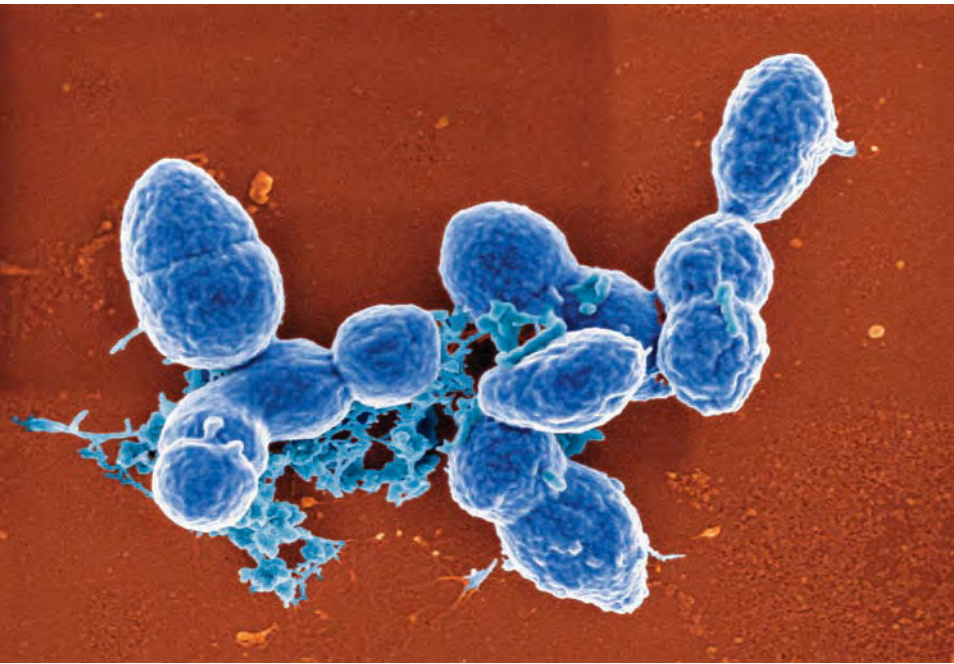


# 16 Control of Gene Expression in Prokaryotes



**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

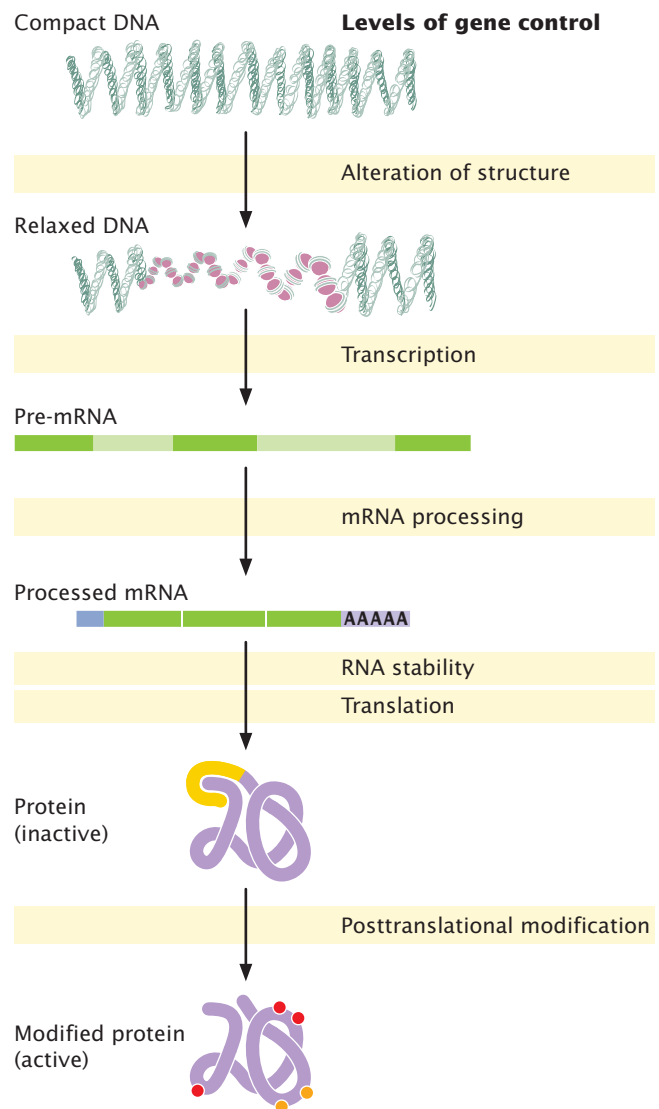
Genes are DNA sequences that are transcribed into RNA. Regulatory elements are DNA sequences that are not transcribed but affect the expression of genes. Positive control includes mechanisms that stimulate gene expression, whereas negative control inhibits gene expression.

#### ✓ 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

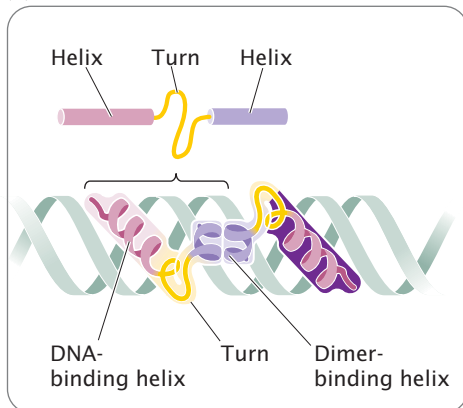
Why is transcription a particularly important level of gene regulation in both bacteria and eukaryotes?

## DNA-Binding Proteins

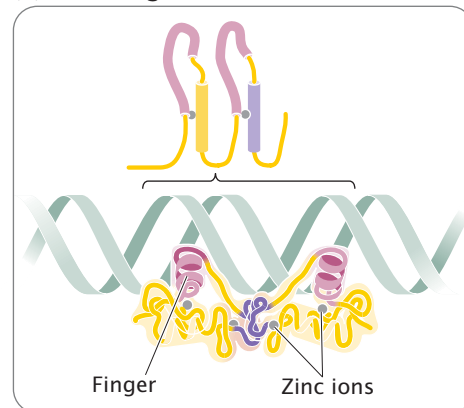
Much of gene regulation in bacteria and eukaryotes is accomplished by proteins that bind to DNA sequences and affect their expression. These regulatory proteins generally have discrete functional parts—called **domains**, typically consisting of 60 to 90 amino acids—that are responsible for binding to 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

