the way in which the base sequence of DNA specifies the amino acid sequences of proteins (the genetic code) was not immediately obvious and remained elusive for another 10 years.

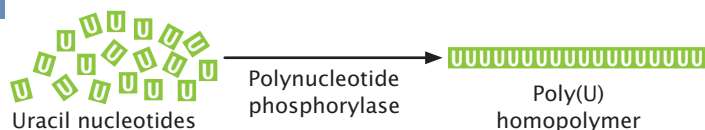
One of the first questions about the genetic code to be addressed was: How many nucleotides are necessary to specify a single amino acid? This basic unit of the genetic code—the set of bases that encode a single amino acid—is a *codon* (see p. 380 in Chapter 14). Many early investigators recognized that codons must contain a minimum of three nucleotides. Each nucleotide position in mRNA can be occupied by one of four bases: A, G, C, or U. If a codon consisted of a single nucleotide, only four different codons (A, G, C, and U) would be possible, which is not enough to encode the 20 different amino acids commonly found in proteins. If codons were made up of two nucleotides each (i.e., GU, AC, etc.), there would be $4 \times 4 = 16$ possible codons—still not enough to encode all 20 amino acids. With three nucleotides per codon, there are $4 \times 4 \times 4 = 64$ possible codons, which is more than enough to specify 20 different amino acids. Therefore, a *triplet code* requiring three nucleotides per codon is the most efficient way to encode all 20 amino acids. Using mutations in bacteriophage, Francis Crick and his colleagues confirmed in 1961 that the genetic code is indeed a triplet code. **TRY PROBLEMS 18 AND 19**



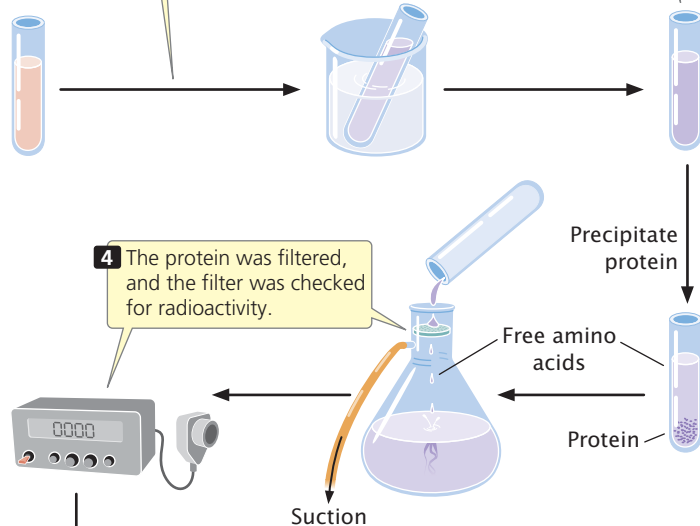
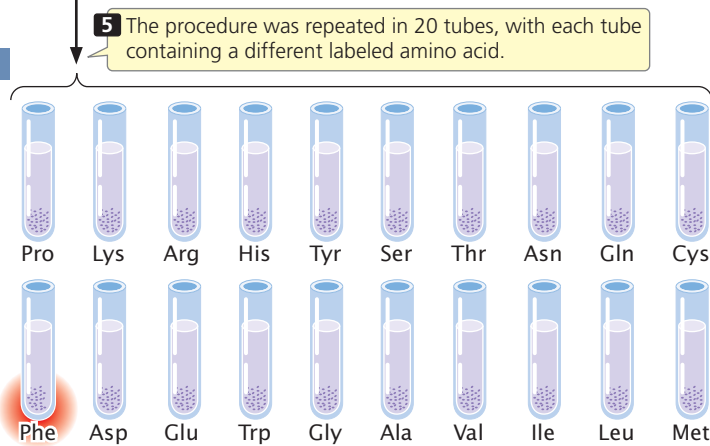
15.7 Proteins have several levels of structural organization.

Experiment

Question: What amino acids are specified by codons composed of only one type of base?

Methods

- 1 A homopolymer—in this case, poly(U) mRNA—was added to a test tube containing a cell-free translation system, 1 radioactively labeled amino acid, and 19 unlabeled amino acids.
- 2 The tube was incubated at 37°C.
- 3 Translation took place.

**Results**

- 6 The tube in which the protein was radioactively labeled contained newly synthesized protein with the amino acid specified by the homopolymer. In this case, UUU specified the amino acid phenylalanine.

Conclusion: UUU encodes phenylalanine; in other experiments, AAA encoded lysine, and CCC encoded proline.

15.8 Nirenberg and Matthaei developed a method for identifying the amino acid specified by a homopolymer.

CONCEPTS

The genetic code is a triplet code, in which three nucleotides encode each amino acid in a protein.

✓ **CONCEPT CHECK 3**

A codon is

- one of three nucleotides that encode an amino acid.
- three nucleotides that encode an amino acid.
- three amino acids that encode a nucleotide.
- one of four bases in DNA.

Breaking the Genetic Code

When it had been firmly established that the genetic code consists of codons that are three nucleotides in length, the next step was to determine which groups of three nucleotides specify which amino acids. Logically, the easiest way to break the code would have been to determine the base sequence of a piece of RNA, add it to a test tube containing all the components necessary for translation, and allow it to direct the synthesis of a protein. The amino acid sequence of the newly synthesized protein could then be determined, and its sequence could be compared with that of the RNA. Unfortunately, there was no way at that time to determine the nucleotide sequence of a piece of RNA; so indirect methods were necessary to break the code.

The use of homopolymers The first clues to the genetic code came in 1961, from the work of Marshall Nirenberg and Johann Heinrich Matthaei. These investigators created synthetic RNAs by using an enzyme called polynucleotide phosphorylase. Unlike RNA polymerase, polynucleotide phosphorylase does not require a template; it randomly links together any RNA nucleotides that happen to be available. The first synthetic mRNAs used by Nirenberg and Matthaei were homopolymers, RNA molecules consisting of a single type of nucleotide. For example, by adding polynucleotide phosphorylase to a solution of uracil nucleotides, they generated RNA molecules that consisted entirely of uracil nucleotides and thus contained only UUU codons (**Figure 15.8**). These poly(U) RNAs were then added to 20 tubes, each containing the components necessary for translation and all 20 amino acids. A different amino acid was radioactively labeled in each of the 20 tubes. Radioactive protein appeared in only one of the tubes—the one containing labeled phenylalanine (see **Figure 15.8**). This result showed that the codon UUU specifies the amino acid phenylalanine. The results of similar experiments using poly(C) and poly(A) RNA demonstrated that CCC

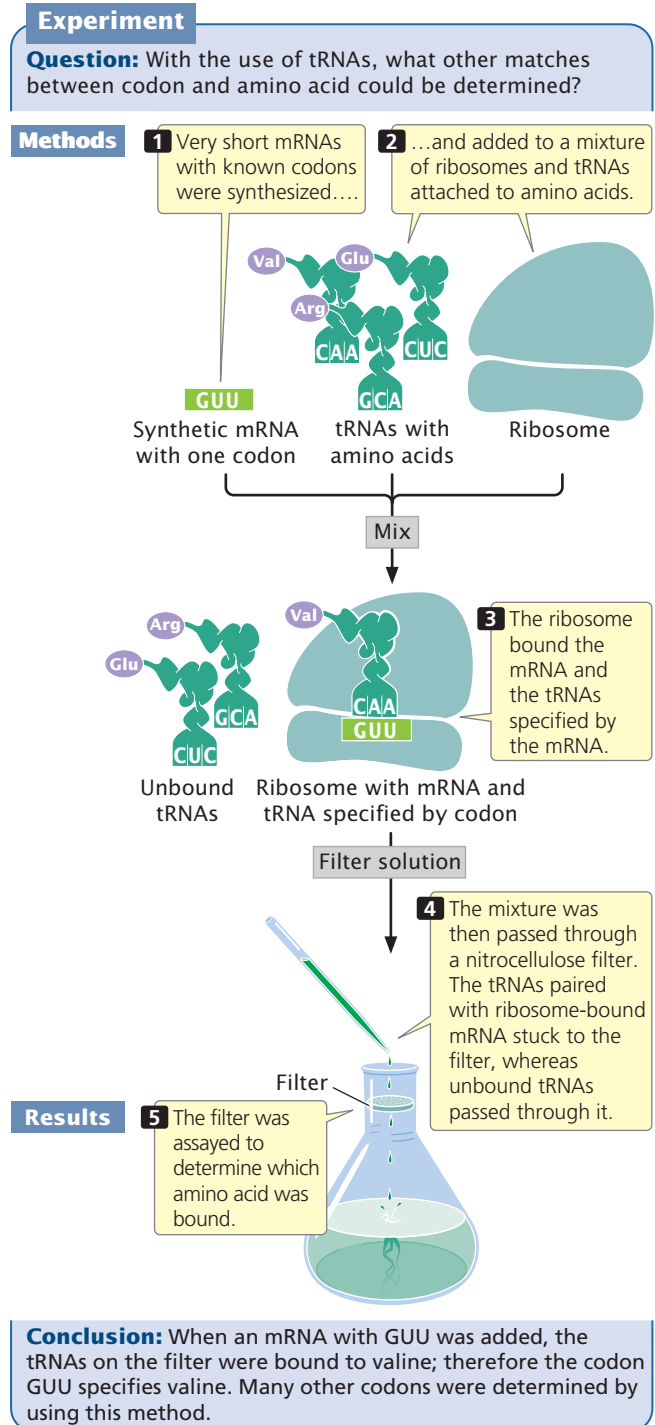
encodes proline and AAA encodes lysine; for technical reasons, the results from poly(G) were uninterpretable.

The use of random copolymers To gain information about additional codons, Nirenberg and his colleagues created synthetic RNAs containing two or three different bases. Because polynucleotide phosphorylase incorporates nucleotides randomly, these RNAs contained random mixtures of the bases and are thus called random copolymers. For example, when adenine and cytosine nucleotides are mixed with polynucleotide phosphorylase, the RNA molecules produced have eight different codons: AAA, AAC, ACC, ACA, CAA, CCA, CAC, and CCC. These poly(AC) RNAs produced proteins containing six different amino acids: asparagine, glutamine, histidine, lysine, proline, and threonine.

The proportions of the different amino acids in the proteins depended on the ratio of the two nucleotides used in creating the synthetic mRNA, and the theoretical probability of finding a particular codon could be calculated from the ratios of the bases. If a 4 : 1 ratio of C to A were used in making the RNA, then the probability of C being at any given position in a codon is $\frac{4}{5}$ and the probability of A being in it is $\frac{1}{5}$. With random incorporation of bases, the probability of any one of the codons with two Cs and one A (CCA, CAC, or ACC) should be $\frac{4}{5} \times \frac{4}{5} \times \frac{1}{5} = \frac{16}{125} = 0.13$, or 13%, and the probability of any codon with two As and one C (AAC, ACA, or CAA) should be $\frac{1}{5} \times \frac{1}{5} \times \frac{4}{5} = \frac{4}{125} = 0.032$, or about 3%. Therefore, an amino acid encoded by two Cs and one A should be more common than an amino acid encoded by two As and one C. By comparing the percentages of amino acids in proteins produced by random copolymers with the theoretical frequencies expected for the codons, Nirenberg and his colleagues could derive information about the *base composition* of the codons. These experiments revealed nothing, however, about the codon base *sequence*; histidine was clearly encoded by a codon with two Cs and one A, but whether that codon was ACC, CAC, or CCA was unknown. There were other problems with this method: the theoretical calculations depended on the random incorporation of bases, which did not always occur, and, because the genetic code is redundant, sometimes several different codons specify the same amino acid.

The use of ribosome-bound tRNAs To overcome the limitations of random copolymers, Nirenberg and Philip Leder developed another technique in 1964 that used ribosome-bound tRNAs. They found that a very short sequence of mRNA—even one consisting of a single codon—would bind to a ribosome. The codon on the short mRNA would then base pair with the matching anticodon on a transfer RNA that carried the amino acid specified by the codon (Figure 15.9). Short mRNAs that were bound to ribosomes were mixed with tRNAs and amino acids, and this mixture was passed through a nitrocellulose filter. The tRNAs that were paired with the ribosome-bound mRNA stuck to the filter,

whereas unbound tRNAs passed through it. The advantage of this system was that it could be used with very short synthetic mRNA molecules that could be synthesized with a known sequence. Nirenberg and Leder synthesized more than 50 short mRNAs with known codons and added them individually to a mixture of ribosomes and tRNAs. They then isolated the tRNAs that were bound to the mRNA and ribosomes and



15.9 Nirenberg and Leder used ribosome-bound tRNAs to provide additional information about the genetic code.

determined which amino acids were present on the bound tRNAs. For example, synthetic RNA with the codon GUU retained a tRNA to which valine was attached, whereas RNAs with the codons UGU and UUG did not. Using this method, Nirenberg and his colleagues were able to determine the amino acids encoded by more than 50 codons.

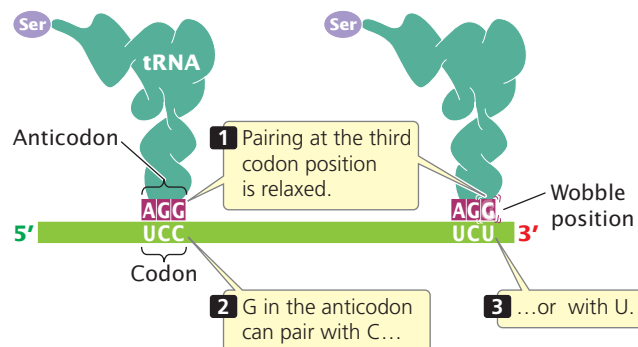
Other experiments provided additional information about the genetic code, and the code was fully deciphered by 1968. In the next section, we will examine some of the features of the code, which is so important to modern biology that Francis Crick compared its place to that of the periodic table of the elements in chemistry.

The Degeneracy of the Code

One amino acid is encoded by three consecutive nucleotides in mRNA, and each nucleotide can have one of four possible bases (A, G, C, and U) at each nucleotide position, thus permitting $4^3 = 64$ possible codons (Figure 15.10). Three of these codons are stop codons, specifying the end of translation. Thus, 61 codons, called **sense codons**, encode amino acids. Because there are 61 sense codons and only 20 different amino acids commonly found in proteins, the code contains more information than is needed to specify the amino acids and is said to be a **degenerate code**. This expression does not mean that the genetic code is depraved; *degenerate* is a term that Francis Crick borrowed from quantum physics, where it describes multiple physical states that have equivalent meaning. The degeneracy of the

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

15.10 The genetic code consists of 64 codons. The amino acids specified by each codon are given in their three-letter abbreviation. The codons are written 5'→3', as they appear in the mRNA. AUG is an initiation codon; UAA, UAG, and UGA are termination (stop) codons.



15.11 Wobble may exist in the pairing of a codon and anticodon. The mRNA and tRNA pair in an antiparallel fashion. Pairing at the first and second codon positions is in accord with the Watson-and-Crick pairing rules (A with U, G with C); however, pairing rules are relaxed at the third position of the codon, and G on the anticodon can pair with either U or C on the codon in this example.

genetic code means that amino acids may be specified by more than one codon. Only tryptophan and methionine are encoded by a single codon (see Figure 15.10). Other amino acids are specified by two codons, and some, such as leucine, are specified by six different codons. Codons that specify the same amino acid are said to be **synonymous**, just as synonymous words are different words that have the same meaning.

As we learned in Chapter 14, tRNAs serve as adapter molecules, binding particular amino acids and delivering them to a ribosome, where the amino acids are then assembled into polypeptide chains. Each type of tRNA attaches to a single type of amino acid. The cells of most organisms possess from about 30 to 50 different tRNAs, and yet there are only 20 different amino acids in proteins. Thus, some amino acids are carried by more than one tRNA. Different tRNAs that accept the same amino acid but have different anticodons are called **isoaccepting tRNAs**. Some synonymous codons specify different isoacceptors.

Even though some amino acids have multiple (isoaccepting) tRNAs, there are still more codons than anticodons, because different codons can sometimes pair with the same anticodon through flexibility in base pairing at the third position of the codon. Examination of Figure 15.10 reveals that many synonymous codons differ only in the third position. For example, alanine is encoded by the codons GCU, GCC, GCA, and GCG, all of which begin with GC. When the codon on the mRNA and the anticodon of the tRNA join (Figure 15.11), the first (5') base of the codon pairs with the third (3') base of the anticodon, strictly according to Watson-and-Crick rules: A with U; C with G. Next, the middle bases of codon and anticodon pair, also strictly following the Watson-and-Crick rules. After these pairs have hydrogen bonded, the third bases pair weakly: there may be flexibility, or **wobble**, in their pairing.

In 1966, Francis Crick developed the wobble hypothesis, which proposed that some nonstandard pairings of bases could take place at the third position of a codon. For example, a G in the anticodon may pair with either a C or a U in

Table 15.2 The wobble rules, indicating which bases in the third position (3' end) of the mRNA codon can pair with bases at the first position (5' end) of the anticodon of the tRNA

First Position of Anticodon	Third Position of Codon	Pairing
C	G	Anticodon 3'-X-Y-C-5' 5'-Y-X-G-3' Codon
G	U or C	Anticodon 3'-X-Y-G-5' 5'-Y-X-U-3' C Codon
A	U	Anticodon 3'-X-Y-A-5' 5'-Y-X-U-3' Codon
U	A or G	Anticodon 3'-X-Y-U-5' 5'-Y-X-A-3' G Codon
I (inosine)	A, U, or C	Anticodon 3'-X-Y-I-5' 5'-Y-X-A-3' U C Codon

the third position of the codon (Table 15.2). The important thing to remember about wobble is that it allows some tRNAs to pair with more than one codon on an mRNA; thus from 30 to 50 tRNAs can pair with 61 sense codons. Some codons are synonymous through wobble. **TRY PROBLEM 24** →

CONCEPTS

The genetic code consists of 61 sense codons that specify the 20 common amino acids; the code is degenerate, meaning that some amino acids are encoded by more than one codon. Isoaccepting tRNAs are different tRNAs with different anticodons that specify the same amino acid. Wobble exists when more than one codon can pair with the same anticodon.

✓ CONCEPT CHECK 4

Through wobble, a single _____ can pair with more than one _____.

- codon, anticodon
- group of three nucleotides in DNA, codon in mRNA
- tRNA, amino acid
- anticodon, codon

The Reading Frame and Initiation Codons

Findings from early studies of the genetic code indicated that the code is generally **nonoverlapping**. An overlapping code is one in which a single nucleotide may be included in more than one codon, as follows:

Nucleotide sequence A U A C G A G U C

Nonoverlapping code A U A C G A G U C
 Ile Arg Val

Overlapping code A U A C G A G U
 Ile
 U A C
 Tyr
 A C G
 Thr

Usually, however, each nucleotide is part of a single codon. A few overlapping genes are found in viruses, but codons within the same gene do not overlap, and the genetic code is generally considered to be nonoverlapping.

For any sequence of nucleotides, there are three potential sets of codons—three ways in which the sequence can be read in groups of three. Each different way of reading the sequence is called a **reading frame**, and any sequence of nucleotides has three potential reading frames. The three reading frames have completely different sets of codons and will therefore specify proteins with entirely different amino acid sequences. Thus, it is essential for the translational machinery to use the correct reading frame. How is the correct reading frame established? The reading frame is set by the **initiation codon**, which is the first codon of the mRNA to specify an amino acid. After the initiation codon, the other codons are read as successive groups of three nucleotides. No bases are skipped between the codons; so there are no punctuation marks to separate the codons.

The initiation codon is usually AUG, although GUG and UUG are used on rare occasions. The initiation codon is not just a sequence that marks the beginning of translation; it specifies an amino acid. In bacterial cells, AUG encodes a modified type of methionine, *N*-formylmethionine; all proteins in bacteria initially begin with this amino acid, but the formyl group (or, in some cases, the entire amino acid) may be removed after the protein has been synthesized. When the codon AUG is at an internal position in a gene, it encodes

unformylated methionine. In archaeal and eukaryotic cells, AUG specifies unformylated methionine both at the initiation position and at internal positions.

Termination Codons

Three codons—UAA, UAG, and UGA—do not encode amino acids. These codons signal the end of the protein in both bacterial and eukaryotic cells and are called **stop codons**, **termination codons**, or **nonsense codons**. No tRNA molecules have anticodons that pair with termination codons.

The Universality of the Code

For many years the genetic code was assumed to be **universal**, meaning that each codon specifies the same amino acid in all organisms. We now know that the genetic code is almost, but not completely, universal; a few exceptions have been found. Most of these exceptions are termination codons, but there are a few cases in which one sense codon substitutes for another. Most exceptions are found in mitochondrial genes; a few nonuniversal codons have also been detected in the nuclear genes of protozoans and in bacterial DNA (Table 15.3). **TRY PROBLEM 20** →

CONCEPTS

Each sequence of nucleotides possesses three potential reading frames. The correct reading frame is set by the initiation codon. The end of a protein-encoding sequence is marked by a termination codon. With a few exceptions, all organisms use the same genetic code.

✓ CONCEPT CHECK 5

Do the initiation and termination codons specify an amino acid? If so, which ones?

Table 15.3 Some exceptions to the universal genetic code

Genome	Codon	Universal Code	Altered Code
Bacterial DNA <i>Mycoplasma capricolum</i>	UGA	Stop	Trp
Mitochondrial DNA			
Human	UGA	Stop	Trp
Human	AUA	Ile	Met
Human	AGA, AGG	Arg	Stop
Yeast	UGA	Stop	Trp
Trypanosomes	UGA	Stop	Trp
Plants	CGG	Arg	Trp
Nuclear DNA			
<i>Tetrahymena</i>	UAA	Stop	Gln
<i>Paramecium</i>	UAG	Stop	Gln

CONNECTING CONCEPTS

Characteristics of the Genetic Code

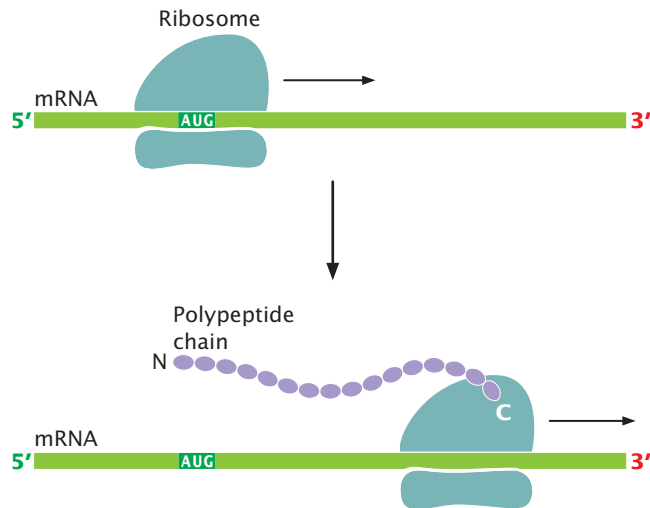
We have now considered a number of characteristics of the genetic code. Let's pause for a moment and review these characteristics.

1. The genetic code consists of a sequence of nucleotides in DNA or RNA. There are four letters in the code, corresponding to the four bases—A, G, C, and U (T in DNA).
2. The genetic code is a triplet code. Each amino acid is encoded by a sequence of three consecutive nucleotides, called a codon.
3. The genetic code is degenerate; that is, of 64 codons, 61 codons encode only 20 amino acids in proteins (3 codons are termination codons). Some codons are synonymous, specifying the same amino acid.
4. Isoaccepting tRNAs are tRNAs with different anticodons that accept the same amino acid; wobble allows the anticodon on one type of tRNA to pair with more than one type of codon on mRNA.
5. The code is generally nonoverlapping; each nucleotide in an mRNA sequence belongs to a single reading frame.
6. The reading frame is set by an initiation codon, which is usually AUG.
7. When a reading frame has been set, codons are read as successive groups of three nucleotides.
8. Any one of three termination codons (UAA, UAG, and UGA) can signal the end of a protein; no amino acids are encoded by the termination codons.
9. The code is almost universal.

15.3 Amino Acids Are Assembled into a Protein Through the Mechanism of Translation

Now that we are familiar with the genetic code, we can begin to study the mechanism by which amino acids are assembled into proteins. Because more is known about translation in bacteria, we will focus primarily on bacterial translation. In most respects, eukaryotic translation is similar, although some significant differences will be noted as we proceed through the stages of translation.

Translation takes place on ribosomes; indeed, ribosomes can be thought of as moving protein-synthesizing machines. Through a variety of techniques, a detailed view of the structure of the ribosome has been produced in recent years, which has greatly improved our understanding of the translational process. A ribosome attaches near the 5' end of an mRNA strand and moves toward the 3' end, translating the codons as it goes (Figure 15.12). Synthesis begins at the amino end of the protein, and the protein is elongated by the addition of new amino acids to the carboxyl end. Protein synthesis includes a series of RNA–RNA interactions: interactions between the mRNA and the rRNA that hold the mRNA in the ribosome, interactions between the codon on the mRNA and the anticodon on the tRNA, and interactions between the tRNA and the rRNAs of the ribosome.

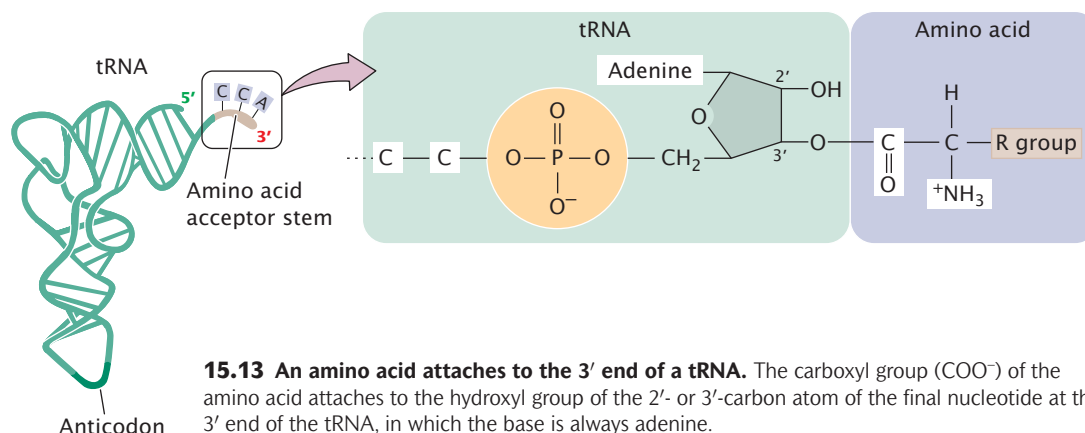


15.12 The translation of an mRNA molecule takes place on a ribosome. The letter N represents the amino end of the protein; C represents the carboxyl end.

Protein synthesis can be conveniently divided into four stages: (1) tRNA charging, in which tRNAs bind to amino acids; (2) initiation, in which the components necessary for translation are assembled at the ribosome; (3) elongation, in which amino acids are joined, one at a time, to the growing polypeptide chain; and (4) termination, in which protein synthesis halts at the termination codon and the translation components are released from the ribosome.

The Binding of Amino Acids to Transfer RNAs

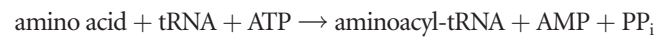
The first stage of translation is the binding of tRNA molecules to their appropriate amino acids, called tRNA charging. Each tRNA is specific for a particular amino acid. All tRNAs have the sequence CCA at the 3' end, and the carboxyl group (COO^-) of the amino acid is attached to the adenine nucleotide at the 3' end of the tRNA (Figure 15.13). If each tRNA is specific for a particular amino acid but all amino acids are attached to the same nucleotide (A) at the 3' end of a tRNA, how does a tRNA link up with its appropriate amino acid?