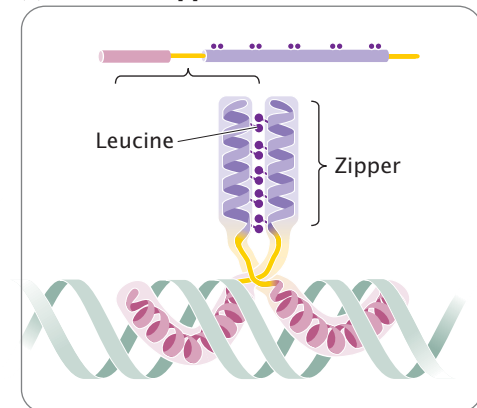
(a) Helix-turn-helix



(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

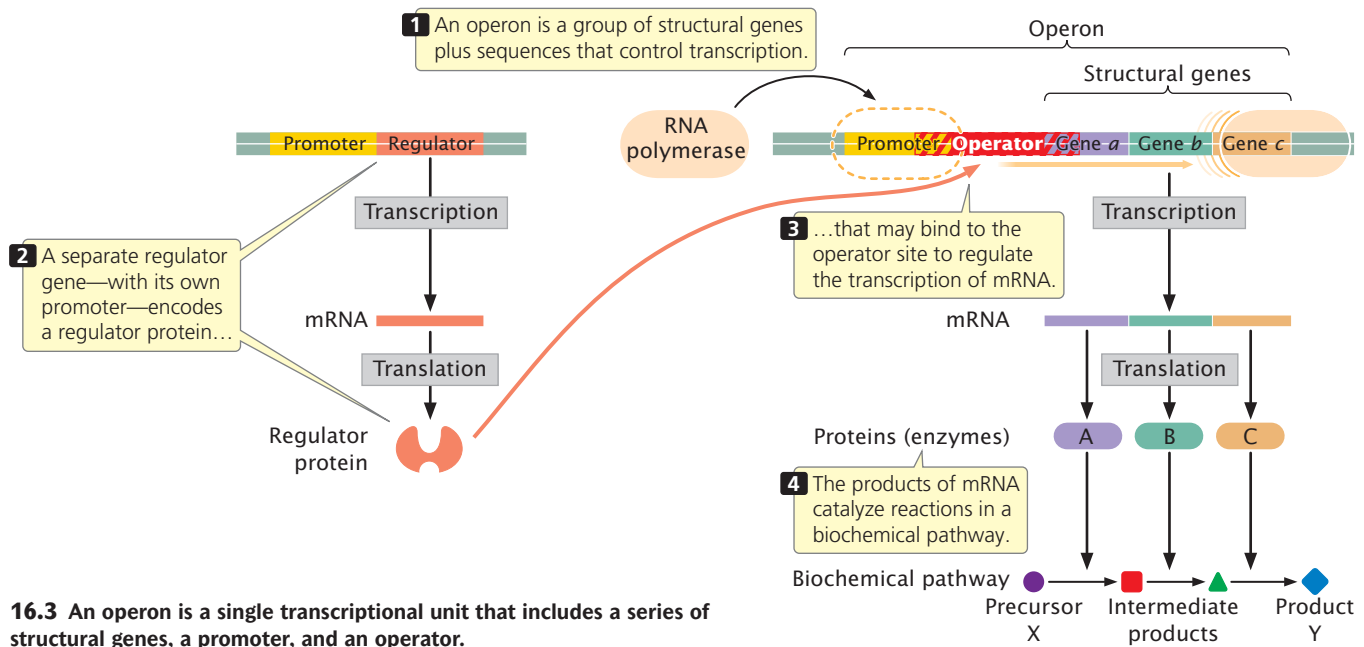
### CONCEPTS

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.



## 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).

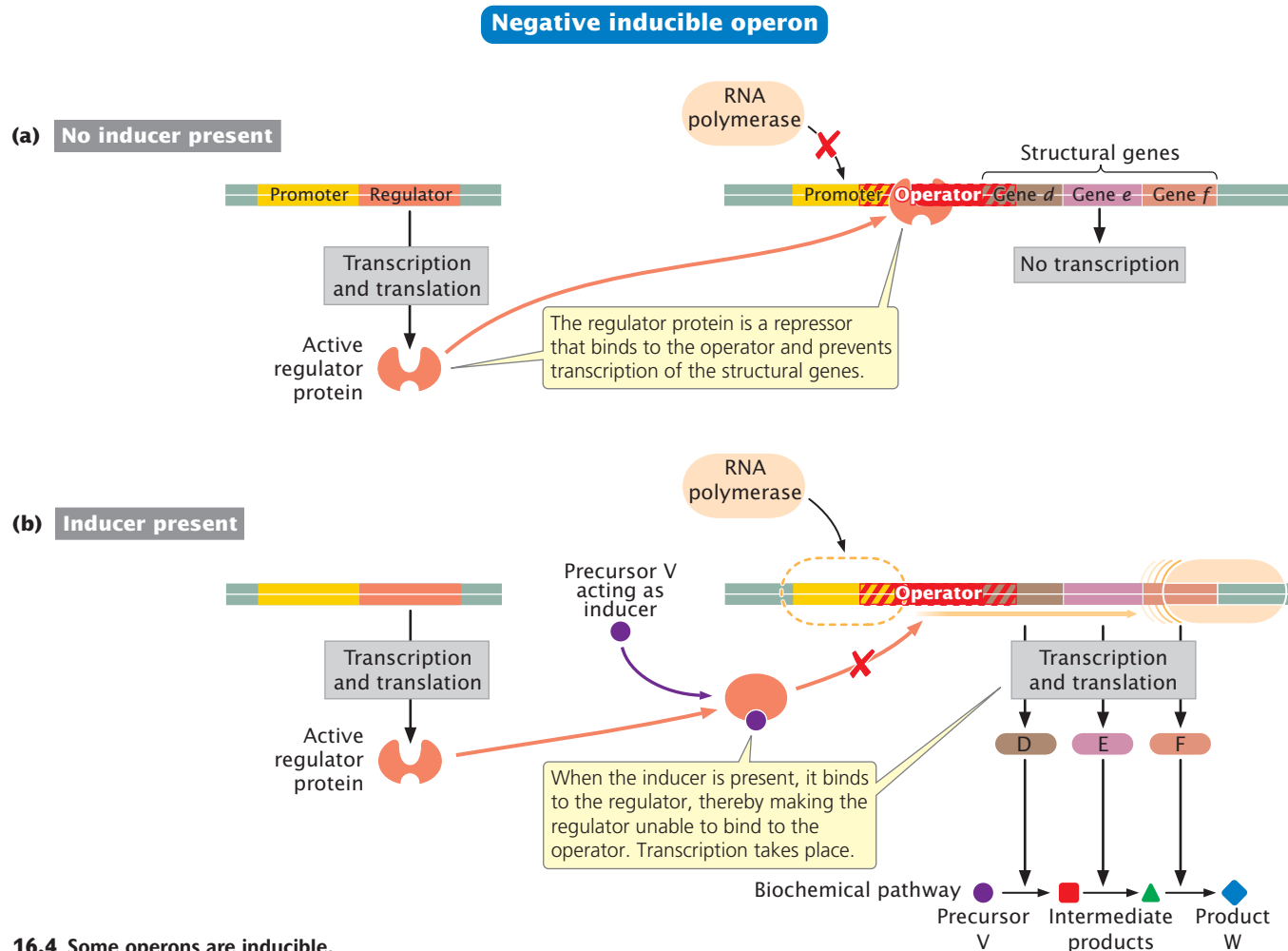
Transcription is turned on when a small molecule, an **inducer**, binds to the repressor (**Figure 16.4b**). Regulatory proteins frequently have two binding sites: one that binds to DNA and another that binds to a small molecule such as an inducer. The binding of the inducer (precursor V in Figure

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

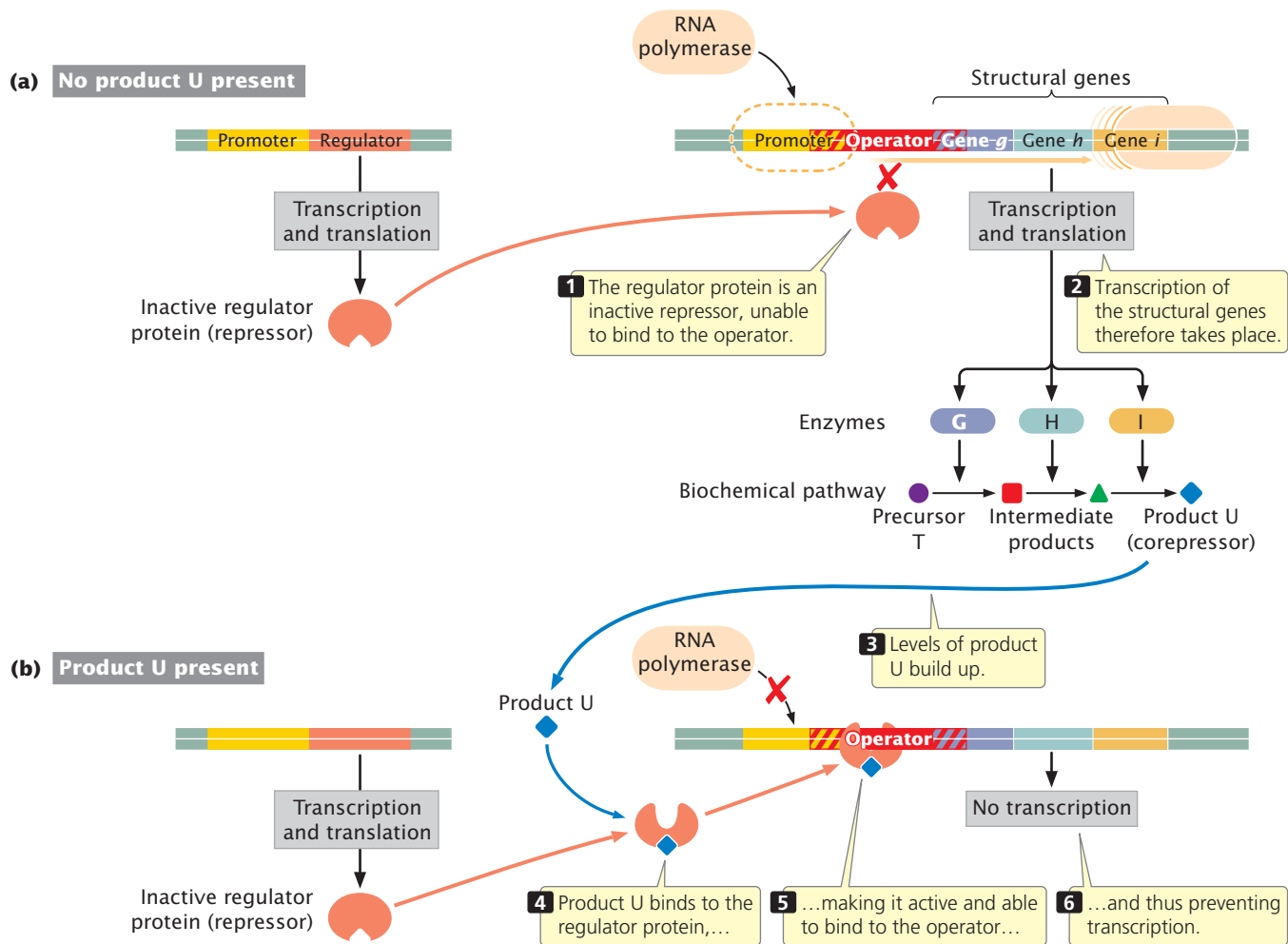
amino acids. For these types of operons, repressible control makes sense because the product produced by the proteins is always needed by the cell. Thus, these operons are normally on and are turned off when there are adequate amounts of the product already present.

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

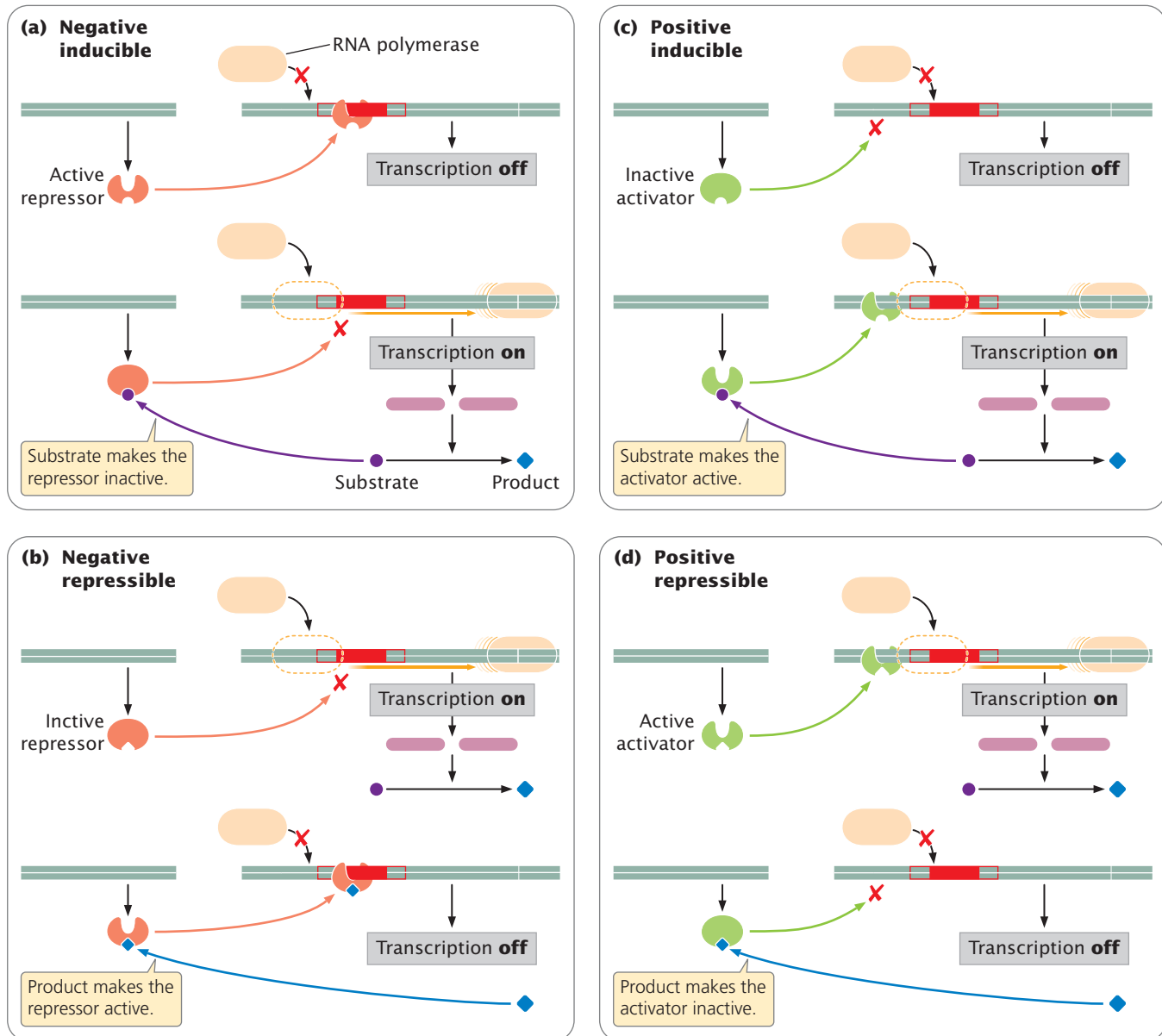
In a negative repressible operon, the regulator protein is synthesized as

- 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





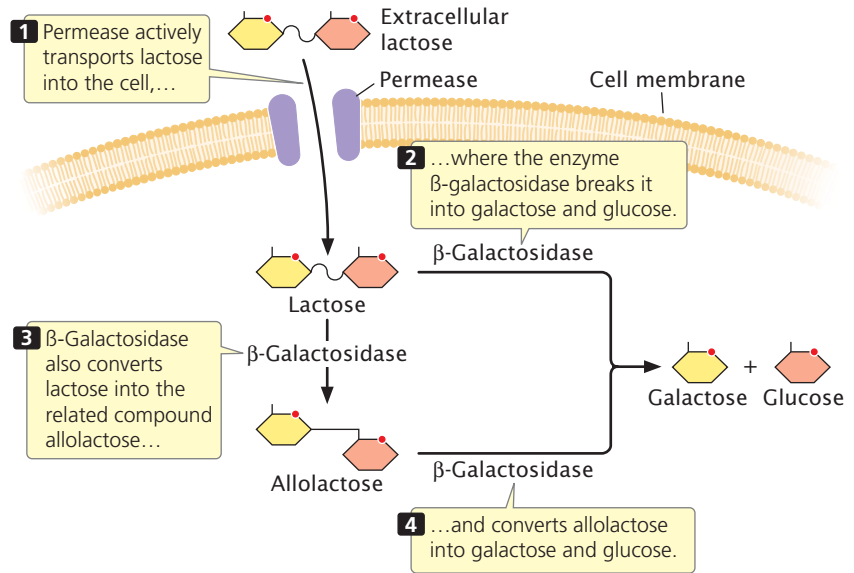
**16.6** A summary of the characteristics of positive and negative control in inducible and repressible operons.

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  $\beta$ -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.

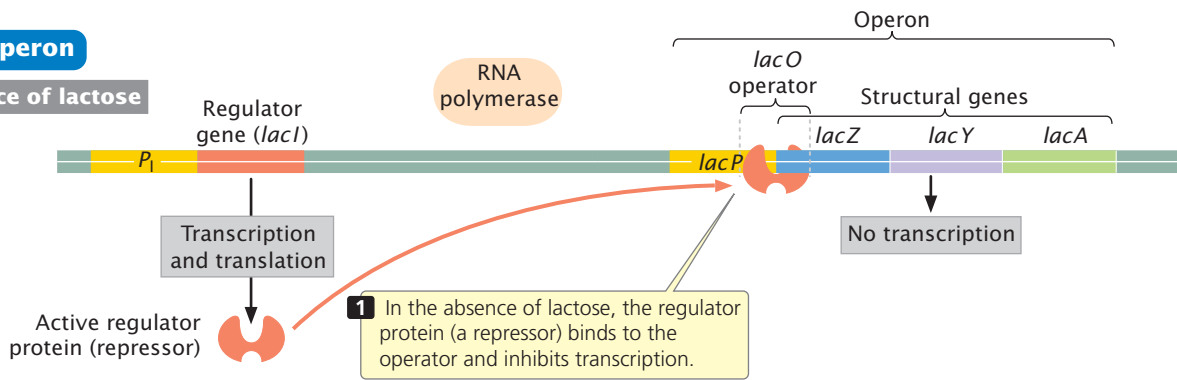
**Regulation of the *lac* operon** The *lac* operon is an example of a negative inducible operon. The enzymes  $\beta$ -galactosidase, permease, and transacetylase are encoded by adjacent structural genes in the *lac* operon of *E. coli*