15.13 An amino acid attaches to the 3' end of a tRNA. The carboxyl group (COO^-) of the amino acid attaches to the hydroxyl group of the 2'- or 3'-carbon atom of the final nucleotide at the 3' end of the tRNA, in which the base is always adenine.

The key to specificity between an amino acid and its tRNA is a set of enzymes called **aminoacyl-tRNA synthetases**. A cell has 20 different aminoacyl-tRNA synthetases, one for each of the 20 amino acids. Each synthetase recognizes a particular amino acid, as well as all the tRNAs that accept that amino acid. Recognition of the appropriate amino acid by a synthetase is based on the different sizes, charges, and R groups of the amino acids. The recognition of tRNAs by a synthetase depends on the differing nucleotide sequences of the tRNAs. Researchers have identified which nucleotides are important in recognition by altering different nucleotides in a particular tRNA and determining whether the altered tRNA is still recognized by its synthetase (Figure 15.14).

The attachment of a tRNA to its appropriate amino acid, termed **tRNA charging**, requires energy, which is supplied by adenosine triphosphate (ATP):



This reaction takes place in two steps (Figure 15.15). To identify the resulting aminoacylated tRNA, we write the three-letter abbreviation for the amino acid in front of the tRNA; for example, the amino acid alanine (Ala) attaches to its tRNA (tRNA^{Ala}), giving rise to its aminoacyl-tRNA ($\text{Ala-tRNA}^{\text{Ala}}$).

Errors in tRNA charging are rare; they occur in only about 1 in 10,000 to 1 in 100,000 reactions. This fidelity is due in part to the presence of editing (proofreading) activity in many of the synthetases. Editing activity detects and removes incorrectly paired amino acids from the tRNAs. Some antifungal chemical agents work by trapping tRNAs in the editing site of the enzyme, preventing their release and thus inhibiting the process of translation in the fungi.

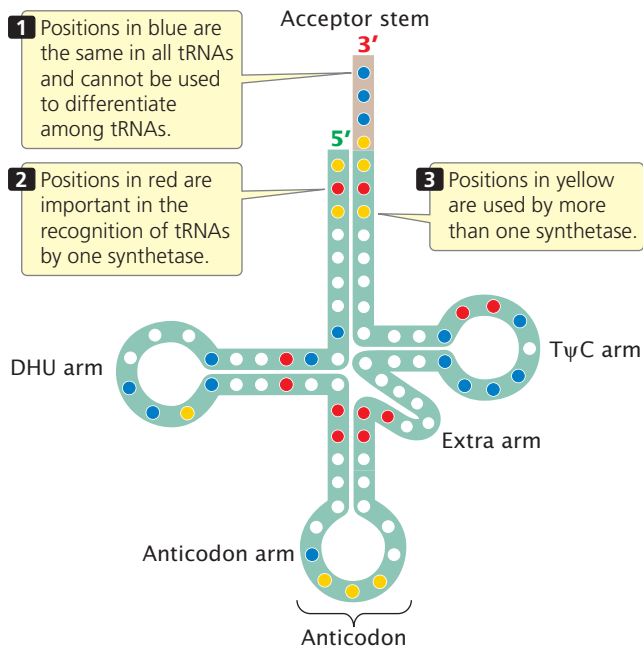
CONCEPTS

Amino acids are attached to specific tRNAs by aminoacyl-tRNA synthetases in a two-step reaction that requires ATP.

✓ CONCEPT CHECK 6

Amino acids bind to which part of the tRNA?

- | | |
|--------------|-----------|
| a. anticodon | c. 3' end |
| b. DHU arm | d. 5' end |



15.14 Certain positions on tRNA molecules are recognized by the appropriate aminoacyl-tRNA synthetase.

The Initiation of Translation

The second stage in the process of protein synthesis is initiation. At this stage, all the components necessary for protein synthesis assemble: (1) mRNA; (2) the small and large subunits of the ribosome; (3) a set of three proteins called initiation factors; (4) initiator tRNA with *N*-formylmethionine attached (fMet-tRNA^{fMet}); and (5) guanosine triphosphate (GTP). Initiation comprises three major steps. First, mRNA binds to the small subunit of the ribosome. Second, initiator tRNA binds to the mRNA through base pairing between the codon and the anticodon. Third, the large ribosome joins the initiation complex. Let's look at each of these steps more closely.

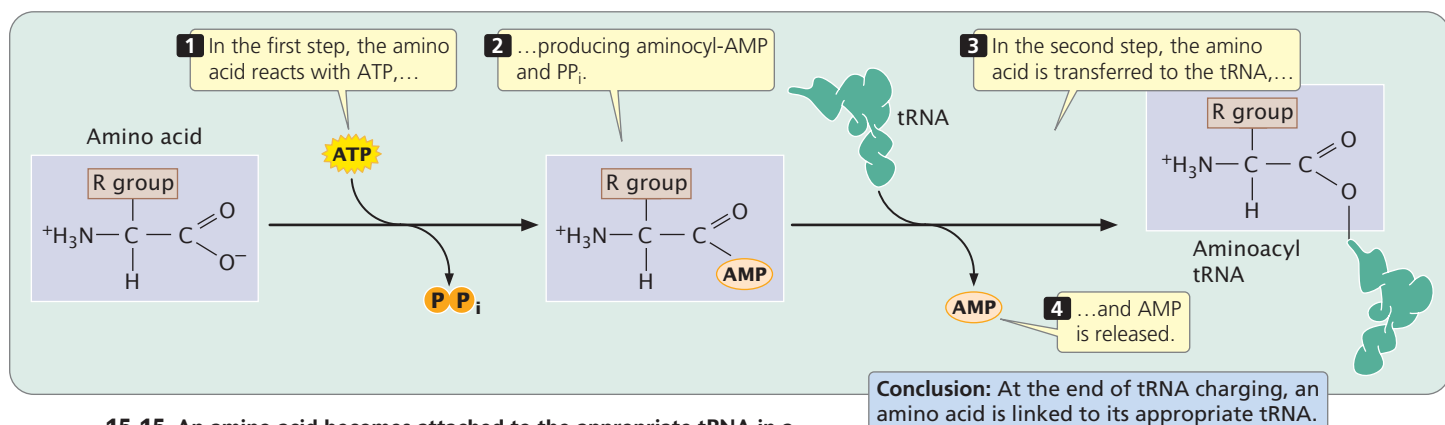
Initiation in bacteria The functional ribosome of bacteria exists as two subunits, the small 30S subunit and the large 50S subunit (Figure 15.16a). An mRNA molecule can

bind to the small ribosome subunit only when the subunits are separate. **Initiation factor 3** (IF-3) binds to the small subunit of the ribosome and prevents the large subunit from binding during initiation (Figure 15.16b). Another factor, **initiation factor 1** (IF-1), enhances the disassociation of the large and small ribosomal subunits.

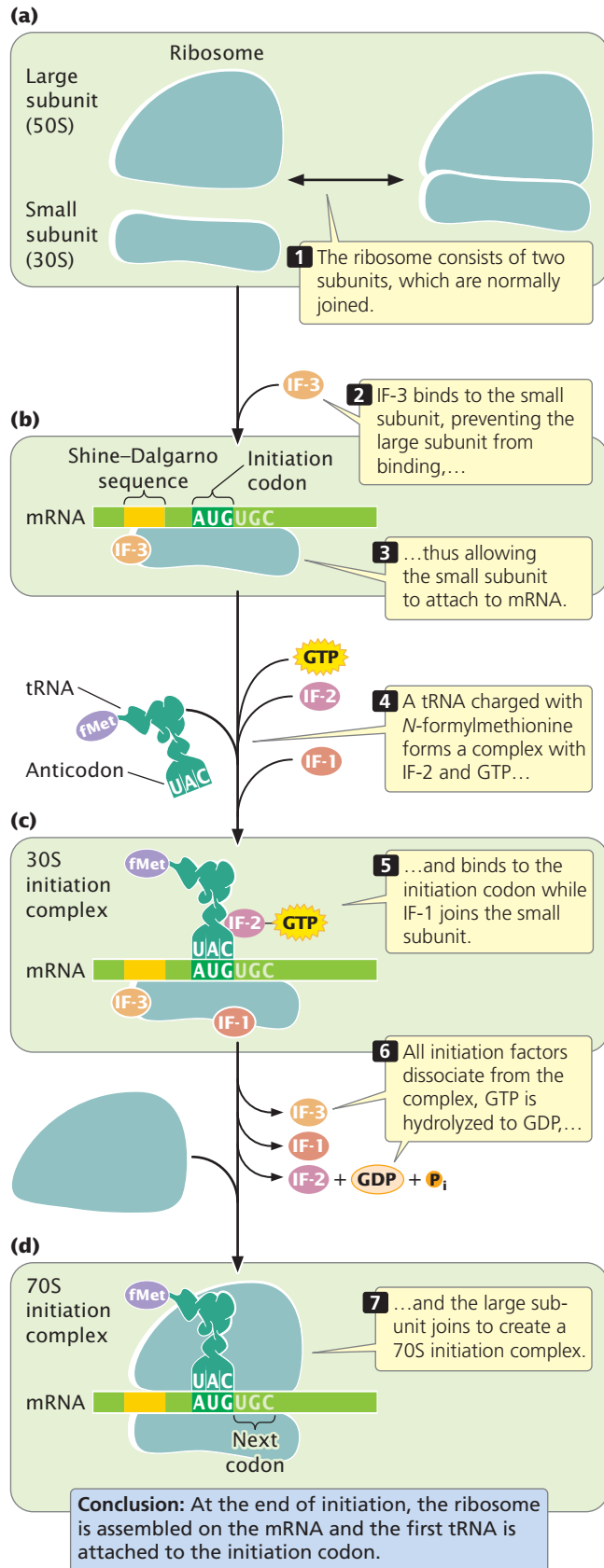
Key sequences on the mRNA required for ribosome binding have been identified in experiments designed to allow the ribosome to bind to mRNA but not proceed with protein synthesis; the ribosome is thereby stalled at the initiation site. After the ribosome is allowed to attach to the mRNA, ribonuclease is added, which degrades all the mRNA except the region covered by the ribosome. The intact mRNA can be separated from the ribosome and studied. The sequence covered by the ribosome during initiation is from 30 to 40 nucleotides long and includes the AUG initiation codon. Within the ribosome-binding site is the Shine–Dalgarno consensus sequence (Figure 15.17; see also Chapter 14), which is complementary to a sequence of nucleotides at the 3' end of 16S rRNA (part of the small subunit of the ribosome). During initiation, the nucleotides in the Shine–Dalgarno sequence pair with their complementary nucleotides in the 16S rRNA, allowing the small subunit of the ribosome to attach to the mRNA and positioning the ribosome directly over the initiation codon.

Next, the initiator tRNA, fMet-tRNA^{fMet}, attaches to the initiation codon (see Figure 15.16c). This step requires **initiation factor 2** (IF-2), which forms a complex with GTP.

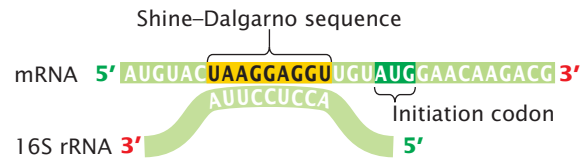
At this point, the initiation complex consists of (1) the small subunit of the ribosome; (2) the mRNA; (3) the initiator tRNA with its amino acid (fMet-tRNA^{fMet}); (4) one molecule of GTP; and (5) several initiation factors. These components are collectively known as the **30S initiation complex** (see Figure 15.16c). In the final step of initiation, IF-3 dissociates from the small subunit, allowing the large subunit of the ribosome to join the initiation complex. The molecule of GTP (provided by IF-2) is hydrolyzed to guanosine diphosphate (GDP), and the initiation factors dissociate (see Figure 15.16d). When the large subunit has joined the initiation complex, the complex is called the **70S initiation complex**.



15.15 An amino acid becomes attached to the appropriate tRNA in a two-step reaction.

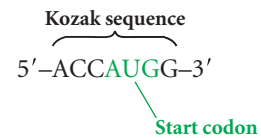


15.16 The initiation of translation requires several initiation factors and GTP.



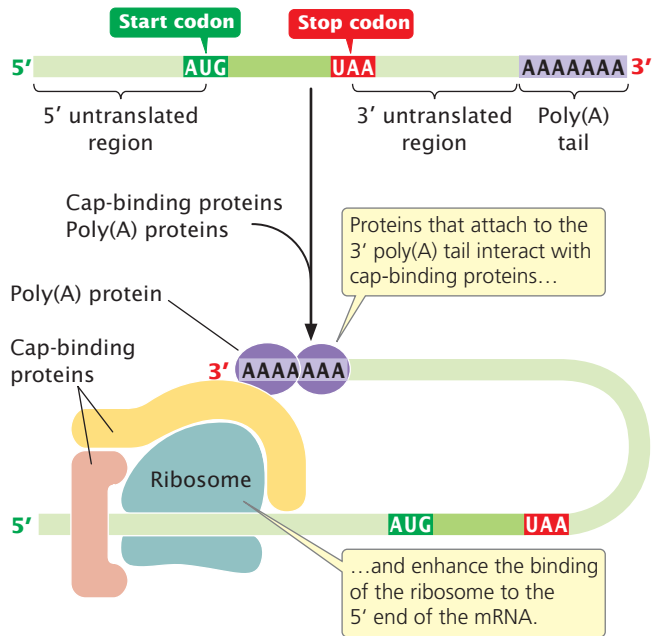
15.17 The Shine-Dalgarno consensus sequence in mRNA is required for the attachment of the small subunit of the ribosome.

Initiation in eukaryotes Similar events take place in the initiation of translation in eukaryotic cells, but there are some important differences. In bacterial cells, sequences in 16S rRNA of the small subunit of the ribosome bind to the Shine-Dalgarno sequence in mRNA. No analogous consensus sequence exists in eukaryotic mRNA. Instead, the cap at the 5' end of eukaryotic mRNA plays a critical role in the initiation of translation. In a series of steps, the small subunit of the eukaryotic ribosome, initiation factors, and the initiator tRNA with its amino acid (Met-tRNA_i^{Met}) form an initiation complex that recognizes the cap and binds there. The initiation complex then moves along (scans) the mRNA until it locates the first AUG codon. The identification of the start codon is facilitated by the presence of a consensus sequence (called the Kozak sequence) that surrounds the start codon:



Another important difference is that eukaryotic initiation requires at least seven initiation factors. Some factors keep the ribosomal subunits separated, just as IF-3 does in bacterial cells. Others recognize the 5' cap on mRNA and allow the small subunit of the ribosome to bind there. Still others possess RNA helicase activity, which is used to unwind secondary structures that may exist in the 5' untranslated region of mRNA, allowing the small subunit to move down the mRNA until the initiation codon is reached. Other initiation factors help bring Met-tRNA_i^{Met} to the initiation complex.

The poly(A) tail at the 3' end of eukaryotic mRNA also plays a role in the initiation of translation. During initiation, proteins that attach to the poly(A) tail interact with proteins that bind to the 5' cap, enhancing the binding of the small subunit of the ribosome to the 5' end of the mRNA. This interaction indicates that the 3' end of mRNA bends over and associates with the 5' cap during the initiation of translation, forming a circular structure known as the closed loop (Figure 15.18). A few eukaryotic mRNAs contain internal ribosome entry sites, where ribosomes can bind directly without first attaching to the 5' cap.



15.18 The poly(A) tail of eukaryotic mRNA plays a role in the initiation of translation.

CONCEPTS

In the initiation of translation in bacterial cells, the small ribosomal subunit attaches to mRNA, and initiator tRNA attaches to the initiation codon. This process requires several initiation factors (IF-1, IF-2, and IF-3) and GTP. In the final step, the large ribosomal subunit joins the initiation complex.

✓ CONCEPT CHECK 7

During the initiation of translation, the small ribosome binds to which consensus sequence in bacteria?

Elongation

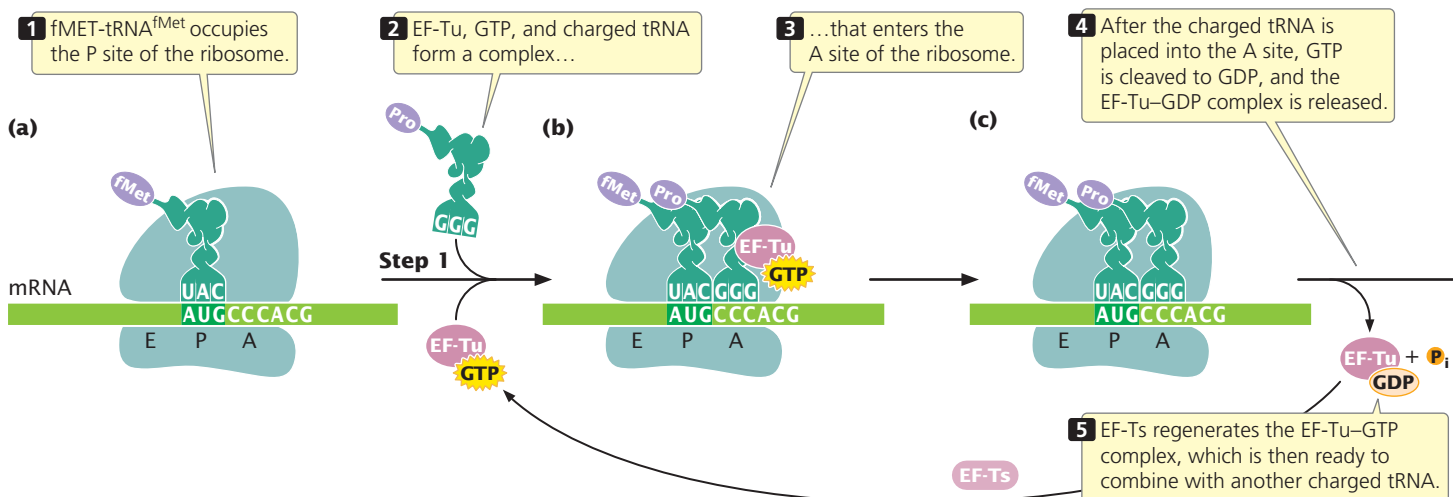
The next stage in protein synthesis is elongation, in which amino acids are joined to create a polypeptide chain. Elongation requires (1) the 70S complex just described; (2) tRNAs charged with their amino acids; (3) several elongation factors; and (4) GTP.

A ribosome has three sites that can be occupied by tRNAs; the **aminoacyl**, or **A, site**, the **peptidyl**, or **P, site**, and the **exit**, or **E, site** (Figure 15.19a). The initiator tRNA immediately occupies the P site (the only site to which the fMet-tRNA^{fMet} is capable of binding), but all other tRNAs first enter the A site. After initiation, the ribosome is attached to the mRNA, and fMet-tRNA^{fMet} is positioned over the AUG start codon in the P site; the adjacent A site is unoccupied (see Figure 15.19a).

Elongation takes place in three steps. In the first step (Figure 15.19b), a charged tRNA binds to the A site. This binding takes place when **elongation factor Tu** (EF-Tu) joins with GTP and then with a charged tRNA to form a three-part complex. This complex enters the A site of the ribosome, where the anticodon on the tRNA pairs with the codon on the mRNA. After the charged tRNA is in the A site, GTP is cleaved to GDP, and the EF-Tu-GDP complex is released (Figure 15.19c). **Elongation factor Ts** (EF-Ts) regenerates EF-Tu-GDP to EF-Tu-GTP. In eukaryotic cells, a similar set of reactions delivers the charged tRNA to the A site.

The second step of elongation is the formation of a peptide bond between the amino acids that are attached to tRNAs in the P and A sites (Figure 15.19d). The formation of this peptide bond releases the amino acid in the P site from its tRNA. Evidence indicates that the catalytic activity is a property of the ribosomal RNA in the large subunit of the ribosome; this rRNA acts as a ribozyme (see p. 352 in Chapter 13).

The third step in elongation is **translocation** (Figure 15.19e), the movement of the ribosome down the mRNA in the 5'→3' direction. This step positions the ribosome over



15.19 The elongation of translation comprises three steps.

the next codon and requires **elongation factor G** (EF-G) and the hydrolysis of GTP to GDP. Because the tRNAs in the P and A sites are still attached to the mRNA through codon–anticodon pairing, they do not move with the ribosome as it translocates. Consequently, the ribosome shifts so that the tRNA that previously occupied the P site now occupies the E site, from which it moves into the cytoplasm where it can be recharged with another amino acid. Translocation also causes the tRNA that occupied the A site (which is attached to the growing polypeptide chain) to be in the P site, leaving the A site open. Thus, the progress of each tRNA through the ribosome in the course of elongation can be summarized as follows: cytoplasm → A site → P site → E site → cytoplasm. As discussed earlier, the initiator tRNA is an exception: it attaches directly to the P site and never occupies the A site.

After translocation, the A site of the ribosome is empty and ready to receive the tRNA specified by the next codon. The elongation cycle (see Figure 15.19b through e) repeats itself: a charged tRNA and its amino acid occupy the A site, a peptide bond is formed between the amino acids in the A and P sites, and the ribosome translocates to the next codon. Throughout the cycle, the polypeptide chain remains attached to the tRNA in the P site. Messenger RNAs, although single stranded, often contain secondary structures formed by pairing of complementary bases on different parts of the mRNA (see Figure 13.1b). As the ribosome moves along the mRNA, these secondary structures are unwound by helicase activity located in the small subunit of the ribosome.

Recently, researchers have developed methods for following a single ribosome as it translates individual codons of an mRNA molecule. These studies revealed that translation does not take place in a smooth continuous fashion. Each translocation step typically requires less than a tenth of a second, but sometimes there are distinct pauses, often lasting a few seconds, between each translocation event when the ribosome moves from one codon to another. Thus, transla-

tion takes place in a series of quick translocations interrupted by brief pauses. In addition to the short pauses between translocation events, translation may be interrupted by longer pauses—lasting from 1 to 2 minutes—that may play a role in regulating the process of translation.

Elongation in eukaryotic cells takes place in a similar manner. Eukaryotes possess at least three elongation factors, one of which also acts in initiation and termination. Another of the elongation factors used in eukaryotes, called elongation factor 2 (EF-2) is the target of a toxin produced by bacteria that causes diphtheria, a disease that, until recently, was a leading killer of children. The diphtheria toxin inhibits EF-2, preventing translocation of the ribosome along the mRNA, and protein synthesis ceases.

CONCEPTS

Elongation consists of three steps: (1) a charged tRNA enters the A site, (2) a peptide bond is created between amino acids in the A and P sites, and (3) the ribosome translocates to the next codon. Elongation requires several elongation factors and GTP.

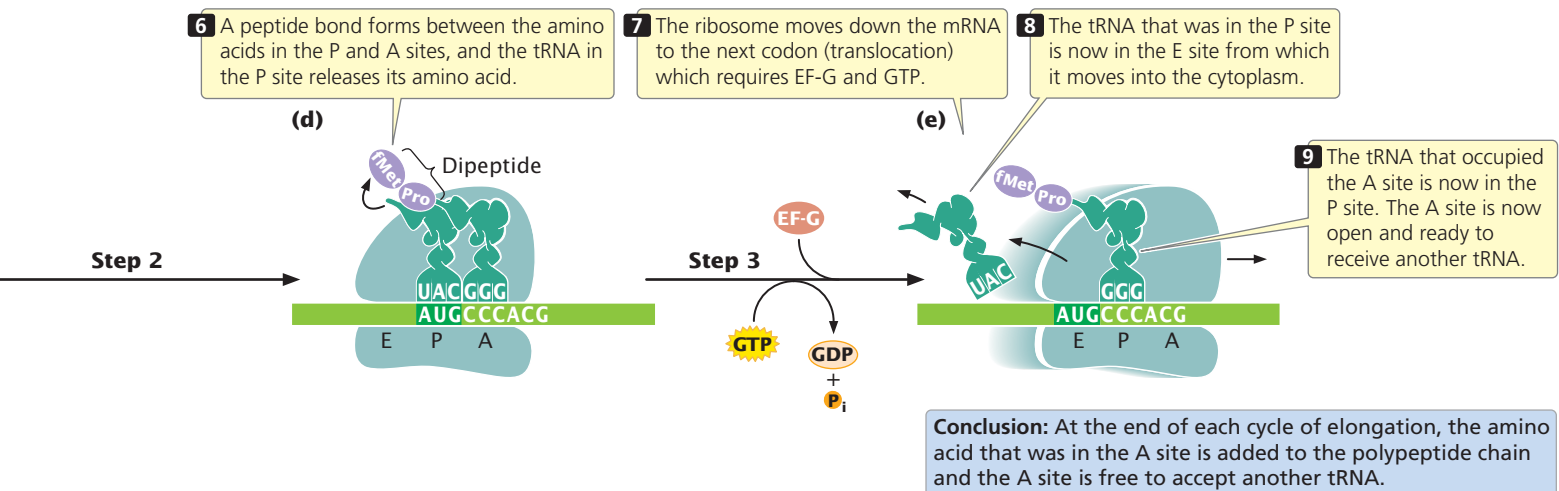
✓ CONCEPT CHECK 8

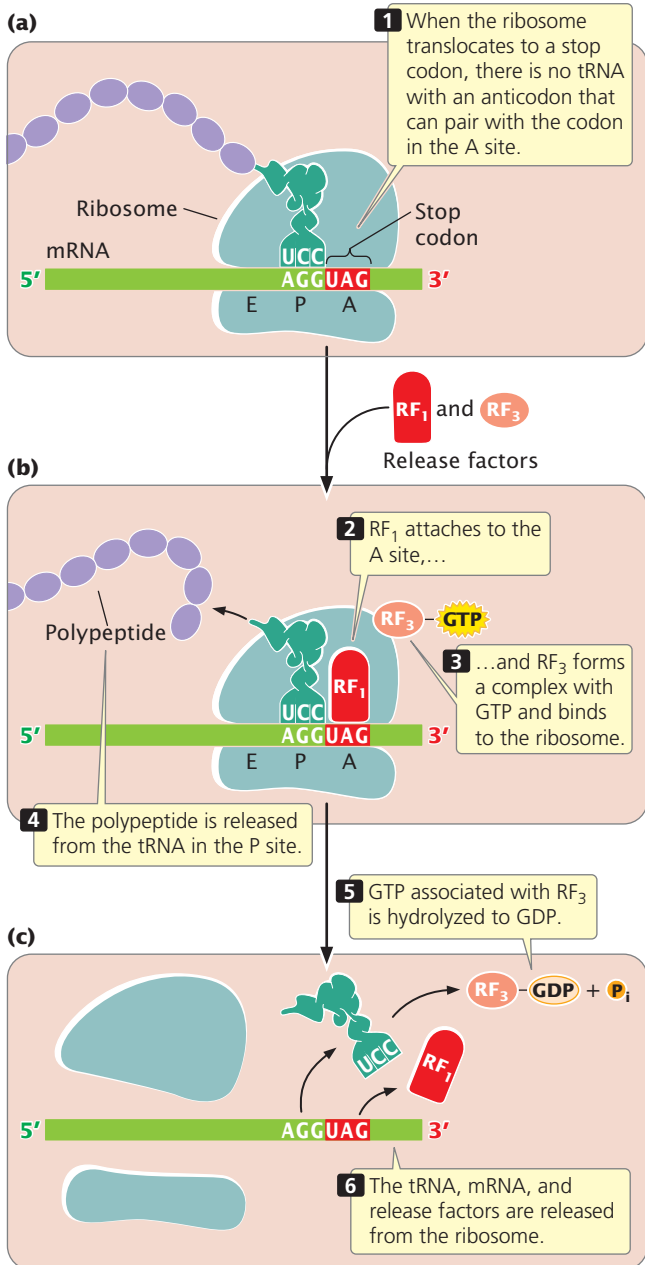
In elongation, the creation of peptide bonds between amino acids is catalyzed by

- a. rRNA.
- b. protein in the small subunit.
- c. protein in the large subunit.
- d. tRNA.