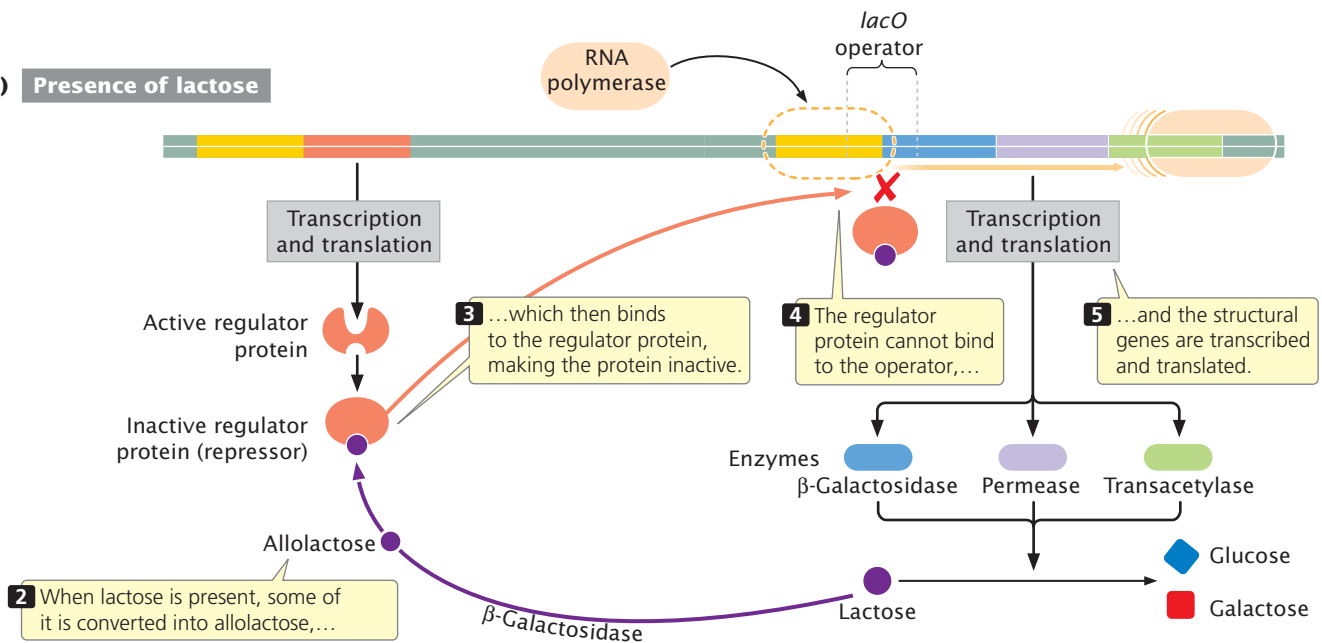
**16.7** Lactose, a major carbohydrate found in milk, consists of 2 six-carbon sugars linked together.

**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).  $\beta$ -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  $\beta$ -galactosidase. If  $\beta$ -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  $\beta$ -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  $\beta$ -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  $\beta$ -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  $\beta$ -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.

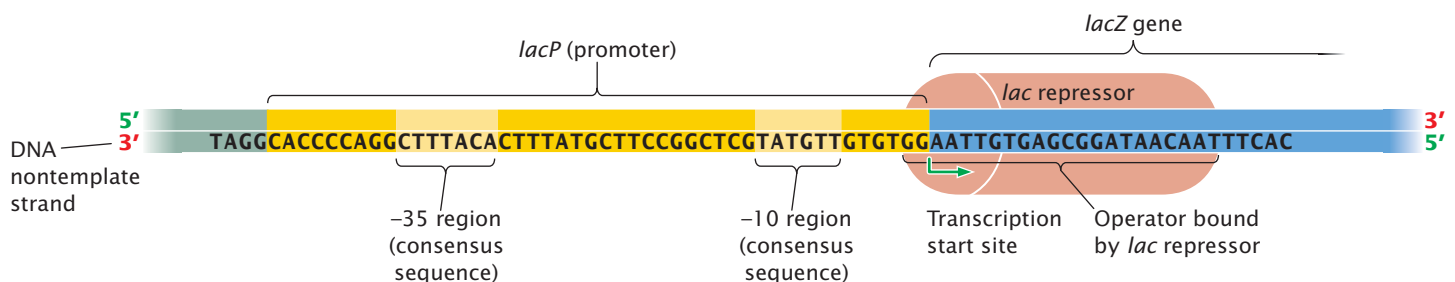
### ✓ CONCEPT CHECK 6

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).

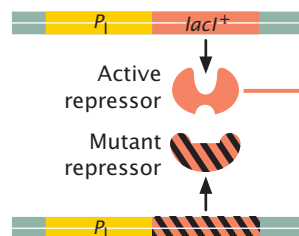
**Structural-gene mutations** Jacob and Monod first discovered some mutant strains that had lost the ability to synthesize either  $\beta$ -galactosidase or permease. (They did not study in detail the effects of mutations on the transacetylase enzyme, and so transacetylase will not be considered here.) 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.

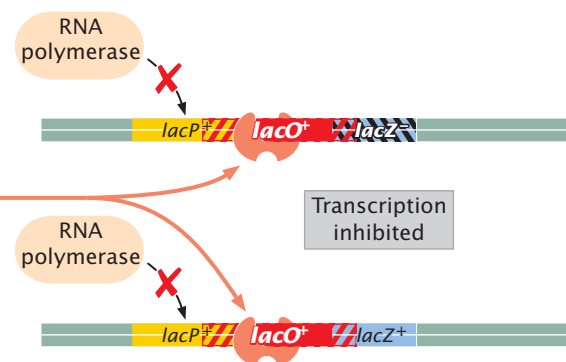
Through the use of partial diploids, Jacob and Monod were able to establish that mutations at the *lacZ* and *lacY* genes were independent and usually affected only the product of the gene in which they occurred. Partial diploids with *lacZ*<sup>+</sup> *lacY*<sup>-</sup> on the bacterial chromosome and *lacZ*<sup>-</sup> *lacY*<sup>+</sup> on the plasmid functioned normally, producing  $\beta$ -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*<sup>+</sup> *lacY*<sup>-</sup>/*lacZ*<sup>-</sup> *lacY*<sup>+</sup>.) In this partial diploid, a single functional  $\beta$ -galactosidase gene (*lacZ*<sup>+</sup>) is sufficient to produce  $\beta$ -galactosidase; whether the functional  $\beta$ -galactosidase gene is coupled to a functional (*lacY*<sup>+</sup>) or a defective (*lacY*<sup>-</sup>) permease gene makes no difference. The same is true of the *lacY*<sup>+</sup> 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  $\beta$ -galactosidase and permease because genes for both proteins are in the same operon and are regulated coordinately.

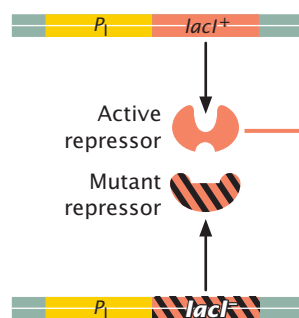
**(a) Absence of lactose**



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

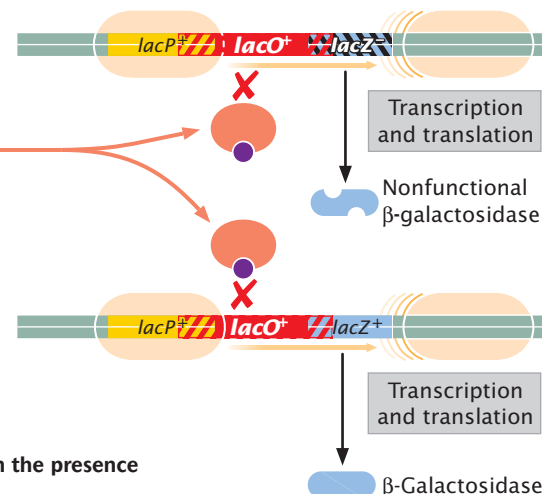


**(b) Presence of lactose**

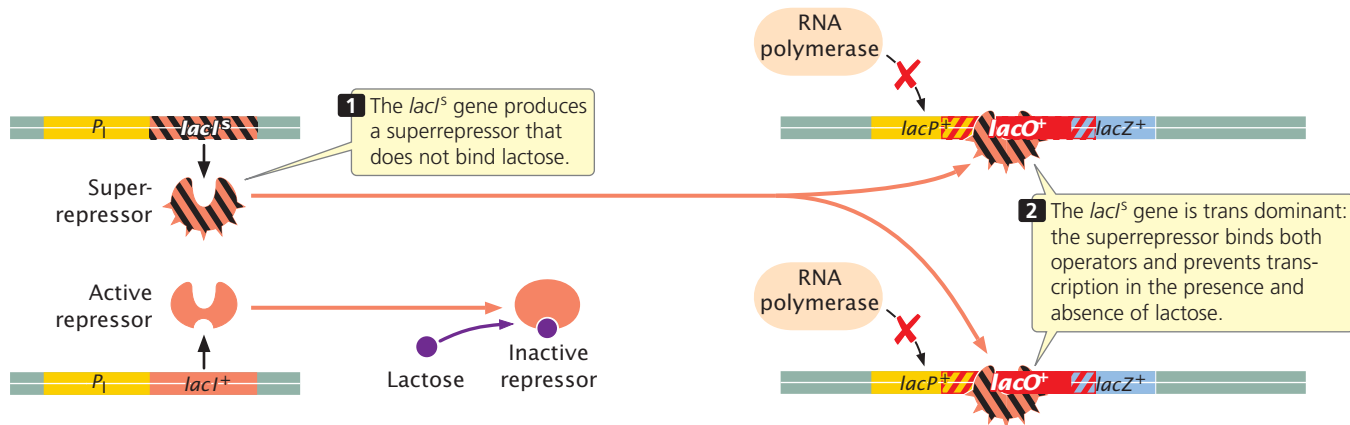


**2** When lactose is present, it inactivates the repressor, and functional  $\beta$ -galactosidase is produced from the *lacZ*<sup>+</sup> gene.

Lactose  
Inactive repressor



**16.10** The partial diploid *lacI*<sup>+</sup> *lacZ*<sup>-</sup>/*lacI*<sup>-</sup> *lacZ*<sup>+</sup> produces  $\beta$ -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  $\beta$ -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  $\beta$ -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  $\beta$ -galactosidase or permease, whether or not lactose was present (Figure 16.11).

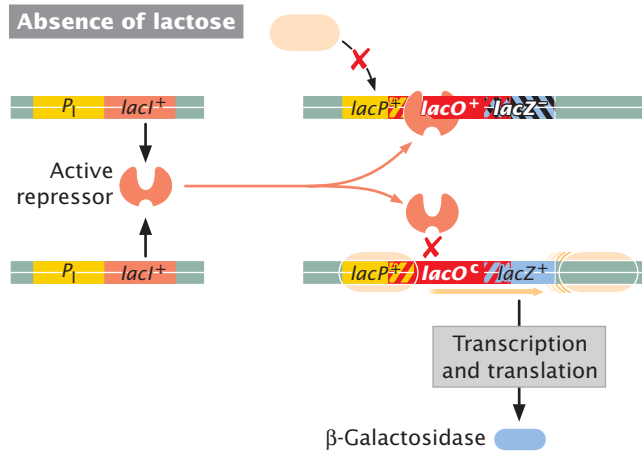
**Operator mutations** Jacob and Monod mapped another class of constitutive mutants to a site adjacent to  $lacZ$ . These mutations occurred at the operator site and were referred

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  $\beta$ -galactosidase, indicating that  $lacO^c$  is dominant over  $lacO^+$ .

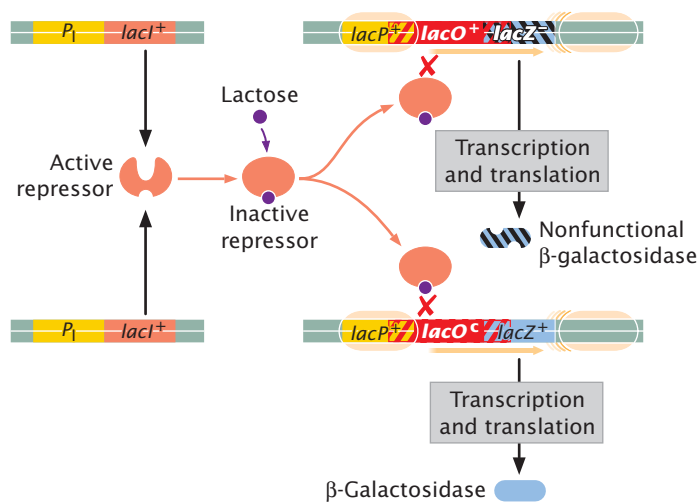
Analysis of other partial diploids showed that the  $lacO$  gene is cis acting, affecting only genes on the same DNA molecule. For example, a partial diploid with genotype  $lacI^+ lacO^+ lacZ^- / lacI^+ lacO^c lacZ^+$  was constitutive, producing  $\beta$ -galactosidase in the presence or absence of lactose (Figure 16.12a), but a partial diploid with genotype  $lacI^+ lacO^+ lacZ^+ / lacI^+ lacO^c lacZ^-$  produced  $\beta$ -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

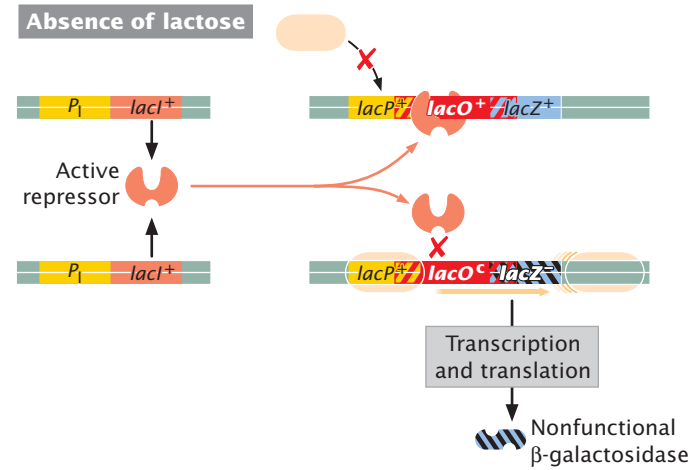
**(a) Partial diploid  $lacI^+ lacO^+ lacZ^- / lacI^+ lacO^c lacZ^+$**



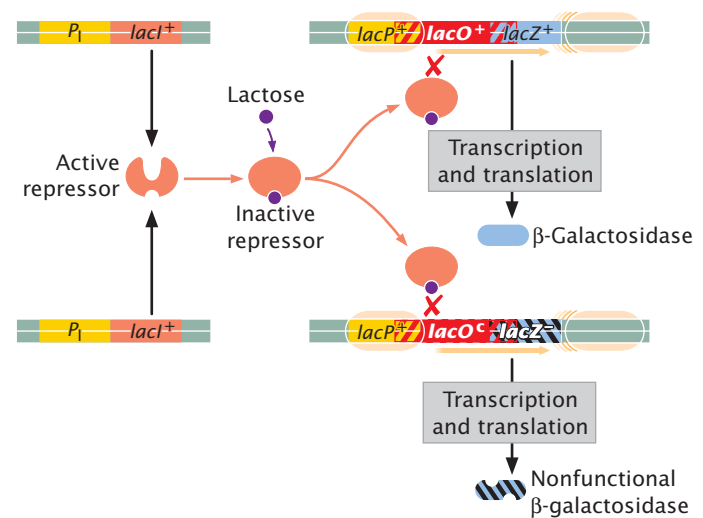
**Presence of lactose**



**(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  $\beta$ -galactosidase in the presence and absence of lactose.

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

 [www.whfreeman.com/pierce4e](http://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  $\beta$ -galactosidase, whereas  $lacI^+ lacP^- lacZ^+ / lacI^+ lacP^+ lacZ^-$  fails to produce  $\beta$ -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  $\beta$ -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  $\beta$ -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  $\beta$ -galactosidase and permease are produced.
- b. The structural *lacZ* gene is mutated; so  $\beta$ -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,  $\beta$ -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  $\beta$ -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  $\beta$ -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  $\beta$ -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.