Termination

Protein synthesis terminates when the ribosome translocates to a termination codon. Because there are no tRNAs with anticodons complementary to the termination codons, no tRNA enters the A site of the ribosome when a termination codon is encountered (Figure 15.20a). Instead, proteins called **release factors** bind to the ribosome (Figure 15.20b). *E. coli* has three release factors—RF₁, RF₂, and RF₃. Release factor

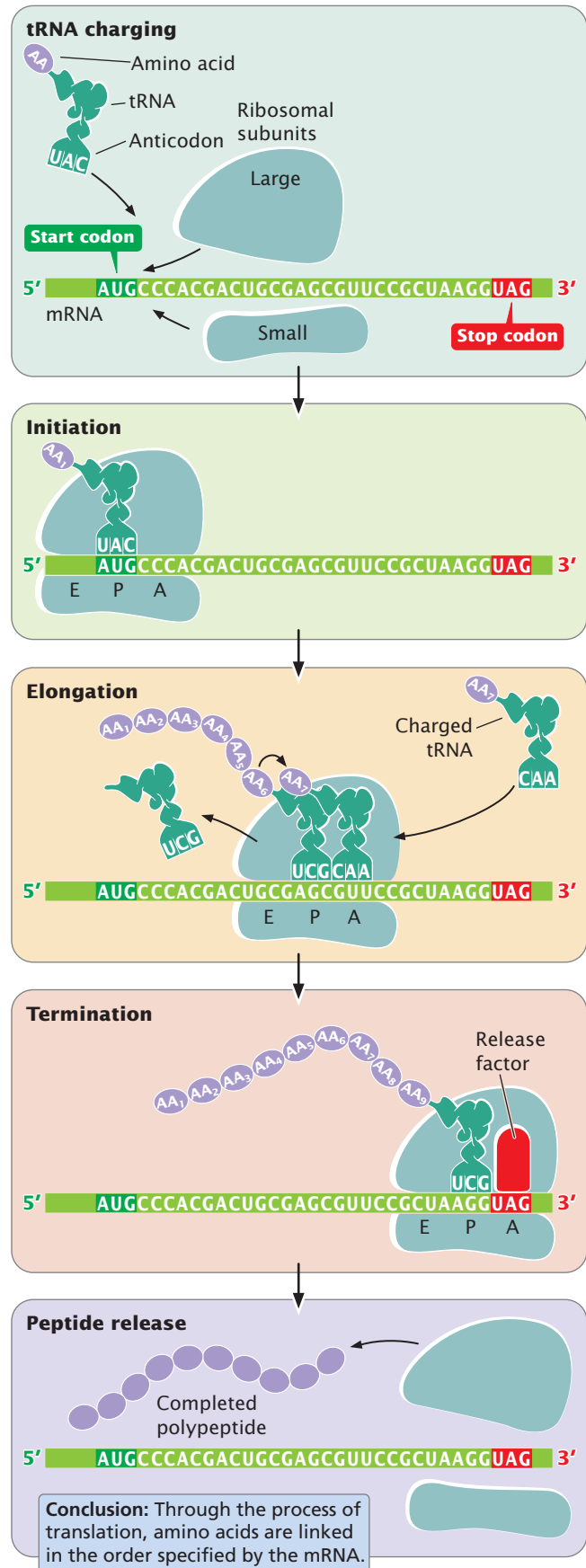




15.20 Translation ends when a stop codon is encountered. Because UAG is the termination codon in this illustration, the release factor is RF₁.

15.21 Translation consists of tRNA charging, initiation, elongation, and termination. In this process, amino acids are linked together in the order specified by mRNA to create a polypeptide chain. A number of initiation, elongation, and release factors take part in the process, and energy is supplied by ATP and GTP.

www.whfreeman.com/pierce4e Explore the process of bacterial translation by examining the consequences of various mutations in the coding region of a gene in Animation 15.1.



1 binds to the termination codons UAA and UAG, and RF₂ binds to UGA and UAA. The binding of release factor RF₁ or RF₂ to the A site of the ribosome promotes the cleavage of the tRNA in the P site from the polypeptide chain and the release of the polypeptide. Release factor 3 binds to the ribosome and forms a complex with GTP. This binding brings about a conformational change in the ribosome, releasing RF₁ or RF₂ from the A site and causing the tRNA in the P site to move to the E site; in the process, GTP is hydrolyzed to GDP. Additional factors help bring about the release of the tRNA from the P site, the release of the mRNA from the ribosome, and the dissociation of the ribosome (**Figure 15.20c**).

Translation in eukaryotic cells terminates in a similar way, except that there are two release factors: eRF₁, which recognizes all three termination codons, and eRF₂, which binds GTP and stimulates the release of the polypeptide from the ribosome. **TRY PROBLEMS 27 AND 30 →**

CONCEPTS

Termination takes place when the ribosome reaches a termination codon. Release factors bind to the termination codon, causing the release of the polypeptide from the last tRNA, of this tRNA from the ribosome, and of the mRNA from the ribosome.

The overall process of protein synthesis, including tRNA charging, initiation, elongation, and termination, is summarized in **Figure 15.21**. The components taking part in this process are listed in **Table 15.4**.

CONNECTING CONCEPTS

A Comparison of Bacterial and Eukaryotic Translation

We have now considered the process of translation in bacterial cells and noted some distinctive differences that exist in eukaryotic cells. Let's take a few minutes to reflect on some of the important similarities and differences of protein synthesis in bacterial and eukaryotic cells.

First, we should emphasize that the genetic code of bacterial and eukaryotic cells is virtually identical; the only difference is in the amino acid specified by the initiation codon. In bacterial cells, the initiation AUG encodes a modified type of methionine, *N*-formylmethionine, whereas, in eukaryotic cells, initiation AUG encodes unformylated methionine. One consequence of the fact that bacteria and eukaryotes use the same code is that eukaryotic genes can be translated in bacterial systems, and vice versa; this feature makes genetic engineering possible, as we will see in Chapter 19.

Another difference is that transcription and translation take place simultaneously in bacterial cells, but the nuclear envelope separates these processes in eukaryotic cells. The physical separation

Table 15.4 Components required for protein synthesis in bacterial cells

Stage	Component	Function
Binding of amino acid to tRNA	Amino acids tRNAs Aminoacyl-tRNA synthetase ATP	Building blocks of proteins Deliver amino acids to ribosomes Attaches amino acids to tRNAs Provides energy for binding amino acid to tRNA
Initiation	mRNA fMet-tRNA ^{fMet} 30S ribosomal subunit 50S ribosomal subunit Initiation factor 1 Initiation factor 2 Initiation factor 3	Carries coding instructions Provides first amino acid in peptide Attaches to mRNA Stabilizes tRNAs and amino acids Enhances dissociation of large and small subunits of ribosome Binds GTP; delivers fMet tRNA ^{fMet} to initiation codon Binds to 30S subunit and prevents association with 50S subunit
Elongation	70S initiation complex Charged tRNAs Elongation factor Tu Elongation factor Ts Elongation factor G GTP Peptidyl transferase center	Functional ribosome with A, P, and E sites where protein synthesis takes place Bring amino acids to ribosome and help assemble them in order specified by mRNA Binds GTP and charged tRNA; delivers charged tRNA to A site Generates active elongation factor Tu Stimulates movement of ribosome to next codon Provides energy Creates peptide bond between amino acids in A site and P site
Termination	Release factors 1, 2, and 3	Bind to ribosome when stop codon is reached and terminate translation

of transcription and translation has important implications for the control of gene expression, which we will consider in Chapter 17, and it allows for extensive modification of eukaryotic mRNAs, as discussed in Chapter 14.

Yet another difference is that mRNA in bacterial cells is short-lived, typically lasting only a few minutes, but the longevity of mRNA in eukaryotic cells is highly variable and is frequently hours or days.

In both bacterial and eukaryotic cells, aminoacyl-tRNA synthetases attach amino acids to their appropriate tRNAs and the chemical process is the same. There are significant differences in the sizes and compositions of bacterial and eukaryotic ribosomal subunits. For example, the large subunit of the eukaryotic ribosome contains three rRNAs, whereas the bacterial ribosome contains only two. These differences allow antibiotics and other substances to inhibit bacterial translation while having no effect on the translation of eukaryotic nuclear genes, as will be discussed later in this chapter.

Other fundamental differences lie in the process of initiation. In bacterial cells, the small subunit of the ribosome attaches directly to the region surrounding the start codon through hydrogen bonding between the Shine–Dalgarno consensus sequence in the 5′ untranslated region of the mRNA and a sequence at the 3′ end of the 16S rRNA. In contrast, the small subunit of a eukaryotic ribosome first binds to proteins attached to the 5′ cap on mRNA and then migrates down the mRNA, scanning the sequence until it encounters the first AUG initiation codon. Additionally, a larger number of initiation factors take part in eukaryotic initiation than in bacterial initiation.

Elongation and termination are similar in bacterial and eukaryotic cells, although different elongation and termination factors are used. In both types of organisms, mRNAs are translated multiple times and are simultaneously attached to several ribosomes, forming polyribosomes, as discussed in Section 15.4.

Much less is known about the process of translation in archaea, but they appear to possess a mixture of eubacterial and eukaryotic features. Because archaea lack nuclear membranes, transcription and translation take place simultaneously, just as they do in eubacterial cells. Archaea utilize unformylated methionine as the initiator amino acid, a characteristic of eukaryotic translation. Some of the initiation and release factors in archaea are similar to those found in eubacteria, whereas others are similar to those found in eukaryotes. Finally, some of the antibiotics that inhibit translation in eubacteria have no effect on translation in archaea, providing further evidence of the fundamental differences between eubacteria and archaea.

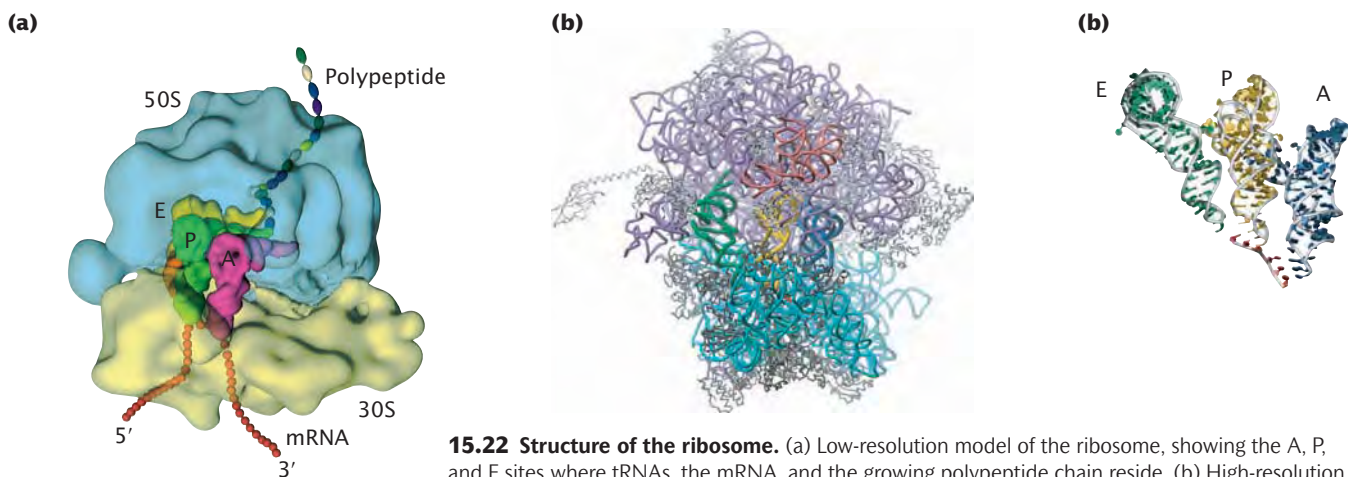
15.4 Additional Properties of RNA and Ribosomes Affect Protein Synthesis

Now that we have considered in some detail the process of translation, we will examine some additional aspects of protein synthesis and the protein-synthesis machinery.

The Three-Dimensional Structure of the Ribosome

The central role of the ribosome in protein synthesis was recognized in the 1950s, and many aspects of its structure have been studied since then. Nevertheless, many details of ribosome structure and function remained a mystery until detailed, three-dimensional reconstructions were completed recently. In 2009, the Nobel Prize in chemistry was awarded to Venkatraman Ramakrishnan, Thomas Steitz, and Ada Yonath for their research on the molecular structure of the ribosome.

Figure 15.22a shows a model of the *E. coli* ribosome at low resolution. A high-resolution image of the bacterial ribosome as determined by X-ray crystallography is represented in the model depicted in **Figure 15.22b**. The mRNA is bound to the small subunit of the ribosome, and the tRNAs are located in the A, P, and E sites (**Figure 15.22c**) that bridge the small and large subunits (see **Figure 15.22a**). Initiation factors 1 and 3 bind to sites on the outside of the small subunit of the ribosome. EF-Tu, EF-G, and other factors complexed with GTP interact with a *factor-binding center*. High-resolution crystallographic images provide information indicating that a *decoding center* resides in the small subunit of the ribosome (the decoding center cannot be seen in **Figure 15.22b**). This center senses the fit between the codon on the mRNA and the anticodon on the incoming charged tRNA. Only tRNAs with the correct anticodon are bound tightly by the ribosome. The structural analyses also indicate that the large subunit of the ribosome contains the *peptidyl transferase center*, where peptide-bond formation takes place. There are no ribosomal proteins in the peptidyl



15.22 Structure of the ribosome. (a) Low-resolution model of the ribosome, showing the A, P, and E sites where tRNAs, the mRNA, and the growing polypeptide chain reside. (b) High-resolution model of the ribosome. (c) Positions of tRNAs in E, P, and A sites of the ribosome shown in part **b**.

transferase center; peptide-bond formation is carried out by the RNA molecule. These analyses also reveal that a tunnel connects the site of peptide-bond formation with the back of the ribosome; the growing polypeptide chain passes through this tunnel to the outside of the ribosome. The tunnel can accommodate about 35 amino acids of the growing polypeptide chain.

Polyribosomes

In both prokaryotic and eukaryotic cells, mRNA molecules are translated simultaneously by multiple ribosomes (Figure 15.23). The resulting structure—an mRNA with several ribosomes attached—is called a **polyribosome**. Each ribosome successively attaches to the ribosome-binding site at the 5' end of the mRNA and moves toward the 3' end; the polypeptide associated with each ribosome becomes progressively longer as the ribosome moves along the mRNA.

In prokaryotic cells, transcription and translation are simultaneous; so multiple ribosomes may be attached to the 5' end of the mRNA while transcription is still taking place at the 3' end, as shown in Figure 15.23.

CONCEPTS

In both prokaryotic and eukaryotic cells, multiple ribosomes may be attached to a single mRNA, generating a structure called a polyribosome.

✓ CONCEPT CHECK 9

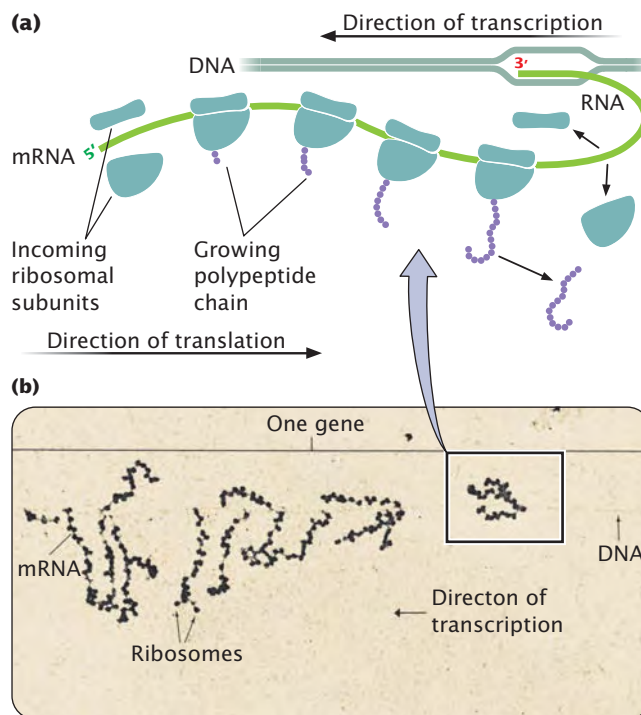
In a polyribosome, the polypeptides associated with which ribosomes will be the longest?

- Those at the 5' end of mRNA
- Those at the 3' end of mRNA
- Those in the middle of mRNA
- All polypeptides will be the same length.

Messenger RNA Surveillance

The accurate transfer of genetic information from one generation to the next and from genotype to phenotype is critical for the proper development and functioning of an organism. Consequently, cells have evolved a number of quality-control mechanisms to ensure the accuracy of information transfer. Protein synthesis is no exception: several mechanisms, collectively termed **mRNA surveillance**, exist to detect and deal with errors in mRNAs that may create problems in the course of translation.

Nonsense-mediated mRNA decay A common mutation is one in which a codon that specifies an amino acid is altered to become a termination codon (called a nonsense mutation, see Chapter 18). A nonsense mutation does not affect transcription, but translation ends prematurely



15.23 An mRNA molecule may be transcribed simultaneously by several ribosomes. (a) Four ribosomes are translating an mRNA molecule; the ribosomes move from the 5' end to the 3' end of the mRNA. (b) In this electron micrograph of a polyribosome, the dark-staining spheres are ribosomes, and the long, thin filament connecting the ribosomes is mRNA. The 5' end of the mRNA is toward the left-hand side of the micrograph. [Part b: O. L. Miller, Jr., and Barbara A. Hamaklo.]

when the termination codon is encountered. The resulting protein is truncated and often nonfunctional. A common way that nonsense mutations arise in eukaryotic cells is when one or more of the exons is skipped or improperly spliced. Improper splicing leads to the deletion or addition of nucleotides in the mRNA, which alters the reading frame and often introduces premature termination codons.

To prevent the synthesis of aberrant proteins resulting from nonsense mutations, eukaryotic cells have evolved a mechanism called **nonsense-mediated mRNA decay (NMD)**, which results in the rapid elimination of mRNA containing premature termination codons. The mechanism responsible for nonsense-mediated mRNA decay is still unclear. In mammals, it appears to entail proteins that bind to the exon–exon junctions. These exon-junction proteins may interact with enzymes that degrade the mRNA. One possibility is that the first ribosome to translate the mRNA removes the exon-junction proteins, thus protecting the mRNA from degradation. However, when the ribosome encounters a premature termination codon, the ribosome does not traverse the entire mRNA, and some of the exon-junction proteins are not removed, resulting in nonsense-mediated mRNA decay. **TRY PROBLEM 33** →

Stalled ribosomes and nonstop mRNAs A problem that occasionally arises in translation is when a ribosome stalls on the mRNA before translation is terminated. This situation can arise when a mutation in the DNA changes a termination codon into a codon that specifies an amino acid. It can also arise when transcription terminates prematurely, producing a truncated mRNA lacking a termination codon. In these cases, the ribosome reaches the end of the mRNA without encountering a termination codon and stalls, still attached to the mRNA. Attachment to the mRNA prevents the ribosome from being recycled for use on other mRNAs and, if such occurrences are frequent, the result is a shortage of ribosomes that diminishes overall levels of protein synthesis.

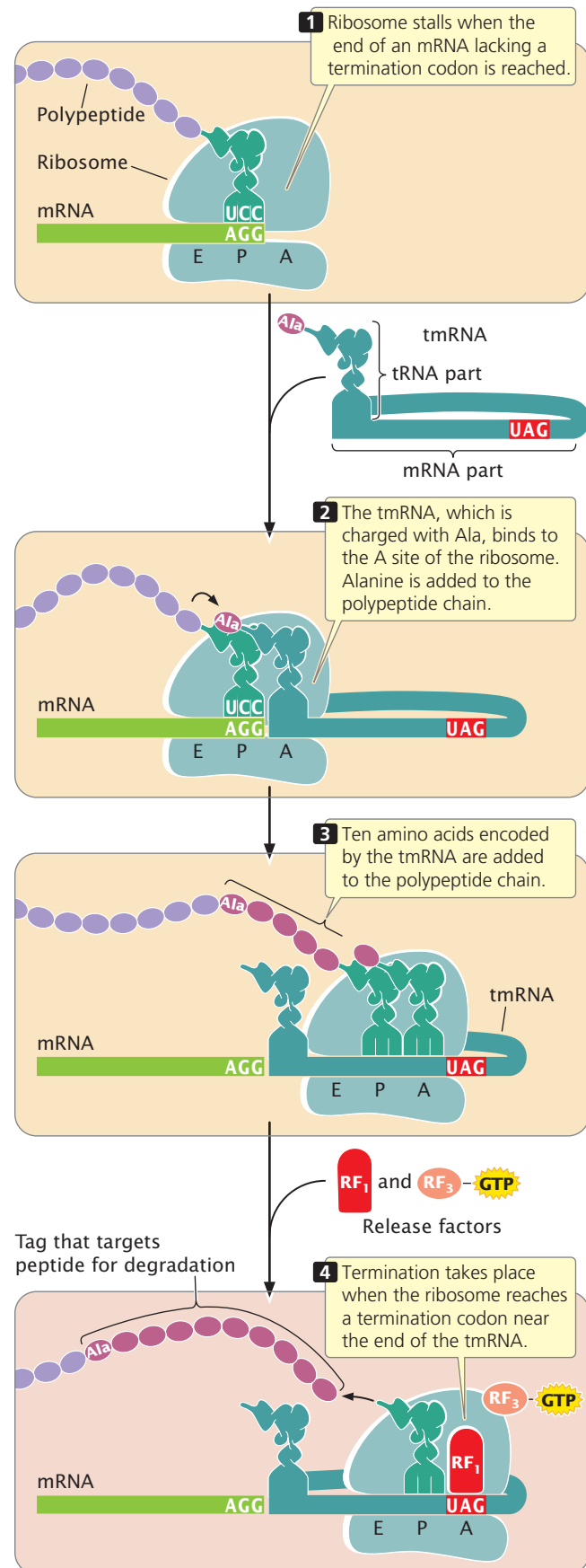
To deal with stalled ribosomes, bacteria have evolved a kind of molecular tow truck called **transfer-messenger RNA (tmRNA)**. This RNA molecule has properties of both tRNA and mRNA; its tRNA component is normally charged with the amino acid alanine. When a ribosome becomes stalled on an mRNA, EF-Tu delivers the tmRNA to the ribosome's A site, where tmRNA acts as surrogate tRNA (Figure 15.24). A peptide bond is created between the amino acid in the P site of the ribosome and alanine (attached to tmRNA) now in the A site, transferring the polypeptide chain to the tmRNA and releasing the tRNA in the P site.

The ribosome then resumes translation, switching from the original, aberrant mRNA to the mRNA part of tmRNA. Translation adds 10 amino acids encoded by the tmRNA, and then a termination codon is reached at the 3' end of the tmRNA, which terminates translation and releases the ribosome. The added amino acids are a special tag that targets the incomplete polypeptide chain for degradation. Some evidence suggests that the tmRNA also targets the aberrant mRNA for degradation. How stalled ribosomes are recognized by the tmRNA is not clear, but this method is efficient at recycling stalled ribosomes and eliminating abnormal proteins that result from truncated transcription.

Eukaryotes have evolved a different mechanism to deal with mRNAs that are missing termination codons. Instead of restarting the stalled ribosome and degrading the abnormal protein that results, eukaryotic cells use a mechanism called **nonstop mRNA decay**, which results in the rapid degradation of abnormal mRNA. In this mechanism, the codon-free A site of the stalled ribosome is recognized by a special protein that binds to the A site and recruits other proteins, which then degrade the mRNA from its 3' end.

CONCEPTS

Cells possess mRNA surveillance mechanisms to detect and eliminate mRNA molecules containing errors that create problems in the course of translation. These mechanisms include tmRNA, which rescues ribosomes stalled on the mRNA, nonstop mRNA decay, which degrades mRNAs lacking a termination codon, and nonsense-mediated mRNA decay, which eliminates mRNAs that possess a premature stop codon.



15.24 The tmRNA in bacteria allows stalled ribosomes to resume translation.

The Posttranslational Modifications of Proteins

After translation, proteins in both prokaryotic and eukaryotic cells may undergo alterations termed posttranslational modifications. A number of different types of modifications are possible. As mentioned earlier, the formyl group or the entire methionine residue may be removed from the amino end of a protein. Some proteins are synthesized as larger precursor proteins and must be cleaved and trimmed by enzymes before the proteins can become functional. For others, the attachment of carbohydrates may be required for activation. Amino acids within a protein may be modified: phosphates, carboxyl groups, and methyl groups are added to some amino acids. In eukaryotic cells, the amino end of a protein is often acetylated after translation.

The functions of many proteins critically depend on the proper folding of the polypeptide chain; some proteins spontaneously fold into their correct shapes, but, for others, correct folding may initially require the participation of other molecules called **molecular chaperones**. Some molecular chaperones are associated with the ribosome and fold newly synthesized polypeptide chains as they emerge from the ribosome tunnel, in which case protein folding takes place during ongoing translation.

A common posttranslational modification in eukaryotes is the attachment of a protein called ubiquitin, which targets the protein for degradation. Another modification of some proteins is the removal of 15 to 30 amino acids, called the **signal sequence**, at the amino end of the protein. The signal sequence helps direct a protein to a specific location within the cell, after which the sequence is removed by special enzymes.

CONCEPTS

Many proteins undergo posttranslational modifications after their synthesis.

Translation and Antibiotics

Antibiotics are drugs that kill microorganisms. To make an effective antibiotic, not just any poison will do: the trick is to kill the microbe without harming the patient. Antibiotics must be carefully chosen so that they destroy bacterial cells but not the eukaryotic cells of their host.

Translation is frequently the target of antibiotics because translation is essential to all living organisms and differs significantly between bacterial and eukaryotic cells. For

example, as already mentioned, bacterial and eukaryotic ribosomes differ in size and composition. A number of antibiotics bind selectively to bacterial ribosomes and inhibit various steps in translation, but they do not affect eukaryotic ribosomes. Tetracyclines, for instance, are a class of antibiotics that bind to the A site of a bacterial ribosome and block the entry of charged tRNAs, yet they have no effect on eukaryotic ribosomes. Neomycin binds to the ribosome near the A site and induces translational errors, probably by causing mistakes in the binding of charged tRNAs to the A site. Chloramphenicol binds to the large subunit of the ribosome and blocks peptide-bond formation. Streptomycin binds to the small subunit of the ribosome and inhibits initiation, and erythromycin blocks translocation. Although chloramphenicol and streptomycin are potent inhibitors of translation in bacteria, they do not inhibit translation in archaea.

The three-dimensional structure of puromycin resembles the 3' end of a charged tRNA, permitting puromycin to enter the A site of a ribosome efficiently and inhibit the entry of tRNAs. A peptide bond can form between the puromycin molecule in the A site and an amino acid on the tRNA in the P site of the ribosome, but puromycin cannot bind to the P site and translocation does not take place, blocking further elongation of the protein. Because tRNA structure is similar in all organisms, puromycin inhibits translation in both bacterial and eukaryotic cells; consequently, puromycin kills eukaryotic cells along with bacteria and is sometimes used in cancer therapy to destroy tumor cells.

Many antibiotics act by blocking specific steps in translation, and different antibiotics affect different steps in protein synthesis. Because of this specificity, antibiotics are frequently used to study the process of protein synthesis.