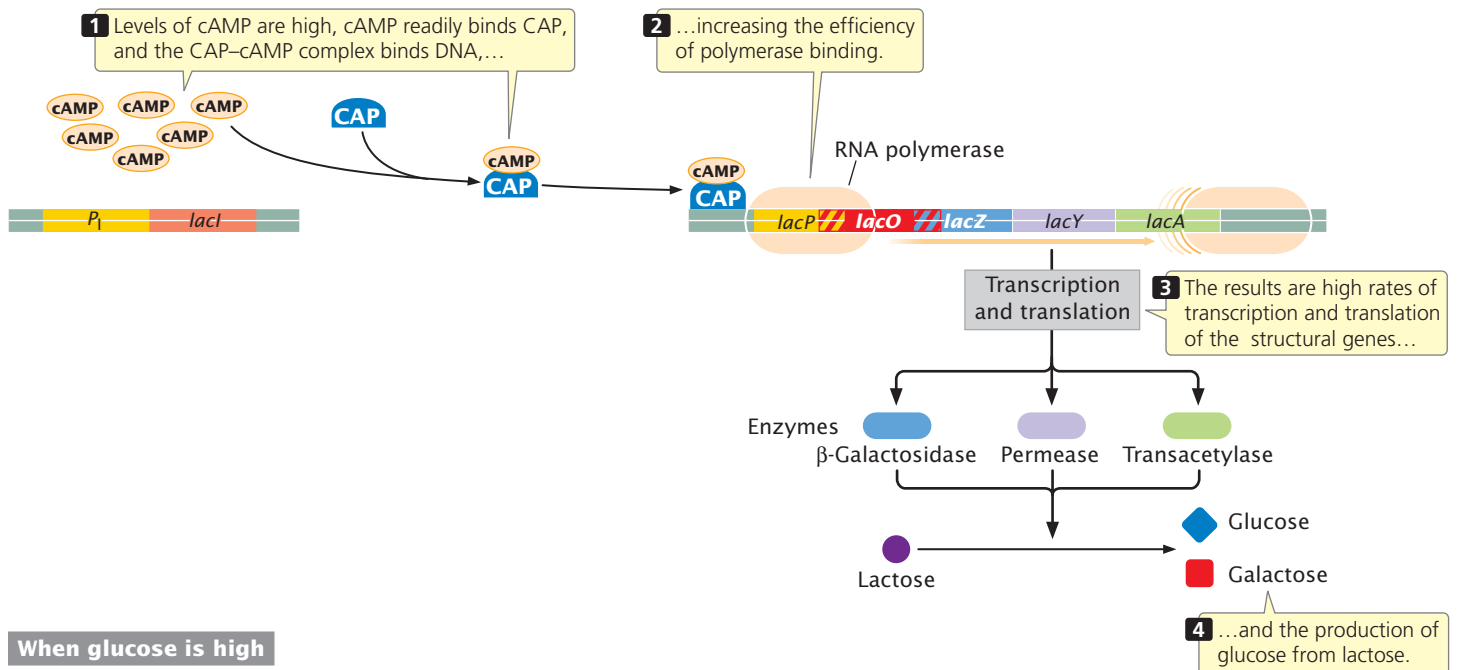
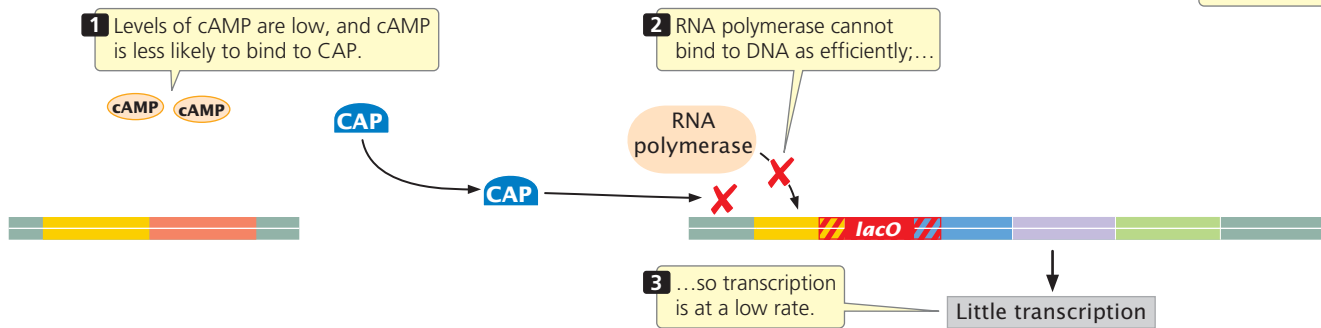
Lactose absent		Lactose present	
$\beta$ -Galactosidase	Permease	$\beta$ -Galactosidase	Permease
–	–	+	+
–	+	–	+
–	–	–	–
–	–	+	+

## 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**

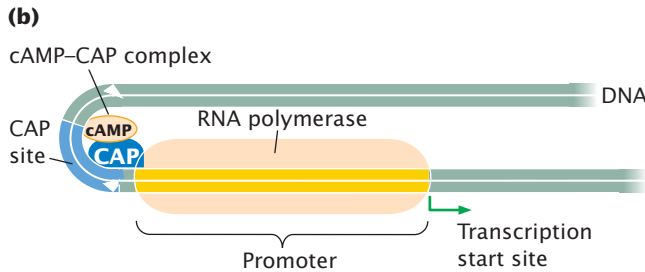
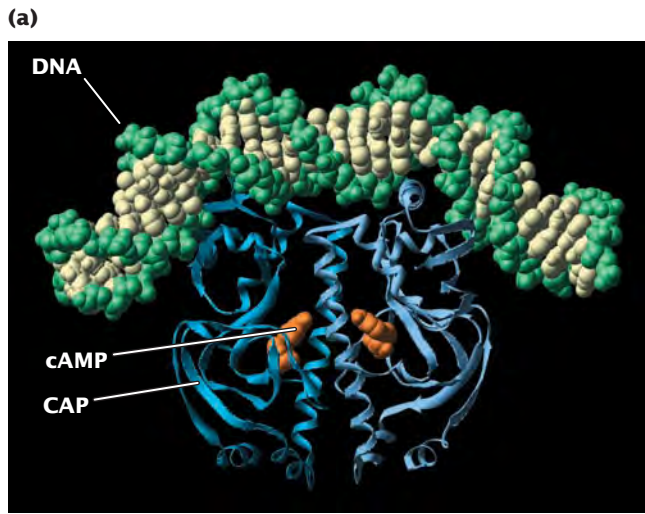
What is the effect of high levels of glucose on the *lac* operon?

- 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





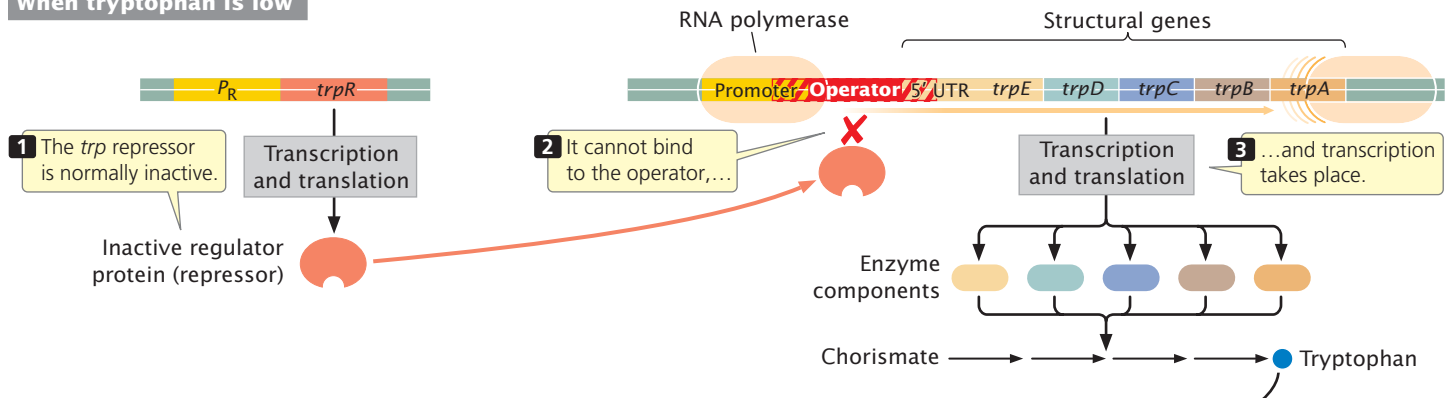
**16.14** The binding of the cAMP-CAP complex to DNA produces a sharp bend in DNA that activates transcription.

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.

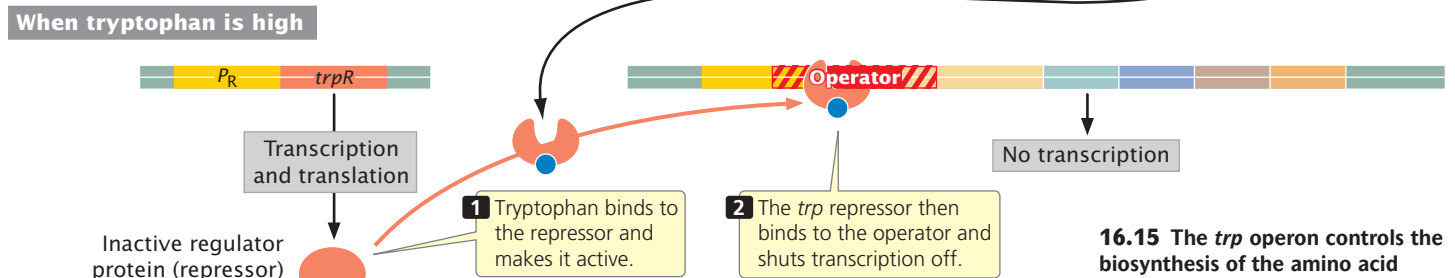
The *trp* operon contains five structural genes (*trpE*, *trpD*, *trpC*, *trpB*, and *trpA*) that produce the components of three enzymes (two of the enzymes consist of two polypeptide chains). These enzymes convert chorismate into tryptophan (Figure 16.15). The first structural gene, *trpE*, contains 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.

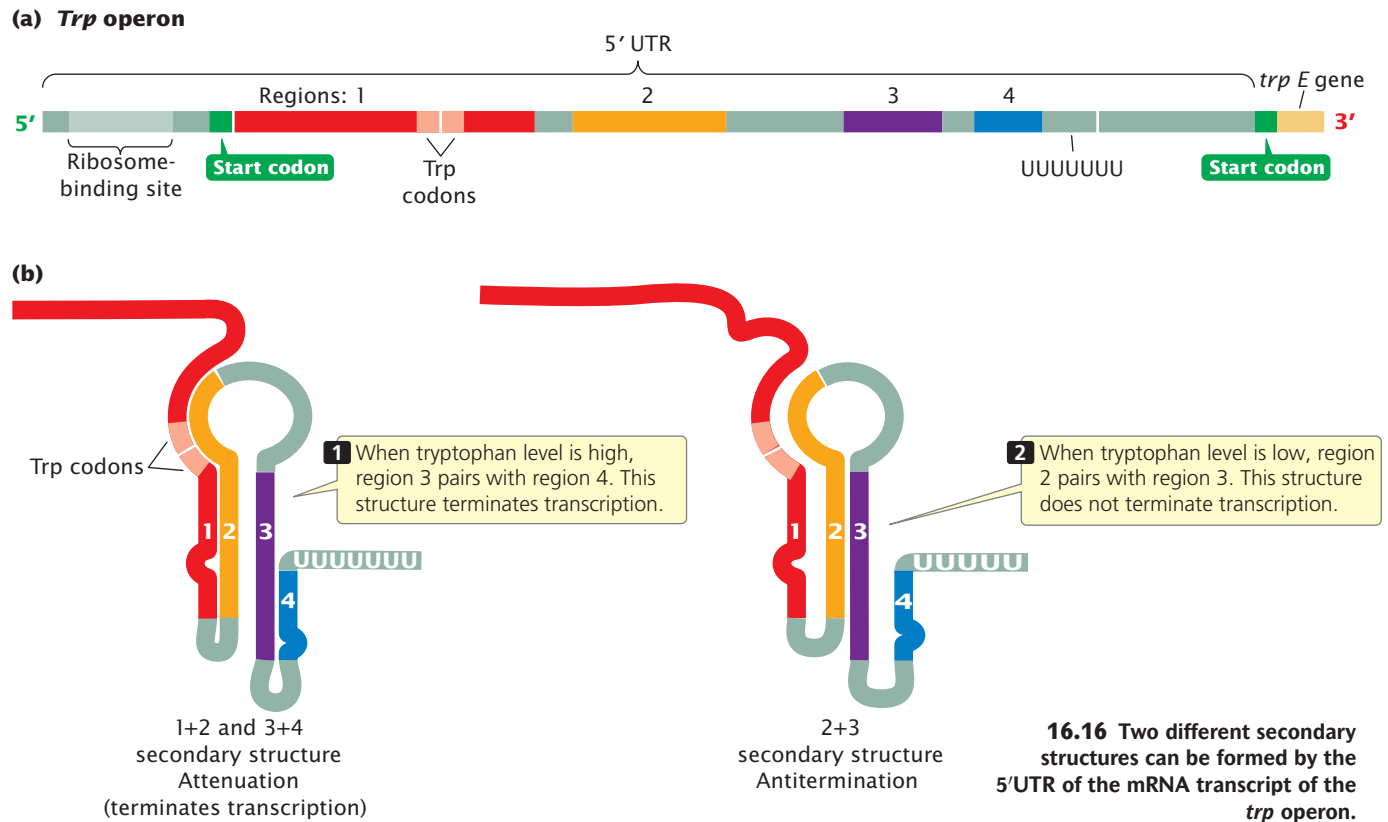
Close examination of the *trp* operon reveals a region of 162 nucleotides that corresponds to the long 5' UTR of the mRNA (mentioned earlier) transcribed from the *trp* operon (**Figure 16.16a**). The 5' UTR (also called a leader) contains four regions: region 1 is complementary to region 2, region 2 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)

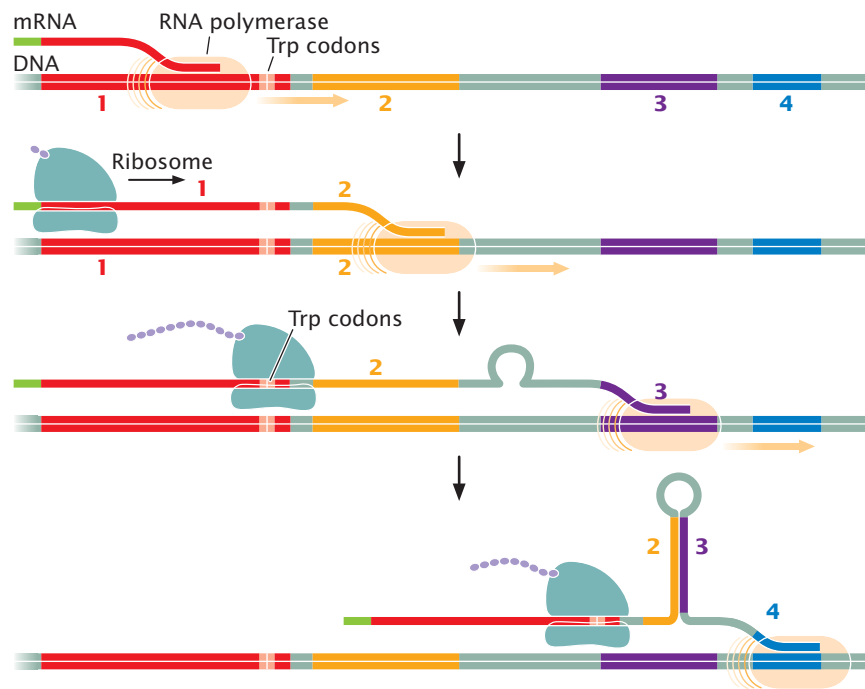
A ribosome attaches to the 5' end of the 5' UTR and translates region 1 while region 2 is being transcribed.

(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)

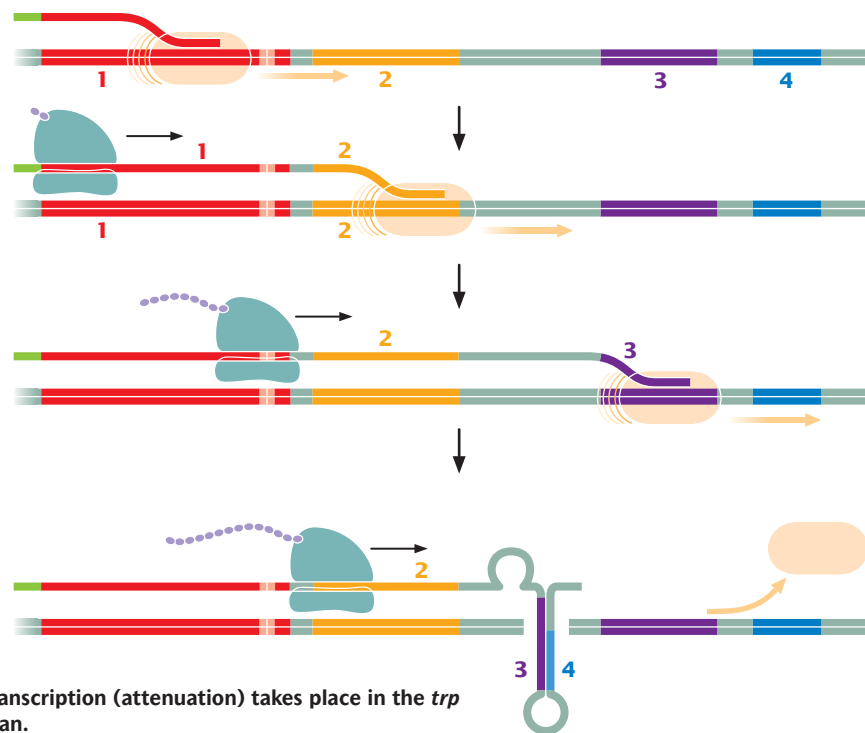
A ribosome binds to the 5' end of the 5' UTR and translates region 1 while region 2 is being transcribed.