Nonstandard Protein Synthesis

The process of translation that we have considered thus far is part of the common genetic machinery used by all organisms. Many bacteria and fungi also possess an alternative protein-synthesis pathway, which they use to synthesize a few short, specialized proteins. Remarkably, this system does not rely on the ribosome; instead, it uses enzymes—some of which are the largest known—to assemble amino acids. Because this pathway does not rely on the ribosome, it is able to produce some proteins with unusual structures. For example, proteins produced by nonribosomal protein synthesis may contain some unusual molecular forms of amino acids and even some molecular relatives of amino acids. Because these proteins have unusual structures, they often resist degradation and may go undetected by pathogens.

CONCEPTS SUMMARY

- George Beadle and Edward Tatum developed the one gene, one enzyme hypothesis, which proposed that each gene specifies one enzyme; this hypothesis was later modified to become the one gene, one polypeptide hypothesis.
- Twenty different amino acids are used in the composition of proteins. The amino acids in a protein are linked together by peptide bonds. Chains of amino acids fold and associate to produce the secondary, tertiary, and quaternary structures of proteins.
- Solving the genetic code required several different approaches including the use of synthetic mRNAs with random sequences and short mRNAs that bind charged tRNAs.
- The genetic code is a triplet code: three nucleotides specify a single amino acid. It is also degenerate (meaning that more than one codon may specify an amino acid), nonoverlapping, and universal (almost).
- Different tRNAs (isoaccepting tRNAs) may accept the same amino acid. Different codons may pair with the same anticodon through wobble, which can exist at the third position of the codon and allows some nonstandard pairing of bases in this position.
- The reading frame is set by the initiation codon. The end of the protein-coding section of an mRNA is marked by one of three termination codons.
- Protein synthesis comprises four steps: (1) the binding of amino acids to the appropriate tRNAs, (2) initiation, (3) elongation, and (4) termination.
- The binding of an amino acid to a tRNA requires the presence of a specific aminoacyl-tRNA synthetase and ATP. The amino acid is attached by its carboxyl end to the 3' end of the tRNA.
- In bacterial translation initiation, the small subunit of the ribosome attaches to the mRNA and is positioned over the initiation codon. It is joined by the first tRNA and its associated amino acid (*N*-formylmethionine in bacterial cells) and, later, by the large subunit of the ribosome. Initiation requires several initiation factors and GTP.
- In elongation, a charged tRNA enters the A site of a ribosome, a peptide bond is formed between amino acids in the A and P sites, and the ribosome moves (translocates) along the mRNA to the next codon. Elongation requires several elongation factors and GTP.
- Translation is terminated when the ribosome encounters one of the three termination codons. Release factors and GTP are required to bring about termination.
- Each mRNA may be simultaneously translated by several ribosomes, producing a structure called a polyribosome.
- Cells possess RNA surveillance mechanisms that eliminate mRNAs with errors that may create problems in translation.
- Antibiotics frequently work by interfering with translation, because many aspects of translation differ in bacteria and eukaryotes.
- Many proteins undergo posttranslational modification. A few peptides are synthesized by nonribosomal pathways that utilize large enzymes.

IMPORTANT TERMS

one gene, one enzyme hypothesis (p. 405)	stop (termination or nonsense) codon (p. 412)	translocation (p. 416)
one gene, one polypeptide hypothesis (p. 405)	universal genetic code (p. 412)	elongation factor G (EF-G) (p. 417)
amino acid (p. 405)	aminoacyl-tRNA synthetase (p. 413)	release factor (RF ₁ , RF ₂ , RF ₃) (p. 417)
peptide bond (p. 405)	tRNA charging (p. 413)	polyribosome (p. 421)
polypeptide (p. 405)	initiation factor (IF-1, IF-2, IF-3) (pp. 414)	mRNA surveillance (p. 421)
sense codon (p. 410)	30S initiation complex (p. 414)	nonsense-mediated mRNA decay (NMD) (p. 421)
degenerate genetic code (p. 410)	70S initiation complex (p. 414)	transfer-messenger RNA (tmRNA) (p. 422)
synonymous codons (p. 410)	aminoacyl (A) site (p. 416)	nonstop mRNA decay (p. 422)
isoaccepting tRNAs (p. 410)	peptidyl (P) site (p. 416)	molecular chaperone (p. 423)
wobble (p. 410)	exit (E) site (p. 416)	signal sequence (p. 423)
nonoverlapping genetic code (p. 411)	elongation factor Tu (EF-Tu) (p. 416)	
reading frame (p. 411)	elongation factor Ts (EF-Ts) (p. 416)	
initiation codon (p. 411)		

ANSWERS TO CONCEPT CHECKS

1. B → A → C
2. The amino acid sequence (primary structure) of the protein
3. b
4. d
5. The initiation codon in bacteria encodes *N*-formylmethionine; in eukaryotes, it encodes methionine. Termination codons do not specify amino acids.
6. c
7. The Shine–Dalgarno sequence
8. a
9. b

WORKED PROBLEMS

1. A series of auxotrophic mutants were isolated in *Neurospora*. Examination of fungi containing these mutations revealed that they grew on minimal medium to which various compounds (A, B, C, D) were added; growth responses to each of the four compounds are presented in the following table. Give the order of compounds A, B, C, and D in a biochemical pathway. Outline a biochemical pathway that includes these four compounds and indicate which step in the pathway is affected by each of the mutations.

Mutation number	Compound			
	A	B	C	D
134	+	+	–	+
276	+	+	+	+
987	–	–	–	+
773	+	+	+	+
772	–	–	–	+
146	+	+	–	+
333	+	+	–	+
123	–	+	–	+

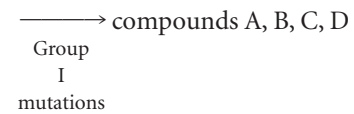
• **Solution**

To solve this problem, we should first group the mutations by which compounds allow growth, as follows:

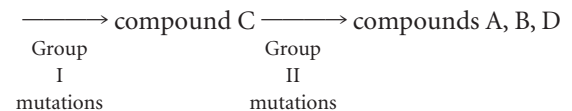
Mutation		Compound			
Group	Number	A	B	C	D
I	276	+	+	+	+
	773	+	+	+	+
II	134	+	+	–	+
	146	+	+	–	+
	333	+	+	–	+
III	123	–	+	–	+
IV	987	–	–	–	+
	772	–	–	–	+

The underlying principle used to determine the order of the compounds in the pathway is as follows: if a compound is added after the block, it will allow the mutant to grow, whereas, if a compound is added before the block, it will have no effect. Applying this principle to the data in the table, we see that mutants in group I will grow if compound A, B, C, or D is added

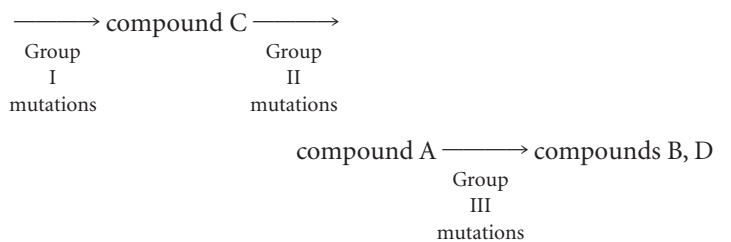
to the medium; so these mutations must affect a step before the production of all four compounds:



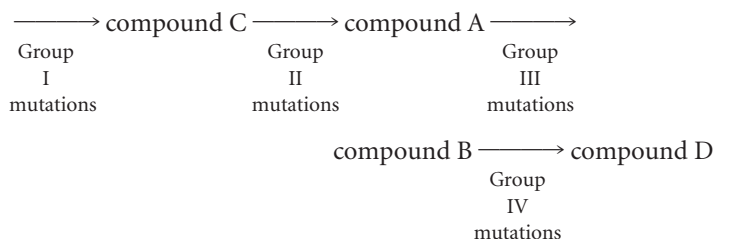
Group II mutants will grow if compound A, B, or D is added but not if compound C is added. Thus, compound C comes before A, B, and D; and group II mutations affect the conversion of compound C into one of the other compounds:



Group III mutants allow growth if compound B or D is added but not if compound A or C is added. Thus, group III mutations affect steps that follow the production of A and C; we have already determined that compound C precedes A in the pathway, and so A must be the next compound in the pathway:



Finally, mutants in group IV will grow if compound D is added but not if compound A, B, or C is added. Thus, compound D is the fourth compound in the pathway, and mutations in group IV block the conversion of B into D:



2. If there were five different types of bases in mRNA instead of four, what would be the minimum codon size (number of nucleotides) required to specify the following numbers of different amino acid types: (a) 4, (b) 20, (c) 30?

• **Solution**

To answer this question, we must determine the number of combinations (codons) possible when there are different numbers of bases and different codon lengths. In general, the number of different codons possible will be equal to:

$$b^{lg} = \text{number of codons}$$

where b equals the number of different types of bases and lg equals the number of nucleotides in each codon (codon length). If there are five different types of bases, then:

$$5^1 = 5 \text{ possible codons}$$

$$5^2 = 25 \text{ possible codons}$$

$$5^3 = 125 \text{ possible codons}$$

The number of possible codons must be greater than or equal to the number of amino acids specified. Therefore, a codon length of one nucleotide could specify 4 different amino acids, a codon length of two nucleotides could specify 20 different amino acids, and a codon length of three nucleotides could specify 30 different amino acids: (a) one, (b) two, (c) three.

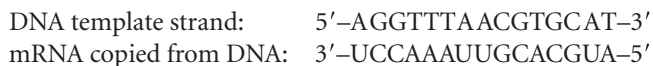
3. A template strand in bacterial DNA has the following base sequence:



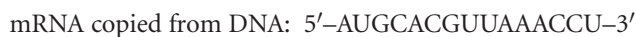
What amino acids are encoded by this sequence?

• **Solution**

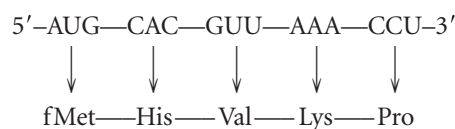
To answer this question, we must first work out the mRNA sequence that will be transcribed from this DNA sequence. The mRNA must be antiparallel and complementary to the DNA template strand:



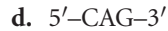
An mRNA is translated $5' \rightarrow 3'$; so it will be helpful if we turn the RNA molecule around with the 5' end on the left:



The codons consist of groups of three nucleotides that are read successively after the first AUG codon; using Figure 15.10, we can determine that the amino acids are

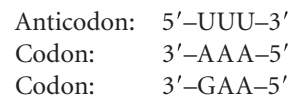


4. The following triplets constitute anticodons found on a series of tRNAs. Name the amino acid carried by each of these tRNAs.

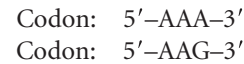


• **Solution**

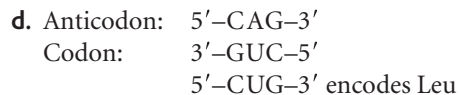
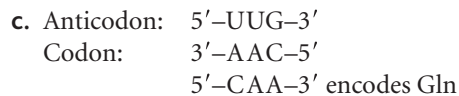
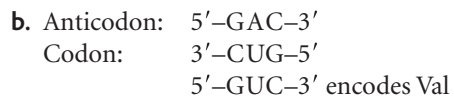
To solve this problem, we first determine the codons with which these anticodons pair and then look up the amino acid specified by the codon in Figure 15.10. The codons are antiparallel and complementary to the anticodons. For part *a*, the anticodon is $5'\text{-UUU-}3'$. According to the wobble rules in Table 15.2, U in the first position of the anticodon can pair with either A or G in the third position of the codon, so there are two codons that can pair with this anticodon:



Listing these codons in the conventional manner, with the 5' end on the right, we have:



According to Figure 15.10, both codons specify the amino acid lysine (Lys). Recall that the wobble in the third position allows more than one codon to specify the same amino acid; so any wobble that exists should produce the same amino acid as the standard base pairings would, and we do not need to figure the wobble to answer this question. The answers for parts *b*, *c*, and *d* are:



Strain with mutation	Minimal medium to which is added		
	Compound I	Compound II	Compound III
a			
b			
c			

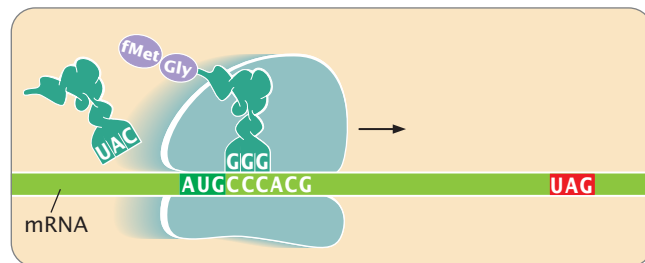
Section 15.2

- *18. Assume that the number of different types of bases in RNA is four. What would be the minimum codon size (number of nucleotides) required if the number of different types of amino acids in proteins were: (a) 2, (b) 8, (c) 17, (d) 45, (e) 75?
19. How many codons would be possible in a triplet code if only three bases (A, C, and U) were used?
- *20. Using the genetic code presented in Figure 15.10, give the amino acids specified by the following bacterial mRNA sequences, and indicate the amino and carboxyl ends of the polypeptide produced.
- 5'-AUGUUUAAAUUUAAAUUUUGA-3'
 - 5'-AUGUAUAUAUAUAUGA-3'
 - 5'-AUGGAUGAAAGAUUUCUCGCUUGA-3'
 - 5'-AUGGGUUAGGGGACAUCAUUUUGA-3'
21. A nontemplate strand on bacterial DNA has the following base sequence. What amino acid sequence will be encoded by this sequence?
- 5'-ATGATACTAAGGCC-3'
- *22. The following amino acid sequence is found in a tripeptide: Met-Trp-His. Give all possible nucleotide sequences on the mRNA, on the template strand of DNA, and on the nontemplate strand of DNA that can encode this tripeptide.
23. How many different mRNA sequences can encode a polypeptide chain with the amino acid sequence Met-Leu-Arg? (Be sure to include the stop codon.)
- *24. A series of tRNAs have the following anticodons. Consider the wobble rules listed in Table 15.2 and give all possible codons with which each tRNA can pair.
- 5'-GGC-3'
 - 5'-AAG-3'
 - 5'-IAA-3'
 - 5'-UGG-3'
 - 5'-CAG-3'

25. An anticodon on a tRNA has the sequence 5'-GCA-3'.
- What amino acid is carried by this tRNA?
 - What would be the effect if the G in the anticodon were mutated to a U?
26. Which of the following amino acid changes could result from a mutation that changed a single base? For each change that could result from the alteration of a single base, determine which position of the codon (first, second, or third nucleotide) in the mRNA must be altered for the change to result.
- Leu → Gln
 - Phe → Ser
 - Phe → Ile
 - Pro → Ala
 - Asn → Lys
 - Ile → Asn

Section 15.3

27. Arrange the following components of translation in the approximate order in which they would appear or be used in protein synthesis:
- 70S initiation complex
 - 30S initiation complex
 - release factor 1
 - elongation factor G
 - initiation factor 3
 - elongation factor Tu
 - fMet-tRNA^{fMet}
28. The following diagram illustrates a step in the process of translation. Sketch the diagram and identify the following elements on it.



- 5' and 3' ends of the mRNA
- A, P, and E sites
- Start codon
- Stop codon

- e. Amino and carboxyl ends of the newly synthesized polypeptide chain
 - f. Approximate location of the next peptide bond that will be formed
 - g. Place on the ribosome where release factor 1 will bind
29. Refer to the diagram in Problem 28 to answer the following questions.
- a. What will be the anticodon of the next tRNA added to the A site of the ribosome?
 - b. What will be the next amino acid added to the growing polypeptide chain?
- *30. A synthetic mRNA added to a cell-free protein-synthesizing system produces a peptide with the following amino acid sequence: Met-Pro-Ile-Ser-Ala. What would be the effect on translation if the following components were omitted from the cell-free protein-synthesizing system? What, if any, type of protein would be produced? Explain your reasoning.
- a. Initiation factor 1
 - b. Initiation factor 2
 - c. Elongation factor Tu
 - d. Elongation factor G
 - e. Release factors RF₁, RF₂, and RF₃
 - f. ATP
 - g. GTP
31. For each of the following sequences, place a check mark in the appropriate space to indicate the process *most immediately* affected by deleting the sequence. Choose only one process for each sequence (i.e., one check mark per sequence).

Sequence deleted	Process most immediately affected by deletion			
	Replication	Transcription	RNA processing	Translation
a. <i>ori</i> site	_____	_____	_____	_____
b. 3' splice-site consensus	_____	_____	_____	_____
c. poly(A) tail	_____	_____	_____	_____
d. terminator	_____	_____	_____	_____
e. start codon	_____	_____	_____	_____
f. -10 consensus	_____	_____	_____	_____
g. Shine-Dalgarno	_____	_____	_____	_____

32. MicroRNAs are small RNA molecules that bind to the 3' end of mRNAs and suppress translation (see Chapter 14). How miRNAs suppress translation is still being investigated. Some eukaryotic mRNAs have internal ribosome-binding sites downstream of the 5' cap, where ribosomes normally bind. In one investigation, miRNAs did not suppress the translation of ribosomes that attach to internal ribosome-binding sites (R. S. Pillai et al. 2005. *Science* 309:1573–1576). What does this finding suggest about how miRNAs suppress translation?

Section 15.4

33. Mutations that introduce stop codons cause a number of genetic diseases. For example, from 2% to 5% of the people who have cystic fibrosis possess a mutation that causes a premature stop codon in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR). This premature stop codon produces a truncated form of CFTR that is nonfunctional and results in the symptoms of cystic fibrosis. One possible way to treat people with genetic diseases caused by these types of mutations is to trick the ribosome into reading through the stop codon, inserting an amino acid into its place. Although the protein produced may have one altered amino acid, it is more likely to be at least partly functional than is the truncated protein produced when the ribosome stalls at the stop codon. Indeed, geneticists have conducted clinical trials on people with cystic fibrosis with the use of a drug called PTC124, which interferes with the ribosome's ability to correctly read stop codons (C. Ainsworth. 2005. *Nature* 438:726–728). On the basis of what you know about the mechanism of nonsense-mediated mRNA decay (NMD), would you expect NMD to be a problem with this type of treatment? Why or why not?



Child with cystic fibrosis [Lisa Eastman/Alamy.]

CHALLENGE QUESTIONS

Section 15.2

34. The redundancy of the genetic code means that some amino acids are specified by more than one codon. For example, the amino acid leucine is encoded by six different codons. Within a genome, synonymous codons are not present in equal numbers; some synonymous codons appear much more frequently than others, and the preferred codons differ among different species. For example, in one species, the codon UUA might be used most often to encode leucine, whereas, in another species, the codon CUU might be used most often. Speculate on a reason for this bias in codon usage and why the preferred codons are not the same in all organisms.

Section 15.3

35. In what ways are spliceosomes and ribosomes similar? In what ways are they different? Can you suggest some possible reasons for their similarities.
- *36. Several experiments were conducted to obtain information about how the eukaryotic ribosome recognizes the AUG start codon. In one experiment, the gene that encodes methionine initiator tRNA ($\text{tRNA}_i^{\text{Met}}$) was located and changed. The nucleotides that specify the anticodon on $\text{tRNA}_i^{\text{Met}}$ were mutated so that the anticodon in the tRNA was 5'-CCA-3' instead of 5'-CAU-3'. When this mutated gene was placed in a eukaryotic cell, protein synthesis took place but the proteins produced were abnormal. Some of the proteins produced contained extra amino acids, and others contained fewer amino acids than normal.
- What do these results indicate about how the ribosome recognizes the starting point for translation in eukaryotic cells? Explain your reasoning.
 - If the same experiment had been conducted on bacterial cells, what results would you expect?

16 Control of Gene Expression in Prokaryotes



The bacterium *Streptococcus pneumoniae* is responsible each year for millions of cases of pneumonia, inner-ear infection, sinus infection, and meningitis. Gene regulation is required for transformation, a process by which many strains of *S. pneumoniae* have become resistant to antibiotics. [Dr. Gary Gaugler/Photo Researchers.]

STRESS, SEX, AND GENE REGULATION IN BACTERIA

Each year, *Streptococcus pneumoniae*, commonly known as pneumococcus, is responsible for millions of cases of pneumonia, inner-ear infection, sinus infection, and meningitis. Before the widespread use of antibiotics, pneumonia resulting from *S. pneumoniae* infection was a leading cause of death, particularly among the young and the elderly. The development of penicillin and other antibiotics, beginning in the 1940s, provided an effective tool against *Streptococcus* infections, and deaths from pneumonia and other life-threatening bacterial infections plummeted. In recent years, however, many strains of *S. pneumoniae* and other pathogenic bacteria have developed resistance to penicillin and other antibiotics. Some bacterial strains have rapidly acquired resistance to multiple antibiotics, making treatment of *S. pneumoniae* difficult. How has antibiotic resistance spread so rapidly among bacteria?

Although bacteria are simple organisms, they engage in complex and ingenious mechanisms to exchange genetic information and rapidly adapt to changing conditions. One such mechanism is transformation, first demonstrated by Fred Griffith in 1928, in which bacteria take up new DNA released by dead cells and integrate the foreign DNA into their own genomes (see pp. 274–275 in Chapter 10). Transformation is considered a type of sexuality in bacteria because it results in the exchange of genetic material between individual cells. It is widespread among bacteria and has been a major means by which antibiotic resistance has spread.

In *S. pneumoniae*, the ability to carry out transformation—called competence—requires from 105 to 124 genes, collectively termed the *com* (for competence) regulon. Genes within the *com* regulon facilitate the passage of DNA through the cell membrane and the DNA's recombination with homologous genes in the bacterial chromosome. For competence to take place, genes in the *com* regulon must be precisely turned on and off in a carefully regulated sequence.

The *com* regulon is activated in response to a protein called competence-stimulating peptide (CSP), which is produced by bacteria and is exported into the surrounding medium. When enough CSP accumulates, it attaches to a receptor on the bacterial cell membrane, which then activates a regulator protein that stimulates the transcription of genes within the *com* regulon and sets in motion a complex cascade of reactions that ultimately results in the development of competence.

Early steps in the development of competence are poorly understood, but most experts formerly assumed that CSP is continually produced and slowly accumulates in the medium; the assumption was that, when CSP reaches a critical threshold, transformation is induced. However, recent research has altered the view that the induction of transformation is a passive process. This work has demonstrated that the production of CSP is itself regulated. When *S. pneumoniae* is subjected to stress, it produces more CSP, which causes transformation and allows the bacteria to genetically adapt to the stress by acquiring new genetic material from the environment. In fact, the presence of antibiotics is one type of stress that stimulates the production of CSP and induces transformation. This reaction to stress is probably a primary reason that bacteria such as *S. pneumoniae* have so quickly evolved resistance to antibiotics. **TRY PROBLEM 10** →

The regulation of gene expression not only is important for the development of transformation, but is critical for the control of numerous life processes in all organisms. This chapter is the first of two chapters about **gene regulation**, the mechanisms and systems that control the expression of genes. In this chapter, we consider systems of gene regulation in bacteria. We begin by considering the necessity for gene regulation, the levels at which gene expression is controlled, and the difference between genes and regulatory elements. We then examine the structure and function of operons, groups of genes that are regulated together in bacteria, and consider gene regulation in some specific examples of operons. Finally, we study several types of bacterial gene regulation that are facilitated by RNA molecules.

16.1 The Regulation of Gene Expression Is Critical for All Organisms

A major theme of molecular genetics is the central dogma, which states that genetic information flows from DNA to RNA to proteins (see Figure 10.16). Although the central dogma provided a molecular basis for the connection between genotype and phenotype, it failed to address a critical question: How is the flow of information along the molecular pathway *regulated*?

Consider *Escherichia coli*, a bacterium that resides in your large intestine. Your eating habits completely determine the nutrients available to this bacterium: it can neither seek out nourishment when nutrients are scarce nor move away when confronted with an unfavorable environment. *E. coli* makes up for its inability to alter the external environment by being internally flexible. For example, if glucose is present, *E. coli* uses it to generate ATP; if there's no glucose, it utilizes lactose, arabinose, maltose, xylose, or any of a number of other sugars. When amino acids are available, *E. coli* uses them to synthesize proteins; if a particular amino acid is absent, *E. coli* produces the enzymes needed to synthesize that amino acid. Thus, *E. coli* responds to environmental

changes by rapidly altering its biochemistry. This biochemical flexibility, however, has a high price. Producing all the enzymes necessary for every environmental condition would be energetically expensive. So how does *E. coli* maintain biochemical flexibility while optimizing energy efficiency?

The answer is through gene regulation. Bacteria carry the genetic information for synthesizing many proteins, but only a subset of this genetic information is expressed at any time. When the environment changes, new genes are expressed, and proteins appropriate for the new environment are synthesized. For example, if a carbon source appears in the environment, genes encoding enzymes that take up and metabolize this carbon source are quickly transcribed and translated. When this carbon source disappears, the genes that encode these enzymes are shut off.

Multicellular eukaryotic organisms face a different dilemma. Individual cells in a multicellular organism are specialized for particular tasks. The proteins produced by a nerve cell, for example, are quite different from those produced by a white blood cell. The problem that a eukaryotic cell faces is how to specialize. Although they are quite different in shape and function, a nerve cell and a blood cell still carry the same genetic instructions.

A multicellular organism's challenge is to bring about the specialization of cells that have a common set of genetic instructions (the process of development). This challenge is met through gene regulation: all of an organism's cells carry the same genetic information, but only a subset of genes are expressed in each cell type. Genes needed for other cell types are not expressed. Gene regulation is therefore the key to both unicellular flexibility and multicellular specialization, and it is critical to the success of all living organisms.

CONCEPTS

In bacteria, gene regulation maintains internal flexibility, turning genes on and off in response to environmental changes. In multicellular eukaryotic organisms, gene regulation brings about cellular differentiation.

The mechanisms of gene regulation were first investigated in bacterial cells, in which the availability of mutants and the ease of laboratory manipulation made it possible to unravel the mechanisms. When the study of these mechanisms in eukaryotic cells began, bacterial gene regulation clearly seemed to differ from eukaryotic gene regulation. As more and more information has accumulated about gene regulation, however, a number of common themes have emerged, and, today, many aspects of gene regulation in bacterial and eukaryotic cells are recognized to be similar. Before examining specific elements of bacterial gene regulation (this chapter) and eukaryotic gene regulation (see Chapter 17), we will briefly consider some themes of gene regulation common to all organisms.

Genes and Regulatory Elements

In considering gene regulation in both bacteria and eukaryotes, we must distinguish between the DNA sequences that are transcribed and the DNA sequences that regulate the expression of other sequences. As discussed in Chapter 14, one definition of a *gene* is any DNA sequence that is transcribed into an RNA molecule. Genes include DNA sequences that encode proteins, as well as sequences that encode rRNA, tRNA, snRNA, and other types of RNA. **Structural genes** encode proteins that are used in metabolism or biosynthesis or that play a structural role in the cell. **Regulatory genes** are genes whose products, either RNA or proteins, interact with other DNA sequences and affect the transcription or translation of those sequences. In many cases, the products of regulatory genes are DNA-binding proteins. Bacteria and eukaryotes use regulatory genes to control the expression of many of their structural genes. However, a few structural genes, particularly those that encode essential cellular functions, are expressed continually and are said to be **constitutive**. Constitutive genes are therefore not regulated.

We will also encounter DNA sequences that are not transcribed at all but still play a role in regulating genes and other DNA sequences. These **regulatory elements** affect the expression of sequences to which they are physically linked. Regulatory elements are common in both bacterial and eukaryotic cells, and much of gene regulation in both types of organisms takes place through the action of proteins produced by regulatory genes that recognize and bind to regulatory elements.

The regulation of gene expression can be through processes that stimulate gene expression, termed *positive control*, or through processes that inhibit gene expression, termed *negative control*. Bacteria and eukaryotes use both positive and negative control mechanisms to regulate their genes. However, negative control is more important in bacteria, whereas eukaryotes are more likely to use positive control mechanisms.