(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](http://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.

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?

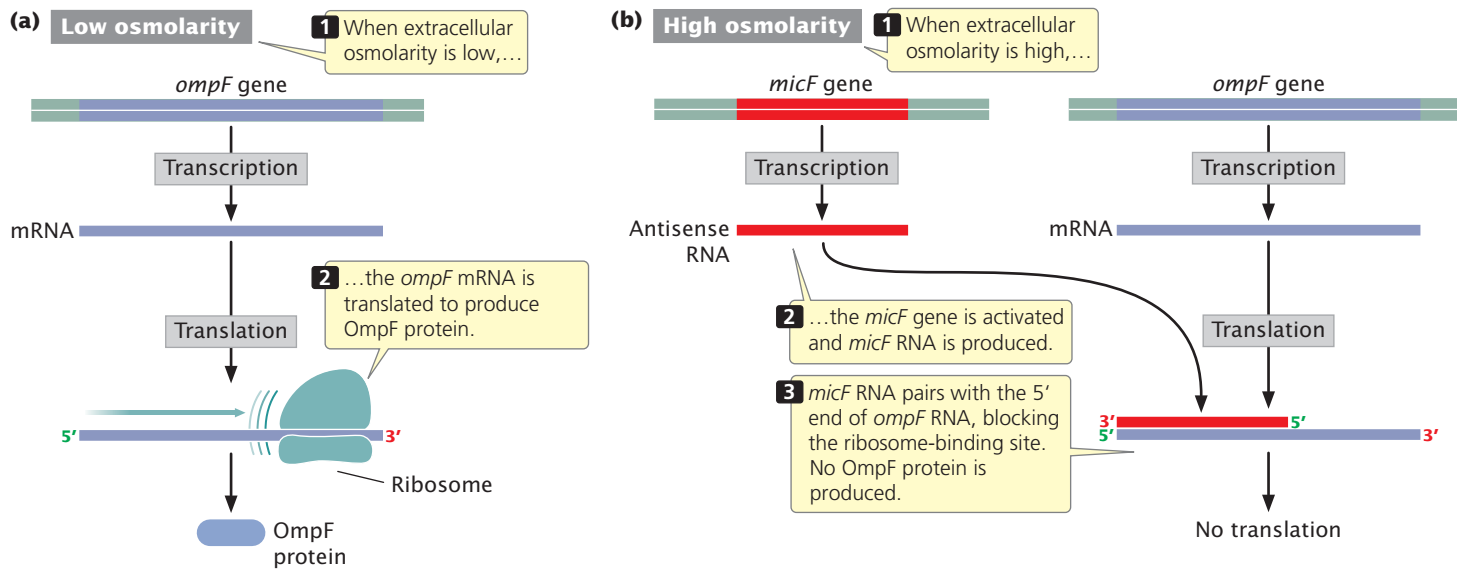
- a. 1 and 3
- b. 2 and 3
- c. 2 and 4
- d. 3 and 4

## 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 →](#)

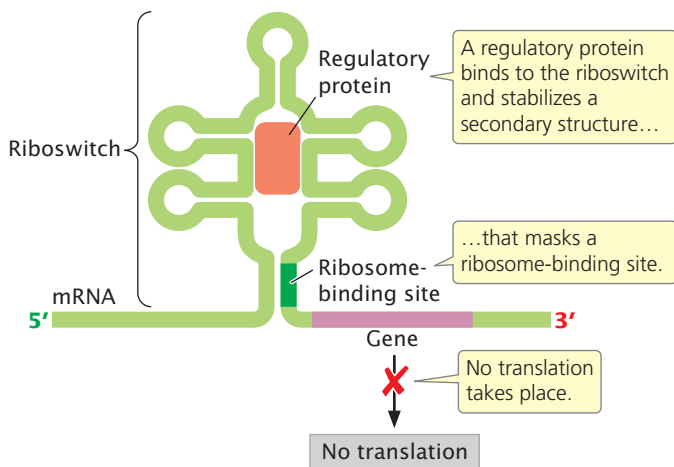
## Riboswitches

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

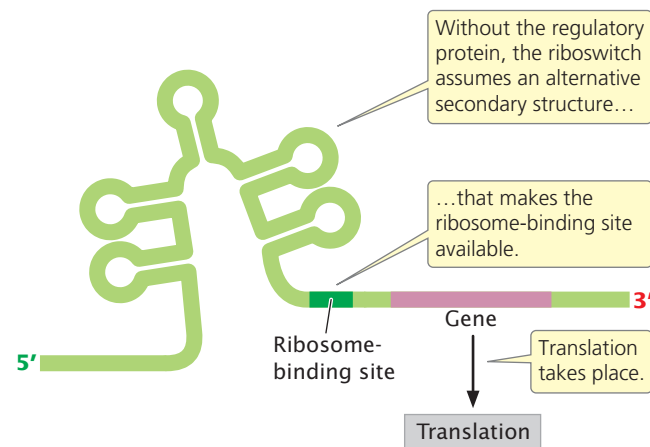
Some mRNA molecules contain regulatory sequences called **riboswitches**, where molecules can bind and affect gene expression by influencing the formation of secondary structures in the mRNA (Figure 16.19). Riboswitches were first discovered in 2002 and now appear to be common in bacteria, regulating 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-

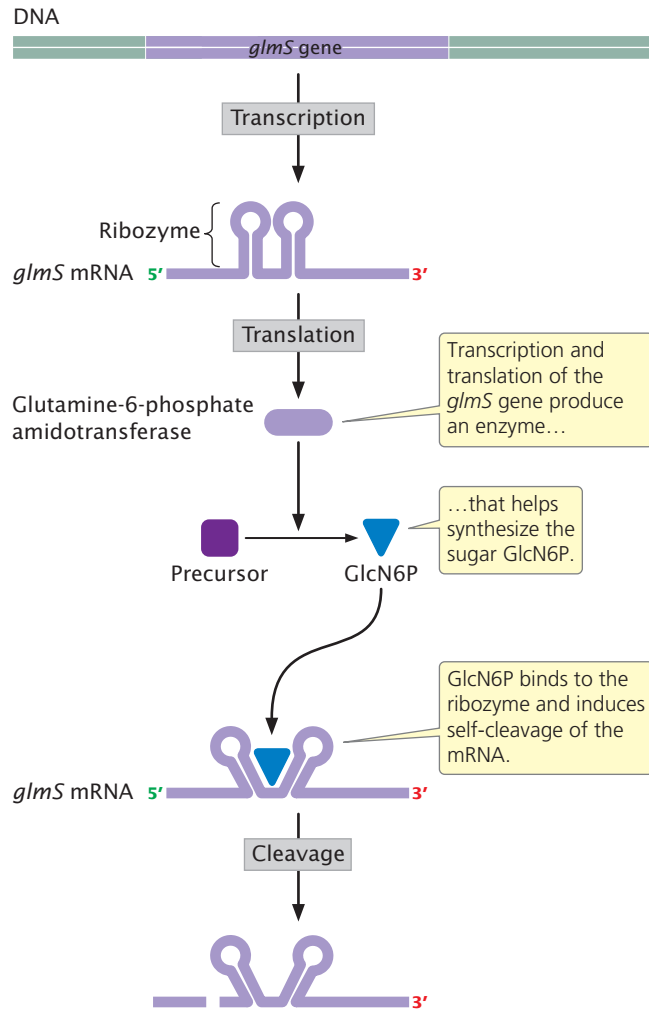
### (a) Regulatory protein present



### (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.