CONCEPTS

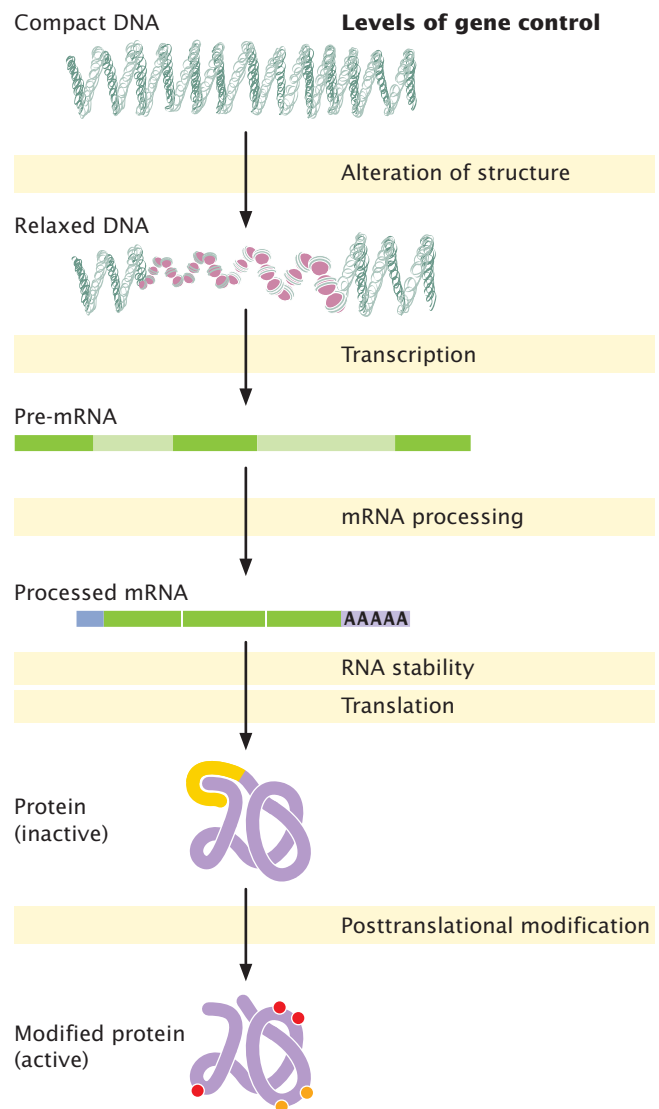
Genes are DNA sequences that are transcribed into RNA. Regulatory elements are DNA sequences that are not transcribed but affect the expression of genes. Positive control includes mechanisms that stimulate gene expression, whereas negative control inhibits gene expression.

✓ CONCEPT CHECK 1

What is a constitutive gene?

Levels of Gene Regulation

In both bacteria and eukaryotes, genes can be regulated at a number of points along the pathway of information flow from genotype to phenotype (Figure 16.1). First, regulation



16.1 Gene expression can be controlled at multiple levels.

can be through the alteration of DNA or chromatin structure; this type of gene regulation takes place primarily in eukaryotes. Modifications to DNA or its packaging can help to determine which sequences are available for transcription or the rate at which sequences are transcribed. DNA methylation and changes in chromatin are two processes that play a pivotal role in gene regulation.

A second point at which a gene can be regulated is at the level of transcription. For the sake of cellular economy, limiting the production of a protein early in the process makes sense, and transcription is an important point of gene regulation in both bacterial and eukaryotic cells. A third potential point of gene regulation is mRNA processing. Eukaryotic mRNA is extensively modified before it is translated: a 5' cap is added, the 3' end is cleaved and polyadenylated, and introns are removed (see Chapter 14). These modifications determine the stability of the mRNA, whether the mRNA can be translated, the rate of translation, and the amino acid sequence of the protein produced. There is growing evidence that a number of regulatory mechanisms in eukaryotic cells operate at the level of mRNA processing.

A fourth point for the control of gene expression is the regulation of RNA stability. The amount of protein produced depends not only on the amount of mRNA synthesized, but also on the rate at which the mRNA is degraded. A fifth point of gene regulation is at the level of translation, a complex process requiring a large number of enzymes, protein factors, and RNA molecules (see Chapter 15). All of these factors, as well as the availability of amino acids, affect the rate at which proteins are produced and therefore provide points at which gene expression can be controlled. Translation can also be affected by sequences in mRNA.

Finally, many proteins are modified after translation (see Chapter 15), and these modifications affect whether the proteins become active; so genes can be regulated through processes that affect posttranslational modification. Gene expression can be affected by regulatory activities at any or all of these points.

CONCEPTS

Gene expression can be controlled at any of a number of points along the molecular pathway from DNA to protein, including DNA or chromatin structure, transcription, mRNA processing, RNA stability, translation, and posttranslational modification.

✓ CONCEPT CHECK 2

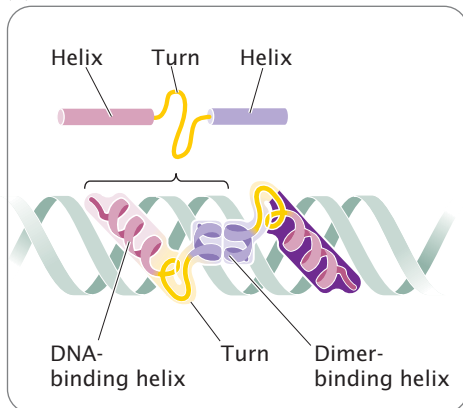
Why is transcription a particularly important level of gene regulation in both bacteria and eukaryotes?

DNA-Binding Proteins

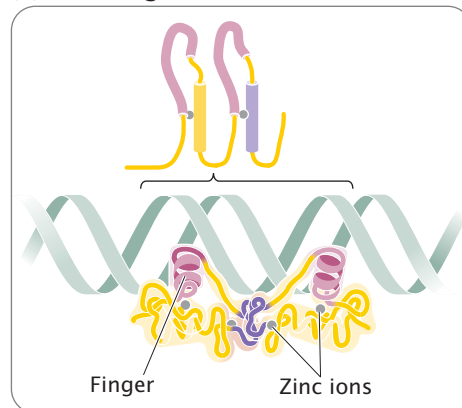
Much of gene regulation in bacteria and eukaryotes is accomplished by proteins that bind to DNA sequences and affect their expression. These regulatory proteins generally have discrete functional parts—called **domains**, typically consisting of 60 to 90 amino acids—that are responsible for binding to DNA. Within a domain, only a few amino acids actually make contact with the DNA. These amino acids (most commonly asparagine, glutamine, glycine, lysine, and arginine) often form hydrogen bonds with the bases or interact with the sugar-phosphate backbone of the DNA. Many regulatory proteins have additional domains that can bind other molecules such as other regulatory proteins. By physically attaching to DNA, these proteins can affect the expression of a gene. Most DNA-binding proteins bind dynamically, which means that they are transiently binding and unbinding DNA and other regulatory proteins. Thus, although they may spend most of their time bound to DNA, they are never permanently attached. This dynamic nature means that other molecules can compete with DNA-binding proteins for regulatory sites on the DNA.

DNA-binding proteins can be grouped into several distinct types on the basis of a characteristic structure, called a motif, found within the binding domain. Motifs are simple structures, such as alpha helices, that can fit into the major groove of the DNA. For example, the helix-turn-helix motif (Figure 16.2a), consisting of two alpha helices connected

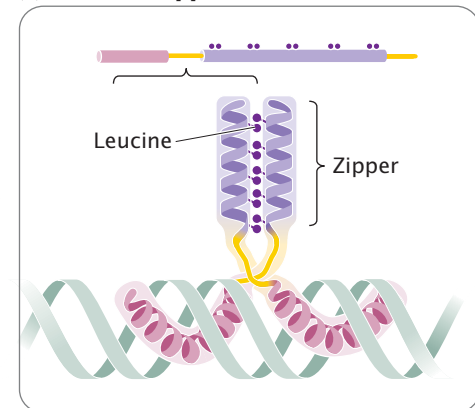
(a) Helix-turn-helix



(b) Zinc fingers



(c) Leucine zipper



16.2 DNA-binding proteins can be grouped into several types on the basis of their structures, or motifs.

Table 16.1 Common DNA-binding motifs

Motif	Location	Characteristics	Binding Site in DNA
Helix-turn-helix	Bacterial regulatory proteins; related motifs in eukaryotic proteins	Two alpha helices	Major groove
Zinc-finger	Eukaryotic regulatory and other proteins	Loop of amino acids with zinc at base	Major groove
Steroid receptor	Eukaryotic proteins	Two perpendicular alpha helices with zinc surrounded by four cysteine residues	Major groove and DNA backbone
Leucine-zipper	Eukaryotic transcription factors	Helix of leucine residues and a basic arm; two leucine residues interdigitate	Two adjacent major grooves
Helix-loop-helix	Eukaryotic proteins	Two alpha helices separated by a loop of amino acids	Major groove
Homeodomain	Eukaryotic regulatory proteins	Three alpha helices	Major groove

by a turn, is common in bacterial regulatory proteins. The zinc-finger motif (**Figure 16.2b**), common to many eukaryotic regulatory proteins, consists of a loop of amino acids containing a zinc ion. The leucine zipper (**Figure 16.2c**) is another motif found in a variety of eukaryotic binding proteins. These common DNA-binding motifs and others are summarized in **Table 16.1**.

CONCEPTS

Regulatory proteins that bind DNA have common motifs that interact with sequences in the DNA.

✓ CONCEPT CHECK 3

How do amino acids in DNA-binding proteins interact with DNA?

- By forming covalent bonds with DNA bases
- By forming hydrogen bonds with DNA bases
- By forming covalent bonds with DNA sugars

structural genes are transcribed into a single mRNA, which is translated to produce enzymes A, B, and C. These enzymes carry out a series of biochemical reactions that convert precursor molecule X into product Y. The transcription of structural genes *a*, *b*, and *c* is under the control of a promoter, which lies upstream of the first structural gene. RNA polymerase binds to the promoter and then moves downstream, transcribing the structural genes.

A **regulator gene** helps to control the transcription of the structural genes of the operon. Although it affects operon function, the regulator gene is not considered part of the operon. The regulator gene has its own promoter and is transcribed into a short mRNA, which is translated into a small protein. This **regulator protein** can bind to a region of the operon called the **operator** and affect whether transcription can take place. The operator usually overlaps the 3' end of the promoter and sometimes the 5' end of the first structural gene (see **Figure 16.3**).

16.2 Operons Control Transcription in Bacterial Cells

A significant difference between bacterial and eukaryotic gene control lies in the organization of functionally related genes. Many bacterial genes that have related functions are clustered and under the control of a single promoter. These genes are often transcribed together into a single mRNA. A group of bacterial structural genes that are transcribed together (along with their promoter and additional sequences that control transcription) is called an **operon**. The operon regulates the expression of the structural genes by controlling transcription, which, in bacteria, is usually the most important level of gene regulation.

Operon Structure

The organization of a typical operon is illustrated in **Figure 16.3**. At one end of the operon is a set of structural genes, shown in **Figure 16.3** as gene *a*, gene *b*, and gene *c*. These

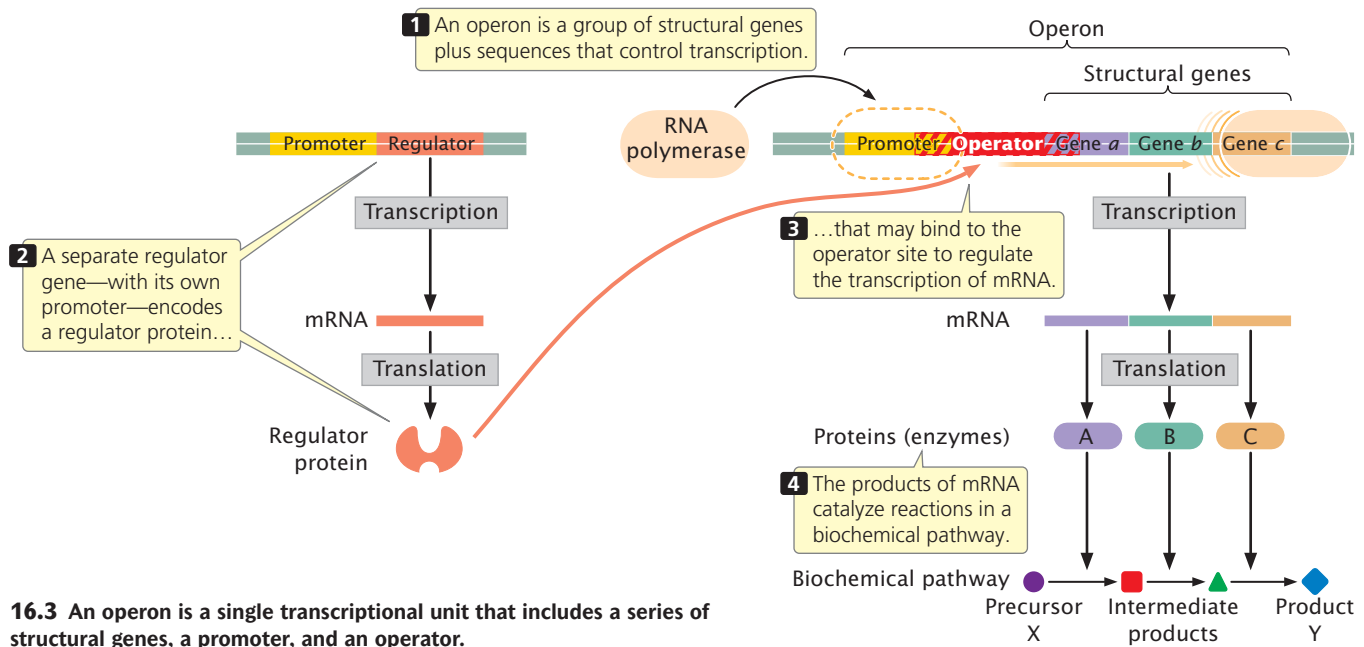
CONCEPTS

Functionally related genes in bacterial cells are frequently clustered together as a single transcriptional unit termed an operon. A typical operon includes several structural genes, a promoter for the structural genes, and an operator site to which the product of a regulator gene binds.

✓ CONCEPT CHECK 4

What is the difference between a structural gene and a regulator gene?

- Structural genes are transcribed into mRNA, but regulator genes aren't.
- Structural genes have complex structures; regulator genes have simple structures.
- Structural genes encode proteins that function in the structure of the cell; regulator genes carry out metabolic reactions.
- Structural genes encode proteins; regulator genes control the transcription of structural genes.



16.3 An operon is a single transcriptional unit that includes a series of structural genes, a promoter, and an operator.

Negative and Positive Control: Inducible and Repressible Operons

There are two types of transcriptional control: **negative control**, in which a regulatory protein is a repressor, binding to DNA and inhibiting transcription; and **positive control**, in which a regulatory protein is an activator, stimulating transcription. Operons can also be either inducible or repressible. **Inducible operons** are those in which transcription is normally off (not taking place); something must happen to induce transcription, or turn it on. **Repressible operons** are those in which transcription is normally on (taking place); something must happen to repress transcription, or turn it off. In the next sections, we will consider several varieties of these basic control mechanisms.

Negative inducible operons In a negative inducible operon, the regulator gene encodes an active *repressor* that readily binds to the operator (**Figure 16.4a**). Because the operator site overlaps the promoter site, the binding of this protein to the operator physically blocks the binding of RNA polymerase to the promoter and prevents transcription. For transcription to take place, something must happen to prevent the binding of the repressor at the operator site. This type of system is said to be *inducible*, because transcription is normally off (inhibited) and must be turned on (induced).

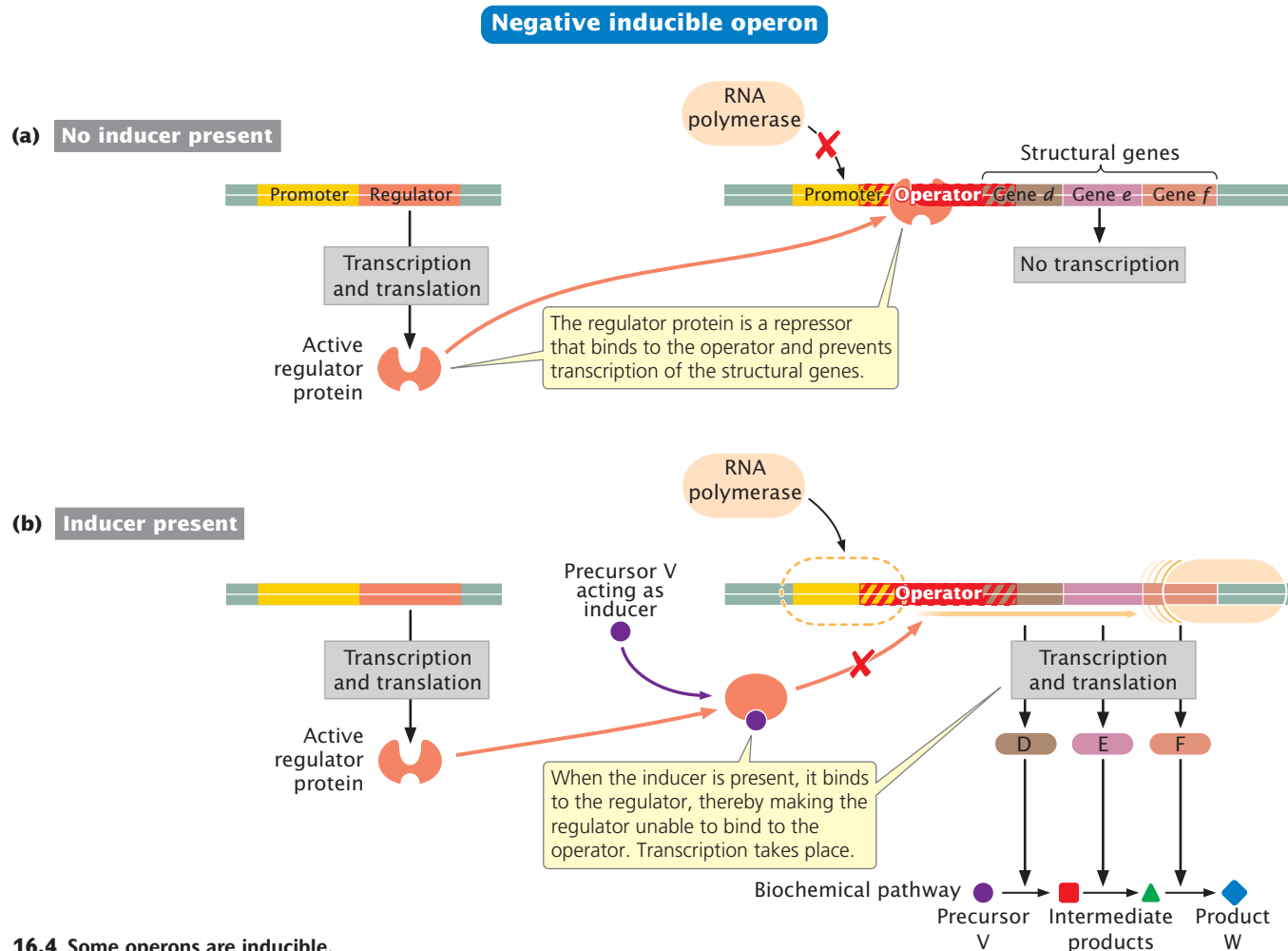
Transcription is turned on when a small molecule, an **inducer**, binds to the repressor (**Figure 16.4b**). Regulatory proteins frequently have two binding sites: one that binds to DNA and another that binds to a small molecule such as an inducer. The binding of the inducer (precursor V in Figure

16.4b) alters the shape of the repressor, preventing it from binding to DNA. Proteins of this type, which change shape on binding to another molecule, are called **allosteric proteins**.

When the inducer is absent, the repressor binds to the operator, the structural genes are not transcribed, and enzymes D, E, and F (which metabolize precursor V) are not synthesized (see **Figure 16.4a**). This mechanism is an adaptive one: because no precursor V is available, synthesis of the enzymes would be wasteful when they have no substrate to metabolize. As soon as precursor V becomes available, some of it binds to the repressor, rendering the repressor inactive and unable to bind to the operator site. RNA polymerase can now bind to the promoter and transcribe the structural genes. The resulting mRNA is then translated into enzymes D, E, and F, which convert substrate V into product W (see **Figure 16.4b**). So, an operon with negative inducible control regulates the synthesis of the enzymes economically: the enzymes are synthesized only when their substrate (V) is available.

Inducible operons usually control proteins that carry out degradative processes—proteins that break down molecules. For these types of proteins, inducible control makes sense because the proteins are not needed unless the substrate (which is broken down by the proteins) is present.

Negative repressible operons Some operons with negative control are *repressible*, meaning that transcription *normally* takes place and must be turned off, or repressed. The regulator protein in this type of operon also is a repressor but is synthesized in an *inactive* form that cannot by itself bind to the operator. Because no repressor is bound to the operator,



16.4 Some operons are inducible.

RNA polymerase readily binds to the promoter and transcription of the structural genes takes place (**Figure 16.5a**).

To turn transcription off, something must happen to make the repressor active. A small molecule called a **corepressor** binds to the repressor and makes it capable of binding to the operator. In the example illustrated (see **Figure 16.5a**), the product (U) of the metabolic reaction is the corepressor. As long as the level of product U is high, it is available to bind to the repressor and activate it, preventing transcription (**Figure 16.5b**). With the operon repressed, enzymes G, H, and I are not synthesized, and no more U is produced from precursor T. However, when all of product U is used up, the repressor is no longer activated by U and cannot bind to the operator. The inactivation of the repressor allows the transcription of the structural genes and the synthesis of enzymes G, H, and I, resulting in the conversion of precursor T into product U. Like inducible operons, repressible operons are economical: the enzymes are synthesized only as needed.

Repressible operons usually control proteins that carry out the biosynthesis of molecules needed in the cell, such as

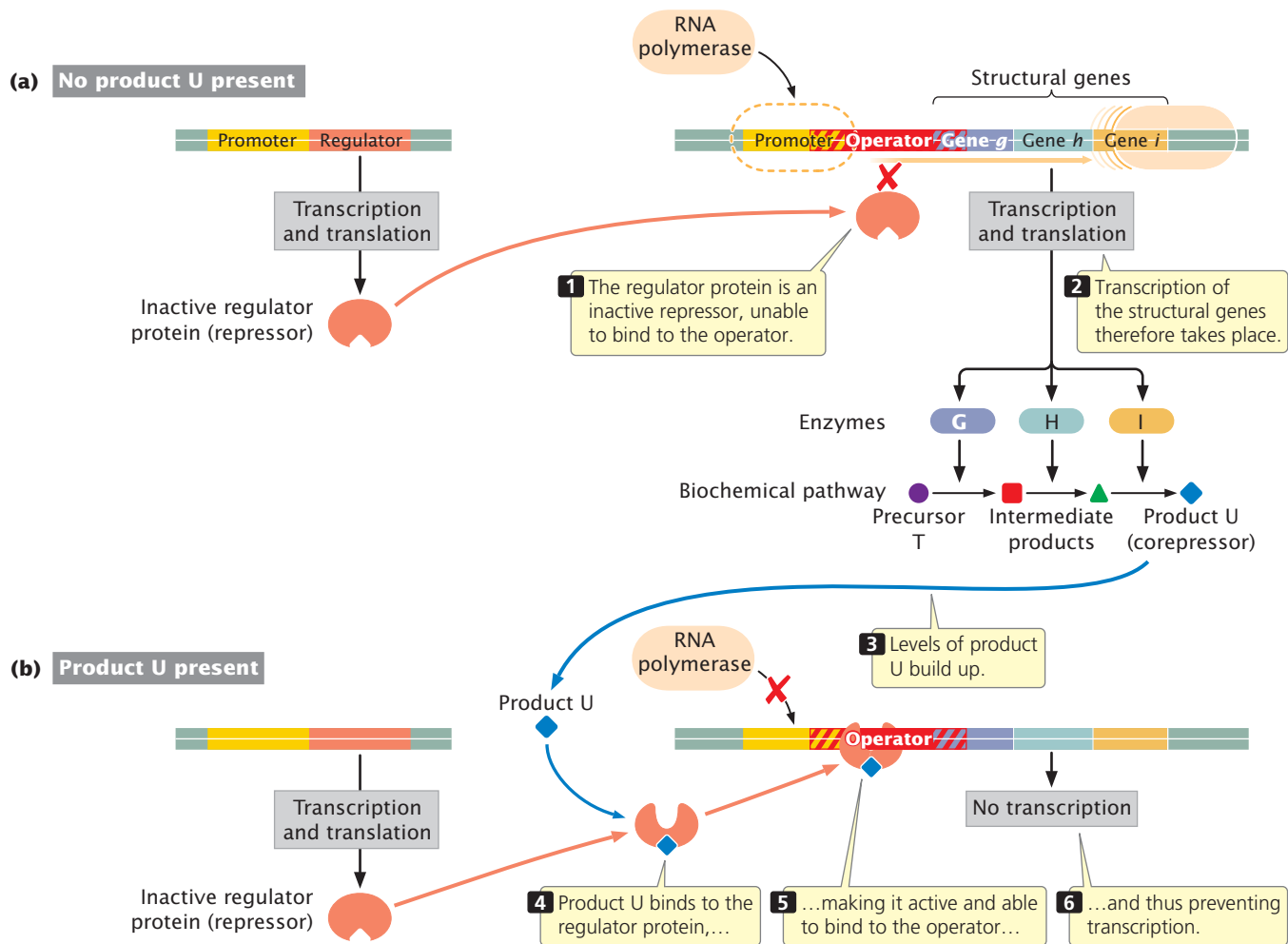
amino acids. For these types of operons, repressible control makes sense because the product produced by the proteins is always needed by the cell. Thus, these operons are normally on and are turned off when there are adequate amounts of the product already present.

Note that both the inducible and the repressible systems that we have considered are forms of negative control, in which the regulatory protein is a repressor. We will now consider positive control, in which a regulator protein stimulates transcription.

Positive control With positive control, a regulatory protein is an activator: it binds to DNA (usually at a site other than the operator) and stimulates transcription. Positive control can be inducible or repressible.

In a positive *inducible* operon, transcription is normally turned off because the regulator protein (an activator) is produced in an inactive form. Transcription takes place when an inducer has become attached to the regulatory protein, rendering the regulator active. Logically, the inducer should be the precursor of the reaction controlled by the

Negative repressible operon



16.5 Some operons are repressible.

operon so that the necessary enzymes would be synthesized only when the substrate for their reaction was present.

A positive operon can also be repressible; the regulatory protein is produced in a form that readily binds to DNA, meaning that transcription normally takes place and has to be repressed. Transcription is inhibited when a substance becomes attached to the activator and renders it unable to bind to the DNA so that transcription is no longer stimulated. Here, the product (P) of the reaction controlled by the operon would logically be the repressing substance, because it would be economical for the cell to prevent the transcription of genes that allow the synthesis of P when plenty of P was already available. The characteristics of positive and negative control in inducible and repressible operons are summarized in **Figure 16.6**. **TRY PROBLEM 11** →

CONCEPTS

There are two basic types of transcriptional control: negative and positive. In negative control, when a regulatory protein (repressor) binds to DNA, transcription is inhibited; in positive control, when

a regulatory protein (activator) binds to DNA, transcription is stimulated. Some operons are inducible; transcription is normally off and must be turned on. Other operons are repressible; transcription is normally on and must be turned off.

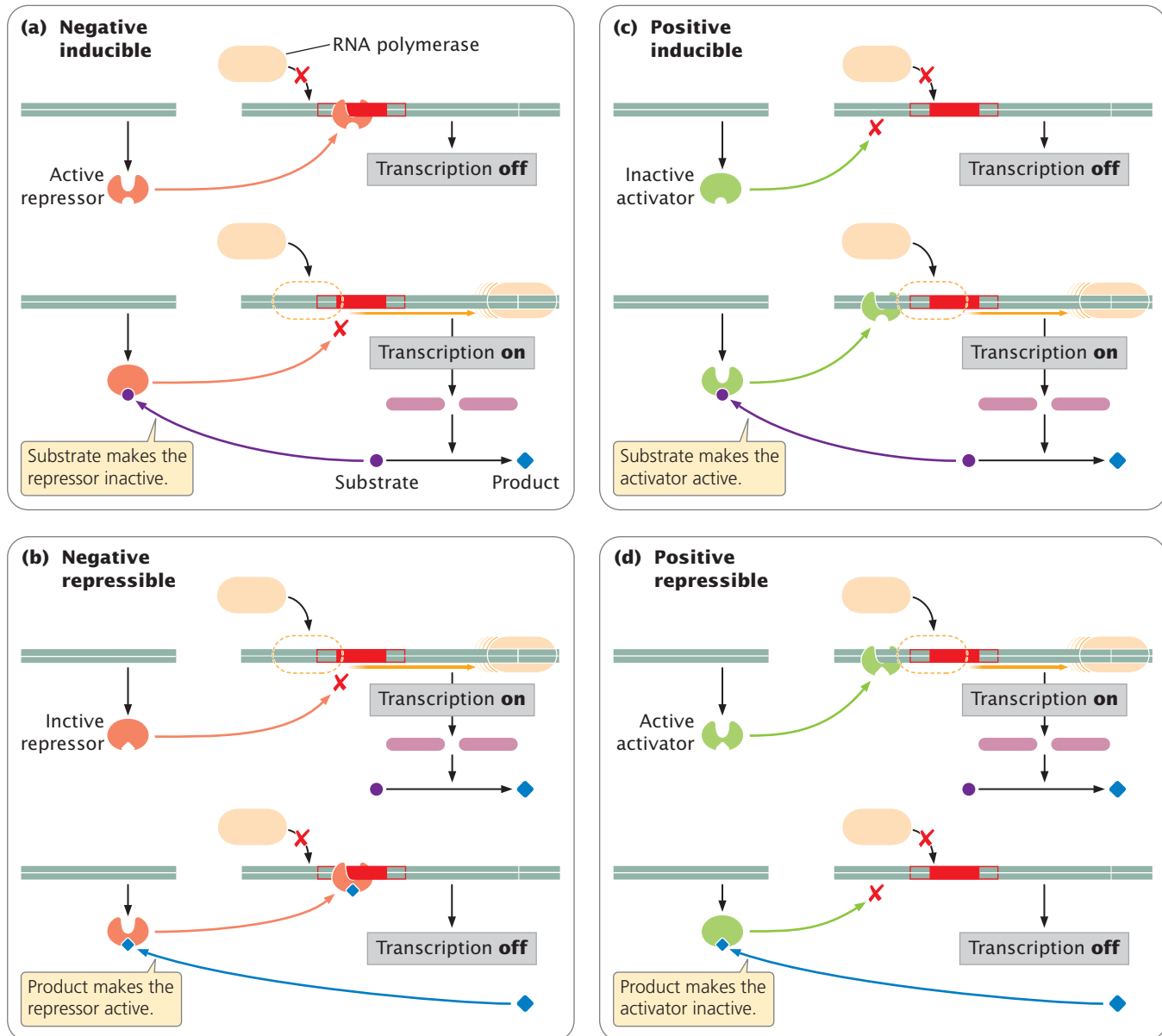
✓ CONCEPT CHECK 5

In a negative repressible operon, the regulator protein is synthesized as

- an active activator.
- an inactive activator.
- an active repressor.
- an inactive repressor.

The *lac* Operon of *E. coli*

In 1961, François Jacob and Jacques Monod described the “operon model” for the genetic control of lactose metabolism in *E. coli*. This work and subsequent research on the genetics of lactose metabolism established the operon as the basic unit of transcriptional control in bacteria. Despite the fact



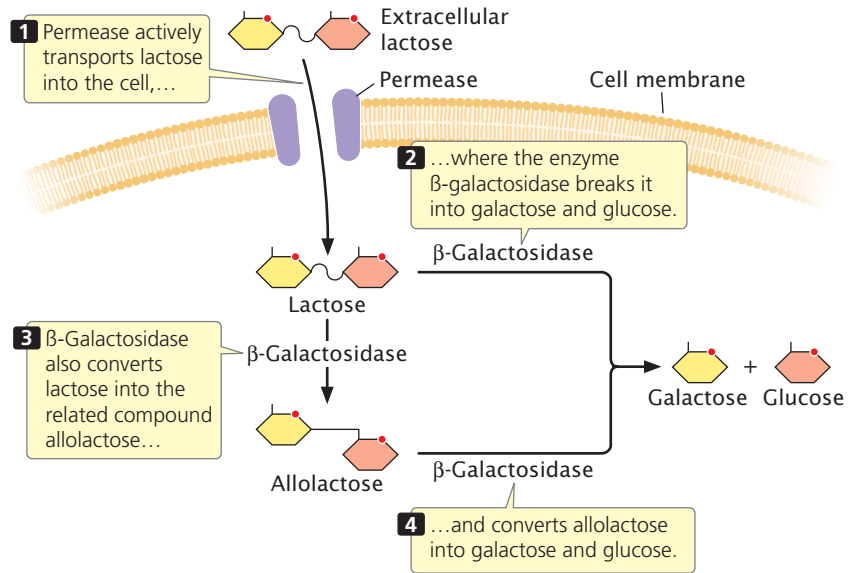
16.6 A summary of the characteristics of positive and negative control in inducible and repressible operons.

that, at the time, no methods were available for determining nucleotide sequences, Jacob and Monod deduced the structure of the operon *genetically* by analyzing the interactions of mutations that interfered with the normal regulation of lactose metabolism. We will examine the effects of some of these mutations after seeing how the *lac* operon regulates lactose metabolism.

Lactose metabolism Lactose is a major carbohydrate found in milk; it can be metabolized by *E. coli* bacteria that reside in the mammalian gut. Lactose does not easily diffuse across the *E. coli* cell membrane and must be actively transported into the cell by the protein permease

(**Figure 16.7**). To utilize lactose as an energy source, *E. coli* must first break it into glucose and galactose, a reaction catalyzed by the enzyme β -galactosidase. This enzyme can also convert lactose into allolactose, a compound that plays an important role in regulating lactose metabolism. A third enzyme, thiogalactoside transacetylase, also is produced by the *lac* operon, but its function in lactose metabolism is not yet known.

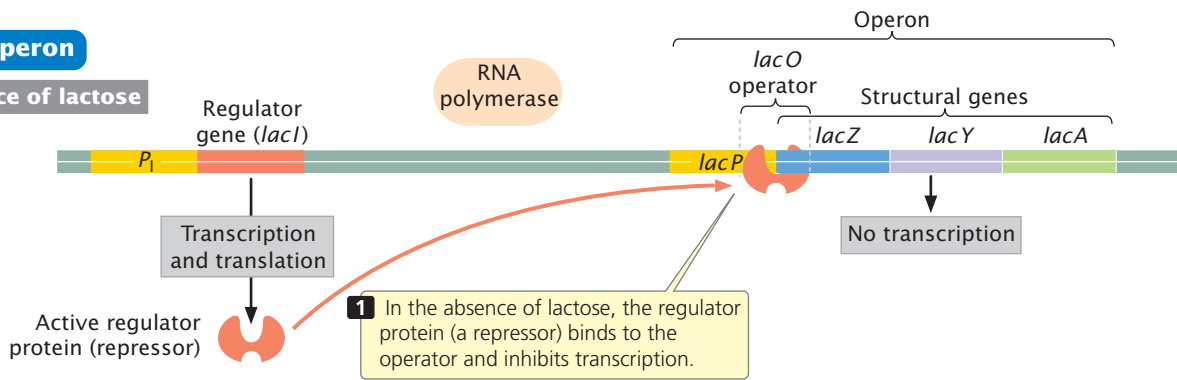
Regulation of the *lac* operon The *lac* operon is an example of a negative inducible operon. The enzymes β -galactosidase, permease, and transacetylase are encoded by adjacent structural genes in the *lac* operon of *E. coli*



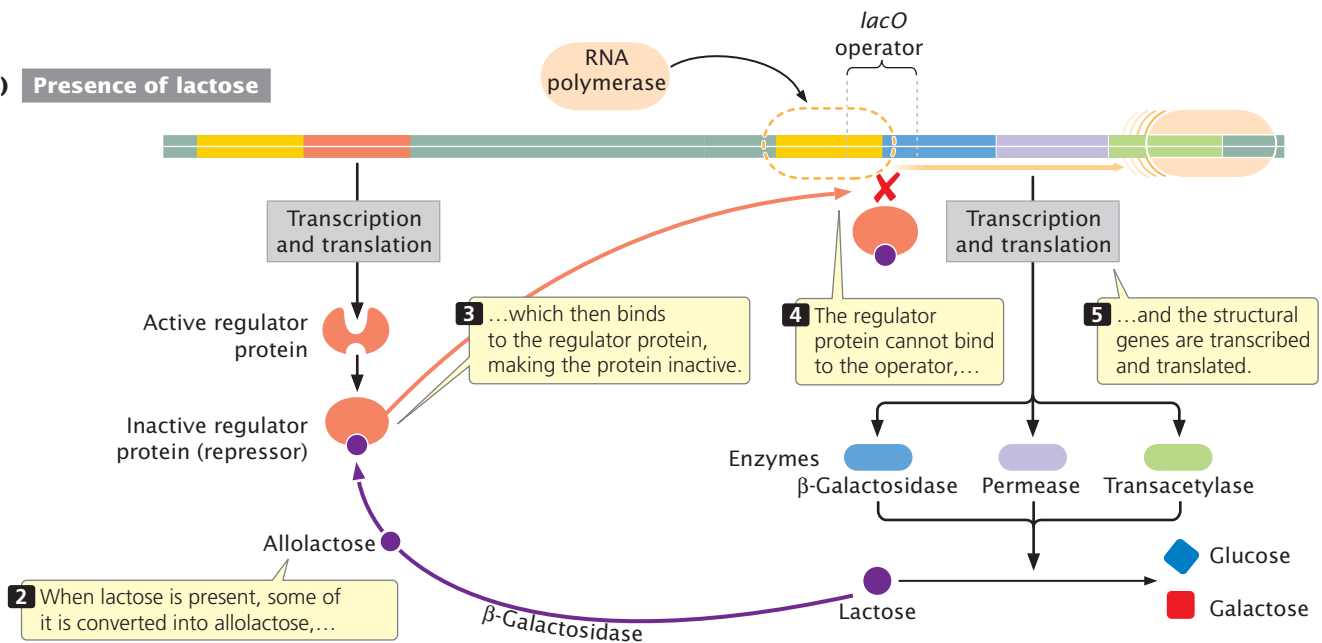
16.7 Lactose, a major carbohydrate found in milk, consists of 2 six-carbon sugars linked together.

The lac operon

(a) Absence of lactose



(b) Presence of lactose



16.8 The *lac* operon regulates lactose metabolism.

(Figure 16.8a) and have a common promoter (*lacP* in Figure 16.8a). β -Galactosidase is encoded by the *lacZ* gene, permease by the *lacY* gene, and transacetylase by the *lacA* gene. When lactose is absent from the medium in which *E. coli* grows, few molecules of each protein are produced. If lactose is added to the medium and glucose is absent, the rate of synthesis of all three proteins simultaneously increases about a thousandfold within 2 to 3 minutes. This boost in protein synthesis results from the transcription of *lacZ*, *lacY*, and *lacA* and exemplifies **coordinate induction**, the simultaneous synthesis of several proteins, stimulated by a specific molecule, the inducer (Figure 16.8b).

Although lactose appears to be the inducer here, allolactose is actually responsible for induction. Upstream of *lacP* is a regulator gene, *lacI*, which has its own promoter (P_1). The *lacI* gene is transcribed into a short mRNA that is translated into a repressor. The repressor consists of four identical polypeptides and has two types of binding sites; one type of site binds to allolactose and the other binds to DNA. In the absence of lactose (and, therefore, allolactose), the repressor binds to the *lac* operator site *lacO* (see Figure 16.8a). Jacob and Monod mapped the operator to a position adjacent to the *lacZ* gene; more-recent nucleotide sequencing has demonstrated that the operator actually overlaps the 3' end of the promoter and the 5' end of *lacZ* (Figure 16.9).

RNA polymerase binds to the promoter and moves down the DNA molecule, transcribing the structural genes. When the repressor is bound to the operator, the binding of RNA polymerase is blocked, and transcription is prevented. When lactose is present, some of it is converted into allolactose, which binds to the repressor and causes the repressor to be released from the DNA. In the presence of lactose, then, the repressor is inactivated, the binding of RNA polymerase is no longer blocked, the transcription of *lacZ*, *lacY*, and *lacA* takes place, and the *lac* proteins are produced.

Have you spotted the flaw in the explanation just given for the induction of the *lac* proteins? You might recall that permease is required to transport lactose into the cell. If the *lac* operon is repressed and no permease is being produced, how does lactose get into the cell to inactivate the repressor and turn on transcription? Furthermore, the inducer is actually allolactose, which must be produced from lactose by β -galactosidase. If β -galactosidase production is repressed, how can lactose metabolism be induced?

The answer is that repression never *completely* shuts down transcription of the *lac* operon. Even with active repressor bound to the operator, there is a low level of transcription and a few molecules of β -galactosidase, permease, and transacetylase are synthesized. When lactose appears in the medium, the permease that is present transports a small amount of lactose into the cell. There, the few molecules of β -galactosidase that are present convert some of the lactose into allolactose, which then induces transcription.

Several compounds related to allolactose also can bind to the *lac* repressor and induce transcription of the *lac* operon. One such inducer is isopropylthiogalactoside (IPTG). Although IPTG inactivates the repressor and allows the transcription of *lacZ*, *lacY*, and *lacA*, this inducer is not metabolized by β -galactosidase; for this reason, IPTG is often used in research to examine the effects of induction, independent of metabolism.

CONCEPTS

The *lac* operon of *E. coli* controls the transcription of three genes needed in lactose metabolism: the *lacZ* gene, which encodes β -galactosidase; the *lacY* gene, which encodes permease; and the *lacA* gene, which encodes thiogalactoside transacetylase. The *lac* operon is negative inducible: a regulator gene produces a repressor that binds to the operator site and prevents the transcription of the structural genes. The presence of allolactose inactivates the repressor and allows the transcription of the *lac* operon.

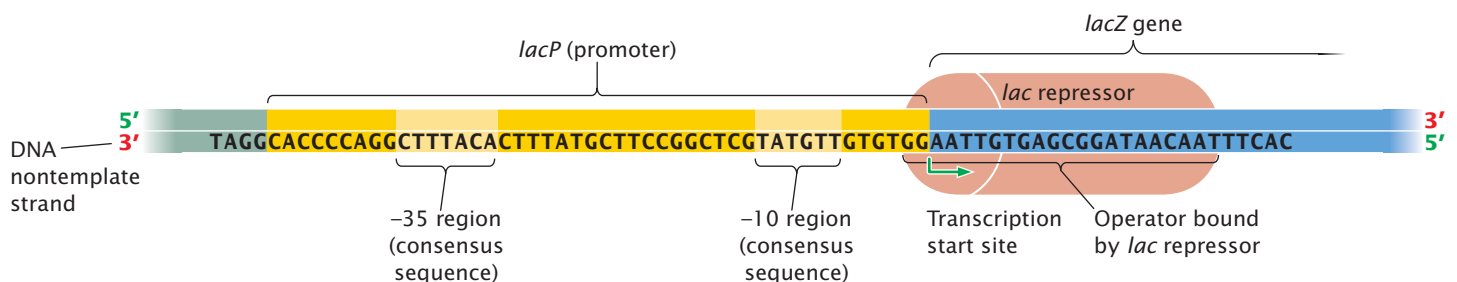
✓ CONCEPT CHECK 6

In the presence of allolactose, the *lac* repressor

- binds to the operator.
- binds to the promoter.
- cannot bind to the operator.
- binds to the regulator gene.

lac Mutations

Jacob and Monod worked out the structure and function of the *lac* operon by analyzing mutations that affected lactose metabolism. To help define the roles of the different components of the operon, they used **partial diploid** strains of *E. coli*. The cells of these strains possessed two different DNA molecules: the full bacterial chromosome and an extra piece



16.9 In the *lac* operon, the operator overlaps the promoter and the 5' end of the first structural gene.

of DNA. Jacob and Monod created these strains by allowing conjugation to take place between two bacteria (see Chapter 8). In conjugation, a small circular piece of DNA (the F plasmid, see Chapter 8) is transferred from one bacterium to another. The F plasmid used by Jacob and Monod contained the *lac* operon; so the recipient bacterium became partly diploid, possessing two copies of the *lac* operon. By using different combinations of mutations on the bacterial and plasmid DNA, Jacob and Monod determined that some parts of the *lac* operon are cis acting (able to control the expression of genes only when on the same piece of DNA), whereas other parts are trans acting (able to control the expression of genes on other DNA molecules).

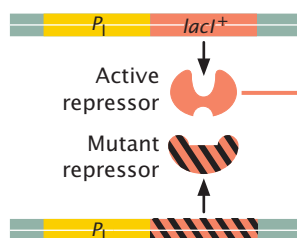
Structural-gene mutations Jacob and Monod first discovered some mutant strains that had lost the ability to synthesize either β -galactosidase or permease. (They did not study in detail the effects of mutations on the transacetylase enzyme, and so transacetylase will not be considered here.) The mutations in the mutant strains mapped to the *lacZ* or *lacY* structural genes and altered the amino acid sequences of the proteins encoded by the genes. These mutations

clearly affected the *structure* of the proteins and not the regulation of their synthesis.

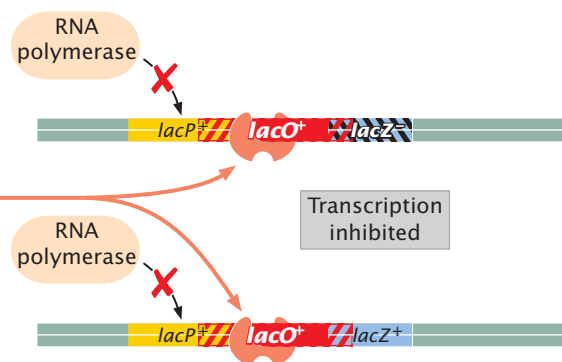
Through the use of partial diploids, Jacob and Monod were able to establish that mutations at the *lacZ* and *lacY* genes were independent and usually affected only the product of the gene in which they occurred. Partial diploids with *lacZ*⁺ *lacY*⁻ on the bacterial chromosome and *lacZ*⁻ *lacY*⁺ on the plasmid functioned normally, producing β -galactosidase and permease in the presence of lactose. (The genotype of a partial diploid is written by separating the genes on each DNA molecule with a slash: *lacZ*⁺ *lacY*⁻/*lacZ*⁻ *lacY*⁺.) In this partial diploid, a single functional β -galactosidase gene (*lacZ*⁺) is sufficient to produce β -galactosidase; whether the functional β -galactosidase gene is coupled to a functional (*lacY*⁺) or a defective (*lacY*⁻) permease gene makes no difference. The same is true of the *lacY*⁺ gene.

Regulator-gene mutations Jacob and Monod also isolated mutations that affected the *regulation* of protein production. Mutations in the *lacI* gene affect the production of both β -galactosidase and permease because genes for both proteins are in the same operon and are regulated coordinately.

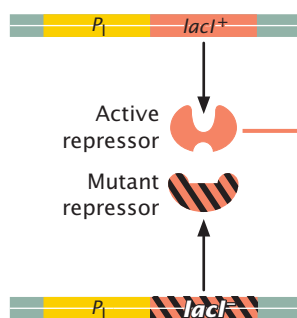
(a) Absence of lactose



1 The *lacI*⁺ gene is trans dominant: the repressor produced by *lacI*⁺ can bind to both operators and repress transcription in the absence of lactose.

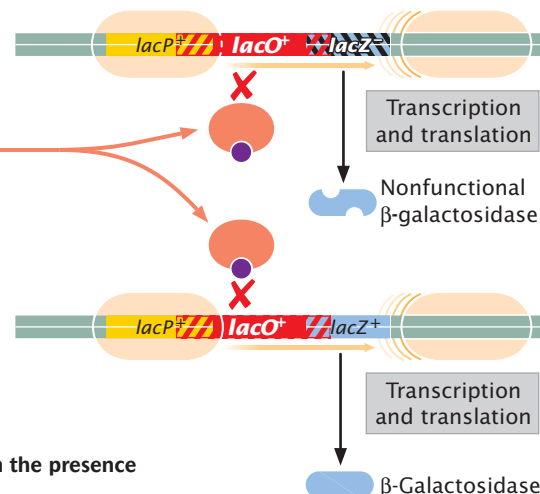


(b) Presence of lactose

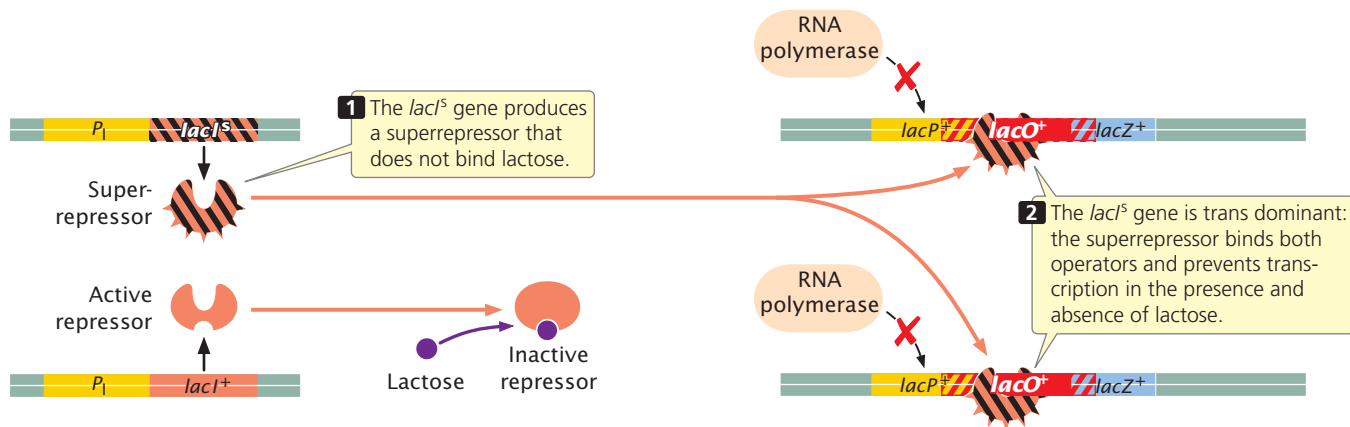


2 When lactose is present, it inactivates the repressor, and functional β -galactosidase is produced from the *lacZ*⁺ gene.

Lactose (purple dot) binds to the active repressor (red shape), forming an inactive repressor (red shape with purple dot).



16.10 The partial diploid *lacI*⁺ *lacZ*⁻/*lacI*⁻ *lacZ*⁺ produces β -galactosidase only in the presence of lactose because the *lacI* gene is trans dominant.



16.11 The partial diploid $lacI^s lacZ^+ / lacI^+ lacZ^+$ fails to produce β -galactosidase in the presence and absence of lactose because the $lacI^s$ gene encodes a superrepressor.

Some of these mutations were constitutive, causing the *lac* proteins to be produced all the time, whether lactose was present or not. Such mutations in the regulator gene were designated $lacI^-$. The construction of partial diploids demonstrated that a $lacI^+$ gene is dominant over a $lacI^-$ gene; a single copy of $lacI^+$ (genotype $lacI^+ / lacI^-$) was sufficient to bring about normal regulation of protein production. Furthermore, $lacI^+$ restored normal control to an operon even if the operon was located on a different DNA molecule, showing that $lacI^+$ can be trans acting. A partial diploid with genotype $lacI^+ lacZ^- / lacI^- lacZ^+$ functioned normally, synthesizing β -galactosidase only when lactose was present (Figure 16.10). In this strain, the $lacI^+$ gene on the bacterial chromosome was functional, but the $lacZ^-$ gene was defective; on the plasmid, the $lacI^-$ gene was defective, but the $lacZ^+$ gene was functional. The fact that a $lacI^+$ gene could regulate a $lacZ^+$ gene located on a different DNA molecule indicated to Jacob and Monod that the $lacI^+$ gene product was able to operate on either the plasmid or the chromosome.

Some *lacI* mutations isolated by Jacob and Monod prevented transcription from taking place even in the presence of lactose and other inducers such as IPTG. These mutations were referred to as superrepressors ($lacI^s$), because they produced defective repressors that could not be inactivated by an inducer. The $lacI^s$ mutations produced a repressor with an altered inducer-binding site, which made the inducer unable to bind to the repressor; consequently, the repressor was always able to attach to the operator site and prevent transcription of the *lac* genes. Superrepressor mutations were dominant over $lacI^+$; partial diploids with genotype $lacI^s lacZ^+ / lacI^+ lacZ^+$ were unable to synthesize either β -galactosidase or permease, whether or not lactose was present (Figure 16.11).

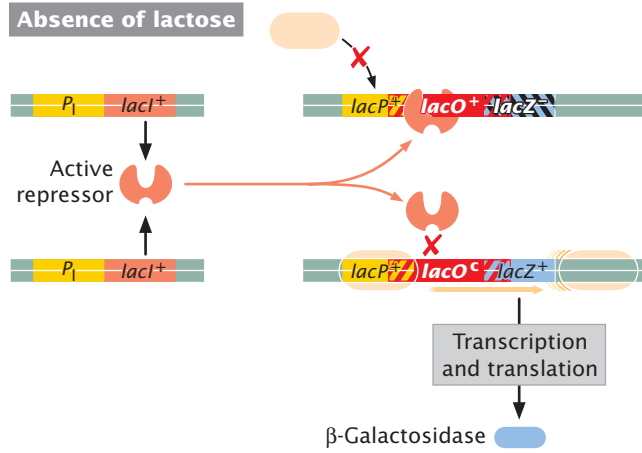
Operator mutations Jacob and Monod mapped another class of constitutive mutants to a site adjacent to *lacZ*. These mutations occurred at the operator site and were referred

to as $lacO^c$ (*O* stands for operator and “c” for constitutive). The $lacO^c$ mutations altered the sequence of DNA at the operator so that the repressor protein was no longer able to bind. A partial diploid with genotype $lacI^+ lacO^c lacZ^+ / lacI^+ lacO^+ lacZ^+$ exhibited constitutive synthesis of β -galactosidase, indicating that $lacO^c$ is dominant over $lacO^+$.

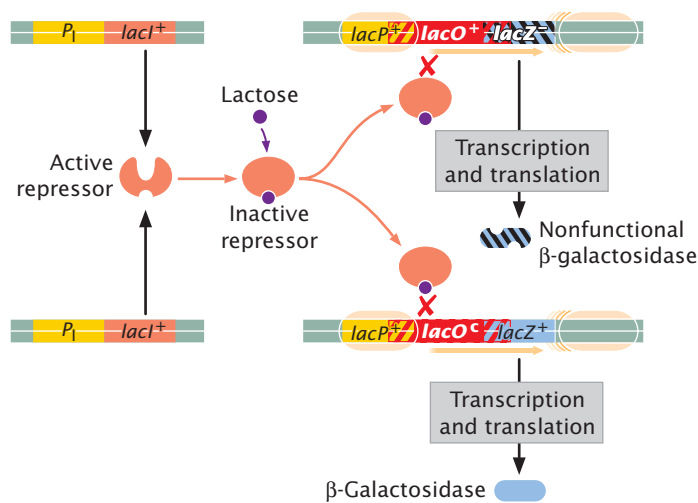
Analysis of other partial diploids showed that the *lacO* gene is cis acting, affecting only genes on the same DNA molecule. For example, a partial diploid with genotype $lacI^+ lacO^+ lacZ^- / lacI^+ lacO^c lacZ^+$ was constitutive, producing β -galactosidase in the presence or absence of lactose (Figure 16.12a), but a partial diploid with genotype $lacI^+ lacO^+ lacZ^+ / lacI^+ lacO^c lacZ^-$ produced β -galactosidase only in the presence of lactose (Figure 16.12b). In the constitutive partial diploid ($lacI^+ lacO^+ lacZ^- / lacI^+ lacO^c lacZ^+$; see Figure 16.12a), the $lacO^c$ mutation and the functional $lacZ^+$ gene are present on the same DNA molecule; but, in $lacI^+ lacO^+ lacZ^+ / lacI^+ lacO^c lacZ^-$ (see Figure 16.12b), the $lacO^c$ mutation and the functional $lacZ^+$ gene are on different molecules. The *lacO* mutation affects only genes to which it is physically connected, as is true of all operator mutations. They prevent the binding of a repressor protein to the operator and thereby allow RNA polymerase to transcribe genes on the same DNA molecule. However, they cannot prevent a repressor from binding to normal operators on other DNA molecules. **TRY PROBLEM 20** →

Promoter mutations Mutations affecting lactose metabolism have also been isolated at the promoter site; these mutations are designated $lacP^-$, and they interfere with the binding of RNA polymerase to the promoter. Because this binding is essential for the transcription of the structural genes, *E. coli* strains with $lacP^-$ mutations don't produce *lac* proteins either in the presence or in the absence of lactose. Like operator mutations, $lacP^-$ mutations are cis acting and thus affect only genes on the same DNA molecule. The partial diploid $lacI^+ lacP^- lacZ^+ / lacI^+ lacP^+ lacZ^+$ exhibits

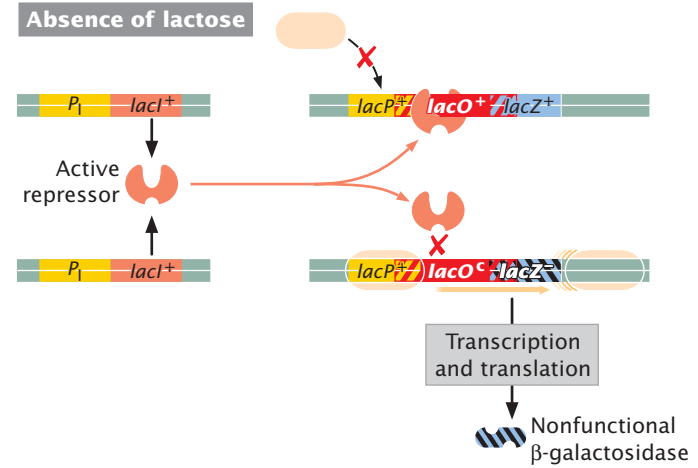
(a) Partial diploid $lacI^+ lacO^+ lacZ^- / lacI^+ lacO^c lacZ^+$



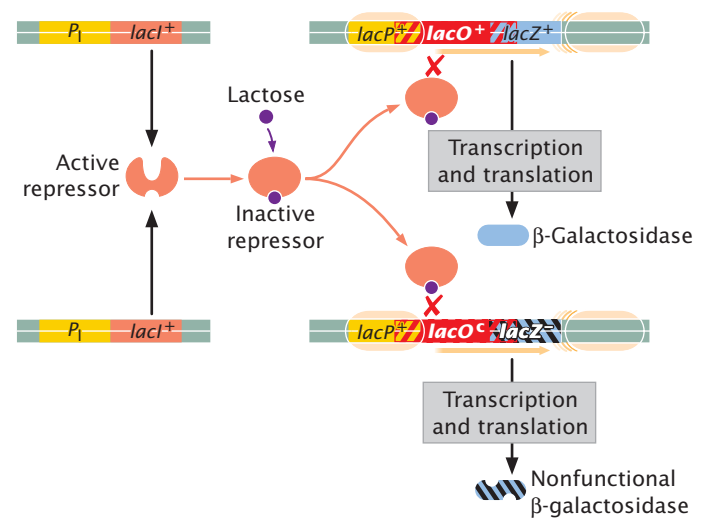
Presence of lactose



(b) Partial diploid $lacI^+ lacO^+ lacZ^+ / lacI^+ lacO^c lacZ^-$




Presence of lactose



16.12 Mutations in $lacO$ are constitutive and cis acting. (a) The partial diploid $lacI^+ lacO^+ lacZ^- / lacI^+ lacO^c lacZ^+$ is constitutive, producing β -galactosidase in the presence and absence of lactose.

(b) The partial diploid $lacI^+ lacO^+ lacZ^+ / lacI^+ lacO^c lacZ^-$ is inducible (produces β -galactosidase only when lactose is present), demonstrating that the $lacO$ gene is cis acting.

 www.whfreeman.com/pierce4e See Animation 16.1 to explore the effects of different combinations of $lacI$ and $lacO$ mutations on the expression of the lac operon.

normal synthesis of β -galactosidase, whereas $lacI^+ lacP^- lacZ^+ / lacI^+ lacP^+ lacZ^-$ fails to produce β -galactosidase whether or not lactose is present.

Worked Problem

For *E. coli* strains with the following lac genotypes, use a plus sign (+) to indicate the synthesis of β -galactosidase and per-

mease and a minus sign (–) to indicate no synthesis of the proteins when lactose is absent and when it is present.

Genotype of strain

- $lacI^+ lacP^+ lacO^+ lacZ^+ lacY^+$
- $lacI^+ lacP^+ lacO^c lacZ^- lacY^+$
- $lacI^+ lacP^- lacO^+ lacZ^+ lacY^-$
- $lacI^+ lacP^+ lacO^+ lacZ^- lacY^- / lacI^- lacP^+ lacO^+ lacZ^+ lacY^+$

• Solution

Genotype of strain
a. $lacI^+ lacP^+ lacO^+ lacZ^+ lacY^+$
b. $lacI^+ lacP^+ lacO^c lacZ^- lacY^+$
c. $lacI^+ lacP^- lacO^+ lacZ^+ lacY^-$
d. $lacI^+ lacP^+ lacO^+ lacZ^- lacY^- / lacI^- lacP^+ lacO^+ lacZ^+ lacY^+$

- a. All the genes possess normal sequences, and so the *lac* operon functions normally: when lactose is absent, the regulator protein binds to the operator and inhibits the transcription of the structural genes, and so β -galactosidase and permease are not produced. When lactose is present, some of it is converted into allolactose, which binds to the repressor and makes it inactive; the repressor does not bind to the operator, and so the structural genes are transcribed and β -galactosidase and permease are produced.
- b. The structural *lacZ* gene is mutated; so β -galactosidase will not be produced under any conditions. The *lacO* gene has a constitutive mutation, which means that the repressor is unable to bind to *lacO*, and so transcription takes place at all times. Therefore, permease will be produced in both the presence and the absence of lactose.
- c. In this strain, the promoter is mutated, and so RNA polymerase is unable to bind and transcription does not take place. Therefore, β -galactosidase and permease are not produced under any conditions.
- d. This strain is a partial diploid, which consists of two copies of the *lac* operon—one on the bacterial chromosome and the other on a plasmid. The *lac* operon represented in the upper part of the genotype has mutations in both the *lacZ* and the *lacY* genes, and so it is not capable of encoding β -galactosidase or permease under any conditions. The *lac* operon in the lower part of the genotype has a defective regulator gene, but the normal regulator gene in the upper operon produces a diffusible repressor (trans acting) that binds to the lower operon in the absence of lactose and inhibits transcription. Therefore, no β -galactosidase or permease is produced when lactose is absent. In the presence of lactose, the repressor cannot bind to the operator, and so the lower operon is transcribed and β -galactosidase and permease are produced.

➔ Now try your own hand at predicting the outcome of different *lac* mutations by working Problem 18 at the end of the chapter.

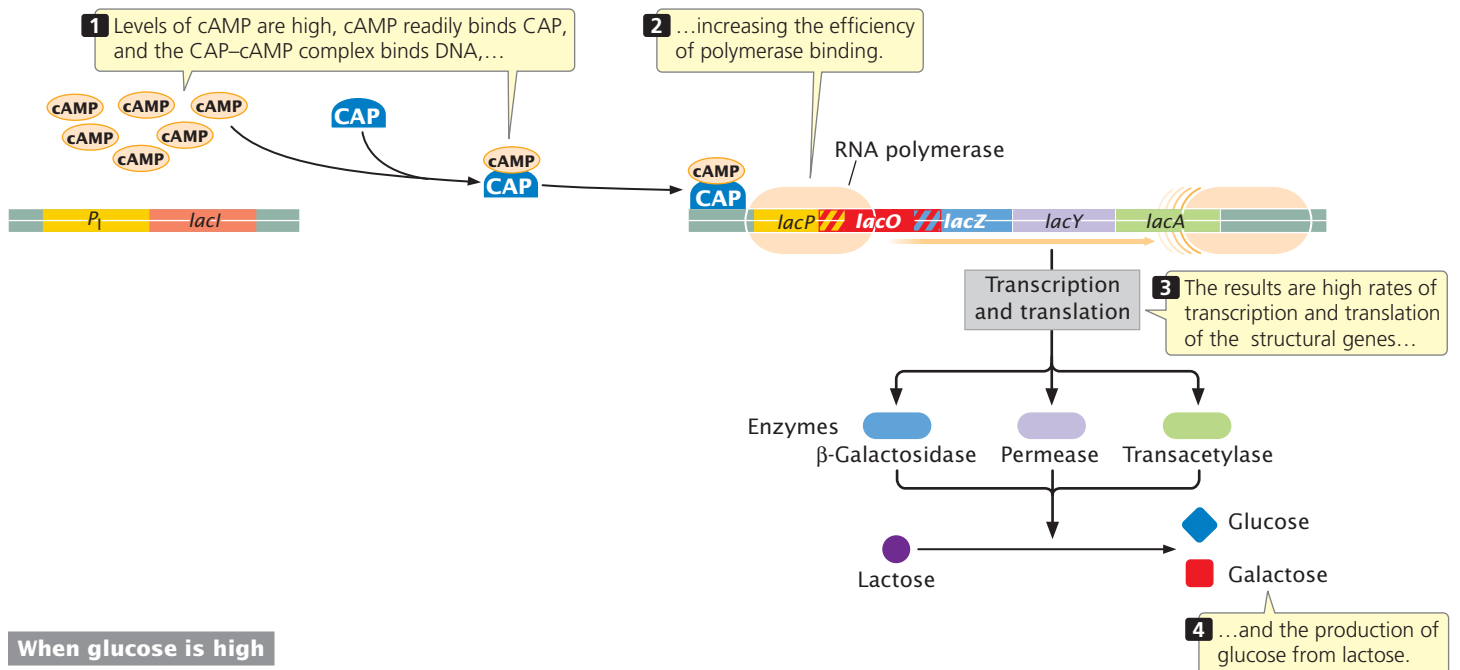
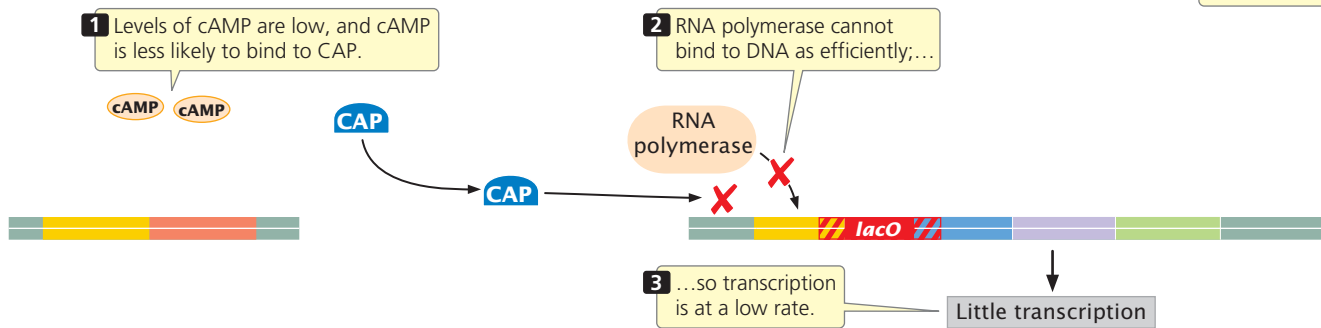
Lactose absent		Lactose present	
β -Galactosidase	Permease	β -Galactosidase	Permease
–	–	+	+
–	+	–	+
–	–	–	–
–	–	+	+

Positive Control and Catabolite Repression

E. coli and many other bacteria metabolize glucose preferentially in the presence of lactose and other sugars. They do so because glucose enters glycolysis without further modification and therefore requires less energy to metabolize than do other sugars. When glucose is available, genes that participate in the metabolism of other sugars are repressed, in a phenomenon known as **catabolite repression**. For example, the efficient transcription of the *lac* operon takes place only if lactose is present and glucose is absent. But how is the expression of the *lac* operon influenced by glucose? What brings about catabolite repression?

Catabolite repression results from positive control in response to glucose. (This regulation is in addition to the negative control brought about by the repressor binding at the operator site of the *lac* operon when lactose is absent.) Positive control is accomplished through the binding of a dimeric protein called the **catabolite activator protein (CAP)** to a site that is about 22 nucleotides long and is located within or slightly upstream of the promoter of the *lac* genes (**Figure 16.13**). RNA polymerase does not bind efficiently to many promoters unless CAP is first bound to the DNA. Before CAP can bind to DNA, it must form a complex with a modified nucleotide called **adenosine-3', 5'-cyclic monophosphate** (cyclic AMP, or cAMP), which is important in cellular signaling processes in both bacterial and eukaryotic cells. In *E. coli*, the concentration of cAMP is regulated so that its concentration is inversely proportional to the level of available glucose. A high concentration of glucose within the cell lowers the amount of cAMP, and so little cAMP–CAP complex is available to bind to the DNA. Subsequently, RNA polymerase has poor affinity for the *lac* promoter, and little transcription of the *lac* operon takes place. Low concentrations of glucose stimulate high levels of cAMP, resulting in increased cAMP–CAP binding to DNA. This increase enhances the binding of RNA polymerase to the promoter and increases transcription of the *lac* genes by some 50-fold.

The catabolite activator protein exerts positive control in more than 20 operons of *E. coli*. The response to CAP varies among these promoters; some operons are activated by low levels of CAP, whereas others require high levels.

When glucose is low**When glucose is high**

16.13 The catabolite activator protein (CAP) binds to the promoter of the *lac* operon and stimulates transcription. CAP must complex with adenosine-3',5'-cyclic monophosphate (cAMP) before binding to the promoter of the *lac* operon. The binding of cAMP–CAP to the promoter activates transcription by facilitating the binding of RNA polymerase. Levels of cAMP are inversely related to glucose: low glucose stimulates high cAMP; high glucose stimulates low cAMP.

CAP contains a helix-turn-helix DNA-binding motif and, when it binds to the CAP site on DNA, it causes the DNA helix to bend (**Figure 16.14**). The bent helix enables CAP to interact directly with the RNA polymerase enzyme bound to the promoter and facilitate the initiation of transcription.

TRY PROBLEM 14 →

CONCEPTS

In spite of its name, catabolite repression is a type of positive control in the *lac* operon. The catabolite activator protein (CAP), complexed with cAMP, binds to a site near the promoter and stimulates the binding of RNA polymerase. Cellular levels of cAMP are controlled by glucose; a low glucose level increases the abundance of cAMP and enhances the transcription of the *lac* structural genes.

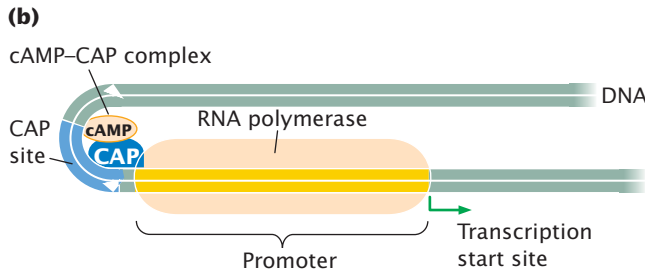
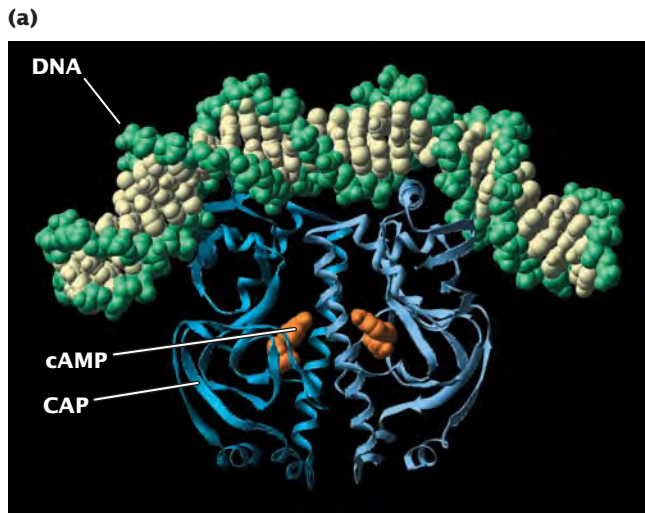
✓ CONCEPT CHECK 7

What is the effect of high levels of glucose on the *lac* operon?

- Transcription is stimulated.
- Little transcription takes place.
- Transcription is not affected.
- Transcription may be stimulated or inhibited, depending on the levels of lactose.

The *trp* Operon of *E. coli*

The *lac* operon just discussed is an inducible operon, one in which transcription does not normally take place and must be turned on. Other operons are repressible; transcription in



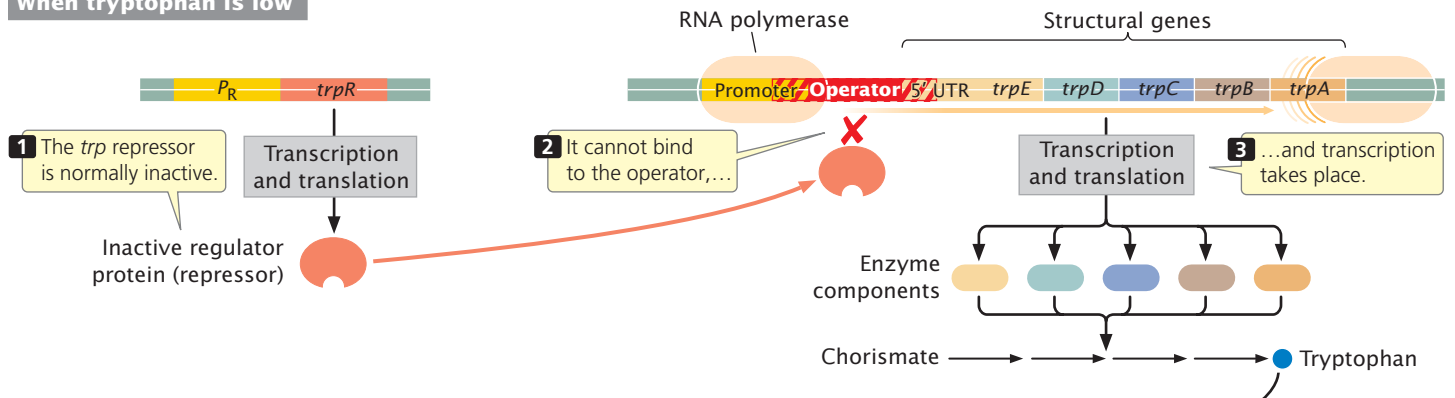
16.14 The binding of the cAMP–CAP complex to DNA produces a sharp bend in DNA that activates transcription.

these operons is normally turned on and must be repressed. The tryptophan (*trp*) operon in *E. coli*, which controls the biosynthesis of the amino acid tryptophan, is an example of a negative repressible operon.

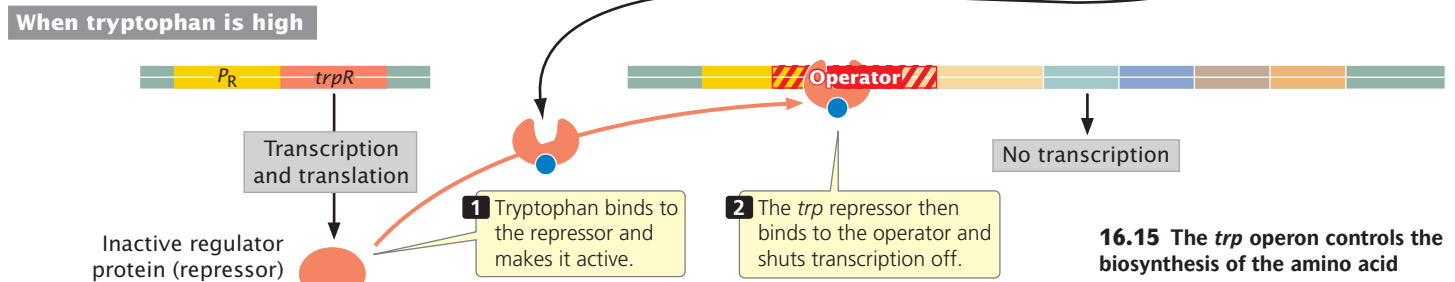
The *trp* operon contains five structural genes (*trpE*, *trpD*, *trpC*, *trpB*, and *trpA*) that produce the components of three enzymes (two of the enzymes consist of two polypeptide chains). These enzymes convert chorismate into tryptophan (Figure 16.15). The first structural gene, *trpE*, contains a long 5' untranslated region (5' UTR) that is transcribed but does not encode any of these enzymes. Instead, this 5' UTR plays an important role in another regulatory mechanism, discussed in the next section. Upstream of the 5' UTR is the *trp* promoter. When tryptophan levels are low, RNA polymerase binds to the promoter and transcribes the five structural genes into a single mRNA, which is then translated into enzymes that convert chorismate into tryptophan.

Some distance from the *trp* operon is a regulator gene, *trpR*, which encodes a repressor that alone cannot bind DNA (see Figure 16.15). Like the *lac* repressor, the tryptophan repressor has two binding sites, one that binds to DNA at the operator site and another that binds to tryptophan (the activator). Binding with tryptophan causes a conformational change in the repressor that makes it capable of binding to DNA at the operator site, which overlaps the promoter. When the operator is occupied by the tryptophan repressor, RNA polymerase cannot bind to the promoter and the structural genes cannot be transcribed. Thus, when cellular levels of tryptophan are low, transcription of the *trp* operon takes place and more tryptophan is synthesized; when cellular levels of tryptophan are high, transcription of the *trp* operon is inhibited and the synthesis of more tryptophan does not take place.

When tryptophan is low



When tryptophan is high



16.15 The *trp* operon controls the biosynthesis of the amino acid tryptophan in *E. coli*.

CONCEPTS

The *trp* operon is a negative repressible operon that controls the biosynthesis of tryptophan. In a repressible operon, transcription is normally turned on and must be repressed. Repression is accomplished through the binding of tryptophan to the repressor, which renders the repressor active. The active repressor binds to the operator and prevents RNA polymerase from transcribing the structural genes.

✓ CONCEPT CHECK 8

In the *trp* operon, what happens to the *trp* repressor in the absence of tryptophan?

- It binds to the operator and represses transcription.
- It cannot bind to the operator and transcription takes place.
- It binds to the regulator gene and represses transcription.
- It cannot bind to the regulator gene and transcription takes place.

16.3 Some Operons Regulate Transcription Through Attenuation, the Premature Termination of Transcription

We've now seen several different ways in which a cell regulates the initiation of transcription in an operon. Some operons have an additional level of control that affects the *continuation* of transcription rather than its initiation. In **attenuation**, transcription begins at the start site, but termination takes place prematurely, before the RNA polymerase even reaches the structural genes. Attenuation takes place in a number of operons that encode enzymes participating in the biosynthesis of amino acids.

Attenuation in the *trp* Operon of *E. coli*

We can understand the process of **attenuation** most easily by looking at one of the best-studied examples, which is found in the *trp* operon of *E. coli*. The *trp* operon is unusual in that it is regulated both by repression and by attenuation. Most operons are regulated by one of these mechanisms but not by both of them.

Attenuation first came to light when Charles Yanofsky and his colleagues made several observations in the early 1970s that indicated that repression at the operator site is not the only method of regulation in the *trp* operon. They isolated a series of mutants that exhibited high levels of transcription, yet control at the operator site was unaffected, suggesting that some mechanism other than repression at the operator site was controlling transcription. Furthermore, they observed that two mRNAs of different sizes were transcribed from the *trp* operon: a long mRNA containing sequences for the structural genes and a much shorter mRNA of only 140 nucleotides. These observations led Yanofsky to propose that a mechanism that caused premature termination of transcription also regulates transcription in the *trp* operon.

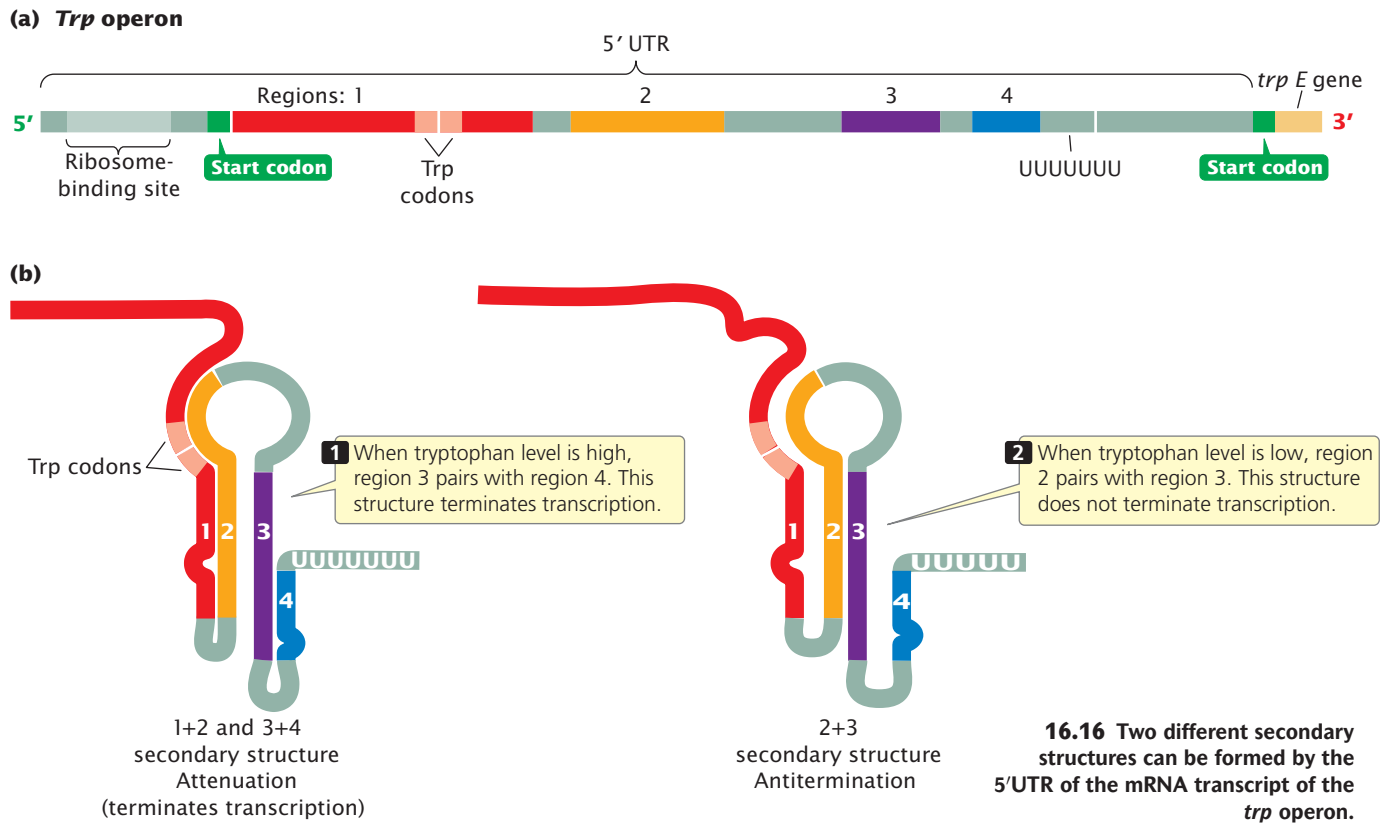
Close examination of the *trp* operon reveals a region of 162 nucleotides that corresponds to the long 5' UTR of the mRNA (mentioned earlier) transcribed from the *trp* operon (**Figure 16.16a**). The 5' UTR (also called a leader) contains four regions: region 1 is complementary to region 2, region 2 is complementary to region 3, and region 3 is complementary to region 4. These complementarities allow the 5' UTR to fold into two different secondary structures (**Figure 16.16b**). Only one of these secondary structures causes attenuation.

One of the secondary structures contains one hairpin produced by the base pairing of regions 1 and 2 and another hairpin produced by the base pairing of regions 3 and 4. Notice that a string of uracil nucleotides follows the 3+4 hairpin. Not coincidentally, the structure of a bacterial intrinsic terminator (see Chapter 13) includes a hairpin followed by a string of uracil nucleotides; this secondary structure in the 5' UTR of the *trp* operon is indeed a terminator and is called an **attenuator**. The attenuator forms when cellular levels of tryptophan are high, causing transcription to be terminated before the *trp* structural genes can be transcribed.

When cellular levels of tryptophan are low, however, the alternative secondary structure of the 5' UTR is produced by the base pairing of regions 2 and 3 (see Figure 16.16b). This base pairing also produces a hairpin, but this hairpin is not followed by a string of uracil nucleotides; so this structure does *not* function as a terminator. RNA polymerase continues past the 5' UTR into the coding section of the structural genes, and the enzymes that synthesize tryptophan are produced. Because it prevents the termination of transcription, the 2+3 structure is called an **antiterminator**.

To summarize, the 5' UTR of the *trp* operon can fold into one of two structures. When the tryptophan level is high, the 3+4 structure forms, transcription is terminated within the 5' UTR, and no additional tryptophan is synthesized. When the tryptophan level is low, the 2+3 structure forms, transcription continues through the structural genes, and tryptophan is synthesized. The critical question, then, is: Why does the 3+4 structure arise when the level of tryptophan in the cell is high, whereas the 2+3 structure arises when the level is low?

To answer this question, we must take a closer look at the nucleotide sequence of the 5' UTR. At the 5' end, upstream of region 1, is a ribosome-binding site (see Figure 16.16a). Region 1 encodes a small protein. Within the coding sequence for this protein are two UGG codons, which specify the amino acid tryptophan; so tryptophan is required for the translation of this 5' UTR sequence. The small protein encoded by the 5' UTR has not been isolated and is presumed to be unstable; its only apparent function is to control attenuation. Although it was stated in Chapter 14 that a 5' UTR is not translated into a protein, the 5' UTR of operons subject to attenuation is an exception to this rule. The precise timing and interaction of transcription and translation in the 5' UTR determine whether attenuation takes place.



Transcription when tryptophan levels are low Let's first consider what happens when intracellular levels of tryptophan are low. Recall that, in prokaryotic cells, transcription and translation are coupled: while transcription is taking place at the 3' end of the mRNA, translation is initiated at the 5' end. RNA polymerase begins transcribing the DNA, producing region 1 of the 5' UTR (Figure 16.17a). Closely following RNA polymerase, a ribosome binds to the 5' UTR and begins to translate the coding region. Meanwhile, RNA polymerase is transcribing region 2 (Figure 16.17b). Region 2 is complementary to region 1 but, because the ribosome is translating region 1, the nucleotides in regions 1 and 2 cannot base pair.

RNA polymerase begins to transcribe region 3, and the ribosome reaches the UGG tryptophan codons in region 1. When it reaches the tryptophan codons, the ribosome stalls (Figure 16.17c) because the level of tryptophan is low and tRNAs charged with tryptophan are scarce or even unavailable. The ribosome sits at the tryptophan codons, awaiting the arrival of a tRNA charged with tryptophan. Stalling of the ribosome does not, however, hinder transcription; RNA polymerase continues to move along the DNA, and transcription gets ahead of translation.

Because the ribosome is stalled at the tryptophan codons in region 1, region 2 is free to base pair with region 3, forming the 2+3 hairpin (Figure 16.17d). This hairpin does not cause termination, and so transcription continues. Because region 3 is already paired with region 2, the 3+4 hairpin (the attenuator) never forms, and so attenuation

does not take place and transcription continues. RNA polymerase continues along the DNA, past the 5' UTR, transcribing all the structural genes into mRNA, which is translated into the enzymes encoded by the *trp* operon. These enzymes then synthesize more tryptophan.

Transcription when tryptophan levels are high Now, let's see what happens when intracellular levels of tryptophan are high. Once again, RNA polymerase begins transcribing the DNA, producing region 1 of the 5' UTR (Figure 16.17e). Closely following RNA polymerase, a ribosome binds to the 5' UTR and begins to translate the coding region (Figure 16.17f). When the ribosome reaches the two UGG tryptophan codons, it doesn't slow or stall, because tryptophan is abundant and tRNAs charged with tryptophan are readily available (Figure 16.17g). This point is critical to note: because tryptophan is abundant, translation can keep up with transcription.

As it moves past region 1, the ribosome partly covers region 2 (Figure 16.17h); meanwhile, RNA polymerase completes the transcription of region 3. Although regions 2 and 3 are complementary, the ribosome physically blocks their pairing.

RNA polymerase continues to move along the DNA, eventually transcribing region 4 of the 5' UTR. Region 4 is complementary to region 3, and, because region 3 cannot base pair with region 2, it pairs with region 4. The pairing of regions 3 and 4 (see Figure 16.17h) produces the

When tryptophan level is low

(a)

RNA polymerase begins transcribing the DNA, producing region 1 of the 5' UTR.

(b)

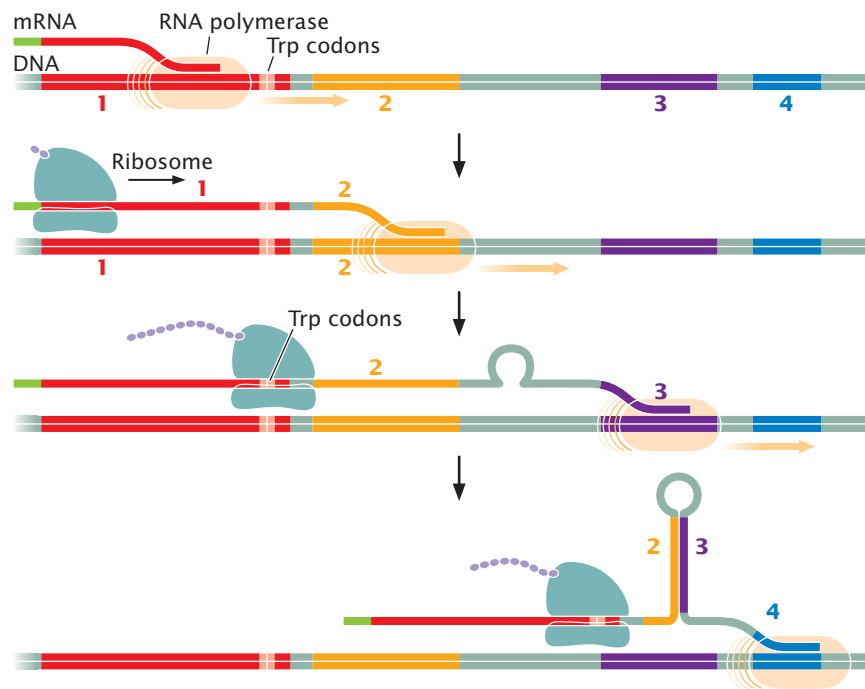
A ribosome attaches to the 5' end of the 5' UTR and translates region 1 while region 2 is being transcribed.

(c)

The ribosome stalls at the Trp codons in region 1 because tryptophan is low. Because the ribosome is stalled, region 2 is not covered by the ribosome when region 3 is transcribed.

(d)

When region 3 is transcribed, it pairs with region 2. When region 4 is transcribed, it cannot pair with region 3, because region 3 is already paired with region 2; the attenuator never forms, and transcription continues.



When tryptophan level is high

(e)

RNA polymerase begins transcribing DNA, producing region 1 of the 5' UTR.

(f)

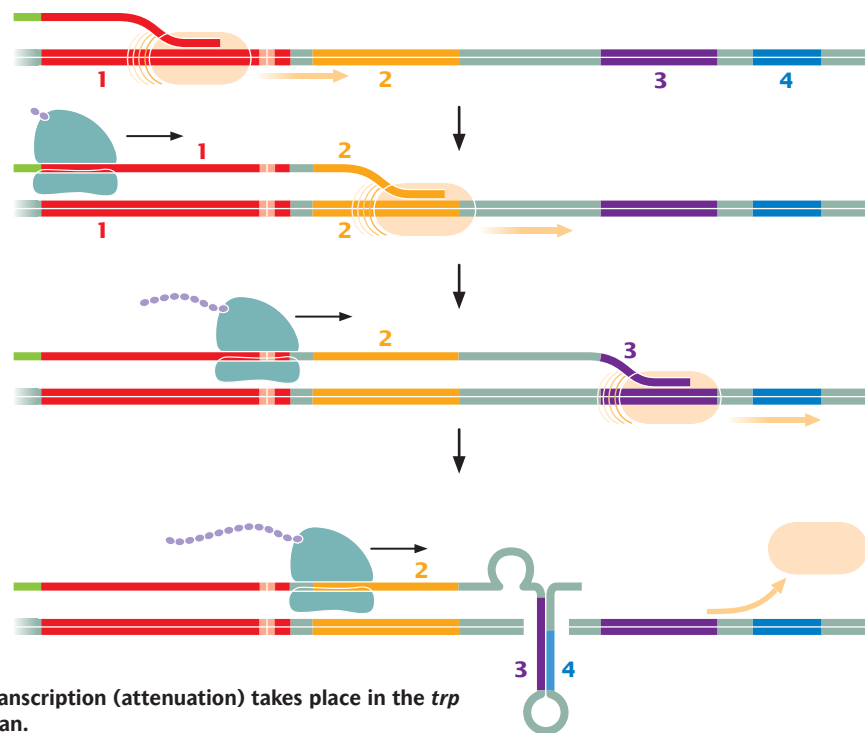
A ribosome binds to the 5' end of the 5' UTR and translates region 1 while region 2 is being transcribed.

(g)

RNA polymerase transcribes region 3. The ribosome does not stall at the Trp codons, because tryptophan is abundant.

(h)

The ribosome covers part of region 2, preventing it from pairing with region 3. Region 4 is transcribed and pairs with region 3, producing the attenuator that terminates transcription.



16.17 Whether the premature termination of transcription (attenuation) takes place in the *trp* operon depends on the cellular level of tryptophan.



www.whfreeman.com/pierce4e

See if you can cause attenuation and antitermination by pausing the ribosome at the *trp* codon for different lengths of time in Animation 16.2.

See if you can cause attenuation and antitermination by pausing the ribosome at the *trp* codon for different lengths of time in Animation 16.2.

attenuator and transcription terminates just beyond region 4. The structural genes are not transcribed, no tryptophan-producing enzymes are translated, and no additional tryptophan is synthesized. Important events in the process of attenuation are summarized in **Table 16.2**.

A key factor controlling attenuation is the number of tRNA molecules charged with tryptophan, because their availability is what determines whether the ribosome stalls at the tryptophan codons. A second factor concerns the synchronization of transcription and translation, which is

Table 16.2 Events in the process of attenuation

Intracellular Level of Tryptophan	Ribosome Stalls at Trp Codons	Position of Ribosome When Region 3 Is Transcribed	Secondary Structure of 5' UTR	Termination of Transcription of <i>trp</i> Operon
High	No	Covers region 2	3+4 hairpin	Yes
Low	Yes	Covers region 1	2+3 hairpin	No

critical to attenuation. Synchronization is achieved through a pause site located in region 1 of the 5' UTR. After initiating transcription, RNA polymerase stops temporarily at this site, which allows time for a ribosome to bind to the 5' end of the mRNA so that translation can closely follow transcription. It is important to point out that ribosomes do not traverse the convoluted hairpins of the 5' UTR to translate the structural genes. Ribosomes that attach to the 5' end of region 1 of the mRNA encounter a stop codon at the end of region 1. New ribosomes translating the structural genes attach to a different ribosome-binding site located near the beginning of the *trpE* gene. **TRY PROBLEM 24** →

Why Does Attenuation Take Place in the *trp* Operon?

Why do bacteria need attenuation in the *trp* operon? Shouldn't repression at the operator site prevent transcription from taking place when tryptophan levels in the cell are high? Why does the cell have two types of control? Part of the answer is that repression is never complete; some transcription is initiated even when the *trp* repressor is active; repression reduces transcription only as much as 70-fold. Attenuation can further reduce transcription another 8- to 10-fold; so together the two processes are capable of reducing transcription of the *trp* operon more than 600-fold. Both mechanisms provide *E. coli* with a much finer degree of control over tryptophan synthesis than either could achieve alone.

Another reason for the dual control is that attenuation and repression respond to different signals: repression responds to the cellular levels of tryptophan, whereas attenuation responds to the number of tRNAs charged with tryptophan. There may be times when a cell's ability to respond to these different signals is advantageous. Finally, the *trp* repressor affects several operons other than the *trp* operon. At an earlier stage in the evolution of *E. coli*, the *trp* operon may have been controlled only by attenuation. The *trp* repressor may have evolved primarily to control the other operons and only incidentally affects the *trp* operon.

Attenuation is a difficult process to grasp because you must simultaneously visualize how two dynamic processes—transcription and translation—interact, and it's easy to confuse the two processes. Remember that attenuation refers to the early termination of *transcription*, not translation (although events in translation bring about the termination of transcription). Attenuation often causes confusion

because we know that transcription must precede translation. We're comfortable with the idea that transcription might affect translation, but it's harder to imagine that the effects of translation could influence transcription, as they do in attenuation. The reality is that transcription and translation are closely coupled in prokaryotic cells, and events in one process can easily affect the other.

CONCEPTS

In attenuation, transcription is initiated but terminates prematurely. When tryptophan levels are low, the ribosome stalls at the tryptophan codons and transcription continues. When tryptophan levels are high, the ribosome does not stall at the tryptophan codons, and the 5' UTR adopts a secondary structure that terminates transcription before the structural genes can be copied into RNA (attenuation).

✓ CONCEPT CHECK 9

Attenuation results when which regions of the 5' UTR pair?

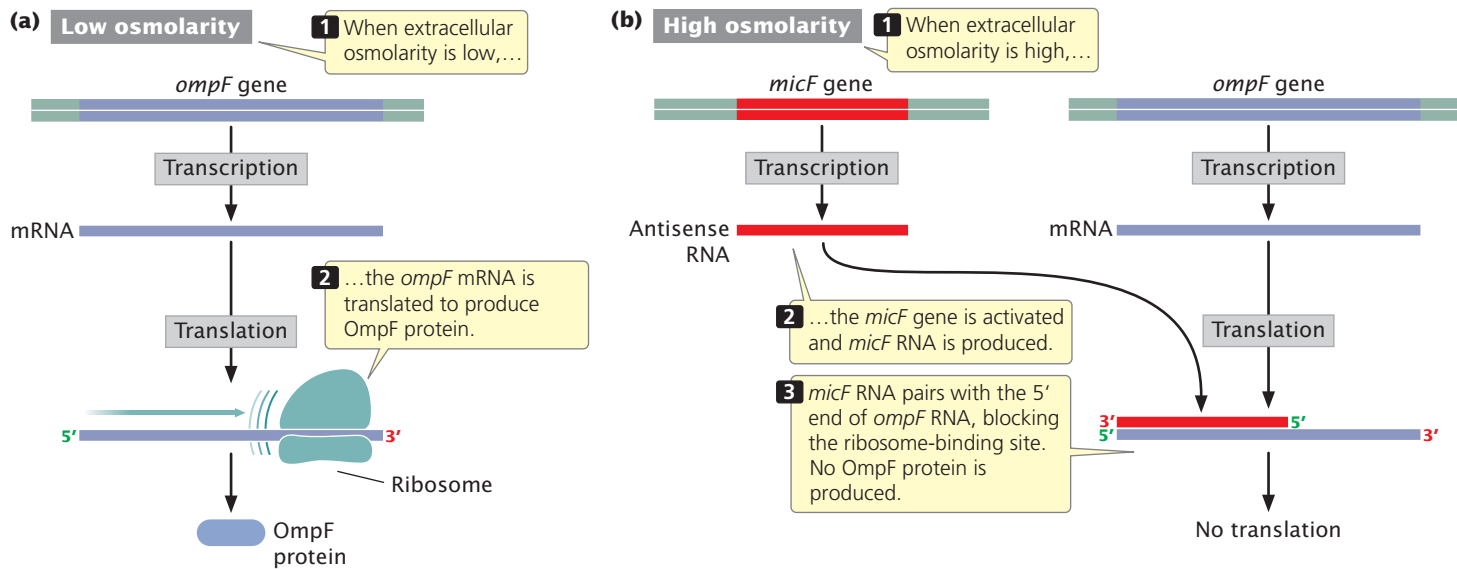
- a. 1 and 3
- b. 2 and 3
- c. 2 and 4
- d. 3 and 4

16.4 RNA Molecules Control the Expression of Some Bacterial Genes

All the regulators of gene expression that we have considered so far have been proteins. Several examples of RNA regulators also have been discovered.

Antisense RNA

Some small RNA molecules are complementary to particular sequences on mRNAs and are called **antisense RNA**. They control gene expression by binding to sequences on mRNA and inhibiting translation. Translational control by antisense RNA is seen in the regulation of the *ompF* gene of *E. coli* (**Figure 16.18a**). This gene encodes an outer-membrane protein that functions as a channel for the passive diffusion of small polar molecules, such as water and ions, across the cell membrane. Under most conditions, the *ompF* gene is transcribed and translated and the OmpF protein is synthesized. However, when the osmolarity of the medium increases, the cell depresses the production of OmpF protein to help maintain cellular osmolarity. A regulator gene named *micF*—for



16.18 Antisense RNA can regulate translation.

mRNA-interfering complementary RNA—is activated and *micF* RNA is produced (Figure 16.18b). The *micF* RNA, an antisense RNA, binds to a complementary sequence in the 5' UTR of the *ompF* mRNA and inhibits the binding of the ribosome. This inhibition reduces the amount of translation (see Figure 16.18b), which results in fewer OmpF proteins in the outer membrane and thus reduces the detrimental movement of substances across the membrane owing to the changes in osmolarity. A number of examples of antisense RNA controlling gene expression have now been identified in bacteria and bacteriophages. [TRY PROBLEM 27 →](#)

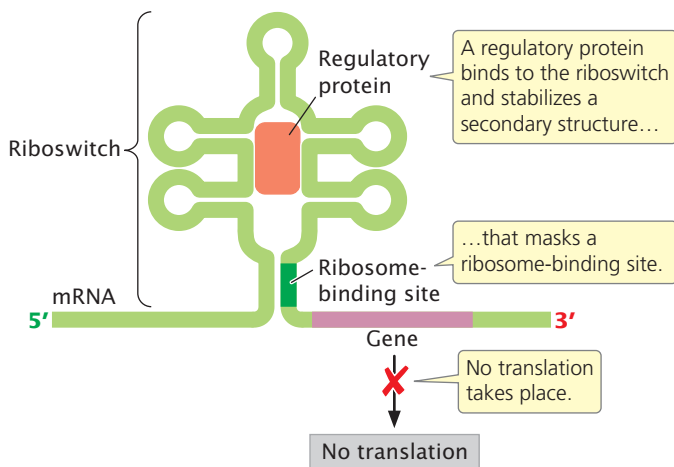
Riboswitches

We have seen that operons of bacteria contain DNA sequences (promoters and operator sites) where the binding of small molecules induces or represses transcription.

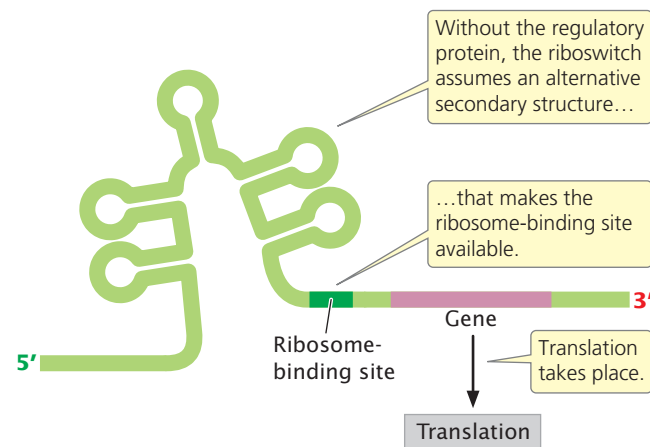
Some mRNA molecules contain regulatory sequences called **riboswitches**, where molecules can bind and affect gene expression by influencing the formation of secondary structures in the mRNA (Figure 16.19). Riboswitches were first discovered in 2002 and now appear to be common in bacteria, regulating about 4% of all bacterial genes. They are also present in archaea, fungi, and plants.

Riboswitches are typically found in the 5' UTR of the mRNA and fold into compact RNA secondary structures with a base stem and several branching hairpins. In some cases, a small regulatory molecule binds to the riboswitch and stabilizes a terminator, which causes premature termination of transcription. In other cases, the binding of a regulatory molecule stabilizes a secondary structure that masks the ribosome-binding site, preventing the initiation of translation. When not bound by the regulatory molecule, the riboswitch assumes an alternative structure that elimi-

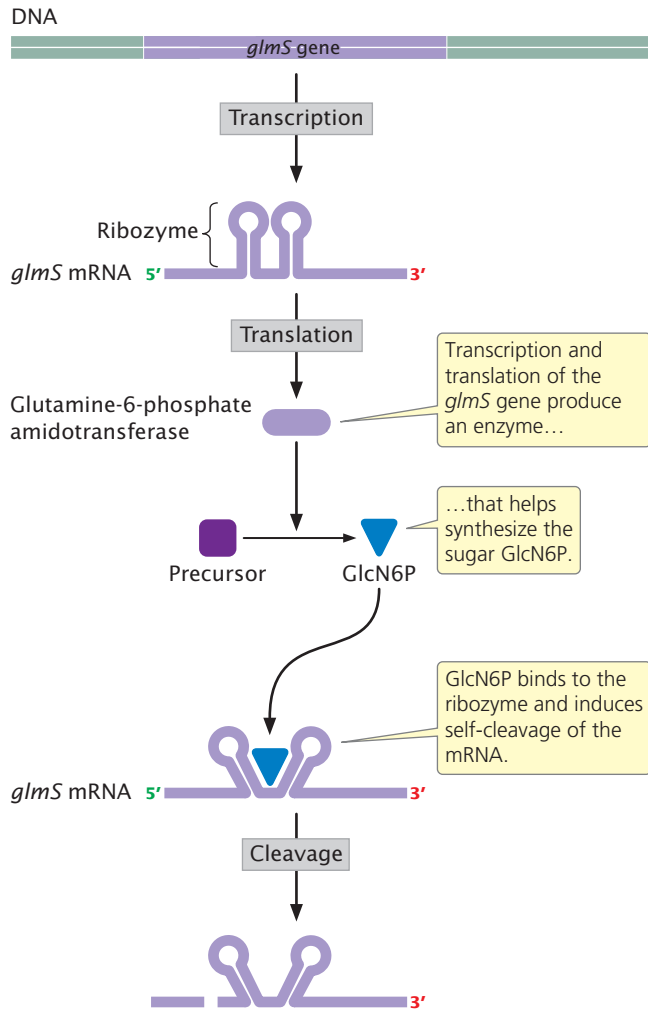
(a) Regulatory protein present



(b) Regulatory protein absent



16.19 Riboswitches are RNA sequences in mRNA that affect gene expression.



16.20 Ribozymes, when bound with small regulatory molecules, can induce the cleavage and degradation of mRNA.

notes the premature terminator or makes the ribosome-binding site available.

A riboswitch and the synthesis of vitamin B12 An example of a riboswitch is seen in bacterial genes that encode enzymes having roles in the synthesis of vitamin B12. The genes for these enzymes are transcribed into an mRNA molecule that has a riboswitch. When the activated form of

vitamin B12—called coenzyme B12—is present, it binds to the riboswitch, the mRNA folds into a secondary structure that obstructs the ribosome-binding site, and no translation of the mRNA takes place. In the absence of coenzyme B12, the mRNA assumes a different secondary structure. This secondary structure does not obstruct the ribosome-binding site, and so translation is initiated, the enzymes are synthesized, and more vitamin B12 is produced. For some riboswitches, the regulatory molecule acts as a repressor (as just described) by inhibiting transcription or translation; for others, the regulatory molecule acts as an inducer by causing the formation of a secondary structure that allows transcription or translation to take place.

Riboswitches That Function As Ribozymes

Another type of gene control is carried out by mRNA molecules called ribozymes, which possess catalytic activity (see Chapter 14). Termed RNA-mediated repression, this type of control has been demonstrated in the *glmS* gene of the bacterium *Bacillus subtilis*. Transcription of this gene produces an mRNA molecule that encodes the enzyme glutamine-fructose-6-phosphate amidotransferase (Figure 16.20), which helps synthesize a small sugar called glucosamine-6-phosphate (GlcN6P). Within the 5' UTR of the *glmS* mRNA are about 75 nucleotides that act as a ribozyme. When GlcN6P is absent, the *glmS* gene is transcribed and translated to produce the enzyme, which synthesizes more GlcN6P. However, when sufficient GlcN6P is present, it binds to the ribozyme part of the mRNA, which then induces self-cleavage of the mRNA and prevents its translation.

CONCEPTS

Antisense RNA is complementary to other RNA or DNA sequences. In bacterial cells, it can inhibit translation by binding to sequences in the 5' UTR of mRNA and preventing the attachment of the ribosome. Riboswitches are sequences in mRNA molecules that bind regulatory molecules and induce changes in the secondary structure of the mRNA that affect gene expression. In RNA-mediated repression, a ribozyme sequence on the mRNA induces self-cleavage and degradation of the mRNA when bound by a regulatory molecule.

CONCEPTS SUMMARY

- Gene expression can be controlled at different levels, including the alteration of DNA or chromatin structure, transcription, mRNA processing, RNA stability, translation, and posttranslational modification. Much of gene regulation is through the action of regulatory proteins binding to specific sequences in DNA.
- Genes in bacterial cells are typically clustered into operons—groups of functionally related structural genes and the sequences that control their transcription. Structural genes in an operon are transcribed together as a single mRNA molecule.
- In negative control, a repressor protein binds to DNA and inhibits transcription. In positive control, an activator protein binds to DNA and stimulates transcription. In inducible operons, transcription is normally off and must be turned on; in repressible operons, transcription is normally on and must be turned off.